UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE PESQUISAS HIDRÁULICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS E SANEAMENTO AMBIENTAL

TESE DE DOUTORADO

REANÁLISE HIDROLÓGICA: ESTUDO DE CASO NA BACIA AMAZÔNICA

Sly Wongchuig Correa

Porto Alegre, RS – Brasil, 2019

SLY WONGCHUIG CORREA

TESE DE DOUTORADO

REANÁLISE HIDROLÓGICA: ESTUDO DE CASO NA BACIA AMAZÔNICA

Tese apresentada ao Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental, do Instituto de Pesquisas Hidráulicas - IPH da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Doutor em Recursos Hídricos e Saneamento Ambiental.

Orientador: Prof. Dr. Rodrigo Cauduro Dias de Paiva (IPH/UFRGS)

Coorientadores: Prof. Dr. Walter Collischonn (IPH/UFRGS), Dr. Sylvain Biancamaria (OMP/LEGOS)

Banca Examinadora:

Prof. Dr. Fernando Mainardi Fan

Prof. Dr. Pedro Luiz Borges Chaffe

Prof. Dr. Jhan Carlo Espinoza Villar

Prof. Dr. Pierre-André Garambois

Porto Alegre, RS – Brasil, 2019

CIP - Catalogação na Publicação

Wongchuig, Sly
Reanálise Hidrológica: Estudo de Caso na Bacia Amazônica / Sly Wongchuig. -- 2019.
236 f.
Orientador: Rodrigo de Paiva.

Coorientadores: Walter Collischonn, Sylvain Biancamaria.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Porto Alegre, BR-RS, 2019.

1. Modelagem Hidrológica-Hidrodinâmica. 2. Retrospectiva Hidrológica. 3. Assimilação de Dados. 4. Sensoriamento Remoto. 5. Secas e Cheias. I. de Paiva, Rodrigo, orient. II. Collischonn, Walter, coorient. III. Biancamaria, Sylvain, coorient. IV. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

SLY WONGCHUIG CORREA

TESE DE DOUTORADO

REANÁLISE HIDROLÓGICA: ESTUDO DE CASO NA BACIA AMAZÔNICA

Tese apresentada ao Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental, do Instituto de Pesquisas Hidráulicas - IPH da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Doutor em Recursos Hídricos e Saneamento Ambiental.

Aprovado em: Porto Alegre, 11 de junho de 2019

Prof. Dr. Rodrigo Cauduro Dias de Paiva – IPH/UFRGS
Orientador

Prof. Dr. Fernando Mainardi Fan – IPH/UFRGS
Examinador

Prof. Dr. Pedro Luiz Borges Chaffe – UFSC
Examinador

Prof. Dr. Jhan Carlo Espinoza Villar – IGE, Grenoble (França)
Examinador

Prof. Dr. Pierre-André Garambois – INSA, Strasbourg (França)
Examinador

Esta tese de doutorado foi desenvolvida na *Universidade Federal do Rio Grande do Sul* (UFRGS) do Brasil, no *Instituto de pesquisas Hidráulicas* (IPH), pelo *Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental* (PPGRHSA) sob orientação do Prof. Dr. Rodrigo Cauduro Dias de Paiva (IPH-UFRGS) e coorientação do Prof. Dr. Walter Collischonn (IPH-UFRGS), e com período de Doutorado sanduiche no *Observatoire Midi-Pyrénées* (OMP) e *Laboratoire d'études en géophysique et océanographie spatiales* (LEGOS) na França sob coorientação do Dr. Sylvain Biancamaria (OMP-LEGOS), com financiamento da *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) do Brasil.

Agradecimentos

Agradeço primeiramente a Deus, aos meus pais Alejandra e Raul, aos meus irmãos Gabriela e Alex, e aos meus sobrinhos Stephano e Diana Alessandra pelo suporte emocional, confiança, conselhos e apoio incondicional em cada passo que dou.

Ao meu professor orientador Dr. Rodrigo Paiva, pela confiança para assumir minha orientação, pela convivência, amizade, ensinamentos e suporte emocional durante este período de quatro anos de doutorado. Também pela motivação transmitida na decisão do tema de pesquisa e durante as diferentes fases do projeto.

Ao Prof. Dr. Walter Collischonn pela experiência transmitida, conselhos, convivência, amizade e inspiração ao longo do desenvolvimento do trabalho.

Ao Dr. Sylvain Biancamaria pela convivência e amizade durante minha estadia na cidade de Toulouse; pela orientação e pelos conselhos na técnica de assimilação de dados ensemble Kalman filter (EnKF) e de localização, assim como na manipulação e uso do simulador da futura missão do satélite SWOT.

Aos meus colegas do grupo de pesquisa de Hidrologia de Grande Escala (HGE) durante o período de doutorado, e especialmente aos meus amigos Ayan, Hugo, João Paulo, Germano, Pedro, Vinícius, Alice, Aline, Fernando, Cléber, Arthur, Rafael, Mino e Paulo, pela convivência excelente, por todos os debates e discussões de ideias, e por sempre me apoiar e me ajudar quando eu precisei. Também pelos conselhos que de alguma maneira colaboraram com o encaminhamento e conclusão deste trabalho. Aos meus amigos do programa de pós-graduação em Porto Alegre e aos que conhecí durante o período do meu doutorado sanduiche em Toulouse, Daniele, Benício, Matheus, Thiago, Franciele, Rômulo, Adrien, Henrique, Manuela, Marielle, Marie, Ahmad, Viet, Dorine, Felix e muitos mais, cujos nomes não caberiam nesta seção de agradecimentos; nunca serão esquecidos.

À Universidade Federal de Rio Grande do Sul (UFRGS), ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental (PPGRHSA) do Instituo de Pesquisas Hidráulicas (IPH) e aos meus professores pela formação acadêmica e profissional oferecida durante o período de doutorado. À Divisão de Alimentação da UFRGS através do restaurante universitário pela alimentação de qualidade.

Esta tese foi possível pelo amparo no fornecimento da bolsa de estudos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) durante o período de formação acadêmica de doutorado no Brasil e durante o período de doutorado sanduiche de seis meses na cidade de Toulouse - França.

Por último, mas não menos importante, quero agradecer ao Brasil por ter me acolhido e me oferecido muitas oportunidades para meu desenvolvimento acadêmico e profissional durante seis anos.

Resumo

Autor: Sly Wongchuig Correa

Orientadores: Rodrigo Cauduro Dias de Paiva, Walter Collischonn e Sylvain Biancamaria

A disponibilidade de registros hidro climátologicos acurados, espacial e temporalmente distribuídos tem sido um desafio no manejo dos recursos hídricos em nível global, especialmente nos países em desenvolvimento onde as observações in-situ são limitadas. Modelos hidrológicos têm sido usados para superar este problema através da estimativa de variáveis hidrológicas e hidráulicas espacialmente distribuídas, dentre elas a vazão (Q), nível d'água (WSE), extensão de áreas inundadas (FWE). Apesar disso, a modelagem ainda apresenta incertezas devido a incompreensão de muitos processos envolvidos, simplificações na conceptualização de alguns processos, erros nas forçantes hidro climáticas e/ou nas observações usadas para calibrar ou validar os modelos, dentre outros. Metodologias como a assimilação de dados (DA) têm sido desenvolvidas nas últimas décadas em áreas da climatologia, oceanografia e mais recentemente implementadas e aprimoradas na hidrologia continental. DA essencialmente utiliza observações como registros in-situ, de sensoriamento remoto e/ou dados sintéticos de futuras missões espaciais como a do satélite SWOT, para atualizar as variáveis de estado do modelo objetivando uma melhor estimativa das simulações. A precipitação é considerada como a principal forçante dos modelos hidrológicos, nesse sentido inúmeras bases de dados globais têm sido disponibilizadas baseadas em modelos climáticos, informação in-situ e/ou derivados de satélites, alguns denominados de reanálises climáticos. Nesse contexto, esta tese objetivou o desenvolvimento e implementação de diversas metodologias para a melhora das simulações na modelagem hidrológica hidrodinâmica, levando o conceito de reanálise climática para a hidrologia denominado aqui como reanálise hidrológica. Como uma primeira aproximação, oito bases de dados de precipitação global foram usadas para forçar o modelo de grandes bacias MGB na bacia Amazônica, e cujos resultados de vazão foram validados com registros de 27 estações in-situ. Resultados indicaram que três destas bases obtiveram um bom desempenho na representação das séries de vazão no período de estudo. Grande porcentagem de eventos hidrológicos extremos foram capturados, e a tendência temporal das cheias e secas se mostrou condizente com estudos anteriores baseados em informação in-situ. A principal limitação destas três bases é sua cobertura temporal (~30 anos), embora bases de precipitação mais extensas possam ser utilizadas conjuntamente com DA de extensos registros de vazão in-situ para melhorar as simulações. A técnica de DA ensemble Kalman filter (EnKF) foi avaliada no MGB na bacia do rio Taquari-Antas. Diversos cenários foram avaliados para determinar a sensibilidade na resposta do MGB devido a perturbação de diferentes variáveis de estado do modelo. Os resultados indicam que o esquema EnKF-MGB foi adequado para simular series de vazões e valores de vazões de referência. A partir destas experiências foi desenvolvida a reanálise hidrológica do século 20 na bacia Amazônica no período de 1900 até 2010 mediante o uso do MGB forçado pela reanálise de chuva ERA-20CM com remoção de viés e a assimilação de dados de inúmeras observações in-situ. O método de localização espacial no impacto da assimilação (LEnKF) foi implementado. Resultados indicam um bom desempenho pelo uso conjunto destas metodologias na estimativa das series de vazão, a partir da validação com inúmeras estações in-situ de vazão e nível d'água. As observações in-situ ainda apresentam uma limitação espacial que podem ser complementadas pelas abundantes informações de sensoriamento remoto das últimas décadas. Os dados sintéticos da futura missão SWOT foram implementados e avaliados no MGB e no esquema de EnKF especificamente na bacia do rio Purus mediante o denominado "twin experiment". O modelo foi corrompido para atingir as incertezas de modelos hidrológicos globais (GHMs), e as pseudo-observações de Q, WSE e FWE foram usadas para recuperar o modelo original. Resultados indicam que considerando os erros esperados das observações do SWOT o modelo conseguiu recuperar a maioria das variáveis de estado do modelo no nível de incerteza dos GHMs. Além disso, quando a metodologia de DA de múltiplas observações foi implementada, os resultados se mostraram ainda mais promissores. Considerando estes resultados é instigante pensar no uso das metodologias desenvolvidas nesta tese mediante o uso de múltiplos produtos de sensoriamento remoto atualmente disponiveis como altimetria por satélite, armazenamento d'água terrestre, extensão de superfícies inundadas, dentre outros. Desta forma incrementando a performance nas simulações, e adicionalmente tendendo a uma representação mais física dos processos hidrológicos devido à diversidade das observações.

Palavras-chave: Modelagem Hidrológica-Hidrodinâmica, Retrospectiva Hidrológica, Assimilação de Dados, Sensoriamento Remoto, Secas, Cheias.

Abstract

Author: Sly Wongchuig Correa

Advisors: Rodrigo Cauduro Dias de Paiva, Walter Collischonn and Sylvain Biancamaria

The availability of accurate, distributed spatially and temporally hydroclimatic records, has been a challenge for water resources management worldwide, mainly in developing countries, where in situ observations are limited. Hydrological models have been used to overcome this problem by estimating spatially distributed hydrological and hydraulic variables, including discharge (Q), water surface level (WSE), flooded water extent (FWE), among others. In spite of this, the models still have uncertainties due to the lack of understanding of many of the processes involved, simplifications in the conceptualization of some processes, errors in the hydroclimatic forcing and/or in the observations used to calibrate or validate the models, among others. In recent decades, methodologies such as data assimilation (DA) have been developed in the areas of climatology, oceanography and, more recently, have been implemented and improved in continental hydrology. DA essentially uses observations such as in situ records, remote sensing and/or synthetic data from future space missions such as the SWOT satellite, to update model state variables in order to better estimate simulations. The precipitation is considered the main forcing for hydrological models, and numerous global databases based on climate models, in situ information and/or satellite derivatives have been made available, some of which are referred to as climatic reanalysis. In this context, this thesis aimed at the development and implementation of various methodologies for the improvement of simulations in hydrological and hydrodynamic modeling, bringing the concept of climate reanalysis to hydrology, here referred to as hydrological reanalysis. As a first approach, eight global rainfall databases were used to force the large-scale model MGB in the Amazon basin, and simulations were validatd against discharge records from 27 in-situ stations. The results indicated that three of these databases performed well in representing the discharge series during the study period. A large percentage of extreme hydrological events were also captured and the temporal trend of floods and droughts were consistent with previous studies based on in situ information. The main limitation of these three datasets is their temporal coverage (~30 years); however, it is possible to use longest rainfall datasets together with methods such as DA of extensive in-situ discharge records to improve the simulations. The DA ensemble Kalman filter (EnKF) technique was assessed into the MGB and the results were evaluated in the Taquari-Antas river basin. Several scenarios were assessed to determine the sensitivity of the MGB model response due to disturbance of different state variables of the model. The results indicate that the EnKF-MGB scheme was suitable for simulating discharge series and reference discharge values. Based on these experiences, the 20th century hydrological reanalysis was developed in the Amazon basin in the period from 1900 to 2010 using the MGB model forced by the ERA-20CM precipitation reanalysis with bias removal and data assimilation by numerous in situ observations. The method of spatial localization in the impact of assimilation (LEnKF) was implemented. The results indicate a good performance by the joint use of these methodologies in the estimation of the series of discharge, from the validation with numerous in-situ stations of discharge and water level. In situ observations still present a spatial limitation that can be complemented by the abundant remote sensing information available for the last few decades. The synthetic data of the future SWOT mission was implemented and assessed in the MGB and the EnKF scheme The Purus basin was chosen to perform the so-called "twin experiment". The model was corrupted to achieve the uncertainties of global hydrological models (GHMs), and pseudo-observations of Q, WSE and FWE were used to recover the original model. The results indicate that considering the expected errors of the SWOT observations, the model was able to recover most of the state variables of the model regarding GHMs uncertainty level. Moreover, when the multiple-observation DA methodology was implemented, the results were even more promising. Considering these results, it is interesting to think about the use of the methodologies developed in this thesis, using multiple current remote sensing products such as satellite altimetry, terrestrial water storage, extension of flooded surfaces, among others. In this manner increasing the performance of the simulations, and additionallyby tending to a more physical representation of the hydrological processes due to the diversity of the observations.

Keywords: Hydrological-Hydrodynamic Modeling, Hydrological Retrospective, Data Assimilation, Remote Sensing, Droughts, Floods.

Resumen

Autor: Sly Wongchuig Correa

Orientadores: Rodrigo Cauduro Dias de Paiva, Walter Collischonn y Sylvain Biancamaria

La disponibilidad de registros hidroclimáticos precisos, distribuidos espacial y temporalmente, ha sido un desafío para la gestión de los recursos hídricos a nivel mundial, especialmente en países en vías de desarrollo, donde las observaciones in situ son limitadas. Modelos hidrológicos has sido usados para superar este problema mediante la estimativa de variables hidrológicas e hidráulicas de manera espacialmente distribuida, entre ellas caudal (Q), nivel del agua (WSE), extensión de áreas inundadas (FWE). A pesar de ello, los modelos aún presentan incertidumbres debido a la falta de comprensión de muchos de los procesos involucrados, simplificaciones en la conceptualización de algunos procesos, errores en las forzantes hidroclimáticas y/o en las observaciones utilizadas para calibrar o validar los modelos, entre otros. En las últimas décadas se han desarrollado metodologías como la asimilación de datos (DA) en las áreas de climatología, oceanografía y más recientemente, se han implementado y mejorado en la hidrología continental. DA utiliza esencialmente observaciones tales como registros in situ, teledetección y/o datos sintéticos de futuras misiones espaciales tales como el satélite SWOT, para actualizar las variables de estado del modelo con el fin de estimar mejor las simulaciones. Se considera que la precipitación es la principal forzante de los modelos hidrológicos, en ese sentido, durante los últimos años, han sido puestos a disposición numerosas bases de datos de lluvia de cobertura global basadas en modelos climáticos, información in situ y/o derivados de satélites, algunos de los cuales se denominan reanálisis climáticos. En este contexto, esta tesis tuvo como objetivo el desarrollo e implementación de varias metodologías para la mejora de las simulaciones mediante el uso de modelos hidrológicos e hidrodinámicos, llevando el concepto de reanálisis climático a la hidrología, denominada aquí como reanálisis hidrológico. Como primera aproximación, se utilizaron ocho bases de datos globales de precipitaciones para forzar el modelo de grandes cuencas MGB en la cuenca Amazónica, cuyas simulaciones fueron validadas con registros de caudal provenientes de 27 estaciones in-situ. Los resultados indicaron que tres de estas bases de datos obtuvieron un buen desempeño en la representación de la serie de caudales durante el periodo de estudio. Se capturó un gran porcentaje de eventos hidrológicos extremos y la tendencia temporal de inundaciones y sequías fue consistente con estudios previos basados en información in situ. La principal limitación de estas tres bases es su cobertura temporal (~30 años), sin embargo, es posible utilizar bases más extensas junto a métodos como DA de registros extensos de caudal in-situ para mejorar las simulaciones. La técnica de DA ensemble Kalman filter (EnKF) fue evaluada usando el MGB en la cuenca del rio Taquari-Antas. Diversos escenarios fueron evaluados para determinar la sensibilidad de la respuesta del modelo MGB debido a la perturbación de diferentes variables de estado del modelo. Los resultados indican que el esquema EnKF-MGB fue adecuado para simular series de caudal y valores de caudal de referencia. Basándose en estas experiencias, el reanálisis hidrológico del siglo 20 se desarrolló para la cuenca del Amazonas en el período de 1900 a 2010 mediante el uso del modelo MGB forzado por el reanálisis de precipitación ERA-20CM con remoción de sesgo y la asimilación de los datos de numerosas observaciones in situ. Se implementó el método de localización espacial en el impacto de la asimilación (LEnKF). Los resultados indican un buen desempeño por el uso conjunto de estas metodologías en la estimativa de las series de caudal, a partir de la validación con numerosas estaciones in-situ de caudal y nivel de agua. Las observaciones in situ siguen presentando una limitación espacial que puede complementarse con abundante información de teledetección disponible para las últimas décadas. Los datos sintéticos de la futura misión SWOT fueron implementados y evaluados en el MGB usando el esquema EnKF, específicamente en la cuenca del río Purus a través del llamado "twin experiment". El modelo fue corrompido para alcanzar las incertidumbres de los modelos hidrológicos globales (GHMs), y las pseudo-observaciones de Q, WSE y FWE fueron utilizadas para recuperar el modelo original. Los resultados indican que considerando los errores esperados de las observaciones del SWOT, el modelo fue capaz de recuperar la mayoría de las variables de estado del modelo considerando el nivel de incertidumbre de los GHMs. Además, cuando se implementó la metodología de DA de observaciones múltiples, los resultados fueron aún más prometedores. Considerando estos resultados, es interesante pensar en el uso de las metodologías desarrolladas en esta tesis a través del uso de múltiples productos actuales de teledetección como altimetría satelital, almacenamiento de agua terrestre, extensión de superficies inundadas, entre otros. de esta forma incrementando la performance de las simulaciones, y adicionalmente tendiendo a una representación más física de los procesos hidrológicos debido a la diversidad de las observaciones.

Keywords: Modelaje Hidrológico-Hidrodinámico, Retrospectiva Hidrológica, Asimilación de Datos, Sensoriamiento Remoto, Sequias, Inundaciones.

Résumé

Autor: Sly Wongchuig Correa

Directeurs: Rodrigo Cauduro Dias de Paiva, Walter Collischonn et Sylvain Biancamaria

La disponibilité de données hydroclimatiques exactes, réparties dans l'espace et dans le temps, a constitué un défi pour la gestion des ressources en eau dans le monde entier, principalement dans les pays en développement, où les observations in situ sont limitées. Des modèles hydrologiques ont été utilisés pour résoudre ce problème en estimant des variables hydrologiques et hydrauliques réparties dans l'espace, y compris le débit (Q), le niveau de la surface de l'eau (WSE) et l'étendue des eaux inondées (FWE). Malgré cela, les modèles comportent encore des incertitudes dues au manque de compréhension de plusieurs des processus impliqués, à des simplifications dans la conceptualisation de certains processus, à des erreurs dans le forçage hydroclimatique et/ou dans les observations utilisées pour calibrer ou valider les modèles, entre autres. Au cours des dernières décennies, des méthodologies telles que l'assimilation de données (DA) ont été mises au point dans les domaines de la climatologie et de l'océanographie et, plus récemment, ont été appliquées et améliorées en hydrologie continentale. La DA utilise essentiellement des observations telles que les enregistrements in situ, la télédétection et/ou les données synthétiques des futures missions spatiales telles que le satellite SWOT, pour mettre à jour les variables d'état du modèle afin de mieux estimer les simulations. Les précipitations sont considérées comme le principal forçage des modèles hydrologiques, et de nombreuses bases de données mondiales basées sur des modèles climatiques, des informations in situ et/ou des dérivés satellitaires ont été mises à disposition, dont certaines sont appelées réanalyses climatiques. Dans ce contexte, cette thèse visait le développement et la mise en œuvre de diverses méthodologies pour l'amélioration des simulations en modélisation hydrologique et hydrodynamique, apportant le concept de réanalyse climatique à l'hydrologie, appelée ici réanalyse hydrologique. Dans un premier temps, huit bases de données mondiales sur les précipitations ont été utilisées pour forcer le modèle à grande échelle de la MGB dans le bassin amazonien. D'après la validation à partir des registres de débit de 27 stations in situ, les résultats indiquent que trois de ces bases de données ont donné de bons résultats pour ce qui est de représenter les séries de débit au cours de la période à l'étude. Un pourcentage élevé d'événements hydrologiques extrêmes a également été capturé et la tendance temporelle des inondations et des sécheresses était conforme aux études antérieures fondées sur des données in situ. La principale limite de ces trois ensembles de données est leur couverture temporelle (~30 ans); cependant, il est possible d'utiliser les ensembles de données sur les précipitations les plus longues avec des méthodes telles que la DA de relevés détaillés du débit in situ pour améliorer les simulations. La technique du filtre de Kalman de l'ensemble DA (EnKF) a été évaluée à l'aide de la MGB dans le bassin du Taquari-Antas. Plusieurs scénarios ont été évalués pour déterminer la sensibilité de la réponse du modèle MGB en raison de la perturbation de différentes variables d'état du modèle. Les résultats indiquent que le schéma EnKF-MGB permettait de simuler des séries de débits et des valeurs de débits de référence. Sur la base de ces expériences, la réanalyse hydrologique du 20ème siècle a été développée dans le bassin amazonien entre 1900 et 2010 en utilisant le modèle MGB forcé par la réanalyse des précipitations ERA-20CM avec élimination des biais et assimilation de données par de nombreuses observations in situ. La méthode de localisation spatiale dans l'impact de l'assimilation (LEnKF) a été mise en œuvre. Les résultats indiquent une bonne performance par l'utilisation conjointe de ces méthodologies dans l'estimation des séries de débits, à partir de la validation avec de nombreuses stations in situ de débit et de niveau d'eau. Les observations in situ présentent encore une limitation spatiale qui peut être complétée par l'abondance des informations de télédétection disponibles depuis quelques décennies. Les données synthétiques de la future mission SWOT ont été mises en œuvre et évaluées dans la MGB et le schéma EnKF, en particulier dans le bassin du Purus par le biais de la "twin experiment". Le modèle a été corrompu pour atteindre les incertitudes des modèles hydrologiques globaux (GHMs), et des pseudo-observations de Q, WSE et FWE ont été utilisées pour récupérer le modèle original. Les résultats indiquent que, compte tenu des erreurs attendues des observations SWOT, le modèle a été en mesure de récupérer la plupart des variables d'état du modèle concernant le niveau d'incertitude des GHMs. De plus, les résultats aient été encore plus prometteurs lorsque la méthodologie DA à observations multiples a été appliquée. Au vu de ces résultats, il est intéressant de réfléchir à l'utilisation des méthodologies développées dans cette thèse, à travers l'utilisation de multiples produits actuels de télédétection tels que l'altimétrie par satellite, le stockage terrestre de l'eau, l'extension des surfaces inondées, entre autres. Non seulement en raison de l'augmentation des performances des simulations, mais aussi en tendant vers une représentation plus physique des processus hydrologiques grâce à la diversité des observations.

Mots clés: Modelisation Hydrologique-Hydrodynamique, Rétrospective Hydrologique, Assimilation de Donnés, Télédétection, Sécheresses, Inondations.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviações

ANA – Agencia Nacional das Aguas

B.L.U.E. - Best Linear Unbiased Estimator

CDF - Cumulative distribution function

CFSR - Climate Forecast System Reanalysis

CHIRPS - The Climate Hazards Group InfraRed Precipitation with Station data

CIRES - Research in Environmental Sciences

CNES - Centre National d'Etudes Spatiales

CPRM - Geological Survey of Brazil

CRU – Climatological dataset of the Climate and Research Unit

CSA/ASC - Canadian Space Agency/Agence Spatiale Canadienne

CSR - University of Texas / Center for Space Research

DA – Data Assimilation

DEM – Digital Elevation Model

DLR - German Aerospace Center

eartH2Observe - Global Earth Observation for Integrated Water Resource Assessment

ECMWF - European Centre for Medium-Range Weather Forecasts

EnKF – Ensemble Kalman Filter

EnSRF - Ensemble Square Root Filter

eQM - Empirical quantile mapping

ERA-20CM - The 20th Century Reanalysis

EROS - Earth Resources Observation and Science

GCM - Global climate models

GFZ - GeoForschungsZentrum Potsdam

GIEMS - Global Inundation Extent from Multi-Satellites

GPCP - Global Precipitation Climatology Project

GSWO - Global Surface Water Occurence

gQM - Parametric quantile mapping gamma distribution

GRACE – Gravity Recovery and Climate Experiment mission

GRACE-FO – Gravity Recovery and Climate Experiment Follow-On mission

GSFC - Goddard Space Flight Center

HBV - Hydrologiska Byrans Vattenavdelning

HOP - HYBAM Observed Precipitation

HR – Hydrological Retrospective

HRU – Hydrological Response Units

HRXX – Hyadrological Reanalysis of the 20th century

HYBAM – Hydrologie, Biogeochimie at Geodynamique du Bassin Amazonien

IPH – Instituto de Pesquisas Hidráulicas

ISRO - Indian Space Research Organisation

JPL - Jet Propulsion Laboratory

JRA - Japanese Meteorological Agency

Q_{90} - 90% exceedance discharge

Q_{LTA} - Long-term average discharge

$Q_{MAX\ RP=5}$ - Maximum discharge with five years of return period

KF - Filtro de Kalman

KGE – Kling and Gupta Efficiency

LEGOS - Laboratoire d'Etudes en Géophysique et Océanographie

LEnKF – Local Ensemble Kalman Filter

LSC - Linear scaling

LSPCS - Large Scale L2 HR Pixel Cloud Simulator

MERRA - Modern-Era Retrospective analysis for Research and Applications

MIKE SHE - Systeme Hydrologique European

MGB – Modelo de Grandes Bacias

MSWEP - Multi-Source Weighted-Ensemble Precipitation

NASA – National Aeronautics and Space Administration

NCEP-NCAR - National Center for Environmental Prediction –National Center for Atmospheric Research

NSE - Nash and Suttcliffe efficiency index

NRC - National Research Council

OL – Open Loop simulation

ORE-HyBam - Environmental Research Observatory SO HYBAM

PCR-GLOBWB - PCRaster GLOBal Water Balance model

PF - Particle Filter

PSS - Peirce Skills Score

PUB - Prediction in Ungauged Basins

PUC - Prediction in Ungauged Climates

RCM - Regional climate models

REnKF - Retrospective Ensemble Kalman Filter

RMSE - Root-mean-square error

RRMSE – Relative root-mean-square error

SENAMHI – Servicio Nacional de Meteorologia e Hidrologia del Perú

SIC - Sea ice cover

SO HYBAM - Environmental Research Observatory

SRTM – Shuttle Radar Topography Mission

SST - Sea Surface Temperature

SWAT - Soil and Water Assessment Tool

SWOT – Surface Waters and Ocean Topography mission

TRMM – Tropical Rainfall Measurement Mission

TWS – Terrestrial Water Storage

UFRGS – Universidade Federal do Rio Grande do Sul

UKSA - United-Kindom Space Agency

USGS - United States Geological Survey

VIC - Variable Infiltration Capacity

Capítulo 1

Introdução

1.1. Introdução geral e justificativa

1.1.1. Contexto

A disponibilidade de registros consistentes e detalhados de eventos hidrológicos tem sido um desafio no manejo hídrico ao redor do mundo; por exemplo a caracterização de eventos hidrológicos extremos passados é uma tarefa difícil principalmente em regiões em desenvolvimento onde observações in-situ são escassas e limitadas apenas às décadas recentes (Sivapalan et al., 2003). Mas por que é importante conhecer o comportamento hidrológico passado? Primeiramente o conhecimento da hidrologia das últimas décadas nos permite analizar e relativizar eventos extremos e raros tais como cheias e secas. Diversos estudos requerem uma amostra observacional suficientemente grande, e estas informações seriam de importância para um adequado planejamento e organização dos tomadores de decisões. Desta maneira a gestão dos recursos hídricos permite estudos de: alocação dos recursos hídricos numa cidade ou bacia hidrográfica; desenvolvimento de projetos de engenharia que reduzam os impactos (e.g. inundações); análises de eventos extremos passados tais como cheias e secas, dentre outros, sendo crítico na implementação de políticas sustentáveis para a gestão do recurso hídrico e contra seus impactos negativos, o que aponta a necessidade de pesquisas voltadas para a compreensão dos eventos ocorridos no passado visando uma maior segurança hídrica. Estas condições demonstram a necessidade por obter registros hidrológicos passados que permitam entender o comportamento hidrológico de uma determinada região assim como o estudo de quão frequentes ou intensos estes foram (e.g. Marengo e Espinoza, 2016; Swierczynski et al. 2017). Claramente, a disponibilidade de series históricas e espacialmente distribuídas traz vantagens tanto na geração do conhecimento cientifico, como nas melhorias na gestão que finalmente se traduziriam em benefícios econômicos para a população.

Estudos recentes têm sido realizados com o objetivo de suprir as limitações na disponibilidade de dados hidro climatológicos mediante o uso de aproximações ou medições indiretas que compreendem períodos mais extensos que os registros dos instrumentos tradicionais (Moss et al., 1988; Jarret, 1991). Áreas do conhecimento como a paleo-hidrologia (Barber e Finney, 2000; Nunnery, 2012) têm focado na geração de registros que podem se estender até 5000 anos (Baker et al., 2002; Benito et al., 2005) como a estimativa de inundações históricas (Luo et al. 2014) associadas a paleo-indicadores (e.g. sedimentos das inundações, registros de sedimentos nos lagos, medições isotópicas, morfologia erosiva, dentre outros), ou estimativas de series de vazão com a finalidade de reconstruir a variabilidade hidrológica e climática (Barber e Finney, 2000; Nunnery, 2012) através do uso de modelos hidrológicos ou abordagens hidráulicas. Estas tentativas têm por objetivo reproduzir as características hidrológicas para períodos em que os registros ainda não encontravam-se disponibilizados, embora estas metodologias ainda apresentam certas limitações como por exemplo relacionados com a escala temporal (Razavi et al., 2016) ou que estas reconstruções estão geralmente restritas espacialmente aos pontos em que as análises foram feitas (U.S. Army Corps of Engineers, 2003), o que equivaleria a limitação que se

tem com a instrumentação in-situ atual em regiões extensas. Nesse contexto, a presente pesquisa propõe o desenvolvimento de diferentes metodologias para a geração destes registros de maneira espacialmente distribuída e temporalmente extensa. Desta forma, é proposta a continuação a fundamentação teórica de diferentes intens que suportam a concepção desta pesquisa.

1.1.2. Fundamentos

a. Modelagem hidrológica de grande escala

Modelos hidrológicos têm sido amplamente empregados para oferecer estas estimativas através da simulação do ciclo hidrológico continental (Wood et al., 2011; Archfield et al., 2015). Dentre os mais conhecidos temos o Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), European Hydrological System (MIKE-SHE) (Abbott et al., 1996a, b), Hydrologiska Byråns Vattenbalansavdelning (HBV) (Bergström, 1976, 1992), TOPMODEL (Beven et al., 1979), Variable Infiltration Capacity (VIC) (Liang et al., 1994, 1996), ORCHIDEE (d'Orgeval et al., 2008), JULES (Best et al., 2011), SURFEX-TRIP (Decharme et al., 2010, 2013), HTESSEL (Balsamo et al., 2009), PCR-GLOBWB (Van Beek et al., 2011), W3RA (van Dijk et al., 2014), Modelo de Grandes Bacias (MGB) (Collischonn et al., 2007), dentre outros. Recentemente houve uma melhora na representação física dos processos e também na eficiência computacional, o que tem permitido oferecer capacidades de simulação em escalas continentais ou até em escala global (Gao et al., 2010; Paiva et al., 2013a; Yamazaki et al., 2013; Sood e Smakhtin, 2015; Siqueira et al., 2018). Estas estimativas têm atingido certa relevância local, como descrito na iniciativa "hyper-resolution" pelo Wood et al. (2011) e Bierkens et al. (2015). Por exemplo, o modelo MGB tem sido usado em diferentes regiões da America do Sul (e.g. Pontes et al., 2017; Siqueira et al., 2018; Fleischman et al., 2019), no rio Niger (e.g. Fleischmann et al., 2017) e amplamente na bacia do rio Amazonas (Paiva et al., 2013a), obtendo resultados adequados nas estimativas diárias da vazão e do nivel'água. As principais aplicações destes modelos têm focado na necessidade de representar fenômenos hidrológicos importantes como previsão de vazões para sistemas de suporte a decisão, identificação dos impactos gerados pelas mudanças no clima e uso do solo, planejamento de recursos hídricos, dentre outros (Gayathri, et al. 2015). Ainda existem incertezas que devem ser tomadas em consideração tais como as forçantes hidro meteorológicas, a estrutura do modelo, o modelo digital de elevação, deficiências na calibração, dentre outras (Döll et al., 2008; Thielen et al., 2010; Kauffeldt et al. 2016; Siqueira et al., 2018). Estimativas a priori das incertezas de três modelos globais (HTESSEL, LISFLOOD e WaterGAP) apresentam valores de 52% até 103% na raiz do erro quadrático médio relativo (RRMSE) da vazão. Por outro lado, avaliações sintéticas de modelos hidrológicos em escala global (e.g. Fleischmann et al., 2019) apresentam valores das incertezas para as estimativas do nível d'água de ~7.4m e ~1.3m na raiz do erro quadrático médio (RMSE) para valores absolutos e anomalias respectivamente.

Observações in-situ (e.g. vazão ou nível d'agua) têm sido tradicionalmente usadas na área dos recursos hídricos para calibrar e validar modelos hidrológicos (Alsdorf et al., 2007; López López et al., 2017). Em contraste, nos últimos anos o sensoriamento remoto tem complementado a rede in-situ devido a i) uma maior cobertura espacial (Engman, 1995; Tang et al., 2009; van Dijk e Renzullo, 2011; Wanders et al. 2014) e ii) acessibilidade a uma ampla gama de resoluções espaciais, radiométricas e espectrais; estas características permitem o uso em aplicações continentais e globais na hidrologia (Smith, 1997; Melesse et al. 2007; López López et al. 2017). Nesse contexto as informações hidro meteorológicas são fundamentais nos modelos hidrológicos para obter razoáveis estimativas das variáveis hidrológicas (e.g. umidade do solo, vazão, etc.) que permitam um adequado entendimento da dinâmica e de sua distribuição global (López López et al., 2017).

b. Assimilação de dados

Técnicas de assimilação de dados (*data assimilation*, DA) foram inicialmente desenvolvidas em áreas como climatologia e oceanografia e posteriormente implementadas e avaliadas na área da hidrologia (Sun et al., 2016). Sua implementação em escala local (e.g. Vrugt et al., 2005; Clark et al., 2008; Xie e Zhang, 2010; Kurtz et al., 2017) e em grande escala (e.g. Paiva, 2013b; van Dijk et al., 2014; López López et al., 2015; Liu et al., 2012a, 2012b, 2016; Andreadis et al., 2017) encontram-se ainda em desenvolvimento, ou ainda não existam diretrizes bem consolidadas de como implementar esta técnica nos diversos cenários hidrológicos (Liu e Gupta, 2007). Desta maneira é necessario desenvolver técnicas capazes de assimilar de forma efetiva as informações relevantes contidas nas observações. Existem atualmente diversos métodos abordados na literatura para a adaptação desta técnica nos processos hidrológicos (Entekhabi et al., 1994; Vrugt et al., 2005; Paiva, 2013b). Técnicas de DA como o filtro de Kalman por ensemble (EnKF) possibilitam a simplificação na sua implementação em modelos hidrológicos além de representar de maneira mais adequada o comportamento não linear dos processos hidrológicos, comparados com versões previas (e.g. Filtro de Kalman-KF, Filtro de Kalman Estendido-EFK). Estudos prévios das técnicas de DA pelo EnKF, como Paiva et al. (2013b), desenvolveram um esquema de DA na Bacia Amazônica através da assimilação de informação in situ de estações hidro meteorológicas e de altimetria por radar, indicando claramente o potencial desta metodologia no campo da hidrologia. Estudos mais recentes (Revilla-Romero et al., 2016) implementaram o mesmo método (EnKF) no modelo chuva-vazão LISFLOOD assimilando extensões de áreas inundadas a partir de sensoriamento remoto em escala continental (Sul América e África). Embora ainda existam muitas questões abertas na área de DA, como por exemplo, estes não foram plenamente utilizados para gerar longos registros hidrologicos; também existem hoje em dia alguns desafios técnicos para a diminuição dos erros nas estimativas, tais como i) métodos de localização ou ii) a assimilação conjunta de múltiplas observações. A primeira delas permite obter um impacto local da DA das observações (Rasmussen et al. 2015; Munier et al., 2015; Zhang et al., 2016) enquanto a segunda técnica permite obter uma representação mais física de diferentes processos do modelo (Khaki et al., 2018, 2019). Nos últimos

anos estas técnicas têm demonstrado ser relevantes na melhora das estimativas hidrológicas no esquema de DA, pelo que ainda existe uma grande oportunidade para serem exploradas.

c. Observações usadas na hidrologia

A medição das variáveis do ciclo hidrológico é indispensável para utilizar ou elaborar qualquer estudo hidrológico devido que as estimativas das variáveis terrestres podem ser melhoradas através da integração de observações na modelagem hidrológica. Segundo Balsamo et al. (2018) esta integração em modelos abrange vários aspectos: (1) a integração dinâmica das observações em modelos através de técnicas de assimilação de dados, (2) o uso de observações para validação e calibração de modelos e (3) o mapeamento dos parâmetros do modelo utilizados para caracterizar a representação das propriedades do terreno dentro do modelo (e.g. propriedades do solo, cobertura do solo, etc.). Existe assim, uma grande oportunidade para melhorar o desenvolvimento atual de estratégias dentro da modelagem hidrológica. Assim, nesta seção são descritas as informações in situ e as obtidas a partir de sensoriamento remoto, sendo as principais fontes de dados do ciclo hidrológico continental.

As informações in situ (e.g. vazão, nível d'agua, umidade do solo, etc.) são tradicionalmente usadas nos recursos hídricos para restringir (e.g. calibrar e validar) modelos hidrológicos (e.g. Gupta et al., 1998; Madsen, 2000; Wanders et al., 2014), modelos climáticos, modelos de superfície terrestre, dentre outros. Apesar da sua cobertura limitada, densidade limitada e de suas incertezas, as observações in situ continuam a ser a principal referência do sistema de observação devido que são consideradas como as informações mais próximas da realidade. Por outro lado, especificamente na hidrologia, o monitoramento da dinâmica das extensões d'água, armazenamento e o fluxo em áreas que não estão controladas pela rede de um canal definido (e.g. wetlands, planícies interfluviais, etc.) são difíceis de ser mesuradas por estações in situ. Adicionalmente, muitas bacias hidrográficas no mundo ainda se mantem pouco monitoradas (Tang, et al., 2009; Hrachowitz et al., 2013; Loukas e Vasiliades, 2014) especialmente em países em desenvolvimento onde a rede é caracterizada por sua irregular e espalhada distribuição. Atualmente uma das principais preocupações que tem sido identificada é uma diminuição global na rede de monitoramento in situ durante as últimas décadas (Alsdorf et al., 2003, 2007; Sivapalan, 2003), especialmente desde os anos 1980s (Fekete, et al., 2012) assim como sua disponibilidade para fins científicos, o que poderia ter sido limitado devido a politicas institucionais ou internacionais, restringindo a troca de informação (Alsdorf et al., 2007). Um claro exemplo em nível regional é a nova política da Agência Nacional de Águas no Brasil (ANA) pela consituição de uma Rede Hidrometeorológica Nacional de Referência (RHNR), deixando inoperantes muitas estações em favor de melhorar a eficiência de uma rede menor. Uma desvantagem seria a perda de informações de longo prazo, implicando um sério desafio devido que se negligenciariam alguns processos como a mudança do clima, porém seu impacto nos sistemas hidrológicos se tornaria muito mais difícil de ser estudado (Hidalgo et al., 2009). Alguns estudos chamam a atenção à importância das observações in situ em

escalas de bacia hidrográfica, regionais e até mesmo globais (Stahl et al., 2010; Zacharias et al., 2011), argumentando que é necessário um esforço comunitário para manter operacional as redes in situ.

Nas últimas décadas, a capacidade de observar a hidrosfera a partir do ponto de vista do espaço redefiniu não só a nossa perspectiva da Terra como um sistema interligado, mas também a forma como descrevemos os processos dinâmicos que ocorrem acima, sobre e sob a sua superfície. Em contraste com as observações in-situ, métodos como o sensoriamento têm sido usados em décadas recentes, de maneira complementar, devido a i) uma maior abrangência espacial (Engman, 1995; Pietroniro e Prowse, 2002; Tang et al., 2009; van Dijk e Renzullo, 2011; Wanders et al. 2014) e ii) acessibilidade a partir de uma grande diversidade de sensores e resoluções espaço-temporais, radiométricas e espectrais. Estes sensores remotos medem indiretamente diferentes parâmetros da superfície através da radiação eletromagnética refletida pela superfície da terra (Schmugge et al. 2002), e que têm oferecido a possibilidade de obter informação sobre variáveis de interesse (e.g. umidade do solo, cobertura vegetal, altimetria de rios, dentre outros) em áreas como a hidrologia (Sandholt et al., 1999; Wagner et al., 2009). Desta maneira o monitoramento através do sensoriamento remoto permite seu uso em aplicações em grande escala (Smith, 1997; Melesse et al., 2007; López López et al. 2017) e/ou regiões pouco accessíveis como a bacia Amazônica (Koblinsky et al., 1993; Hess et al., 2015) ou a bacia do Congo (Alsdorf et al., 2016; Becker et al., 2018), além que provee uma melhor representatividade dos processos físicos da bacia (Yu, 2015).

Por outro lado, nas últimas décadas modelos hidrológicos foram adaptados para o uso mais eficiente destas novas informações (e.g. Pietroniro e Prowse, 2002; Getirana, 2010; Paiva et al., 2013a). Desta maneira, informações de sensoriamento remoto podem ser utilizadas para (i) a caracterização de hidrosistemas, (ii) desenvolvimento de novas bases teóricas baseadas em observações, (iii) validação de modelos de simulação ou (iv) serem utilizadas integradas a modelos de simulação buscando a melhor representação dos processos físicos. Informações de sensoriamento remoto também são usadas mediante técnicas de DA para a melhora nas estimativas das variáveis de estado do modelo. Algumas destas variáveis são umidade do solo (e.g. Massari et al., 2015; Baguis e Roulin, 2017; Crow e Ryu, 2009), variação no armazenamento d'água terrestre (e.g. Tapley et al., 2004; Khaki et al., 2018), extensão de área inundada (e.g. Lai et al., 2014), assim como informações sintéticas de missões futuras como a missão SWOT (Surface Water and Ocean Topography). A missão SWOT promete suprir informações para um melhor entendimento da dinâmica d'água continental em comparação com missões atuais (Andreadis et al., 2007; García-Pintado et al., 2013; Pedinotti et al. 2014; Oubanas et al. 2018) assim que apresentam uma melhor acurácia (~0.1m) em comparação com missões atuais de altimetria por radar (0.28m até 1.2m) (O'Loughlin et al., 2016).

Diversos estudos têm mostrado um aumento na eficiência das simulações hidrológicas pelo uso conjunto das observações de sensoriamento remoto no esquema de DA (e.g. Khaki et al., 2018, 2019).

Em consequência, existe uma grande oportunidade para aproveitar esta crescente disponibilidade de informação (Corato et al., 2014; Pietroniro e Prowse, 2002; Michailovsky et al., 2013) para seu uso dentro dos sistemas de assimilação de dados em hidrologia (Liu e Gupta, 2007), ou especificamente através do esquema EnKF (Reichle et al. 2002; Crow e Wood, 2003; Paiva et al., 2013b).

d. Reanálises climáticas

Recentemente em áreas da ciência como a climatologia e/ou oceanografía têm sido desenvolvido métodos denominados de reanálises climáticas (*climatic reanalysis*, CR) (e.g. Gibson et al., 1997; Kistler et al., 2001; Compo et al., 2011). As CR são definidas como o estado da arte de variáveis climáticas, produzidas através da combinação ótima de modelos matemáticos e inúmeras observações mediante o uso de técnicas de assimilação de dados (data assimilation, DA) (Saha et al., 2010; Dee et al. 2014). As CR permitem estimar os estados dos diferentes parâmetros da atmosfera, do mar e da superfície da terra. As CRs têm sido aplicadas principalmente para i) melhorar sistemas de previsão devido a atualização das condições iniciais dos modelos o que permite realizar uma previsão mais acurada, ii) gerar registros consistentes de variáveis climáticas em escala global. As reanálises climáticas globais estão sendo produzidas por diversas instituições entre as quais encontram-se o European Centre for Medium-Range Weather Forecasts (ECMWF), National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP-NCAR), Goddard Space Flight Center (GSFC) da NASA, dentre outras. Segundo Reichle e Liu (2015), vários conjuntos de dados de reanálise tem sido usado para estudar a hidrologia continental, principalmente forçando os modelos hidrológicos com a saída de precipitação das CR, o que nesta pesquisa é denominado de "Retrospectiva Hidrológica". Tradicionalmente a reanálise da precipitação continental era gerada por um modelo atmosférico de circulação geral, embora atualmente reanálises como MERRA-Land, ERA-Interim Land, Climate Forecast System Reanalysis (CFSR), dentre outras, usam assimilação da precipitação a partir de observações de satélite e/ou estações in situ (Reichle e Liu, 2015). De maneira geral, os diferentes produtos de reanálise diferem entre eles principalmente nos locais em que os dados observados são escassos como por exemplo na Ásia, no norte da África ou na América do Sul (Pohlmann e Greatbach, 2006). Portanto, deve-se avaliar a qualidade dos diferentes produtos existentes caso estes requeiram ser usados num local especifico.

Iniciativas em escala global, usualmente chamadas de reanálises hidrológicas, têm por objetivo oferecer uma base de dados de variáveis hidrológicas, principalmente a partir dos anos 1979, tais como PCR-GLOBWB (van Beek et al., 2011; Wanders et al., 2014), MERRA Land (Reichle e Liu, 2015), ERA-Interim/Land (Balsamo et al., 2015), eartH2Observe (Schellekens et al., 2017) ou ERA-40 (Uppala et al., 2005; Dutra et al., 2008) desde 1958. A principal deficiência destas iniciativas, como produtos globais, mediante o uso de modelos hidrologicos globais (GHM) ou modelos globais de superfície (LSM), é a simplificação de alguns processos, por exemplo, estes últimos proveem apenas informação de escoamento, enquanto muitos GHMs utilizam equações simplificadas para a propagação na rede de

drenagem. Desta maneira, não é possível representar os processos hidrodinâmicos que ocorrem em regiões de baixa declividade ou locais com planícies de inundação como nas bacias do rio Amazonas ou do rio Niger (Paiva et al., 2013a; Fleischmann et al., 2018). Numa estimativa geral, informações de reanálises hidrológica globais proveem alto nível de incerteza na estimativa da vazão, como mostrado na secção "Modelagem hidrológica". Isto torna as simulações destas bases de dados pouco relevantes localmente.

1.1.3. Bacia Amazônica

Apesar das escassas informações in-situ em muitas regiões, existe a oportunidade de usar diversas ferramentas como modelos hidrológicos, metodologias como DA e informações de sensoriamento remoto que proveem informações espacialmente distribuídas. Desta maneira, a bacia Amazônica se mostra promissória para sua avaliação devido à escassa rede de monitoramento hidrológico in-situ devido a suas dimensões, pouca acessibilidade e complexidade nos processos hidrológicos e hidráulicos a fazem ideal para sua avaliação. A bacia Amazônica é a maior do mundo, possui uma área de drenagem aproximada de 6 milhões de km^2 e descarga aproximadamente 15% da água continental que chega aos oceanos. Adicionalmente possui um importante bioma que tem grande influência no ciclo do carbono e no clima global (Junk, 1993; Dirzo e Raven, 2003; Phillips et al., 2009); possui também um papel importante na sociedade por exemplo no transporte fluvial, agricultura, pesca e produção de energia. Eventos hidrológicos extremos têm sido reportados nos últimos anos na bacia Amazônica mostrando um incremento na frequência de ocorrência desde os anos 1980s (Espinoza et al., 2009; 2011; Sena et al., 2012; Satyamurty et al., 2013; Gloor et al., 2013; Marengo et al., 2013). Eventos de secas na Amazônia têm sido geralmente associados com valores positivos da temperatura superficial (SST pelo acrônimo em inglês) do oceano Atlântico tropical e com os eventos de El Niño; enquanto eventos de cheias tem sido relacionado principalmente com os eventos de La Niña (Aragão et al., 2007; Zeng et al., 2008; de Linage et al., 2014; Marengo e Espinoza, 2016). Desta maneira, a caracterização da variabilidade multidecadal e dos eventos hidrológicos extremos é de alguma maneira difícil considerando que a maioria dos registros de observações in-situ estão disponíveis apenas para décadas recentes e possuem uma baixa densidade.

1.1.4. Reanálise hidrológica

Resumindo, as reanálises são métodos científicos muito comuns nas áreas de meteorologia e oceanografia (e.g. Gibson et al., 1997; Kistler et al., 2001; Liu e Gupta 2007), entretanto, o estudo retrospectivo em áreas da hidrologia usando essas novas tecnologias ainda é incomum. Desta maneira a presente tese propõe o desenvolvimento de diversas técnicas descritas ao longo do documento, dentre elas a modelagem hidrológica e a DA, mediante o uso de extensas bases de dados como reanálises climáticas, informações in-situ e de sensoriamento remoto que permitam melhorar as estimativas de diferentes variáveis hidrológicas-hidráulicas tanto do passado (~100 anos) como do presente. O produto principal destas técnicas é denominado nesta tese de **reanálise hidrológica**. A finalidade é obter um

melhor entendimento dos processos hidrológicos, permitindo por exemplo, caracterizar ou avaliar: i) o comportamento hidrológico histórico (e.g. multi-decadal, anos recentes, etc.) de certa região como os seus eventos extremos; ii) as tendências devido a variabilidade climática; iii) melhor estimativa de vazões de referência para gerenciamento dos recursos hídricos, como outorgas para o consumo d'água, planejamento do setor hidrelétrico, iv) fornecer melhores estimativas de outras variáveis hidrológicas como nível d'água, áreas de inundação, umidade do solo, dentre outros.

1.2. Objetivos

O objetivo principal da presente pesquisa é propor, desenvolver e implementar metodologias para melhorar as estimativas de simulação de modelos hidrológicos em grande escala como o desenvolvimento de reanálises hidrológicas.

Objetivos específicos:

- Avaliar o desempenho de diversas bases de dados de chuva global na bacia Amazônica através da modelagem hidrológica.
- Avaliar detalhadamente a sensibilidade na resposta do modelo hidrológico de grande escala MGB aos parâmetros do método de DA EnKF.
- Comparar o desempenho do esquema de EnKF e métodos de regionalização usados na área da gestão dos recursos hídricos.
- Avaliar o desempenho das simulações desenvolvidas na geração de um registro de vazões para o século 20 (reanalise hidrológica) para a bacia Amazônica.
- Estimar o desempenho do esquema de DA de dados sintéticos de WSE, FWE e Q provenientes da futura missão SWOT num modelo representativo das incertezas de modelos globais e continentais.
- Estabelecer as bases técnicas e perspectivas futuras para desenvolver a reanálise hidrológica nos anos recentes mediante a DA de múltiplos produtos de sensoriamento remoto.

1.3. Organização da Tese

O manuscrito da tese está dividido em seis capítulos, focados em diferentes objetivos específicos propostos e através do uso das diferentes metodologias e bases de dados correspondentes. A Figura 1.3.1 apresenta um esquema global da estrutura da tese, a interrelação entre os capítulos e as metodologias propostas para cada um. No início de cada capitulo é apresentado um breve prólogo sobre o que é

desenvolvido em cada um deles e o que representa cada artigo cientifico (capítulos 2 até 5) desenvolvido no contexto da tese.

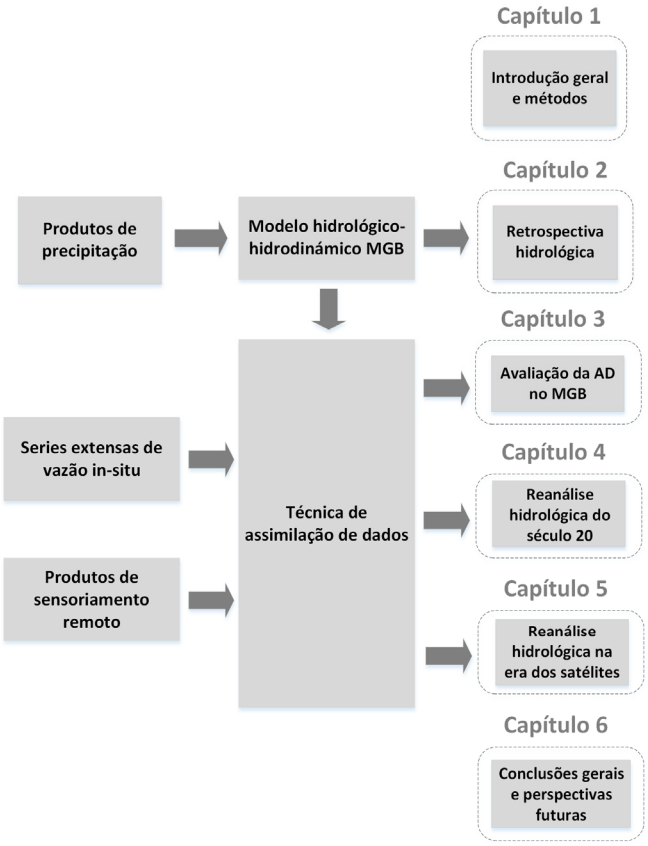

Figura 1.3.1. Esquema geral da estrutura da tese.

No **capítulo 1** se apresenta uma descrição das principais técnicas que serão usadas ao longo deste documento como o modelo hidrológico de grande escala MGB e a metodologia de assimilação de dados ensemble Kalman filter.

O **capítulo 2** consiste na proposta e avaliação de um tipo de reanálise hidrológica baseada na simulação hidrológica forçada por séries extensas de precipitação denominada aqui de retrospectiva hidrológica (HR). Desta maneira, as seguintes questões científicas foram aboradadas para o desenvolvimento deste capitulo: bases de precipitação global e/ou quase-global são adequados para seu uso em modelos hidrológicos em escala regional? Qual é o desempenho destas bases de dados na Bacia Amazônica? As HRs são capazes de capturar eventos hidrológicos extremos como cheias e secas? Aplicou-se o modelo MGB na bacia Amazônica e avaliou-se, com este tipo de retrospectiva hidrológica, a variabilidade hidro climática multi-decadal da vazão, assim como a captura de eventos hidrológicos extremos registrados por observações in-situ e/ou reportados na literatura. A estimativa da tendência temporal das vazões máximas e mínimas anuais, produto das simulações das HRs é comparada com

outros estudos. O objetivo principal é avaliar o desempenho das HRs na Bacia Amazônica, enquanto os objetivos específicos são: i) validar bases de dados globais de precipitação que possuem series relativamente extensas, usadas como input na modelagem hidrológica; ii) avaliar a capacidade das HRs em capturar eventos específicos de cheias e secas registrados por observações in-situ e reportes da literatura; iii) estimar a tendência temporal das vazões máximas e mínimas anuais (relacionadas com a intensidade das cheias e secas) simuladas pelas HRs e comparadas com outros estudos. Este capitulo teve como resultado o a publicação do seguinte artigo cientifico: ***Wongchuig, C.S., de Paiva, R.C.D., Espinoza, J.C., Collischonn, W. 2017. Multi-decadal hydrological retrospective: case study of Amazon floods and droughts. Journal of Hydrololgy, 549:667-684. https://doi.org/10.1016/j.jhydrol.2017.04.019***

A partir dos resultados do capítulo 2, foi observado que as séries mais extensas (>30 anos) apresentaram as performances mais baixas. Não obstante, técnicas como a assimilação de dados podem ajudar na melhoria dessas simulações. Desta maneira, no **capítulo 3** foi avaliado o esquema de assimilação de dados pelo método Ensemble Kalman Filter. Este esquema foi implementado no modelo hidrológico MGB por Paiva, et al. (2013b) a qual é uma versão do modelo com propagação hidrodinâmica usando as equações completas de Saint Venant 1D. Atualmente o modelo MGB utiliza um outro esquema numérico para a propagação da vazão nos rios que é denominado de método inercial. O equacionamento do método inercial implica uma solução explicita e com variáveis de estado apenas diferentes que a versão anterior mas foi necessário um desenvolvimento técnico no código do modelo para a compatibilidade com o novo esquema numérico. Desta maneira, neste capitulo foi adaptada a técnica EnKF na versão mais atual do MGB, por tanto se visa avaliar de maneira detalhada diferentes testes de sensibilidade do MGB tanto pelos diferentes parâmetros do EnKF como pela geração dos membros do ensemble. Para a avaliação desta adaptação, foi estabelecida uma área de estudo numa região com informação suficiente para a validação da simulação de séries de vazão e valores de vazão de referência. Dentre as questões científicas estabelecidas nesse capítulo: quais são os parâmetros do método de DA mais adequados para a área de estudo? Qual é a sensibilidade da resposta do modelo MGB à perturbação da precipitação e de diferentes variáveis de estado do modelo? Métodos de DA são adequados para seu uso na gestão dos recursos hídricos? Os principais objetivos deste capítulo são: i) avaliar a sensibilidade do modelo com diferentes esquemas (número de elementos do ensemble, perturbação das variáveis de estado, assimilação de postos no rio principal ou afluentes, etc.); ii) avaliar a sensibilidade na resposta do MGB à perturbação da precipitação e das variáveis de estado para geração dos membros do ensemble; iii) estimar series de vazões e valores de vazões de referência em comparação com métodos tradicionais (e.g. regionalização). Este capítulo teve como resultado o a submissão do seguinte artigo cientifico: ***Wongchuig, C.S., Fleischman, A., de Paiva, R.C.D., Fadel, W.A. 2019.***

Towards the Discharge Estimation for Water Resources Management Uses with a Semi-Distributed Model and Ensemble Kalman Filter Data Assimilation. Hydrological Science Journal.

As simulações realizadas no capítulo 2 com o desenvolvimento da retrospectiva hidrológica, baseada principalmente na simulação hidrológica, ainda apresentam algumas incertezas na estimativa das variáveis hidrológicas, como por exemplo, pela simplificação dos processos, erros nas forçantes hidro climatológicas, erros nas estimativas dos parâmetros hidráulicos como largura, profundidade e rugosidade, dentre outros. Com a finalidade de diminuir estas incertezas nas estimativas da vazão e de tomar vantagem das bases de dados de chuva mais extensas (~100 anos) avaliadas no capítulo 2, é proposto o **capítulo 4**. Neste capítulo se pretende utilizar tanto extensas séries de forçantes de precipitação como de observações in-situ através da metodologia de DA avaliada no capítulo 3, para a geração de um adequado registro espaço-temporal das variáveis hidrológicas em escala regional o qual é denominada nesta pesquisa de reanálise hidrológica do século 20 (HRXX). As principais questões são: é possível obter resultados adequados na simulação de series de vazão nos últimos 100 anos? Será capaz o HRXX de identificar anomalias na vazão como eventos de secas e cheias nos últimos 100 anos?

O desenvolvimento da HRXX é avaliado na Bacia Amazônica, e devido a sua escala, será implementada a técnica de localização no esquema de DA, cujo objetivo é melhorar as estimativas da DA ao evitar correlações espúrias por causa do número de membros do ensemble e/ou devido à distante localização entre os pontos observados e assimilados. Para mais detalhes desta metodologia ver seção 1.4.2.3. Validou-se a capacidade da HRXX em representar as séries históricas da vazão na bacia Amazônica assim como a captura de eventos hidrológicos extremos como cheias e secas históricas ocorridas no último século. Os principais objetivos deste capítulo são: i) desenvolver uma base de dados para o século 20 mediante o uso de assimilação de extensas observações in-situ (RHXX); ii) implementar e avaliar a técnica de localização no esquema de DA EnKF, assim como a sensibilidade ao raio de influência; iii) validar as series hidrológicas na bacia Amazônica durante o século 20; iv) avaliar a capacidade de representar eventos hidrológicos extremos como cheias e secas desde inicios do século XX em toda a bacia Amazônica. Este capítulo teve como resultado o a publicação do seguinte artigo científico: ***Wongchuig, C.S., de Paiva, R.C.D., Siqueira, V., Collischonn, W. 2019. Hydrological Reanalysis Across the 20th Century: A Case Study of the Amazon Basin. Journal of Hydrology, 570:755-773. https://doi.org/10.1016/j.jhydrol.2019.01.025.***

Outra versão proposta da reanálise hidrológica é aquela dos anos recentes (30 anos), em que inúmeras informações oriundas de sensoriamento remoto são disponíveis, desta maneira o **capítulo 5** apresenta alguns resultados da avaliação da assimilação de dados sintéticos da futura missão espacial SWOT no modelo MGB. A missão SWOT promete uma melhor capacidade tanto em resolução espacial, temporal e acurácia que missões atuais na observação de corpos hídricos continentais, embora esta

missão ainda não se encontra em funcionamento, é possível estimar sua eficiência mendiante simulações sintéticas. Devido às principais aplicações do SWOT estarem focadas na escala continental e/ou global, a delimitação deste capítulo parte das seguintes questões cientificas: como a assimilação de observações provenientes do SWOT melhoram a capacidade de simulação de modelos globais? Quão sensível é o esquema de DA à inceteza dos parâmetros do modelo? Qual é o impacto da DA de uma observação especifica nas outras variaveis do modelo? As assimilações de múltiplas observações proveem uma melhora nas estimativas da simulação?

Técnicas como a DA de múltiplas observações são adicionalmente implementadas como perspectiva futura para o uso de múltiplos produtos de missões atuais, desta maneira se espera uma caracterização mais detalhada de vários processos hidrológicos devido à melhora dos campos espaciais e temporais de diferentes variáveis. O desenvolvimento de um caso estudo da reanálise hidrológica na era dos satélites é avaliada na bacia do rio Purus. Os principais objetivos desse capítulo são: i) implementar e avaliar uma primeira aproximação da DA de dados sintéticos de missões futuras (SWOT) e como estes são relevantes na escala de modelos globais e continentais; ii) avaliar a sensibilidade da DA a diferentes cenários de incertezas do modelo; iii) desenvolver e avaliar as bases técnicas necessárias dentro do modelo MGB para a assimilação de dados de múltiplos produtos de sensoriamento remoto. Este capítulo está baseado em resultados preliminares do artigo: ***Wongchuig, C.S., de Paiva, R.C.D., Biancamaria, S., Collischonn, W. 2019. Assessing data assimilation technique for integrate SWOT data into a large-scale hydrologic-hydrodynamic model: Case study of Purus basin***, e do artigo que está sendo desenvolvido, produto da perspectiva futura desta tese: ***Wongchuig, C.S., de Paiva, R.C.D., Biancamaria, S., Parrens, M., Ahmad, A., Papa, F., Paris, A., Collischonn, W. 2019. Towards multi-mission remote sensing data assimilation into a hydrological-hydrodynamic model for the Amazon basin.***

Finalmente, no **capítulo 6** se apresenta um resumo das conclusões gerais em base aos objetivos gerais da tese e dos capítulos assim como as perspectivas de trabalhos futuros em base aos resultados e/ou métodos propostos ao longo da tese. Assim, se propõe a estimativa das variáveis hidrológicas através da assimilação de observações oriundas de sensoriamento remoto como altimetria de rios (ENVISAT, Jason-1 e Jason-2), extensão de áreas inundadas (GIEMS-D3, SWAF-HR), variação do armazenamento d'água na terra (TWS) via gravimetria do sensor GRACE, dentre outros.

1.4. Proposta metodológica e bases técnicas

Nesta sub-seção é detalhada a metodologia a ser empregada na tese, tendo em vista a abordagem geral proposta. A Fig. 1.4.1 mostra de maneira geral os métodos propostos, as bases técnicas implementadas e os resultados esperados. A metodologia de reanálise hidrológica é dividida

principalmente em duas partes, a reanálise dos últimos 100 anos e da época atual ou era dos satélites. Para isto são empregados métodos de modelagem hidrológica e assimilação de dados; adicionalmente outras técnicas são implementadas para a melhora das simulações que serão detalhadas ao longo desta parte da tese.

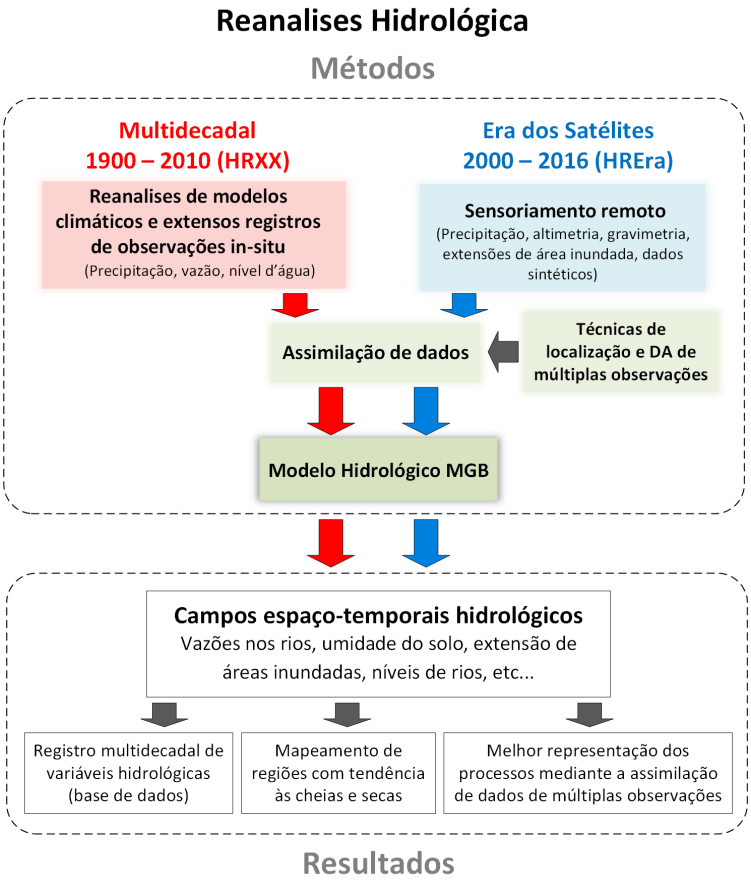

Figura 1.4.1. Representação esquemática da proposta metodológica para reanálises hidrológica e produtos esperados.

Em resumo, para atingir os diferentes objetivos da tese, o seguinte conjunto de métodos e técnicas são utilizados: (i) simulação hidrológica regional com o modelo MGB; (ii) forçantes de precipitação histórica oriunda de reanálises climática, ou produtos gerados a partir da combinação de reanálises, observações in situ e informação de satélites; (iii) métodos de remoção de viés de precipitação; (iv) métodos de assimilação de dados para a fusão das diversas informações, como extensos registros de observações hidrológicas in situ e produtos de sensoriamento remoto; (v) técnicas de localização para aprimorar os resultados do método de DA.

1.4.1. Modelo Hidrológico Hidrodinâmico MGB

Neste subcapitulo é apresentada uma breve descrição do modelo hidrológico-hidrodinâmico MGB (Modelo de Grandes Bacias), assim como os principais processos e as variáveis de estado que serão consideradas com maior ênfase ao longo dos capítulos da tese.

O MGB é um modelo semi-distribuído do tipo chuva-vazão para simulação contínua de grandes bacias hidrográficas (Collischonn, 2001; Collischonn et al., 2007). No modelo, a bacia é dividida em mini bacias as quais são subdivididas em unidades de resposta hidrológica (hydrological response units, HRU), onde processos hidrológicos verticais como interceptação vegetal, evapotranspiração e infiltração são considerados para modelar a geração dos escoamentos superficial, sub-superficial e subterrâneo (Fig. 1.4.2). O balanço vertical da agua é da seguinte forma:

$$W_{i,j}^{t} = W_{i,j}^{t-1} + \left(P_i - ET_{i,j} - Dsup_{i,j} - Dint_{i,j} - Dbas_{i,j}\right)\Delta t \qquad (1.1)$$

onde os índices i e j são relacionados as mini bacias e as HRU respectivamente, enquanto o índice t corresponde ao tempo, W corresponde ao armazenamento do solo (mm), P é a precipitação (mm Δt^{-1}), ET a evapotranspiração (mm Δt^{-1}), Dsup, Dint e Dbas correspondem ao escoamento superficial, sub superficial e subterrâneo (mm Δt^{-1}), e finalmente Δt representa o passo de tempo do balanço, cujo valor corresponde a um dia na presente pesquisa.

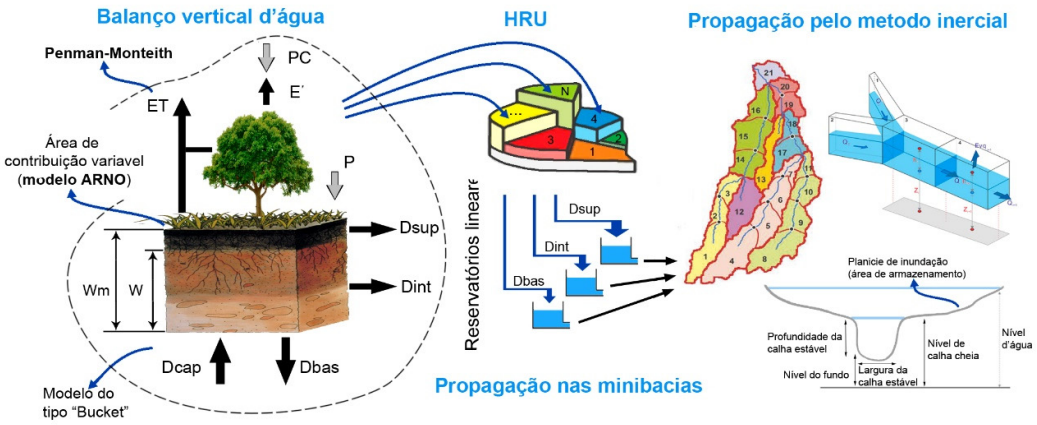

Figura 1.4.2. Esquema do modelo hidrológico MGB.

As forçantes do modelo são séries temporais de dados de precipitação e clima (velocidade do vento, umidade relativa do ar, temperatura do ar, insolação e pressão atmosférica). Em princípio a precipitação é interceptada pela vegetação até uma capacidade máxima de armazenamento que depende do índice de área foliar estimada para cada HRU, logo a evaporação desde o solo, vegetação e água interceptada é calculada pelas equações de Penman-Monteith (Shuttleworth, 1993; Wigmosta et al.,

1994), seguindo uma ordem de prioridade. A energia disponível para a evapotranspiração é primeiramente utilizada para evaporar a água interceptada e o restante da energia, se ainda houver, fica disponível para atender a transpiração da vegetação.

A geração do escoamento superficial (Dsup) é calculado usando o conceito de área de contribuição variável do modelo ARNO (Todini, 1996), enquanto o escoamento sub superficial é estimado a partir das equações de condutividade hidráulica não-saturada Brooks and Corey (Rawls et al., 1993) e o fluxo gerado no reservatório subterrâneo é estimado a partir de uma relação linear simples entre o armazenamento d'água no solo e sua máxima capacidade de armazenamento. O escoamento que é gerado em cada reservatório (Dsup, Dint e Dbas) para cada HRU de cada mini bacia é somado, assim se obtém um valor total de Dsup, Dint e Dbas para cada minibacia. Considerando o tempo de retardo de cada reservatório estes valores são propagados independentemente através de cada minibacia até a rede de drenagem principal usando o conceito dos reservatórios lineares. A vazão propagada na rede principal de cada minibacia é a soma da vazão da minibacia a montante e aquela que é gerada na propria minibacia. Para a propagação da vazão é usada uma simplificação das equações de Saint-Venant 1D em que o termo da inercia advectiva é desprezada (método inercial) (Bates et al., 2010; Fan et al., 2014; Pontes et al., 2017), ficando as Eq. 1.2 e 1.3 correspondentes com a de continuidade e da conservação da quantidade de movimento respectivamente, da seguinte maneira:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \qquad (1.2)$$

$$\frac{\partial Q}{\partial t} + gA\frac{\partial h}{\partial x} = gAS_0 - gAS_f \qquad (1.3)$$

em que Q representa a vazão ($m^3\ s^{-1}$), A é a área da seção transversal (m^2), h é a profundidade do fluxo (m), n o coeficiente de rugosidade Manning. Por outro lado, t e x representam as dimensões temporal e espacial das equações diferenciais.

A Eq. 1.3 Também pode se expressar considerando o componente da fricção em termos da equação de Manning:

$$\frac{\partial Q}{\partial t} + gA\frac{\partial (h+z)}{\partial x} + g\frac{Q|Q|n^2}{AR^{4/3}} = 0 \qquad (1.4)$$

em que z é a cota do fundo (m), R o raio hidráulico (m) e g é a aceleração da gravidade ($m\ s^{-1}$).

Apesar da simplificação deste método de propagação, o método inercial é capaz de representar a propagação da vazão de maneira semelhante ao modelo hidrodinâmico 1D completo, tais como a não-linearidade da celeridade das ondas de cheia, efeitos de remanso e atenuação devido a planícies de inundação, dentre outros (Fan et al., 2014). Para obter mais detalhes dos processos do MGB, revisar Collischonn et al. (2007) e Pontes et al. (2017).

1.4.2. Metodologia de assimilação de dados do filtro de Kalman por ensemble

As observações hidro meteorológicas podem ser usadas nos recursos hídricos de várias maneiras. Para o caso de um modelo chuva-vazão de maneira geral estes usos podem ser, por exemplo: i) como forçantes (e.g. precipitação, temperatura da superfície); ii) com fins de calibração e validação (e.g. vazão, nível d'água) e iii) para combiná-las de maneira ótima com as saídas do modelo para gerar um resultado mais acurado. Esta última aplicação é também conhecida como assimilação de dados (data assimilation, DA) e pode ser usada para melhorar o comportamento de um sistema, seja otimizando os seus parâmetros ou "corrigindo" (atualizando) as suas variáveis de estado, ou uma combinação sinérgica de ambos (Qin et al., 2008); isto é feito mediante a ponderação dos erros do modelo e das observações, considerando que tanto as predições do modelo como as observações são imperfeitas.

Os métodos de DA foram primeiramente usados no modelamento numérico de previsão climática, estes combinam dados atuais e passados num modelo dinâmico explícito, através do uso das equações preditivas do modelo para prover um acoplamento entre suas variáveis (Daley, 1991; Charney et al., 1969). A DA também teve ampla aplicação em áreas da oceanografia (Bennett, 1992) para melhorar as previsões da dinâmica nos oceanos. Embora foi nos anos 1990s que estas técnicas começaram a ser aplicadas com mais frequência na área da hidrologia (Evensen, 1994; McLaughlin, 1995), e teve significativas melhoras na última década (Walker e Houser, 2005). Uma das primeiras aplicações da DA na hidrologia foi através do uso do conteúdo de umidade no solo (Jackson et al., 1981; Bernard et al., 1981; Prevot et al., 1984; Bruckler e Witono, 1989; Ottle e Vidal-Madjar, 1994; Georgakakos e Baumer, 1996).

Uma das questões no sucesso das técnicas de DA são as fontes dos erros envolvidos (Liu e Gupta, 2007). Por um lado, encontram-se os erros das observações (e.g. amostragem, instrumentação), por outro, existem os erros do modelo que estão associados principalmente com a estrutura, os parâmetros e as forçantes do mesmo; o primeiro desses são aqueles derivados da composição conceitual do modelo e/ou as simplificações que foram assumidas na sua formulação. Os erros devido aos parâmetros estão associados com a representação do sistema real, por exemplo no caso de um modelo conceitual, estes parâmetros não são mesuráveis na realidade ou são estimados de maneira indireta e porem possuem uma incerteza associada. Finalmente, os erros nas forçantes estão associados com as incertezas dos dados que foram usados, como por exemplo dados derivados de sensoriamento remoto como precipitação derivada por satélite ou temperatura da superfície proveniente de modelos meteorológicos. Desta maneira, a DA pode ser definida como uma técnica que incorpora observações dentro de um modelo dinâmico para melhorar as suas estimativas. Para atingir estes objetivos os métodos de DA encontram-se principalmente divididos em dois métodos numéricos, métodos sequenciais e métodos variacionais. Alguns destes métodos são descritos neste subcapítulo.

Os métodos variacionais foram desenvolvidos principalmente para previsão meteorológica (Bertino et al., 2003), estes estão baseados na teoria de controle ótima (optimal control theory) (Todorov,

2006), que consiste no uso de variáveis de controle que permitem minimizar uma função objetivo sujeita a certas restrições. Então as variáveis de estado são usadas na otimização como medida de controle, e no processo, será conhecida a trajetória ótima das variáveis de estado a partir da equação que as une (e.g. função do modelo). Desta maneira deverão ser determinadas as variáveis de controle que influenciam nas variáveis de estado no tempo. Finalmente o objetivo da teoria do controle ótimo é resolver problemas de otimização de sistemas que evoluem no tempo e que são sensíveis a forças externas, estabelecendo uma trajetória ótima considerando a relação entre as variáveis de controle e as variáveis de estado. Os métodos de DA variacionais trabalham principalmente em janelas de tempo que contém uma sequência de observações, porem são adequados por exemplo para resolver problemas de suavizado quando existem poucas observações. Este método é usualmente eficaz, embora é difícil de ser implementado em modelos complexos.

Por outra parte, os métodos sequencias usam uma estrutura probabilística necessária para quantificar as incertezas associadas com os resultados; e fornecem estimativas de todo o estado do sistema de maneira sequencial através da propagação da informação apenas para a frente no tempo (Bertino et al., 2003). As técnicas sequencias são amplamente usadas para aplicações em tempo real, realizando uma atualização de maneira sequencial das variáveis de estado do modelo usando as observações uma vez que elas se encontram disponíveis.

Dentre os métodos de DA sequencial mais populares encontram-se aqueles baseados nos filtros de kalman (Kalman filter, KF) (Qin et al., 2008), estes são aproximações com base no método de análise dos mínimos quadrados (Kalman, 1960) utilizados em sistemas dinâmicos. De maneira intuitiva poderia se entender que as variáveis de estado de um sistema dinâmico contem informação quantitativa (e.g. função do modelo) que provê dados sobre o comportamento passado do sistema com a finalidade de predizer o seu comportamento futuro. A versão padrão assume principalmente a linearidade do sistema uma vez que toma em consideração de maneira explicita a natureza dinâmica do modelo e dos erros das observações (Gelb, 1974), os que evoluem no tempo para assim produzir estimativas estatisticamente ótimas para sistemas lineares.

O método de KF desenvolvido por Kalman (1960) há mais de meio século está baseado em alguns supostos básicos, dentre eles, a linearidade do sistema dinâmico estudado. Diversas modificações da versão clássica do KF, assim como outros métodos sequenciais, têm sido desenvolvidas nos últimos anos para serem aplicados em sistemas não-lineares como na área dos recursos hídricos. Alguns destes métodos serão descritos de maneira geral neste subcapítulo, dando ênfase no método de filtro de Kalman por ensemble (*ensemble Kalman filter*, EnKF).

Uma das modificações do KF é o filtro de Kalman estendido (Extended Kalman Filter, EKF) que se apresenta como uma solução do KF para o caso de sistemas dinâmicos com comportamento não linear, utilizando uma abordagem estatística de linearização a cada passo de tempo (Bertino et al., 2003;

Sun et al., 2004). Embora, o método de EKF tem algumas imperfeições na aproximação linear no caso de sistemas com um comportamento altamente não linear. Anos mais tarde Evensen (1994) propôs um novo esquema de assimilação que considerava o ensemble de possíveis trajetórias das variáveis de estado do modelo para uma aplicação na oceanografia, o que derivou no método que é denominado de EnKF. Desta maneira com o método de EnKF podem ser evitadas as imprecisões devido ao processo de linearização que afeta o desempenho da DA devido a fortes características não-lineares de um determinado sistema (Evensen, 1994).

O EnKF está baseado num ensemble de possíveis realizações do modelo, baseado numa aproximação de Monte-Carlo (Evensen, 2003), desta maneira as estimativas dos erros são determinadas em lugar da linearização do modelo usado no EKF. Os métodos de Monte-Carlo tem sido uma ferramenta de pesquisa cujos origens provem do trabalho na bomba atômica no projeto Manhattan durante a segunda guerra mundial, e na atualidade é referida para os métodos que utilizam uma sequência de números aleatórios para realizar simulações estatísticas (Rollett e Manohar, 2005).

Métodos semelhantes ao EnKF como o filtro de partículas (particle filter, PF) também usam uma simulação sequencial de Monte-Carlo além de considerar as incertezas nas forçantes, na estrutura do modelo e nas observações. Segundo Weerts e El Serafy (2006), o método de PF se baseia na atualização da densidade de probabilidade das variáveis de estado do modelo quando novas observações são disponíveis. Para isto se realiza uma amostragem da função de densidade de probabilidade a priori mediante o uso de partículas de diferentes tamanhos que fazem referência ao peso associado a cada uma delas. Considerando este ensemble é realizada uma nova amostragem aleatória das partículas, mas desta vez as novas ponderações atribuídas tornarão mais provável que seja escolhido o valor que está sendo rastreado. Finalmente, é feita uma ligeira modificação nos estados (posição) das novas partículas, para fazer a previsão no instante seguinte. Assim, fica claro que uma diferença entre os métodos de KF e PF é que enquanto nos primeiros são atualizadas as variáveis de estados do modelo, no segundo são os pesos das partículas que são atualizados. Embora o PF seja mais flexível em algumas suposições como o fato de não ser necessário que as distribuições de probabilidade sejam gaussianas como ocorre com o EnKF (Hreinsson, 2008). Weerts e El Serafy (2006) demonstraram num estudo comparativo que o PF é mais sensível com a escolha dos erros do modelo e das observações, enquanto o EnKF é mais adequado para a previsão de vazões.

Outras variações do EnKF como o filtro retrospectivo de Kalman por ensemble (*retrospective ensemble Kalman filter*, REnKF) ou o filtro por ensemble da raiz quadrada (*ensemble square root filter*, EnSRF) tem sido empregados também na área dos recursos hídricos. O primeiro deles foi concebido a partir do critério da associação temporal entre as variáveis de estado, um exemplo é quando a vazão está influenciada pela umidade antecedente do solo e por forçantes meteorológicas recentes, desta maneira Pauwels e De Lannoy (2006) propuseram um esquema de assimilação que permitisse atualizar as condições de umidade presente e passadas. Por outro lado, o método de EnSRF evita a perturbação das

observações no esquema de DA (Whitaker e Hamill, 2002; Clark et al., 2008), desta maneira se evitam os ruídos adicionados na análise, e por outro lado isto diminui a sua demanda computacional.

Um amplo trabalho de revisão das aplicações do método de EnKF foi presentado por Evensen (2003), propondo a formulação desta técnica bem como sua implementação numérica. Dentre as primeiras implementações do EnKF em modelos hidrológicos encontram-se os trabalhos de Clark et al. (2006), Vrugt e Robinson (2007) e Neal et al. (2007). O esquema de KF principalmente é dividido em duas etapas denominadas de forecasting e updating (Evensen, 2004) expressas no vetor das variáveis de estado como $\mathbf{x}^f$ e $\mathbf{x}^a$ respectivamente.

Na figura 1.4.3 é apresentada um esquema geral do processo da técnica de EnKF na etapa de forecasting e updating. Uma breve descrição dos processos é apresenta neste capitulo, enquanto maiores detalhes do equacionamento são apresentados no Anexo A.

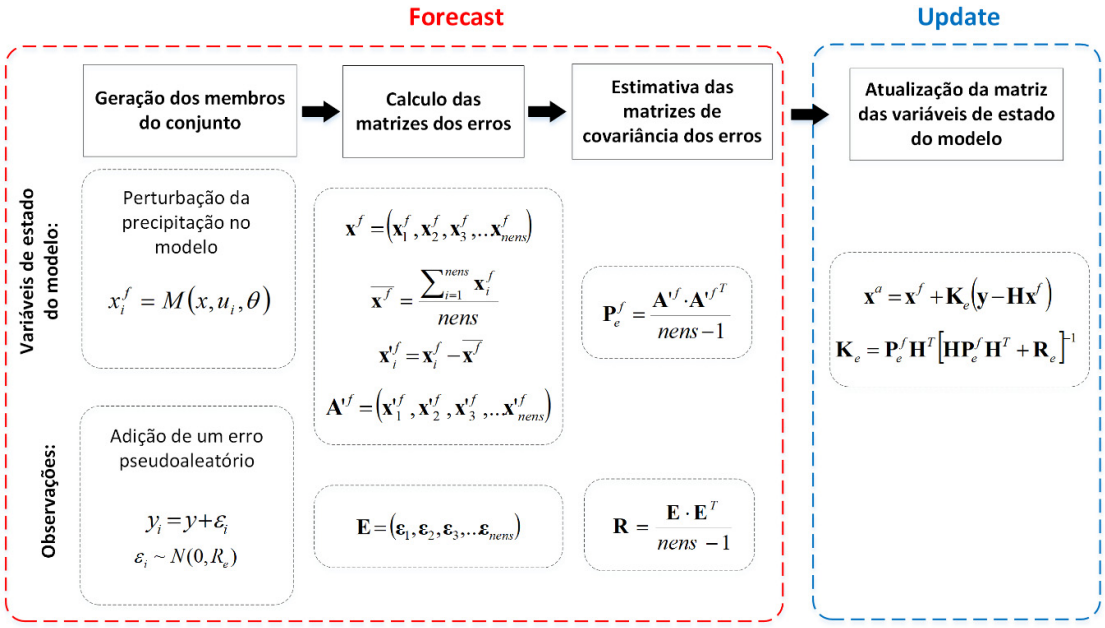

Figura 1.4.3. Esquema geral da técnica de assimilação de dados por EnKF, dividido nos processos forecast (azul) e update (vermelho).

Primeiramente, o esquema do modelo e das observações em função das variáveis de estado pode ser considerado como segue:

$$\mathbf{x}_{k+1} = M(\mathrm{x}_k, u_k, \theta) + \boldsymbol{q}_k \qquad \boldsymbol{y}_k = \boldsymbol{H}(x_k) + \varepsilon_k \qquad (1.5)$$

em que k representa o índice do tempo, **x** representa o vetor das variáveis de estado do modelo com dimensão *nsta*; u e θ representam as forçantes e os parâmetros do modelo respectivamente, *M* é a função do Modelo que relaciona as variáveis de estado ($\mathrm{x_k}$) com as do passo de tempo

posterior (x_{k+1}), e q_k indica as incertezas do modelo. Por outro lado, **y** representa o vetor que contém as observações com dimensão *nobs*, onde nobs representa o número de observações num instante k; **H** representa a função da observação que relaciona as variáveis de estado do modelo com o vetor de observações, **H** é representada por uma matriz de dimensões *nobs*×*nsta*, onde nsta representa o número de variáveis de estado do modelo. Finalmente **ε** é o vetor que contém as incertezas da observação.

Por outro lado, a matriz de covariância de erros das observações e das variáveis de estado do modelo são expressas da seguinte maneira:

$$\mathbf{R}_k = \overline{\varepsilon_k \varepsilon_k{}^T} = \overline{(\mathrm{y}_k - \mathrm{y}_k^t)(\mathrm{y}_k - \mathrm{y}_k^t)^T} \qquad \boldsymbol{P}_k^f = \overline{q_k q_k{}^T} = \overline{(x_k - x_k^t)(x_k - x_k^t)^T} \qquad (1.6)$$

em que y_k^t representa a observação verdadeira (medição sem erros, superíndice t) no instante k.

Uma das suposições do KF é que as matrizes das covariâncias dos erros das observações e das variáveis de estado são independentes:

$$\overline{(\mathrm{y}_k - \mathrm{y}_k^t)(\mathrm{x}_k - \mathrm{x}_k^t)^T} = 0 \qquad (1.7)$$

Logo a equação que contém o vetor das variáveis de estado do modelo atualizadas no tempo k dado que se tem uma observação no mesmo instante de tempo pode ser descrita como uma combinação linear do vetor das observações (ou também chamado vetor de inovação) com o vetor de estado no prognostico (ou chamado de "forecast"). Esta análise é chamada de BLUE (Best Linear Unbiased Estimator), assim obtém-se:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right) \qquad (1.8)$$

$$\mathbf{K} = \mathbf{P}_k^f \mathbf{H}^T \left[\mathbf{H}\mathbf{P}_k^f \mathbf{H}^T + \mathbf{R}_k\right]^{-1} \qquad (1.9)$$

em que $\mathbf{x}^a$ é a variável de estado estimada, **K** é o chamado ganho de Kalman, $\mathbf{P}_k^f$ e $\mathbf{R}_k$ as matrizes de covariância dos erros do modelo (q) e das observações (ε), com dimensões nsta×nsta e nobs×nobs respectivamente.

A estimativa do ganho de Kalman é um passo importante na técnica do KF, e matematicamente se pode indicar que se o ganho for muito pequeno, o processo de assimilação não terá muito efeito, pelo contrário se o ganho for muito grande, o modelo perderá a evolução temporal original (Qin et al., 2008).

O método de EnKF proposto por Evensen (2004) utiliza a mesma equação de atualização (Eq. 1.8) da versão padrão do KF, embora no EnKF o ganho de Kalman é calculado a partir da matriz de covariância dos erros fornecida pelo ensemble das variáveis de estado do modelo. Na formulação estocástica do EnKF, como primeiro passo no método de Monte Carlo, deve se obter a melhor condição inicial para o vetor das variáveis de estado de modelo, utilizando a informação disponível (o que pode

ser chamado de forecast ou background), então a incerteza associada a melhor condição inicial está representada pela variância inicial, sendo que esta dispersão é geralmente estimada pela perturbação das forçantes, parâmetros ou variáveis de estado do modelo (Clark et al. 2008; Biancamaria et al. 2011; Liu et al., 2012; Paiva et al., 2013b). Neste ensemble de variáveis de estado gerado, a média é a melhor condição inicial e, porém, as incertezas do modelo estão associadas a variância deste ensemble (Evensen, 1994). Desta maneira, no método do EnKF, as matrizes de covariância são geradas por um ensemble de trajetórias ou membros a partir das perturbações no modelo, baseados na variância inicial que pode ser por exemplo a incerteza na precipitação quando é perturbada. Segundo Burgers et al. (1998) a escolha do tamanho do número de membros do ensemble é importante para garantir que os erros simulados sejam dominados pelos ruídos estadísticos.

A metodologia de DA pelo esquema de EnKF requere certo custo computacional para sistemas mais complexos como o modelo hidrológico-hidrodinâmico distribuído MGB. Embora a versão do modelo com propagação inercial (Pontes et al., 2017) foi utilizada para esta parte da pesquisa devido a que o esquema numérico explicito permitiu a simplificação na adaptação e a possibilidade da paralelização do código. Este desenvolvimento foi baseado nos algoritmos desenvolvidos por Evensen (2003) e previamente implementados no modelo MGB por Paiva et al. (2013b).

Neste capítulo foi apresentada uma introdução geral da tese a qual aborda uma revisão dos tópicos principais e as lacunas no conhecimento que justificam e motivam o desenvolvimento da presente pesquisa. Desta forma, foram delimitados o objetivo geral e os objetivos específicos os quais estão bastante conectados com os artigos científicos apresentados em cada capítulo. Adicionalmente, são apresentados a descrição do modelo hidrológico MGB, o conceito de DA, alguns métodos existentes, seu equacionamento matemático e a descrição de algumas técnicas que visam melhorar ainda mais as simulações. Algumas técnicas como a geração dos elementos do ensemble e outras técnicas como a localização e o uso de múltiplas observações são descritas em detalhe no Anexo A. Os próximos capítulos desenvolvem a implementação dos métodos apresentados neste capitulo em aplicações práticas principalmente focadas na bacia Amazônica.

Referências

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connel, P. E. and Rasmussen, J. 1986a. An introduction to the European Hydrological System--Systeme Hydrologique Europeen, "SHE," 1, History and philosophy of a physically based distributed modelling system, J. Hydrol., 87, 45-59.

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. and Rasmussen, J. 1986b. An introduction to the European Hydrological System--Syst•me Hydrologique Europ•en "SHE," 2, Structure of a physically based distributed modelling system, J. Hydrol., 87, 61-77.

Aires, F., Miolane, L., Prigent, C., Pham, B., Fluet-Chouinard, E., Lehner, B., Papa, F. 2017. A Global dynamic and long-term inundation extent dataset at high spatial resolution derived through downscaling of satellite observations, J. Hydrometeor., in press.

Alsdorf, D.E., Lettenmeier, D., Vorosmarty C. 2003. The Need for Global, Satellite-based Observations of Terrestrial Surface Waters. Eos, Vol. 84(29):269-280.

Alsdorf, D.E., Rodríguez, E. and Lettenmaier, D.P., 2007. Measuring surface water from space. Rev. Geophys., 45, RG2002, doi:10.1029/2006RG000197.

Alsdorf, D., Beighley, E., Laraque, A., Lee, H., Tshimanga, R., O'Loughlin, F. G., et al. 2016. Opportunities for hydrologic research in the Congo Basin. Reviews of Geophysics, 54:378–409. https://doi.org/10.1002/2016RG000517.

Andreadis, K.M., Clark, E.A., Lettenmaier, D.P., Alsdorf D.E. 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. Geophys Res Lett 34: L10403

Andreadis, K.M., Das, N., Stampoulis, D., Ines, A., Fisher, J.B. Granger, S., Kawara, J., Han, E., Behrangi, A. 2017. The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation. PLoS ONE 12(5): e0176506.https://doi.org/10.1371/journal.pone.0176506

Aragão, L.E.O.C., Malhi, Y., Roman-Cuesta, R.M., Saatchi, S., Anderson, L.O., and Shimabukuro, Y.E., 2007. Spatial patterns and fire response of recent Amazonian droughts. Geophysical Research Letters, 34(7), n/a–n/a. L07701.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., et al. (2015). Accelerating advances in continental domain hydrologic. Water Resources Research, 51:10,078–10,091.

Arnesen, A. S., Silva, T. S., Hess, L. L., Novo, E. M., Rudorff, C. M., Chapman, B. D., and McDonald, K. C., 2013. Monitoring flood extent in the lower Amazon River floodplain using ALOS/PALSAR scansar images. Remote Sensing of Environment, 130, 51 – 61.

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R. 1998. Large area hydrologic modeling and assessment part I: Model development. J Am Water Resour Assoc 34(1): 73– 89.

AVISO User Handbook., Merged TOPEX/Poseidon Products (GDRMs)., 3a Ed., AVI-NT-02-101-CN, Toulouse, 1996.

Baguis, P. and Roulin, E. 2017. Soil Moisture Data Assimilation in a Hydrological Model: A Case Study in Belgium Using Large-Scale Satellite Data. Remote Sens. 2017, 10, 820; doi:10.3390/rs9080820.

Baker, V.R., Webb, R.H., House, P.K., 2002. The scientific and societal value of paleoflood hydrology. In: House, P.K., Webb, R.H., Baker, V.R., Levish, D.R. (Eds.), Ancient Floods, Modern Hazards: Principles and Applications of Paleoflood Hydrology. Water Science and Application Series, 5, 127–146.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A.K. 2009. A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the integrated forecast system, J. Hydrometeorol., 10, 623–643.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F., 2015. Era Interim/ land, a global land surface reanalysis data set. Hydrology and Earth System Sciences, 19(1), 389–407.

Balsamo, G.; Agustì-Parareda, A.; Albergel, C.; Arduini, G.; Beljaars, A.; Bidlot, J.; Bousserez, N.; Boussetta, S.; Brown, A.; Buizza, R.; et al. 2018. Satellite and In Situ Observations for Advancing Global Earth Surface Modelling: A Review. Remote Sens., 12, 2038.

Barber, V.A., Finney, B.P. 2000. Late quaternary paleoclimatic reconstructions for interior Alaska based on paleolake-level data and hydrologic models. Journal of Paleolimnology. 24(1): 29-41.

Becker, M., Papa, F. Frappart, F., Alsdorf, D., Calmant, S., Da Silva, J. S., Prigent, C., Seyler, F. 2018. Satellite-based estimates of surface water dynamics in the Congo River Basin. International Journal of Applied Earth Observation and Geoinformation, 66: 196-209.

Benito, G., Thorndycraft, V.R. 2005. Palaeoflood hydrology and its role in applied hydrological sciences. Journal of Hydrology, 313: 3–15.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments. SMHI Reports RHO, No. 7, Norrkoping, 1976.

Bergström, S.: The HBV model – its structure and applications. SMHI Reports RH, No. 4, Norrkoping, 1992.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. .L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J. 2011. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677–699, doi: 10.5194/gmd-4-677-2011.

Beven, K. J., Kirkby, M. J., 1979. A physically based variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24: 43-69.

Biancamaria, S.; Durand, M.; Andreadis, K. M.; Bates, P. D.; Boone, A.; Mognard, N. M.; Rodriguez, E.; Alsdorf, D. E.; Lettenmaier, D. P.; Clark, E. A. 2011. Assimilation of virtual wide swath altimetry to improve Arctic river modelling. Remote Sensing of Environment, 115, 373–381.

Bierkens MFP, Bell VA, Burek P, Chaney N, Condon LE, David CH, de Roo A, Do¨ll P, Drost N, Famiglietti JS, Flo¨rke M, Gochis DJ, Houser P, Hut R, Keune J, Kollet S, Maxwell RM, Reager JT, Samaniego L, Sudicky E, Sutanudjaja EH, van de Giesen N, Winsemius H, Wood EF. 2015. Hyper-resolution global hydrological modelling: what is next? Everywhere and locally relevant. Hydrol Process 29(2):310–320. doi:10.1002/hyp.10391

Brisco, B., Short, N., van der Sanden, J., Landry, R., Raymond, D. 2009. A semi-automated tool for surface water mapping with RADARSAT-1. Can. J. Remote Sens., 35, 336–344.

Brivio, P.A., Colombo, R., Maggi, R., and Tomasoni, R. 2002. Integration of remote sensing data and GIS for accurate mapping of flooded areas. International Journal of Remote Sensing, 23(3), 429–441.

Engman, E.T. 1995. Recent advances in remote sensing in hydrology. Reviews of Geophysics, Supplement, 967-975.

Clark, Martyn P.; Slater, Andrew G. 2006. Probabilistic Quantitative Precipitation Estimation in Complex Terrain. Journal of Hydrometeorology, Vol. 7, pp. 3 – 22.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., Uddstrom, M. J. 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31, 1309–1324.

Collischonn, W., Allasia, D., da Silva, B.C., and Tucci, C.E.M., 2007. The mgb-iph model for large-scale rainfall—runoff modelling. Hydrological Sciences Journal, 52(5), 878–895.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin,X., Gleason, B.E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, ., Ross, T. F., Trigo, R.M., Wang, X.L., Woodruff, S.D., and Worley, S. J., 2011. The twentieth century reanalysis project. Quarterly Journal of the Royal Meteorological Society, 137(654), 1–28.

Corato, G., Matgen, P., Fenicia, F., Schlaffer, S., Chini, M., 2014. Assimilating satellite-derived soil moisture products into a distributed hydrological model. 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, 2014, 3315-3318.

Crow, W.T. and Wood, E.F. 2003. The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97. Adv. Water Resour., 26(2), 137-149.

Crow, W.T. and Ryu, D. 2009. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. Hydrol. Earth Syst. Sci., 13, 1–16.

Decharme, B., Alkama, R., Douville, H., Becker, M., and Cazenave, A. 2010. Global evaluation of the ISBA-TRIP continental hydrological system, Part II: Uncertainties in river routing simulation related to flow velocity and groundwater storage, J. Hydrometeorol., 11, 601–617.

Decharme, B., Martin, E., and Faroux, S. 2013. Reconciling soil thermal and hydrological lower boundary conditions in land surface models, J. Geophys. Res.-Atmos., 118, 7819–7834.

Dee, D.P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A.J., and Thépaut, J.-N., 2014. Toward a consistent reanalysis of the climate system. Bulletin of the American Meteorological Society, 95(8, 1235–1248.

Dirzo, R., Raven, P.H., 2003. Global State of Biodiversity and Loss. Annual Review of Environment and Resources, 28, 137-167.

Döll, P., Berkhoff, K., Bormann, H., Fohrer, N., Gerten, D., Hagemann, S. and Krol. M. 2008. Advances and visions in large-scale hydrological modelling: findings from the 11th Workshop on Large-Scale Hydrological Modelling. Adv. Geosci., 18: 51–61.

Dutra, E., P. Viterbo, and P.M.A. Miranda. 2008. ERA-40 reanalysis hydrological applications in the characterization of regional drought, Geophys. Res. Lett., 35, L19402, doi: 10.1029/2008GL035381.

d'Orgeval, T., Polcher, J., and de Rosnay, P. 2008. Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes, Hydrol. Earth Syst. Sci., 12, 1387–1401, doi:10.5194/hess-12-1387-2008.

Entekhabi, D. 1994. A simple model of hydrologic cycle and climate: 1. Model construct and sensitivity to landsurface boundary, Advances in Water Resources, 17, 79-91.

Espinoza J.C., Ronchail, J., Guyot, J.L., Cocheneau G., Filizola, N., Lavado, W., de Oliveira, E., Pombosa, R., Vauchel, P. 2009. Spatio – Temporal rainfall variability in the Amazon Basin Countries (Brazil, Peru, Bolivia, Colombia and Ecuador. International Journal of Climatology, 29, 1574-1594.

Espinoza, J.C., Ronchail, J., Guyot, J. L., Junquas, C., Vauchel, P., Lavado, W., Drapeau, G., Pombosa, R. 2011. Climate variability and extreme drought in the upper Solimões River (western Amazon Basin): understanding the exceptional 2010 drought. Geophys. Res. Lett. 38(13): L13406.

Fekete, B. M., Looser, U., Pietroniro, A., and Robarts, R. D. 2012. Rationale for Monitoring Discharge on the Ground, J. Hydrometeorol., 13, 1977–1986, https://doi.org/10.1175/jhm-d-11-0126.1.

Fleischmann, A.S., Siqueira, V.A., Paris, A., Collischonn, W., Paiva, R.C.D., Pontes, P.R.M., Cretaux, J., Berge-Nguyen, M., Biancamaria, S., Gosset, M., Calmant, S., Tanimoune, B. 2018. Modelling hydrologic and hydrodynamic processes in basins with large semi-arid wetlands. Journal of Hydrology, 561: 943-959.

Fleischmann, A., Paiva, R.C.D., Collischonn, W. 2019. Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison. Journal of Hydrology X, 3(2019): 100027.

Fluet-Chouinard, E., Lehner, B., Rebelo, L.-M., Papa, F., Hamilton, S.K., 2015. Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. Remote Sensing of Environment, 158, 348–361.

Gao, H., Tang, Q., Shi, X., Zhu, C., Bohn, T.J., Su, F., Sheffield, J., Pan, M., Lettenmaier, D.P. and Wood, E.F., 2010. Water budget record from Variable Infiltration Capacity (VIC) model. Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records.

García-Pintado, J., Neal, J.C., Mason, D.C., Dance, S., Bates, P.D. 2013. Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling. J. Hydrol., 495:252–266. doi:10.1016/j.jhydrol.2013.03.050

Gayathri, K.D., Ganasri, B.P., Dwarakish, G.S. 2015. A Review on Hydrological Models. International Conference on Water Resources, Coastal and Ocean Engineering (ICWRCOE 2015). Aquatic Procedia, 4: 1001-1007.

Georgakakos, K.P. and Baumer, O.W. 1996. Measurement and utilization of on-site soil moisture data. Journal of Hydrology, 184(1-2): 131-152.

Getirana, A. C. V., Bonnet, M.-P., Rotunno Filho, O. C., Collischonn, W., Guyot, J.-L., Seyler, F. and Mansur, W. J. 2010. Hydrological modelling and water balance of the Negro River basin: evaluation based on in situ and spatial altimetry data. Hydrol. Process., 24: 3219–3236.

Gibson, J. and for Medium Range Weather Forecasts, E. C., 1997. ECMWF Re-analysis Project Report Series, ERA description. Number v. 1. European Centre for Medium-Range Weather Forecasts.

Gloor, M., Brienen, R.J.W., Galbraith, D., Feldpausch, T.R., Schöngart, J., Guyot, J.-L., Espinoza, J.C., Lloyd, J., and Phillips, O.L., 2013. Intensification of the amazon hydrological cycle over the last two decades. Geophysical Research Letters, 40(9), 1729–1733.

Gupta, H.V., Sorooshian, S., Yapo, P.O. 1998. Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751-763.

GRACE-FO. 2018. Gravity Recovery and Climate Experiment Follow-On: Tracking Earth's Mass in Motion.

Hess, L. L., Melack, J. M., Novo, E. M., Barbosa, C. C., and Gastil, M., 2003. Dual-season mapping of wetland inundation and vegetation for the central amazon basin. Remote Sensing of Environment, 87(4), 404 – 428.

Hess, L.L., J.M. Melack, A.G. Affonso, C. Barbosa, M. Gastil-Buhl, and E.M.L.M. Novo, 2015, Wetlands of the Lowland Amazon Basin: Extent, Vegetative Cover, and Dual-season Inundated Area as Mapped with JERS-1 Synthetic Aperture Radar, Wetlands 35: 745-756. doi:10.1007/s13157-015-0666-y

Hidalgo, H. G., Das, T., Dettinger, M. D., Cayan, D. R., Pierce, D. W., Barnett, T. P., Bala, G., Mirin, A., Wood, A. W., Bonfils, C., Santer, B. D., and Nozawa, T. 2009. Detection and attribution of streamflow timing changes to climate change in the Western United States, J. Climate, 22, 3838–3855, https://doi.org/10.1175/2009JCLI2470.1.

Horritt, M.S., Mason, D.C., and Luckman, A.J., 2001. Flood boundary delineation from Synthetic Aperture Radar imagery using a statistical active contour model. International Journal of Remote Sensing, 22(13), 2489–2507.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., Cudennec, C. 2013. A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrological Sciences Journal, 58(6), 1198-1255.

Jarrett, R.D., 1991, Paleohydrology and its value in estimating floods and droughts, in Paulson, R.W., Chase, E.B., Roberts, R.S., and Moody, D.W., Compilers, National Water Summary 1988-89--Hydrologic Events and Floods and Droughts: U.S. Geological Survey Water-Supply Paper 2375, 105-116.

Junk, W. J., 1993. Wetlands of Tropical South America, Kluwer, Dordrecht, Netherlands, 15-2, 679-739.

Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. 82: 35. doi:10.1115/1.3662552

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., Thielen, J. 2016. Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. Environmental Modelling & Software, 75: 68-76.

Khaki, M., Forootan, E., Kuhn, M., Awange, J., Longuevergne, L., Wada, Y. 2018. Efficient basin scale filtering of GRACE satellite products. Remote Sensing of Environment, (204):76–93.

Khaki, M., Hoteit, I., Kuhn, M., Forootan, E., Awange, J. 2019. Assessing data assimilation frameworks for using multi-mission satellite products in a hydrological context. Science of The Total Environment, (647): 1031-1043.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M., 2001. The ncep–ncar 50–year reanalysis, Monthly means cd–rom and documentation. Bulletin of the American Meteorological Society, 82(2), 247–267.

Koblinsky, C. J., Clarke, R. T., Brenner, A. C., and Frey, H. 1993. Measurement of river level variations with satellite altimetry, Wat. Resour. Res., 29(6), 1839-1848

Kurtz, W., Lapin, A., Schilling, O.S., Tang, Q., Schiller, E., Braun, T., Jumkeler, D., Vereecken, H., Sudicky, E., Kropf, P., Franssen, H-J.H., Brunner, P. 2017. Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. Environmental Modelling & Software, 93: 418-435.

Lai, X., Liang, Q., Yesou, H., Daillet, S., 2014. Variational assimilation of remotely sensed flood extents using a 2-D flood model. Hydrol. Earth Syst. Sci. 18, 4325–4339. http://dx.doi.org/10.5194/hess-18-4325-2014.

Lee, H., Jung, H. C., Yuan, T., Beighley, R. E., and Duan, J., 2014. Controls of Terrestrial Water Storage Changes Over the Central Congo Basin Determined by Integrating PALSAR ScanSAR, Envisat Altimetry, and GRACE Data, pages 115–129. John Wiley & Sons, Inc.

Li, J. and Wang, S., 2015. An automated method for mapping inland surface waterbodies with Radarsat-2 imagery. Int. J. Remote Sens., 36, 1367–1384.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99, 14415–14415.

Liang, X., Wood, E. F., and Lettenmaier, D. P. 1996. Surface soil moisture parameterization of the VIC-2L model: evaluation and modification, Global Planet. Change, 13, 195–206.

de Linage, C., Famiglietti, J.S., and Randerson, J.T., 2014. Statistical prediction of terrestrial water storage changes in the amazon basin using tropical pacific and north atlantic sea surface temperature anomalies. Hydrology and Earth System Sciences, 18(6, 2089–2102.

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research, 43, W07401.

Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P. 2012a Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16, 3863–3887, doi:10.5194/hess-16-3863-2012

Liu S, Shao Y, Yang C, Lin Z, Li M. 2012b. Improved regional hydrologic modelling by assimilation of streamflow data intoaregional hydrologic model. Environ Model Softw. 31:141-149. https://doi.org/10.1016/j.envsoft.2011.12.005

Liu, Y., Wang, W., Hu, Y., Cui, Wei. 2016. Improving the Distributed Hydrological Model Performance in Upper Huai River Basin: Using Streamflow Observations to Update the Basin States via the Ensemble Kalman Filter. Advances in Meteorology, Article ID 4921616.2016: 14p. http://dx.doi.org/10.1155/2016/4921616

López López, P., Wanders, N., Schellekens, J., Renzullo, L.J., Sutanudjaja, E.H. Bierkens, F.P. 2015. Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations. Hydrol. Earth Syst. Sci. Discuss., 12: 10559–10601.

López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., and Bierkens, M. F. P. 2017. Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. Hydrol. Earth Syst. Sci., 21, 3125-3144, https://doi.org/10.5194/hess-21-3125-2017, 201

Loukas, A. and Vasiliades, L. 2014. Streamflow simulation methods for ungauged and poorly gauged watersheds. Nat. Hazards Earth Syst. Sci., 14:1641–1661.

Luthcke, S. B., Sabaka, T. J. Loomis, B. D., Arendt, A.A., McCarthy, J.J., and Camp, J. 2013. Antarctica, Greenland and Gulf of Alaska land ice evolution from an iterated GRACE global mascon solution. J. Glac., 59(216), 613-631.

Luo, P., Takara, K., He, B., Duan, W., Apip, Nover, D., Tsugihiro, W., Nakagami, K., Takamiya, I. 2014. Assessment of Paleo-hydrology and Paleo-inundation Conditions: The Process. Procedia Environmental Sciences. 20: 747-752.

Madsen, H. 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. J. Hydrol., 235, 276–288.

Marengo, J.A., Borma, L.S., Rodriguez, D.A., Pinho, P., Soares, W.R., and Alves, L.M. 2013. Recent extremes of drought and flooding in Amazonia, vulnerabilities and human adaptation. American Journal of Climate Change, 2, 87–96.

Marengo, J.A. and Espinoza, J. C., 2016. Extreme seasonal droughts and floods in Amazonia, causes, trends and impacts. International Journal of Climatology, 36(3), 1033–1050.

Massari, C., Brocca, L., Tarpanelli, A., Moramarco, T. 2015. Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe? Remote Sens. 2015, 7, 11403-11433; doi:10.3390/rs70911403.

Melesse, A.M., Weng, Q., Thenkabail, P.S., and Senay, G.B. 2007. Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. Sensors, 7: 3209-3241.

Michailovsky, C.I., Milzow, C., Bauer-Gottwein, P. 2013. Assimilation of radar altimetry to a routing model of the Brahmaputra River. Water Resour. Res., 49, 4807-4816.

Moss M.E., Hindall, S.M., Lins, H.F., Eberle, M. 1988. Water Resources in the Twenty-first Century: A Study of the Implications of Climate Uncertainty. Department of the Interior, U.S. Geological Survey, p. 17.

Munier, S., Polebistki, A., Brown, C., Belaud, G., Lettenmaier, D.P. 2015. SWOT data assimilation for operational reservoir management on the upper Niger River Basin. Water Resour Res. doi:10.1002/2014WR016157

Neal, Jeffrey; Atkinson, Peter; Hutton, Craig. 2007. Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. Journal of Hydrology, Vol. 336, pp. 401 – 415.

Nunnery, A.; Baker, P. A.; Coe, M. T.; Fritz, S. C.; Rigsby, C. A. 2011. A quantitative history of precipitation and hydrologic variability for the last 45 ka: Lake Titicaca, Salar de Coipasa and Salar de Uyuni, Peru and Bolivia. In AGUGeneral AssemblyConference Abstracts #PP41B-1769

Nunnery, A. 2012. A Reconstruction of Precipitation and Hydrologic Variability on the Peruvian and Bolivian Altiplano During the Late. Thesis for the degree of Doctor of Philosophy in Earth and Ocean Sciences in the Graduate School of Duke University.

O'Loughlin, F. E., Neal, J., Yamazaki, D., and Bates, P. D. 2016. ICESat-derived inland water surface spot heights. Water Resour. Res., 52, 3276–3284, doi:10.1002/2015WR018237.

Oubanas, H. Gejadze, I., Malaterre, P-O., Durand, M., Wei, R., Frasson, R.M.P., Domeneghetti, A. 2018. Discharge estimation in ungauged basins through variational data assimilation: The potential of the SWOT mission. Water Resources Research, https://doi.org/10.1002/2017WR021735

Paiva, R.C.D., Buarque, D.C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., and Bulhões Mendes, C.A., 2013a. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resources Research, 49(3), 1226–1243.

Paiva, R.C.D. Collischonn, W., Bonnet, M.P., de Goncalves, L.G.G., Calmant Stéphane, Getirana, A., da Silva J. S., 2013b. Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. Hydrol. Earth Syst. Sci., 17, 2929-2946.

Parrens, M., Al-Bitar, A., Frappart, F., Paiva, R.C.D., Wongchuig, C.S., Papa, F., Yamazaki, D., Kerr, Y. 2019. High resolution mapping of inundation area in the Amazon basin from a combination of L-band passive microwave, optical and radar datasets. Int. J. Appl. Earth. Obs. Geoinformation, 81:58–71. https://doi.org/10.1016/j.jag.2019.04.011

Pedinotti, V., Boone, A., Ricci, S., Biancamaria, S., Mognard, N.M. 2014. Assimilation of satellite data to optimize large-scale hydrological model parameters: a case study for the SWOT mission. Hydrol Earth Syst Sci 18(11):4485–4507. doi:10.5194/hess-18-4485-2014

Pham, B., Prigent, C., Aires, F., 2016. Surface water monitoring in the Mekong delta over a year, with Sentinel-1 SAR observations. Remote Sensing of Environment, 34, 1–13.

Phillips, O.L., Aragão, L.E.O.C., Lewis, S.L., Fisher, J.B., Lloyd, J., L´opez-Gonz´alez, G., Malhi, Y., Monteagudo, A., Peacock, J., Quesada, C.A., van der Heijden, G., Almeida, S., Amaral, I., Arroyo, L., Aymard, G., Baker, T.R., B´anki, O., Blanc, L., Bonal, D., Brando, P., Chave, J., de Oliveira, A.C.A., Cardozo, N.D., Czimczik, C.I., Feldpausch, T.R., Freitas, M.A., Gloor, E., Higuchi, N., Jiménez, E., Lloyd, G., Meir, P., Mendoza, C., Morel, A., Neill, D. A., Nepstad, D., Pati˜no, S., Pe˜nuela, M.C., Prieto, A., Ram´ırez, F., Schwarz, M., Silva, J., Silveira, M., Thomas, A. S., Steege, H. t., Stropp, J., V´asquez, R., Zelazowski, P., Dávila, E.A., Andelman, S., Andrade, A., Chao, K.-J., Erwin, T., Di Fiore, A., C., E. H., Keeling, H., Killeen, T.J., Laurance, W.F., Cruz, A.P., Pitman, N.C.A., Vargas, P.N., Ram´ırez-Angulo, H., Rudas, A., Salamão, R., Silva, N., Terborgh, J., and Torres-Lezama, A., 2009. Drought sensitivity of the amazon rainforest. Science, 323(5919), 1344–1347.

Pietroniro, A., Prowse, T.D. 2002. Applications of remote sensing in hydrology. Hydrological Processes, 16:1537–1541.

Pohlmann, H. and Greatbatch, R.J., 2006. Discontinuities in the late 1960's in different atmospheric data products. Geophysical Research Letters, 33(22, n/a–n/a. L22803.

Pontes, P. R. M., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., Buarque, D. C., Siqueira, V. A., Jardim, P. F., Sorribas, M. V., Collischonn, W. 2017. MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS. Environmental Modelling & Software, 94, 1-20.

Prigent, C., Papa, F., Aires, F., Rossow, W.B., Matthews, E., 2007. Global inundation dynamics inferred from multiple satellite observations, 1993–2000. Journal of Geophysical Research: Atmospheres (1984–2012), 112 (D12).

Qin, Changbo; Jia, Yangwen; Su, Z. (Bob); ZHOU, Zuhao; QIU, Yaqin; SUHUI, Shen. 2008. Integrating Remote Sensing Information Into a Distributed Hydrologycal Model for Improving Water Budget Predictions in Large-scale Basins through Data Assimilation. Sensors, Vol. 8, pp. 4441 – 4465.

Prigent, C., Papa, F., Aires, F., Jimenez, C., Rossow, W.B., Matthews, E., 2012. Changes in land surface water dynamics since the 1990s and relation to population pressure. Geophys. Res. Lett., 39 (8).

Rasmussen, J., Madsen, H., Jensen, K. H., & Refsgaard, J. C. 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydrology and Earth System Sciences, 19(7), 2999-3013.

Razavi, T., Coulibaly, P. 2016. Improving streamflow estimation in ungauged basins using a multi-modelling approach. Hydrological Sciences Journal, 61(15), 2668-2679.

Reichle, R.H., McLaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. Monthly Weather Rev., 130 (1), 103–114.

Reichle, R. and Liu, Q. 2015. Precipitation and global land surface hydrology in the MERRALand and MERRA-2 reanalysis datasets. In EGU General Assembly Conference Abstracts, volume 17 of EGU General Assembly Conference Abstracts, page 1838.

Revilla-Romero, B., Wanders, N., Burek, P., Salamon, P., de Roo, A., 2016. Integrating remotely sensed surface water extent into continental scale hydrology. J. Hydrol. 543, 659–670.

Rosenqvist, A., Birkett, C.M., 2002. Evaluation of JERS-1 SAR mosaics for hydrological applications in the Congo river basin. International Journal of Remote Sensing, 23, 1283-1302.

Rosenqvist, A., Shimada, M., Watanabe, M., 2004. ALOS PALSAR: Technical outline and mission concepts. 4th International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land ApplicationsInnsbruck, Austria, November 16-19.

Saatchi, S.S., Nelson, B., Podest, E., Holt, J., 2000. Mapping land cover types in the Amazon Basin using 1 km JERS-1 mosaic. International Journal of Remote Sensing, 21, 1201-1234.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins,W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R., Rutledge, G., and Goldberg, M., 2010. The ncep climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8), 1015–1057.

Sandholt, I., Andersen, J., Dybkjær, G., Lo, M., Rasmussen, K., Refsgaard, J.C., Høgh-Jensen, K., 1999. Use of remote sensing data in distributed hydrological models: applications in the Senegal River basin. Geografisk Tidsskrift, Danish Journal of Geography, 99, 47-57.

Satyamurty, P., da Costa, C.P.W., Manzi, A.O., Candido, L.A., 2013. A quick look at the 2012 record flood in the amazon basin. Geophys. Res. Lett. 40 (7), 1396–1401.

Save, H., Bettadpur, S., & Tapley, B. 2016. High-resolution CSR GRACE RL05 mascons. Journal of Geophysical Research: Solid Earth, 121, 7547– 7569.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G.P. 2017. A global water resources ensemble of

hydrological models: the eartH2Observe Tier-1 dataset, Earth Syst. Sci. Data, 9, 389-413, https://doi.org/10.5194/essd-9-389-2017

Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J., Rango, A. 2002. Remote sensing in hydrology. Advances in Water Resources, 25:1367–138.

Schwatke, C., D. Dettmering, E. Börgens, and W. Bosch. 2015. Potential of SARAL/AltiKa for Inland Water Applications, Mar. Geod., 38(sup1), 626–643.

Sena, J.A., Beser de Deus, L.A., Freitas, M.A.V., and Costa, L., 2012. Extreme events of droughts and floods in amazonia, 2005 and 2009. Water Resources Management, 26(6), 1665–1676.

Siqueira, V.A., Paiva, R.C.D., Fleischmann, A.S., Fan, F.M., Ruhoff, A.L., Pontes, P.R.M., Paris, A., Calmant, S., and Collischonn, W. 2018. Toward continental hydrologic–hydrodynamic modeling in South America, Hydrol. Earth Syst. Sci., 22, 4815-4842.

Sivapalan, M. 2003. Prediction in ungauged basins: a grand challenge fortheoretical hydrology. Hydrological Processes, 17: 3163-3170.

Smith, L.C. 1997. Satellite Remote Sensing of River Inundation Area, Stage, and Discharge: A Review. Hydrological Processes, 11:1427-1439.

Sood, A. and Smakhtin, V., 2015. Global hydrological models: a review. Hydrol. Sci. J. 60(4), 549e565. http://dx.doi.org/10.1080/02626667.2014.950580.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J. 2010. Streamflow trends in Europe: evidence from a dataset of nearnatural catchments, Hydrol. Earth Syst. Sci., 14, 2367–2382, https://doi.org/10.5194/hess-14-2367-2010.

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348-2366.

Swierczynski, T., Ionita, M., Pino, D., 2017, Using archives of past floods to estimate future flood hazards, Eos, 98, doi, 10.1029/2017EO066221.

Tang, Q., Gao, H., Lu, H., Lettenmaier, D.P. 2009. Remote sensing: hydrology. Progress in Physical Geography 33(4): 490–509.

Tapley B.D., Bettadpur S., Watkins M., Reigber C. 2004. The gravity recovery and climate experiment: mission overview and early results. Geophys. Res. Lett., 31, p. L09607, 10.1029/2004GL019920

Thielen, J., Pappenberger, F., Salamon, P., Bogner, K., Burek, P., de Roo, A., 2010. The State of the Art of Flood Forecasting-Hydrological Ensemble Prediction Systems, p. 145.

Townsend, P.A., 2002. Relationship between forest structure and the detection of flood inundation in forested wetlands using C-band SAR. International Journal of Remote Sensing, 23(3), 443–460.

Uppala, S.M., KÅllberg, P.W., Simmons, A.J., Andrae, U. , Bechtold, V.D., Fiorino, M. , Gibson, J.K., Haseler, J. , Hernandez, A. , Kelly, G.A., Li, X. , Onogi, K. , Saarinen, S. , Sokka, N. , Allan, R.P., Andersson, E. , Arpe, K. , Balmaseda, M.A., Beljaars, A.C., Berg, L.V., Bidlot, J. , Bormann, N., Caires, S. , Chevallier, F. , Dethof, A. , Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L. , Janssen, P.A., Jenne, R., Mcnally, A.P., Mahfouf, J., Morcrette, J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. 2005. The ERA-40 re-analysis. Q.J.R. Meteorol. Soc., 131: 2961-3012. doi:10.1256/qj.04.176

Urban, T.J., Schutz, B.E. and Neuenschwander, A.L., 2008. A survey of ICESat coastal altimetry applications: Continental coast, open ocean island, and inland river, Terr. Atmos. Oceanic Sci., 19(1-2), 1–19.

Van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P. 2011. Global monthly water stress: I. Water balance and water availability, Water Resour. Res., 47, W07517, doi: 10.1029/2010WR009791.

van Dijk, A.I.J.M., Renzullo, L.J. 2011. Water resource monitoring systems and the role of satellite observations. Hydrology and Earth System Sciences 15(1), 39–55.

van Dijk, A. I. J. M., Renzullo, L. J., Wada, Y., and Tregoning, P. 2014. A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, Hydrol. Earth Syst. Sci., 18, 2955–2973, doi: 10.5194/hess-18-2955-2014.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., Verstraten, J. M. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41(1).

Vrugt, J.A.; Robinson, B.A. 2007. Treatment of uncertainty using Ensemble methods: Comparison of secuential data assimilation and Bayesian model averaging. Water Resources Research, Vol. 43, W01411, doi: 10.1029/2005WR004838.

U.S. Army Corps of Engineers. 2003. Application of Paleohydrology to Corps Flood Frequency Analysis. Report Document RD-47.

Wagner, W., Verhoest, N.E.C., Ludwing, R., Tedesco, M. 2009. Remote sensing in hydrological sciences. Hydrol. Earth Syst. Sci., 13:813–817.

Wanders, N., Bierkens, M.F.P., de Jong, S.M., de Roo, A., Karssenberg, D. 2014. The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. Water Resour. Res. 50, 6874–6891. http://dx.doi.org/10.1002/2013WR014639.

Westerhoff, R.S., Kleuskens, M.P.H., Winsemius, H.C., Huizinga, H.J., Brakenridge, G.R., Bishop, C., 2013. Automated global water mapping based on wide-swath orbital synthetic-aperture radar. Hydrol. Earth Syst. Sci., 17, 651–663.

White, L., Brisco, B., Pregitzer, M., Tedford, B., Boychuk, L., 2014. RADARSAT-2 Beam Mode Selection for Surface Water and Flooded Vegetation Mapping. Can. J. Remote Sens., 40, 135–151.

Wiese, D. N., Yuan, D.-N., Boening, C., Landerer, F. W., Watkins, M. M. 2018. JPL GRACE Mascon Ocean, Ice, and Hydrology Equivalent Water Height Release 06 Coastal Resolution Improvement (CRI) Filtered Version 1.0. Ver. 1.0. PO.DAAC, CA, USA.

Wood EF, Roundy JK, Troy TJ, van Beek LPH, Bierkens MFP, Blyth E, de Roo A, Doell P, Ek M, Famiglietti J, Gochis D, van de Giesen N, Houser P, Jaffe P, Kollet S, Lehner B, Lettenmaier DP, Peters-Lidard C, Sivapalan M, Sheffield J, Wade A, Whitehead P. 2011. Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water. Water Resources Research 47: W05301. DOI: 10.1029/2010WR010090.

Xie, X., Zhang, D. 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Advances in Water Resources, 33: 678–690.

Yamazaki, D., De Almeida, G. A. M., and Bates, P. D. 2013. Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector based river network map, Water Resour. Res., 49, 7221–7235, https://doi.org/10.1002/wrcr.20552.

Yoon, Y., Durand, C. J. Merry, Clark, E.A., Andreadis, K., Alsdorf, D. 2012. Estimating river bathymetry from data assimilation of synthetic SWOT measurements. Journal of Hydrology, v. 464–465, p. 363–375.

Yu, Z. 2015. Modeling and Prediction. College of Hydrology and Water Resources, Hohai University, Nanjing, China; and University of Nevada Las Vegas, Las Vegas, NV, USA

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann,

H., Klotz, S., Munch, J. C., Papen, H., Priesack, E., Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H. 2011. A network of terrestrial environmental observatories in Germany, Vadose Zone J., 10, 955–973, https://doi.org/10.2136/vzj2010.0139.

Zeng, N., Yoon, J.-H., Marengo, J. A., Subramaniam, A., Nobre, C. A., Mariotti, A., and Neelin, J. D., 2008. Causes and impacts of the 2005 amazon drought. Environmental Research Letters, 3(1), 014002.

Zhang, D., Madsen, H., Ridler, M.E., Kidmose, J., Jensen, K.H., and Refsgaard, J.C. 2016. Multivariate hydrological data assimilation of soil moisture and groundwater head, Hydrol. Earth Syst. Sci., 20, 4341-4357, https://doi.org/10.5194/hess-20-4341-2016.

Capítulo 2

Retrospectiva Hidrológica Multidecadal: Estudo de caso das cheias e secas na Amazônia

Prólogo

Diferentes produtos de precipitação global, sobretudo aqueles oriundos de reanálises climáticas, diferem entre eles principalmente nos locais em que os dados observados são escassos. Alguns destes produtos já foram testados na Ásia, no norte da África ou na América do Sul (e.g. Betts e Ball, 2005; Pohlmann e Greatbach, 2006; Prakash et al., 2014; Seyyedi et al., 2015). Desta maneira, existe uma necessidade pela avaliação de diferentes produtos existentes para regiões como a bacia Amazônica. Muitos produtos como reanálises climáticas, ainda não proporcionam uma adequada representação da chuva nas últimas décadas, se prevê que as melhoras nos Sistemas Integrados de Previsão (IFS) e, desta maneira, as reanálises baseadas não somente em modelos climáticos, mas complementarmente com modelos de superfície terrestre (e.g. ERA-Interim Land) consigam diminuir os erros. Neste estudo se desenvolve a metodologia denominada de retrospectiva hidrológica que basicamente é a avaliação de produtos de precipitação gerados a partir de reanálises climáticas, sensoriamento remoto e precipitação in-situ ou a combinação destes, como forçantes do modelo hidrológico. A eficiência dos resultados do modelo hidrológico é avaliada através da validação com observações in-situ e diferentes métricas estatísticas. Adicionalmente foram avaliados os eventos hidrológicos extremos passados assim como a sua tendência temporal na Bacia do rio Amazonas.

Multi-decadal hydrological retrospective: case study of Amazon floods and droughts

Wongchuig, C.S., de Paiva, R.C.D., Espinoza, J.C., Collischonn, W.

Paper published in Journal of Hydrology https://doi.org/10.1016/j.jhydr ol.2017.04.019

Abstract

Recently developed methodologies such as climate reanalysis make it possible to create a historical record of climate systems. This paper proposes a methodology called Hydrological Retrospective (HR), which essentially simulates large rainfall datasets, using this as input into hydrological models to develop a record of past hydrology, making it possible to analyze past floods and droughts. We developed a methodology for the Amazon basin, where studies have shown an increase in the intensity and frequency of hydrological extreme events in recent decades. We used eight large precipitation datasets (more than 30 years) as input for a large scale hydrological and hydrodynamic model (MGB-IPH). HR products were then validated against several in situ discharge gauges controlling the main Amazon sub-basins, focusing on maximum and minimum events. For the most accurate HR, based on performance metrics, we performed a forecast skill of HR to detect floods and droughts, comparing the results with in-situ observations. A statistical temporal series trend was performed for intensity of seasonal floods and droughts in the entire Amazon basin. Results indicate that HR could represent most past extreme events well, compared with in-situ observed data, and was consistent with many events reported in literature. Because of their flow duration, some minor regional events were not reported in literature but were captured by HR. To represent past regional hydrology and seasonal hydrological extreme events, we believe it is feasible to use some large precipitation datasets such as i) climate reanalysis, which is mainly based on a land surface component, and ii) datasets based on merged products. A significant upward trend in intensity was seen in maximum annual discharge (related to floods) in western and northwestern regions and for minimum annual discharge (related to droughts) in south and central-south regions of the Amazon basin. Because of the global coverage of rainfall datasets, this methodology can be transferred to other regions for better estimation of future hydrological behavior and its impact on society.

Keywords: Hydrological retrospective, Climatic reanalysis, Hydrological modeling, Droughts, Floods, Amazon basin.

2.1. Introduction

The development of detailed and consistent hydrological records of past extreme events, such as floods and droughts, is key for better assessment of future hydrological risks and to understand the effects of climate variability and change. Reports of increased floods and droughts in the Amazon in recent years, for example, indicate a need for tools to better access past hydrological databases to understand how frequent or extreme the recent events have been (e.g. Marengo and Espinoza, 2016; Swierczynski et al. 2017). Regional-scale hydrological models and climate reanalysis have improved in recent years, creating opportunities for development of a multi-decadal Hydrological Retrospective to characterize past floods and droughts.

The Amazon basin is the world largest basin; it drains about 6 million km^2 and discharges ~15% of the freshwater that reaches the world's oceans. It is also an important large biome with a major influence on the carbon cycle and global climate (Junk, 1993; Dirzo and Raven, 2003; Phillips et al., 2009), and its water resources meet many human needs, such as fluvial transportation, agriculture, fisheries and energy production.

Hydrological extreme events have been reported in the Amazon basin with increased frequency since the 1980s (Espinoza et al., 2009b; 2011; Sena et al., 2012; Satyamurty et al., 2013; Gloor et al., 2013; Marengo et al., 2013). Amazon droughts generally have been associated with positive sea surface temperature (SST) anomalies in the tropical Atlantic and with El Niño events, while floods have been linked mainly to La Niña events (Aragão et al., 2007; Zeng et al., 2008; de Linage et al., 2014; Marengo and Espinoza, 2016).

Recent floods, such as those in in 1989, 1999 and 2009, mainly in the central and eastern regions (Marengo et al., 2012), and in 1986, 1993, 1999, 2009, 2012 and 2014 in the western regions (Espinoza et al., 2013; 2014), caused high impacts on ecosystems, the local population and socioeconomic activities such as fishing, farming, transportation and human health (Schöngart and Junk, 2007; Marengo et al., 2013; Satyamurty et al., 2013; Espinoza et al., 2014; Castello et al., 2015; Alho et al., 2015; Ovando et al., 2016). Droughts in 1964, 1980, 2005 and 2010 (Zeng et al., 2008; Marengo et al., 2008; 2011; Espinoza et al., 2011; Saatchi et al., 2013) caused persistent disturbances in forest and difficulties in river transportation, farming and fishing, which disrupted the food supply to residents, interrupted hydroelectric power generation and affected human health, especially because of smoke from forest fires (Saleska et al., 2007; Aragão et al., 2008; Phillips et al., 2009; Asner and Alencar, 2010; Lewis et al., 2011; Frolking et al., 2011; Xu et al., 2011; Fernandes et al., 2011). Recent studies have detected an increase in maximum discharge over the last century in the main river (e.g. Callède et al., 2004; Gloor et al., 2013), while a decrease in dry-season rainfall and discharge has been detected since the 1970s (Espinoza et al., 2009a; 2011; Marengo et al., 2011). Projections to the end of 21st century show a

probable increase in maximum discharge in the west (main river) and a persistent decrease in the east (e.g. Guimberteau et al., 2013; Boisier et al., 2015; Sorribas et al., 2016).

It is difficult to characterize past hydrological extremes because in situ observation records generally are available only for recent decades and only for a few river locations. Improvements have been made, however, in hydrological models capable of simulating surface hydrological and hydraulic processes and variability and changes related to land use and climate (Nijssen et al., 2001; Coe et al., 2008; Beighley et al., 2009; Neal et al., 2012; Yamazaki et al., 2012; Alfieri et al., 2013; Paiva et al, 2013; Sorribas et al., 2016; Karlsson et al., 2016). Hydrologic remote sensing observations (Lettenmaier et al., 2015) and climate reanalysis have also evolved in recent years. Merging hydrological models with these new datasets could provide information with better spatio-temporal coverage to explore past hydrological behavior (Nogués-Paegle et al., 2002).

These datasets could be coupled with distributed hydrological models to create a hydrological retrospective representing historical hydrological fields (e.g. discharge, soil moisture, water levels) to study past floods and droughts. Several studies using climate reanalysis of hydrology have focused on validation and analysis of rainfall fields, but few were aimed at validation of hydrological modelling (e.g. Reichle and Liu, 2015; Xu et al., 2016), and this method still is not commonly used to explore past extreme events.

In this study, the fusion of hydrological modeling and climate reanalysis or large rainfall datasets is called Hydrological Retrospective (HR), a methodology developed to better understand past hydrological processes, in this case given focus on extreme events. Studying these events is critical for implementing sustainable water management policies and mitigating negative impacts. Understanding past extreme events is also important for estimating likely future patterns, because of close correlations (Knox and Kundzewicz, 1997; Scaife et al. 2008; Bovolo et al., 2012; Smith and Lawson, 2012; Phipps et al., 2013; Lee et al., 2016).

The aims of this study are: i) to develop a Hydrological Retrospective methodology for representing past floods and droughts ii) to validate several precipitation datasets with extensive data series, from both climate reanalysis and merged products, using them as input for hydrological and hydrodynamic modeling in the Amazon basin and comparing the results with in situ discharge observation, using by statistical metrics, iii) to determine whether these datasets and modeling can capture seasonal flood and drought events registered by in-situ discharge and reported in literature, and iv) to estimate the trend of maximum, mean and minimum annual values for historical simulated discharge time series in Amazon basin, using a statistical trend test.

2.2. Data and Methods

2.2.1. Hydrological Retrospective (HR)

To understand the concept of HR, it is essential to describe similar efforts in other scientific areas such as oceanography or meteorology, where reanalysis has a longer history of use (Gibson et al., 1997; Kistler et al., 2001). According to Saha et al. (2010) and Dee et al. (2014), climate reanalysis is the constant reprocessing of climatic observations using weather forecasting systems and data assimilation methods. The goal is to describe the climate using as much information as possible. Information is estimated for several atmosphere, ocean and land surface parameters, then a state of the art is created for each parameter, usually on a global scale. Climate reanalysis is now being done by research institutions such as ECMWF (European Centre for Medium-Range Weather Forecasts), NCEP-NCAR (National Center for Environmental Prediction – National Center for Atmospheric Research) and the NASA (National Aeronautics and Space Administration), Goddard Space Flight Center (GSFC).

There are three types of climate reanalysis: atmospheric, ocean-atmospheric, and reanalysis based on land surface models (Reichle and Liu, 2015). Early reanalysis of precipitation variables was generated using atmospheric models; currently some reanalysis (e.g. MERRA-Land, ERA-Interim Land) uses a land surface component through data assimilation of satellite observation and/or in-situ gauges, which could result in better prediction systems, especially in areas such as continental hydrology (Koster et al., 2011; Dee et al., 2014; Reichle and Liu, 2015). Reanalysis datasets differ mainly in places with low coverage of in situ observations (e.g. Asia, North Africa and South America) (Adler et al. 2001; Pohlmann and Greatbatch, 2006; Garreaud et al. 2009; Bovolo et al. 2012), although products of reanalysis have improved over South America since the 1980s because of availability of satellite data (Kalnay et al. 1996). It is therefore important to evaluate the different datasets available for each study area.

Some examples of evaluation of the performance of precipitation reanalysis at a global scale and for South America and Amazon basin are presented below:

Several studies have been conducted around the world. Betts et al. (2009), Shah and Mishra (2014), Prakash et al. (2015) and Gao et al. (2016) evaluated precipitation reanalysis datasets such as MERRA, ERA-Interim, CFSR, ERA-40, NCEP1, NCEP2, GLDAS and ERA-20CM, finding that performance differed depending on factors such as location, catchment area size, and topographic and climate characteristics. Many precipitation datasets for the South American Monsoon System (SAMS) have also been evaluated, including, in the Bolivian Andes and the Amazon basin, CFSR, MERRA, ERA-40 (Betts et al., 2005; Carvalho et al., 2012; Bovolo et al., 2012; Blacutt et al., 2015). These studies generally concluded that these datasets are skillful for representing the average annual cycle of precipitation, although they underestimate the observed record in rainy seasons and overestimate it in the dry season. Because previous studies show that many rainfall reanalyses based only on atmospheric

models do not provide a robust representation of precipitation, including the land surface component in the reanalysis could be expected to improve performance. Details of various climate reanalyses that are currently available can be found at <reanalyses.org>. Other available sources of rainfall data include ground-based gauge observations, ground-based radar remote sensing and satellite remote sensing (Ebert et al., 2007; Michaelides et al., 2009, Kidd and Huffman 2011, Lettenmaier et al., 2015). In general, the drawback of these datasets is that they typically provide shorter records (1980 or 2000 to present) compared to climate reanalysis (1900s to present). The largest temporal dataset produces better simulation of the evolution of large-scale climate systems (Beck et al. 2016).

HR can be defined as a retrospective study of the evolution of hydrology in a specific region, through an optimal combination of hydrological modeling and other available information (in-situ, remote sensing and/or climate reanalysis). This methodology allows better estimation of spatial and temporal fields of past hydrological variables, such as discharge, water level, soil moisture, evapotranspiration, etc.

HR makes it possible to understand typical hydrological behavior in a particular region, which can then be used to estimate past extreme events. Few studies have used large precipitation datasets as input into hydrological models (e.g. Bastola and Misra, 2014; Reichle and Liu, 2015; Seyyedi et al., 2015; Xu et al., 2016; Werner and Cannon, 2016; Zhu et al., 2016) or to develop global products such as Earth2Observe or PCR-GLOBWB (Wanders et al., 2014). And almost none have been evaluated in detail in South America or used to explore hydrologic extreme events in that region.

2.2.2. The hydrological model MGB-IPH

Several past studies developed hydrologic models for the Amazon basin (Coe et al. 2008; Paiva et al. 2013; Beighley et al. 2009; Yamazaki et al. 2012). This version of the MGB-IPH model for the Amazon basin (Paiva et al. 2013), was selected because of (i) its ability to represent physical processes, such as water balance components, river hydrodynamics and large-scale inundation in the Amazon, and (ii) its performance, as demonstrated by previous validation against observations.

The MGB-IPH model (Collischonn et al. 2007) is a large-scale, distributed, process-based hydrological model. It uses catchment-based discretization and an approach based on hydrological response units. Simulation methods include energy budget and evapotranspiration using the Penman Monteith approach, soil water budget using a bucket model, and surface runoff based on the variable contributing area concept. River flow routing is performed using a one-dimensional river hydraulic module with a floodplain storage model (Paiva et al. 2011). The MGB-IPH has been used in several Amazonian and South American basins (e.g. Collischonn et al., 2008, Pontes et al. 2015, Fan et al. 2015, Fan et al. 2016).

In this MGB-IPH implementation, the Amazon basin was discretized into 5765 catchments, ranging from 100 to 5,000 km^2. The model was forced using TRMM 3B42 precipitation estimates (Huffman et al. 2007), with spatial resolution of 0.25° x 0.25° and a daily time step for a period spanning 12 years (1998 - 2009), and using monthly meteorological data obtained from the CRU CL 2.0 dataset (New et al. 2002). The model parameters related to soil water budget were calibrated using discharge data from in situ stream gauges (47 stations). The model was validated against discharge and water level data from stream gauge stations (111 and 69 sites, respectively), water level satellite altimetry, Terrestrial Water Storage from the GRACE mission, and remote sensing inundation extent. Comparisons between simulations and observations showed relatively high Nash and Suttcliffe index (NSE) values (NSE > 0.6 in ~70% of discharge gauges).

2.2.3. Precipitation datasets

Eight freely available data sets of daily precipitation were used to represent the evolution of temporal large-scale hydrology: five climate reanalysis (ERA-Interim, ERA-20CM, CFSR, JAR-55 and NOAAv2c) and three merged datasets consisting of climate reanalysis, satellite and in-situ gauges (CHIRPS v2.0, ERA-Interim Land and MSWEP); the latter had the shortest temporal data. All have global coverage and more than 30 years of data (Table 1). A brief comparison with in-situ observed precipitation from HYBAM (HOP) was also performed. A short description for each dataset is presented below.

CFSR v1: The Climate Forecast System Reanalysis (CFSR) was developed by the National Center for Environmental Prediction (NCEP). The CSRF was designed as a global coupled atmosphere-ocean-land surface-sea ice system. Dataset available at <http://rda.ucar.edu/pub/cfsr.html>.

CHIRPS v2.0: The Climate Hazards Group InfraRed Precipitation with Station data was developed at United States Geological Survey (USGS) and the Earth Resources Observation and Science (EROS). It is based on merged precipitation models by interpolation of rainfall in situ data. It was mainly designed to provide data for drought alert in regions with scarce in-situ information (Funk et al., 2015). Dataset available at <ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/>.

ERA- Interim: This dataset was developed provisionally to replace the ERA-40 reanalysis. It was based on the European Centre for Medium-Range Weather Forecasts (ECMWF) surface climate model. Dataset available at <http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>.

ERA-Interim Land: This climate reanalysis was developed using precipitation land surface forcing. It is based on the ERA-Interim climate reanalysis, but uses precipitation data forcing from the Global Precipitation Climatology Project (GPCP v2.1), which consists of monthly precipitation means derived from satellite and gauge measurements. Unlike ERA-Interim, this dataset has temporal coverage only from 1979 to 2010 (Balsamo et al., 2015). Dataset available at <http://apps.ecmwf.int/datasets/data/interim-land/type=fc/>.

ERA-20CM: The 20th Century Reanalysis was developed by the European Centre for Medium-Range Weather based on the Integrated Forecast System Cy38r1, sea-surface model HadSST2 and radiative forcing CMIP5. The spatial resolution of the final product is available from 0.125° and it presents 10 members of the ensemble data. Dataset available at <http://apps.ecmwf.int/datasets/data/era20cm-edmo/levtype%3Dsfc/>.

JRA-55: This climate reanalysis was developed by the Japanese Meteorological Agency (JRA), as an extension and improvement of the earlier JRA-25 reanalysis. The dataset initially covered the years 1958 to 2012, but it recently was updated to provide near-real-time data. The main purpose was to develop a dataset for research on multi-decadal variability and climate change (Kobayashi et al., 2015). Dataset available at <http://rda.ucar.edu/datasets/ds628.0/>.

MSWEP: The Multi-Source Weighted-Ensemble Precipitation was developed to be used in hydrologic models (Beck et al., 2016). It was made using the long-term mean from CHPclim, and temporal variability was determined by weighted averaging of precipitation anomalies from seven datasets, which include two global databases of in situ precipitation measurements (CPC Unified and GPCC), three remote sensing datasets (CMORPH, GSMaP-MVK and TMPA 3B42RT) and two climate reanalyses (ERA-Interim and JAR-55). Dataset available at <http://dap.nci.org.au/thredds/remoteCatalogService?catalog=http://dapds00.nci.org.au/thredds/catalog/ub8/global/climate/MSWEP_v1.0/daily/catalog.xml>.

NOAA v2c: Full name: "20th Century Reanalysis version 2". The NOAA climate reanalysis was developed by the Cooperative Institute for Research in Environmental Sciences (CIRES) jointly with the National Oceanic and Atmospheric Administration (NOAA), and was the first climate reanalysis to cover dates earlier than 1900 (Compo, et al., 2011). Dataset available at <http://rda.ucar.edu/datasets/ds131.2/>.

HOP: This is a gridded daily precipitation dataset based on 752 rain gauges over the Amazon basin, covering a period from 1980 to 2009. This grid has 1°x1° spatial resolution and was interpolated through the ordinary kriging geostatistical method. For more details, see Guimberteau et al., 2012 and Espinoza et al., 2016. Dataset available at <www.ore-hybam.org>.

Table 1. Summary of the products used on Hydrological Retrospective

Short name	Institution	Reference	Data sources(s)	Temporal coverage	Temporal resolution *	Spatial resolution used
CFSR	NCEP	Saha et al. (2010)	Reanalysis	1979/01 - 2010/12	SD and M	0.5° x 0.5°
CHIRPS v2.0	CHG	Funk et al. (2015)	Gauge, satellite	1981/01 - present	D and M	0.25° x 0.25°
ERA-Interim	ECMWF	Dee et al. (2011)	Reanalysis	1979/01 - present	SD, D and M	0.75° x 0.75°
ERA-Interim Land	ECMWF	Balsamo et al. (2015)	Gauge, satellite, reanalysis	1979/01 - 2010/12	SD, D and M	0.75° x 0.75°
ERA 20CM	ECMWF	Hersbach et al. (2013)	Reanalysis	1900/01 - 2010/12	D and M	0.125° x 0.125°
JRA-55	JMA	Kobayashi et al. (2015)	Reanalysis	1958/01 - present	SD and M	0.56° x 0.56°

MSWEP	JRC-EC	Beck et al. (2016)	Gauge, satellite, reanalysis	1979/01 - 2014/12	SD and D	0.25° x 0.25°
NOAA V2c	NOAA-CIRES	Compo et al. (2011)	Reanalysis	1851/01 - 2011/12	SD, D and M	2.0° x 2.0°

* SD (Sub-daily), D (Daily) and M (Monthly)

2.2.4. Validation

2.2.4.1. In situ Discharge data

Daily discharge products of the eight HR were validated against in situ discharge data from 27 gauges distributed over the Amazon main river and its tributaries (Fig. 1). Gauges with temporal coverage greater than 30 years and drainage areas larger than 1,000 km^2 were selected. Three gauge stations with drainage areas greater than 900,000 km^2, and where extreme events were usually reported in literature, were also selected for detailed analysis of extreme events. These datasets were available from the Environmental Research Observatory SO HYBAM (www.ore-hybam.org) and the Brazilian Water Agency (Agência Nacional de Águas – ANA; www.hidroweb.ana.gov.br). A table with coordinates and data availability for all gauges can be found in the supplementary material.

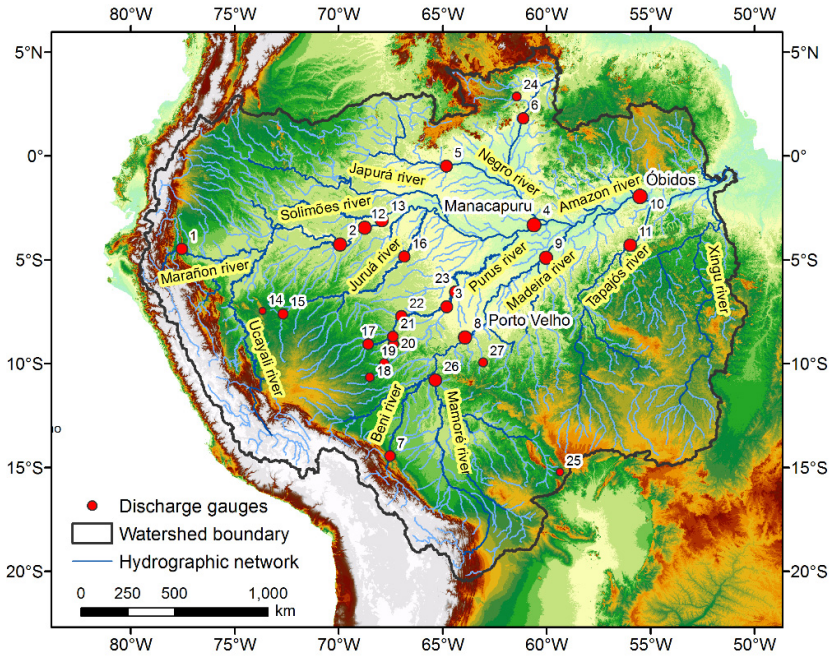

Fig. 1. Amazon River basin with river network, international limits, relief form SRTM DEM, and gauged stations (circle's size is related with catchment area) used for validation. Number corresponds to name of stations as described in Table S1.

2.2.4.2. Performance metrics

Daily discharge outputs from Hydrological Retrospective were compared to observations, based on a few performance metrics. The Nash and Suttcliffe efficiency index (NSE) is a normalized statistic that determines the relative magnitude of the residual variance compared to the observed data variance.

$$NSE = 1 - \frac{\sum_{i=1}^{nt}(Q_{obs}(t) - Q_{sim}(t))^2}{\sum_{i=1}^{nt}(Q_{obs}(t) - \overline{Q_{obs}})^2};$$

where NSE range between -∞ to 1 (perfect fit), values lower than 0 (zero) indicates that the observed mean is better predictor than the model, Q_{obs} represent observed runoff and $\overline{Q_{obs}}$ its mean value, Q_{sim} simulated runoff and nt the total number of observations (e.g. days).

The absolute value of streamflow volume error, similar to the bias, measures the average tendency of the simulated data to be higher or smaller than observed data (Gupta et al., 1999; Singh, 2005).

$$|\Delta V| = \left| \frac{\sum_{i=1}^{nt}(Q_{sim}(t)) - \sum_{i=1}^{nt}(Q_{obs}(t))}{\sum_{i=1}^{nt}(Q_{obs}(t))} \right|$$

where Q_{obs} and Q_{sim} represent observed and simulated runoff respectively, and nt the total number of observations.

The KGE' statistic proposed by Gupta et al. (2009) and modified according Kling et al. (2012), could be offers interesting results in model performance because uses most components such as correlation, variability and bias terms.

$$KGE' = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2}$$

$$a = \frac{CV_{sim}}{CV_{obs}} = \frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}; \qquad b = \frac{\mu_{sim}}{\mu_{obs}}$$

where r is the Pearson linear correlation coefficient between Q_{sim} and Q_{obs}, a and b are ratios of the coefficient of variation and means respectively, between simulated an observed runoff values. KGE' range between -∞ to 1 (perfect fit).

To verify the ability of HR to capture extreme events, compared to observed data, we tested the Pearson linear correlation coefficient for maximum, mean and minimum annual values (rMax, rMean, rMin), comparing simulated and observed runoff.

2.2.5. The intensity of duration events

To estimate the occurrence and intensity of flood and drought events, the duration of the events was computed as the number of days per year that streamflow values exceeded or fell below a certain threshold. These thresholds were calculated for high (flood) and low (drought) streamflow values as the mean of maximum/minimum annual values plus/minus one standard deviation. This methodology was suggested by the Geological Survey of Brazil (CPRM Brazil) and tested by Marengo et al. (2013) for the three largest rivers in the Amazon, and described as Flow Durations. For this study, thresholds were computed using the years 1981 to 2010 as the reference period.

Based on flow duration events from 1981 to 2010 we quantify: i) the potential of HR to detect events not reported in literature and ii) successes and failures in detecting extreme events using HR, compared to in situ observations. A contingency table (Wilks, 2011), a Miss Rate (M) statistic and a simple Peirce Skills Score (PSS) were used to quantify i) and ii). According to Stephenson (2000), PSS is a simple, reliable measure of skill and is obtained by taking the difference between hit rates (H) and false alarm rates (F). This skill score was also adopted because is easy to interpret with ranges between +1 and –1. A positive score indicates that the hit rate exceeds the false alarm rate; this can then be used to infer that there is some forecast skill.

2.2.6. Statistical temporal trend

To analyze the temporal trend of the maximum, mean and minimum annual values of the discharge series, we use the non-parametric Mann Kendall trend test (Mann, 1945). This trend test has been widely used for hydrological and climatic temporal series analysis (Cunderlik and Burn, 2004; Birsan et al., 2005; Espinoza et al., 2009b), and its use usually has been recommended with a confidence interval (CI) of 95% (Marengo, 2007; Karmeshu, 2012). However, past research shows that the presence of autocorrelation influences test results (Marengo et al., 1998; Yue et al., 2002; Hamed, 2008; Clarke, 2010), thus increasing (decreasing) the probability of trend detection if the autocorrelation is positive (negative). We therefore perform a statistical analysis to confirm the occurrence of autocorrelation in the series, and the Mann Kendall trend test was performed for each HR and for a standard period from 1981 to 2010. A more detailed explanation of calculation of the temporal series trend using the Mann Kendall test can be found in the supplementary material.

2.3. Results and discussions

2.3.1. Hydrological Retrospective vs in situ observations

The Hydrological Retrospective was compared to in situ discharge, and performance metrics are presented in Fig. 2 for an average-basin analysis (considering all stations) within a common period (1981 – 2010). The HR based on CFSR, ERA Interim, ERA-20CM, JRA-55 and NOAA datasets show low NSE (< ~0.50) and KGE' (< ~0.65) coefficients, indicating poor representation of observed hydrographs. These HR also present low correlations (< ~0.40) for maximum, mean and minimum annual discharge and are not significant ($p>0.05$) in most cases, which indicates their inability to capture extreme floods and droughts.

The best performances were found for the HRs based on ERA Interim Land, MSWEP and CHIRPS v2.0 HR. NSE and KGE' values were larger than 0.75 and volume errors were lower than 11%, indicating that these HR products adequately represent hydrograph peaks and lowest discharge values. Correlations for annual max, mean and min discharges are usually larger than 0.5 and significant ($p<0.01$) in most cases. ERA Interim Land, MSWEP and CHIRPS v2.0 products are not only based on atmospheric or ocean-atmospheric models, as the other climate reanalyses are, but they also use in situ and/or remote sensing precipitation data. This may be the main reason for the better performance of the HR based on these products.

The boxplots (Fig. 2) of performance metrics for each HR show that ERA Interim Land, CHIRPS and MSWEP display higher values and lower dispersion for NSE, $|\Delta V|$ and KGE'. For rMax, rMean and rMin, ERA 20CM shows the lowest dispersion but poor results for the median of gauges. Interquartile Range (IQR) for mean annual of ERA Interim Land, CHIRPS and MSWEP HR shows similar values and, in general, better results than other HR, while the same pattern is seen for correlations of maximum and minimum annual values.

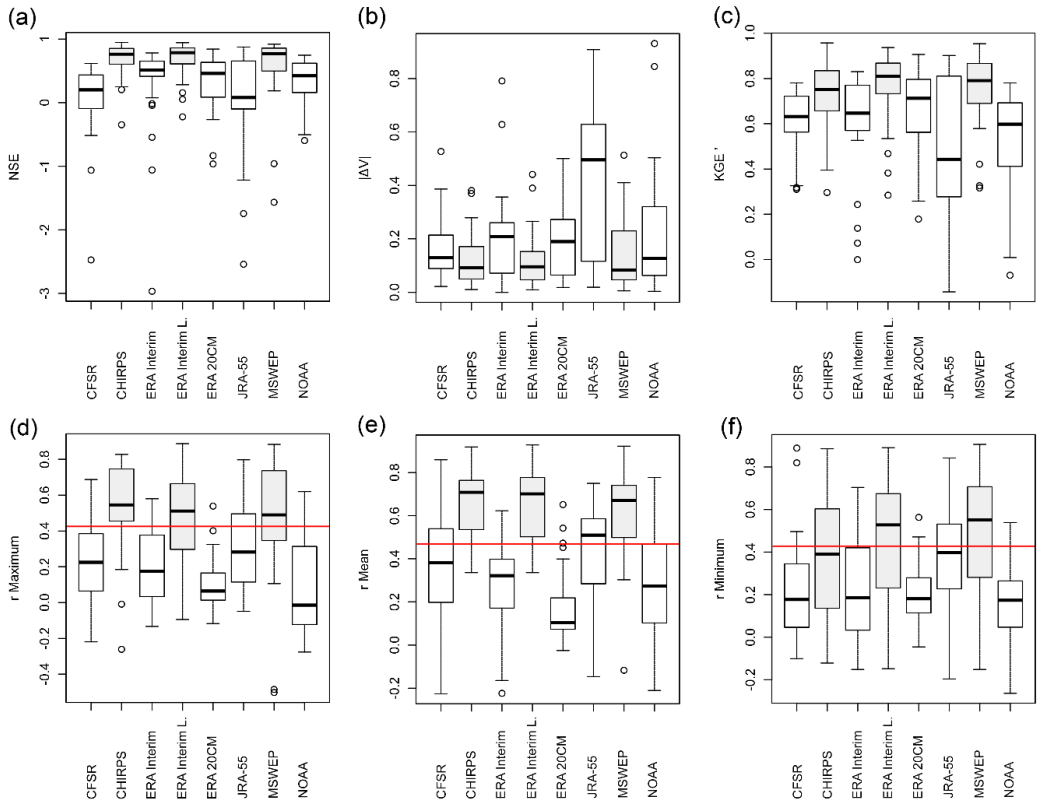

Fig. 2. Boxplot of performance metrics (a) NSE, (b) |ΔV|, (c) KGE' and Pearson linear correlation for (d) maximum, (e) mean and (f) minimum (red line means r critical at significance level of 0.01) for each HR based in climatic reanalysis (white) and HR based in merged products (gray).

A spatial distribution of these coefficients indicates that better results are obtained for the central and eastern regions (Fig. 3), while the worst results are in the western and south-western regions (near to the Andes Mountains), especially for maximum and minimum values. A clear correlation between coefficients and catchment contribution area was not found. Meanwhile, according to Fig. S1 (in supplementary material), the worst representation of precipitation by HR was found in the Amazon headwaters. For the three best HR products (ERA Interim Land, CHIRPS and MSWEP), almost 40% of gauge stations show a Pearson coefficient greater than 0.6 for maximum and minimum annual discharge values and approximately 70% for mean values.

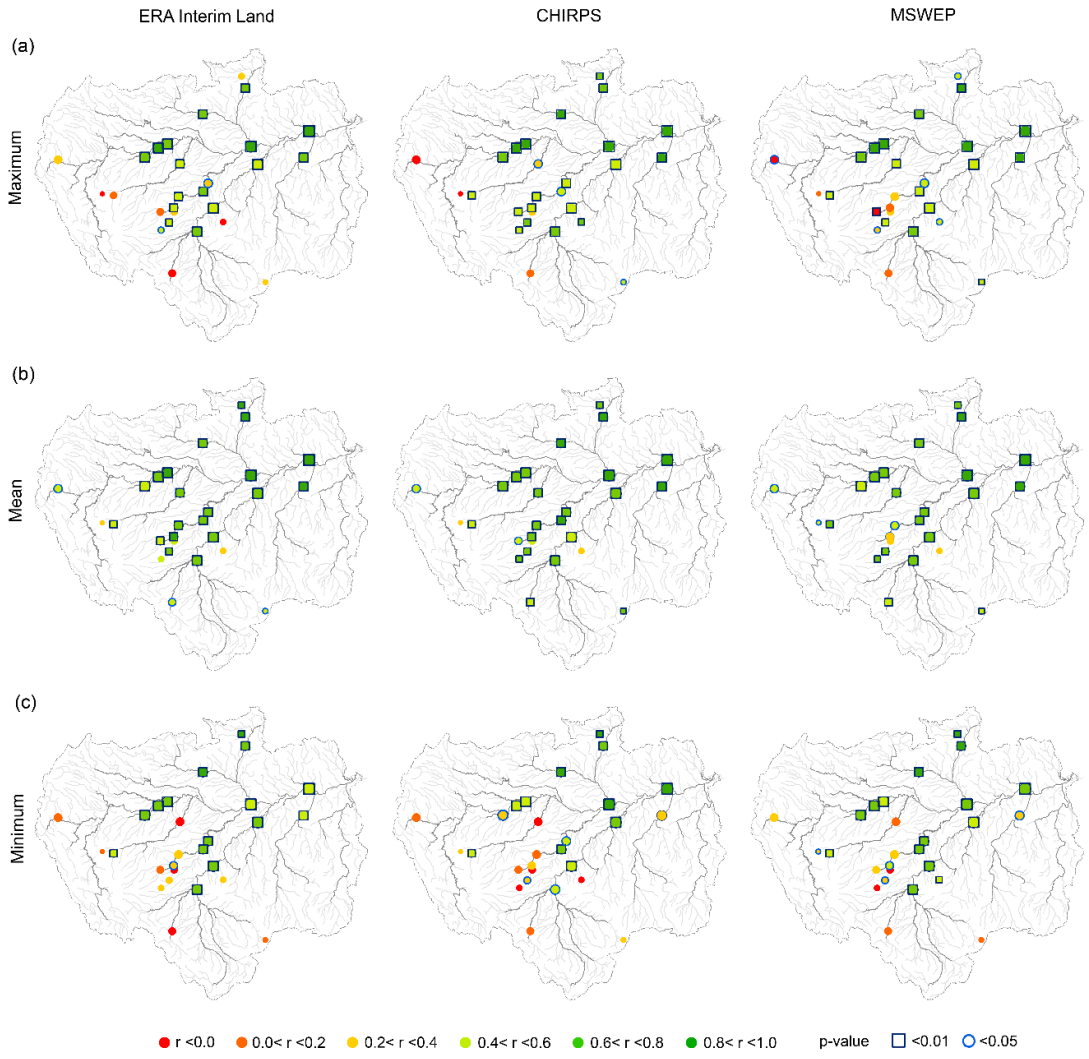

Fig. 3. Spatial distribution of Pearson linear Correlation Coefficient for (a) maximum, (b) mean, and (c) minimum annual discharge for ERA Interim land, CHIRPS and MSWEP HR against in-situ discharge data for each gauge. Significance level is indicated by square ($p<0.01$), circle ($p<0.05$) and without borders ($p>0.05$ considered not significant).

Fig. 4 shows long-term monthly mean (LTMM) discharge (1981-2010 period), indicating seasonal variability of observations and the three best HR (ERA Interim land, CHIRPS and MSWEP). In general, these HR products performed well with seasonal variability, although, discharge values were overestimated for gauges located in the northern and west-central regions and underestimated for the southern gauges.

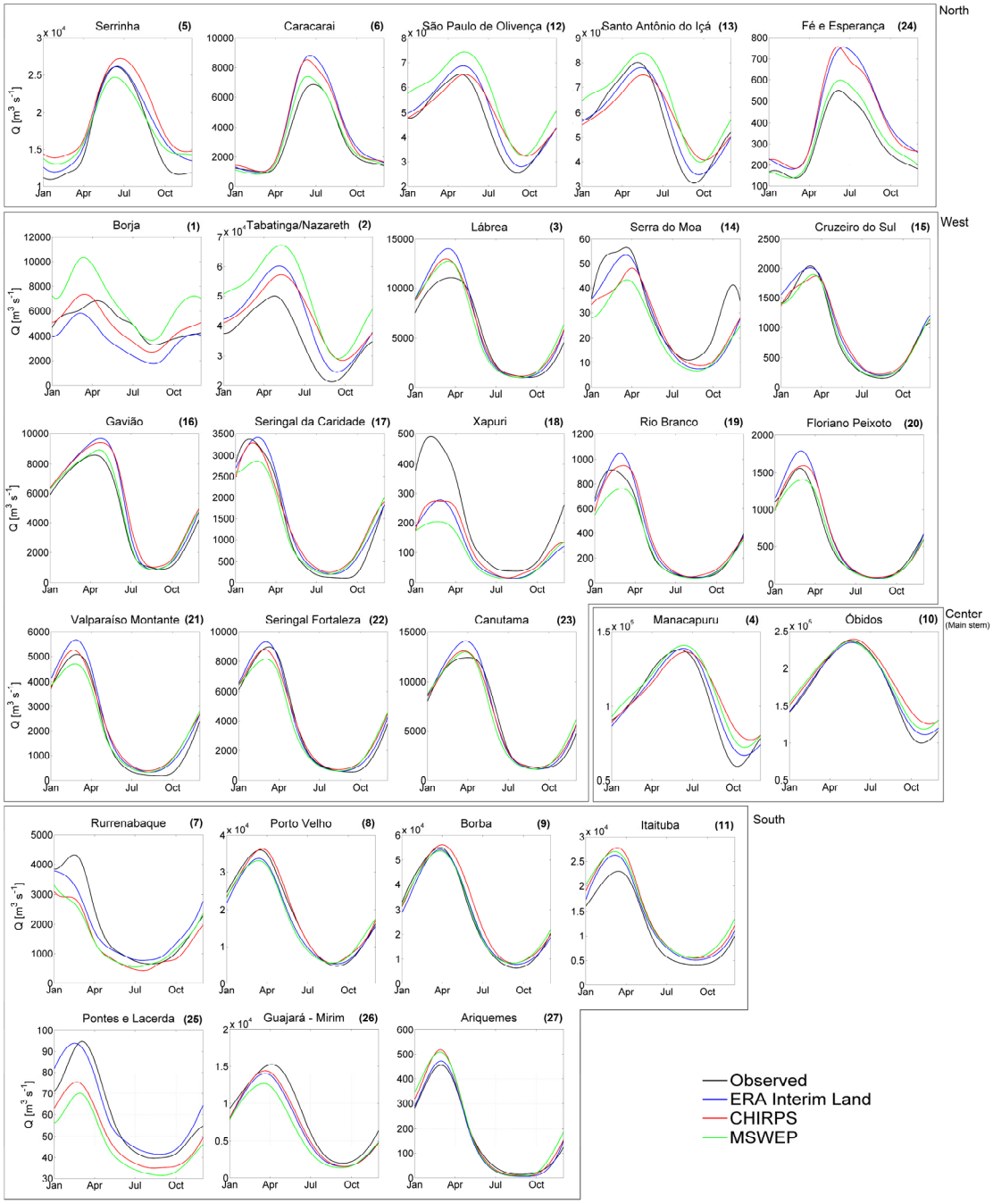

Fig. 4. Long term monthly mean of in-situ discharge data (black), ERA Interim land (blue), CHIRPS (red) and MSWEP (green) for each one of the 27 gauge stations, over the period from 1981 to 2010. Localization labels are included as North, West, Central (main stem) and South.

For historical variability of maximum, mean and minimum annual discharge, Fig. 5 shows the best HR products (according to performance metrics) for three gauge stations located at large Amazon rivers with representative drainage areas, where most extreme events are recorded in literature: Óbidos

at the Amazon River (main stem), Porto Velho at the Madeira River (at the south), and Manacapuru at the central Amazon. Correlation between observed and simulated annual discharges generally was high, indicating the ability of HR to capture the interannual variability of extreme events.

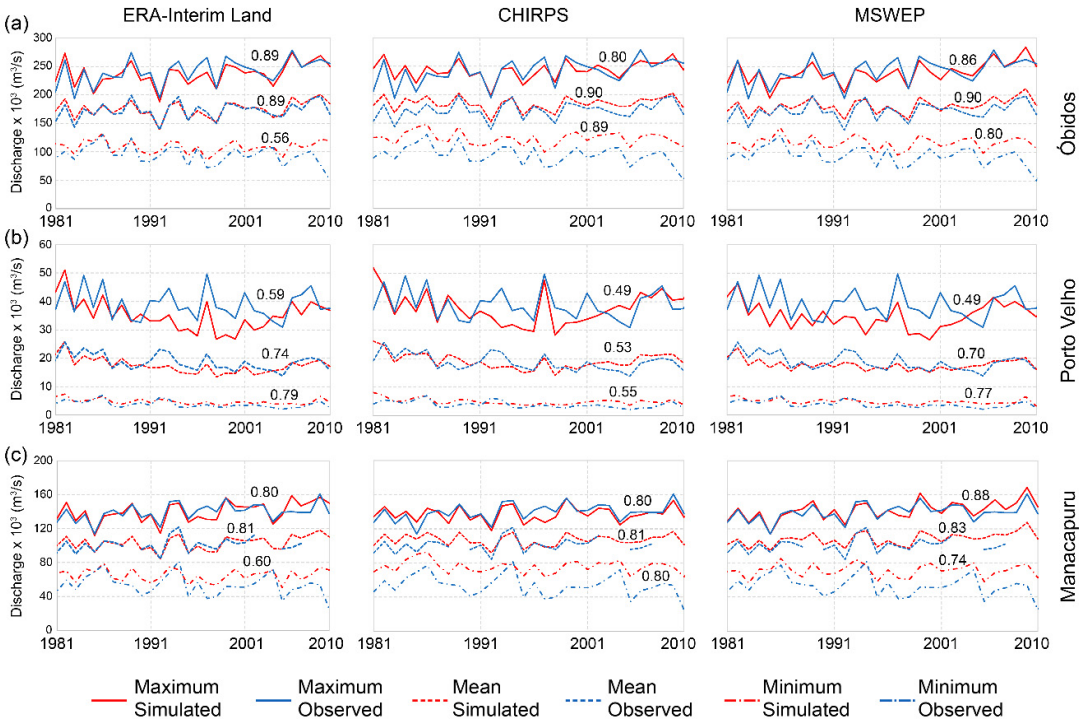

Fig. 5. Comparison of maximum, mean and minimum annual values discharge from the best Hydrological Retrospective (red lines) in the localization of some gauge station (a) Óbidos (Amazon river), (b) Porto Velho (Madeira river) and (c) Manacapuru (Solimões river), against observed discharge data (blue lines). The coefficient of correlation between HR and observed data is indicated in each panel.

Maximum annual discharge was well represented by HR at almost all gauges, although it was slightly underestimated at Porto Velho. This is consistent with Fig. S1 (in supplementary material), where precipitation datasets (ERA Interim Land, CHIRPS and MSWEP) were underestimated compared to HYBAM Observed Precipitation (Guimberteau et al., 2012) for the upstream portion of the Madeira River (central-southern region). For minimum annual discharge, HR shows slight persistent overestimation at these gauges.

2.3.2. Extreme events from literature, in situ data and Hydrological Retrospective

As a first attempt at representing extreme events, Fig. 6 shows the standard anomaly for maximum and minimum annual discharge for three gauge stations (Óbidos, Porto Velho and Manacapuru) where extreme events were commonly reported by literature. They were chosen so as to avoid a false "Miss rate." In general, the anomaly is well represented by HR, especially for ERA Interim

Land and MSWEP. Floods and droughts reported by literature were captured as largest positive and negative values of standard anomaly respectively.

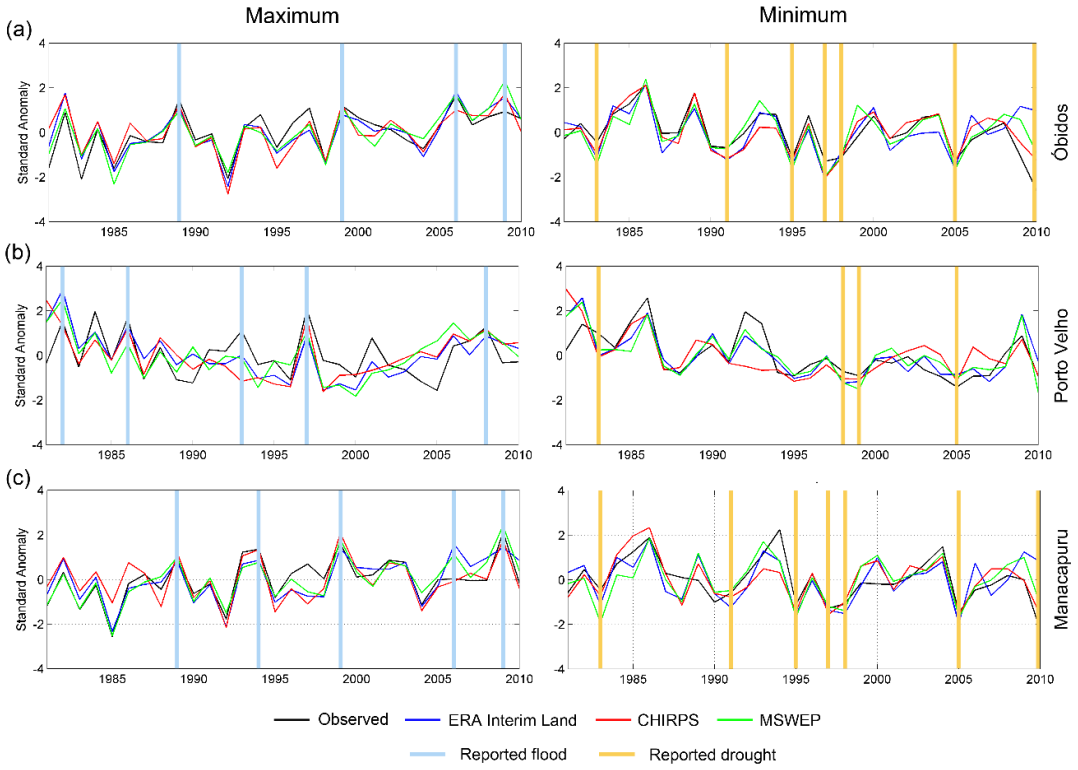

Fig. 6. Standard Anomaly for maximum and minimum annual discharge of in-situ discharge data (black), ERA Interim land (blue), CHIRPS (red) and MSWEP (green), at (a) Óbidos, (b) Porto Velho and (c) Manacapuru gauges, from 1981 to 2010. Vertical bars indicate floods (sky blue) and droughts (orange) events reported by literature.

To examine the intensity of extreme events based on flow duration, the best HR products were compared with observed data from 1981 to 2010. Extreme events detected by HR and observations were presented in temporal series, and events reported by literature were highlighted in specific years. We then presented a spatial distribution, considering specific years when extreme events were reported by literature. Finally, the HR skill score for predicting observations (considered closest to real), based on PSS, was shown for all gauge stations.

The best HR could capture almost all events observed at gauges and reported by literature. For instance, 1989, 1999, 2006 and 2009 flood events detected by observations were identified by HR at the Amazon main stem in the central-eastern region (Fig. 7a). These events were also reported by literature (Ronchail et al., 2006; Marengo et al., 2012; Espinoza et al., 2013; Satyamurty et al., 2013; Pereira et al., 2016). The 1982 flood event detected by HR (~58 days by ERA Interim land and CHIRPS and 10 days by MSWEP) at Óbidos, however, was not reported by literature.

Regarding droughts in the same region, 1995, 1997-1998, 2005 and 2010 events detected by observations and reported by literature (Sombroek, 2001; Espinoza et al., 2011; Marengo et al., 2011; Saatchi et al. 2013) were also captured by HR. Of the four floods and six droughts detected by observations, HR captured 2-3, 3-5 and 2-4 by ERA Interim land, CHIRPS and MSWEP, respectively.

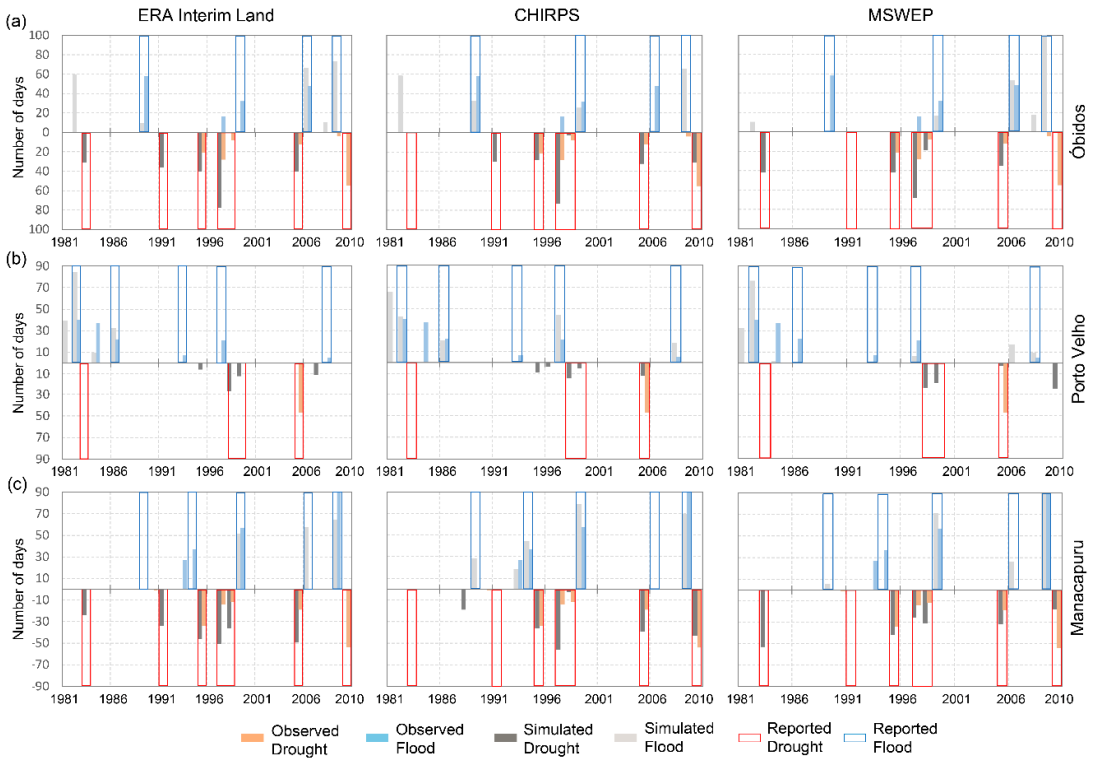

Fig. 7. Duration in days of discharge above or below certain threshold, for simulated/observed low discharge in black/orange and high discharge gray/sky-blue, and events reported by literature such as floods (blue bars) and droughts (red bars), at (a) Óbidos (Amazon river), (b) Porto Velho (Madeira river) and (c) Manacapuru (Solimões river) gauge stations.

At the gauge station on the Madeira River at Porto Velho (Fig. 7b), 1982, 1986, 1997 and 2008 flood events were captured well by HR. The 1993 flood event was also reported by literature (Ronchail et al., 2006; Pereira et al., 2016; Espinoza et al., 2014; Molina-Carpio et al. 2017). Flood events detected by HR and observations in 1981 (~37 days by HR) and 1984 (~35 days by observations) were not reported by literature.

In examining droughts, only CHIRPS and MSWEP HR captured the 2005 drought (12 and 3 days, respectively), compared to observed data (47 days). According Molina-Carpio et al. (2017), the southern region shows unusual behavior, for example, 2010 was not reported as a drought event in the upper Madeira River, unlike the main stem of the Amazon; a similar inconsistency occurs for the 1999 flood reported in Óbidos, while dry conditions were reported in the southern region. At this station, of

the six floods and one drought detected by observations, HR captured 3-0, 4-1 and 4-1 using ERA Interim land, CHIRPS and MSWEP, respectively.

In the central Amazon River at the Manacapuru gauge (Fig. 7c), 1999 and 2009 flood events were captured well by all HR, compared to observations; , CHIRPS HR also captured 1993 and 1994 flood events. Besides these events, 1989, 1994 and 2006 flood events were reported by literature (Ronchail et al., 2006; Marengo et al., 2012; Espinoza et al., 2013; Satyamurty et al., 2013; Pereira et al., 2016). HR captured drought events in 1995, 1997-1998, 2005 and 2010, compared to observed data. ERA Interim Land HR also captured events in 1983 and 1991 (24 and 34 days, respectively) that were not detected by observations, but which were reported by literature (Sombroek, 2001; Marengo et al., 2008; Sheffield, 2011; Sena et al., 2012; Marengo et al., 2012; Marengo and Espinoza, 2016). At Manacapuru, of the four floods and six droughts detected by observations, HR captured 2-4, 4-5 and 2-5 using ERA Interim land, CHIRPS and MSWEP, respectively.

For extreme events detected by observations, each year was compared against the long-term monthly mean (LTMM) (Fig. 8) of HR. The monthly mean of HR for each year was also plotted, so Fig. 8 shows how extreme events detected by observations for each year are correctly positioned or not with respect to LTMM of HR. For instance, for observed data, floods (blue) and droughts (red) are very consistent, as shown in Fig. 8. Meanwhile, the Óbidos gauge shows that floods events in 1989, 1997, 1999 and 2006 were consistently positioned above LTMM, while drought events in 1995, 1997, 1998, 2005 and 2010 were positioned below LTMM. A slight (four-day) drought detected in 2009 by observations at Óbidos station is shown in Fig. 7a and Fig. 8a as a red curve above LTMM, while a major flood was reported by literature for that year (Marengo et al., 2012; Filizola et al., 2014). This detection is explained because Fig. 8 shows events qualitatively, regardless of the duration of any single event. At Porto Velho and Manacapuru stations (Fig. 8a, 8c), year events registered by observations were consistent with HR.

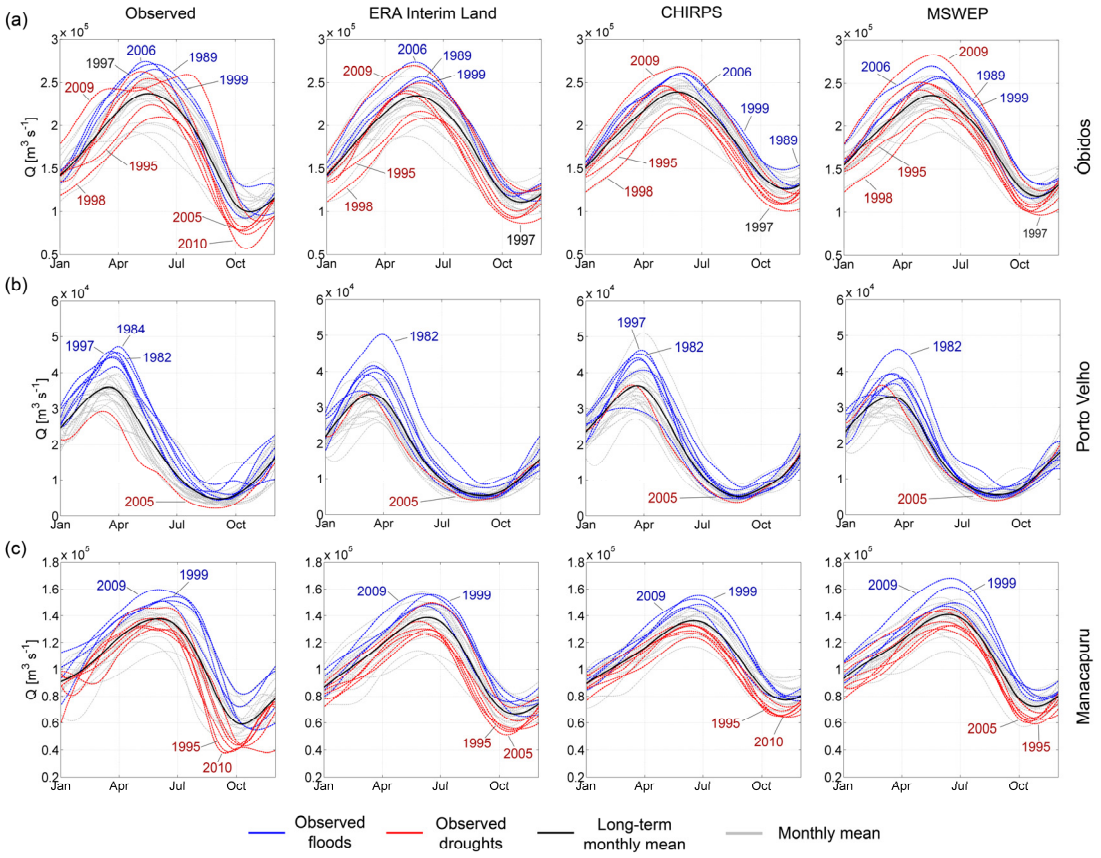

Fig. 8. Long-term monthly mean (black) and monthly means for each year (gray) for Observed, ERA Interim land, CHIRPS and MSWEP HR; flood (blue) and droughts (red) are years registered by in-situ observed discharge data, at (a) Óbidos, (b) Porto Velho and (c) Manacapuru gauge stations, from 1981 to 2010. Particular years are indicated in some hydrographs.

To better compare the performance of HR products (represented by reach river), they were compared spatially against observed discharge data (represented by circles), considering the flow duration of flood events (Fig. 9) and droughts (Fig. 10) in specific years. For flood events, HR products are spatially consistent with observed data for five extreme floods: 1982 in the central-southern Amazon, 1989 in the central-eastern Amazon, 1997 in the central-southern Amazon, and 1999 and 2009 in the central-western Amazon at the main basin.

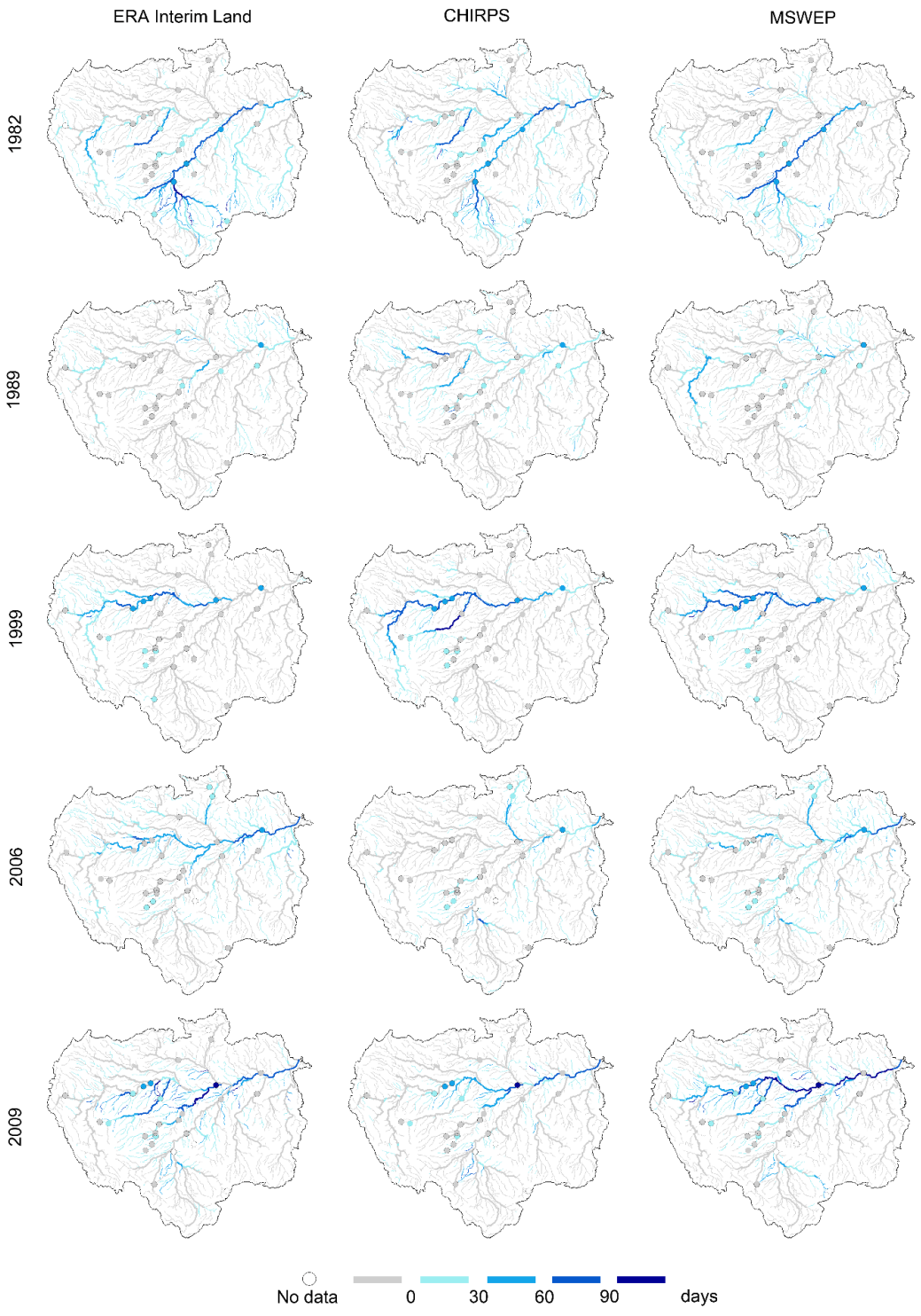

Fig. 9. Spatial distribution of Extreme Intensity of duration Floods (number of days above a threshold) from 1981 to 2010, for Hydrological Retrospective products in each reach river and Observed data (circles).

For drought events, HR products were consistent with observed data for 1983 and 1992 events in the northern region at the Serrinha gauge station; the 1995 drought was captured well by HR at the

Manacapuru gauge station. In 1998, event results were consistent at the Branco River (north region) and Purus River (central region). HR captured the severe 2005 drought in the eastern region of the main river and at the Purus River. Fig. 9 and Fig. 10 show how Hydrological Retrospective can be useful for providing a detailed picture of regional aspects during specific droughts and floods.

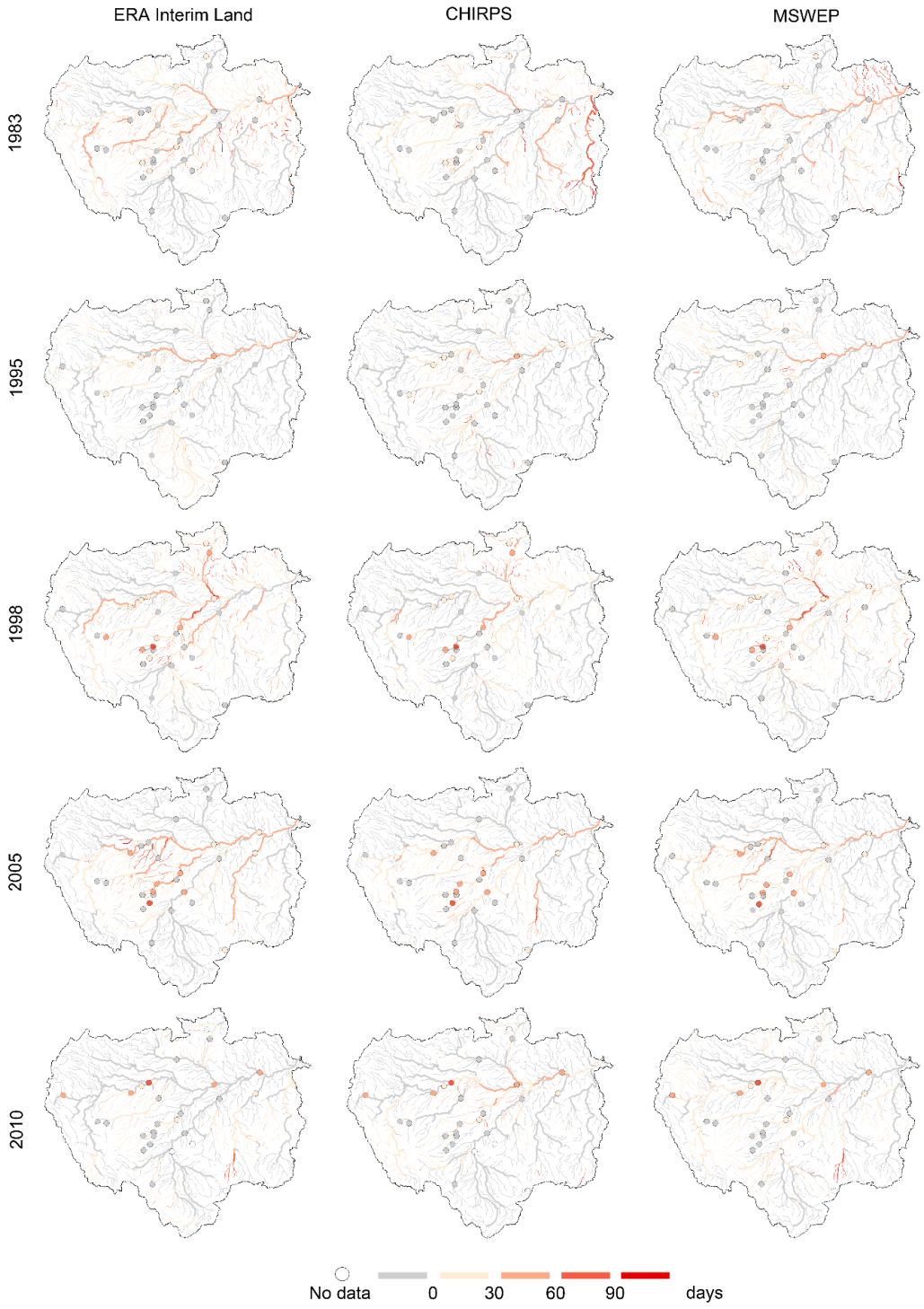

Fig. 10. Spatial distribution of Extreme Intensity of duration Droughts (number of days below a threshold) from 1981 to 2010, for Hydrological Retrospective products in each reach river and Observed data (circles).

To determine whether there were extreme events not reported by literature but detected by HR and observations in those gauge stations, the Miss Rate statistic (M) was calculated (Fig. 11). A value of zero means all events were reported in literature. Our results were analyzed by comparison with observed events (assumed to be closest to reality); in other words, an HR M value higher/lower than observations means that detection of events was overestimated/underestimated. In general, floods and droughts have been well recorded by literature; HR, however, showed slight overestimations and underestimations compared with observations.

Fig. 11 shows that there were flood events not reported in literature (positive values for observed) at the three stations, as mentioned above. Overestimation of floods was detected by ERA Interim Land and MSWEP at Óbidos and Porto Velho, the opposite occurred at the Manacapuru station. For droughts, there apparently were events not reported by literature at Óbidos and Manacapuru (Fig. 11a and 11c) in brief (fewer than five days) events in 2009 and 1990, respectively (showed in Figs. 7a and 7c), but the intensity of flow duration probably was small to be reported. At the Porto Velho station, HR overestimated the detection of droughts in some years.

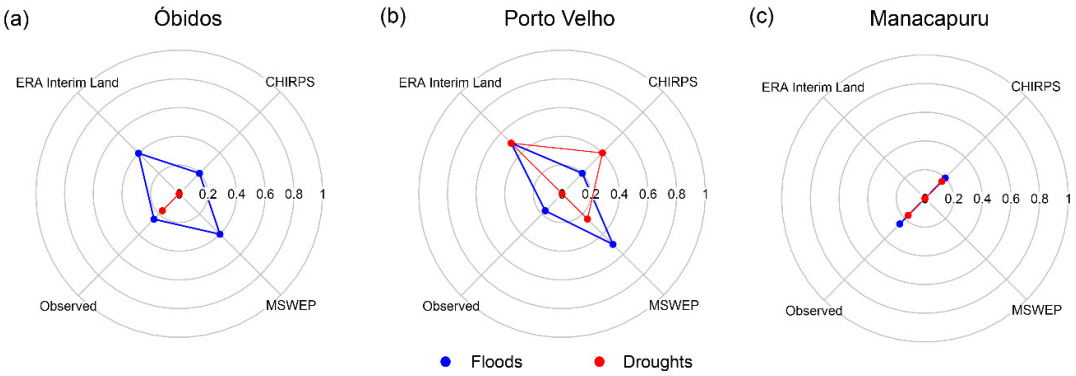

Fig. 11. Miss rate (from 0 to 1) of events reported by literature (forecast) against HR and observations at (a) Óbidos, (b) Porto Velho and (c) Manacapuru gauge stations.

HR's accuracy in forecasting extreme events, compared to in-situ discharge data, is shown in Fig. 12, using the Peirce skill score (PSS) for floods and droughts detected by flow duration for all gauge stations. For floods (Fig. 12a), many gauges have values above zero. The most accurate HR was CHIRPS, with 25/27 (93%) stations above zero. MSWEP and ERA Interim land yielded 24/27 (89%) and 20/27 (74%) stations, respectively. It is interesting that CHIRPS HR at Lábrea and Manacapuru (code numbers 3 and 4, respectively, in Table S1 in supplementary material) had a PSS value of almost "1," which means a perfect forecast (almost no Miss Rate and False Alarm Rate).

For droughts (Fig. 12b), few gauge stations presented negative PSS values. CHIRPS HR was also the best HR for forecasting droughts, with 18/27 (67%) of gauge stations above a PSS value of zero, while ERA Interim land and MSWEP HR showed 16/27 (59%) and 15/27 (60%) of stations, respectively, above zero. Porto Velho (station code number 8) was the gauge station where droughts were forecast best by CHIRPS and MSWEP HR.

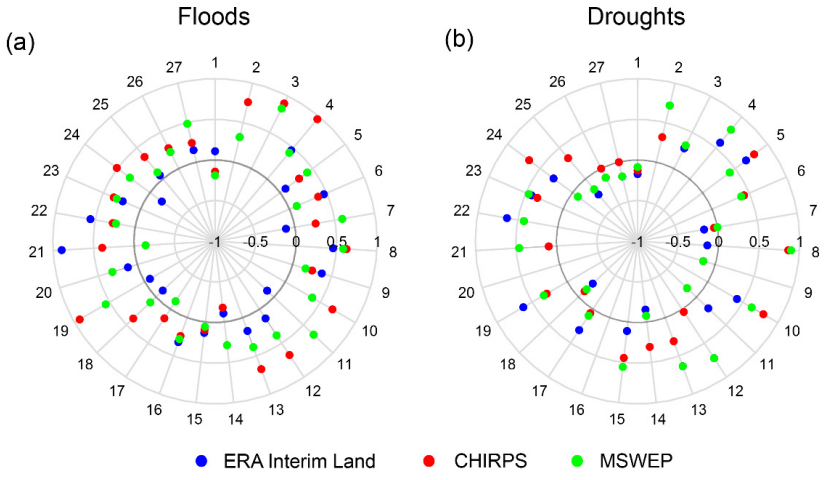

Fig. 12. Peirce score skill (from -1 to 1) of HR ERA Interim Land (blue), CHIRPS (red) and MSWEP (green) to forecast (a) floods and (b) droughts detected by observations in all gage stations from 1981 to 2010.. Code number of stations (see Table S1) were also plotted.

2.3.3. Trend analysis

Trend analysis was performed for maximum, mean and minimum annual discharge values for the years 1981-2010. These series did not show autocorrelation. The spatial distribution of these trends (Fig. 13) shows a positive trend (blue), negative trend (red) or no trend (gray). These results show an upward trend in the north-northwestern region (left bank) and on the main stem, especially for maximum and mean values, representing increasing floods in these regions during the 1981-2010 period. In southern regions and mainly in the Madeira basin, minimum values show a downward trend, which corresponds to an increase in the intensity of droughts during that period.

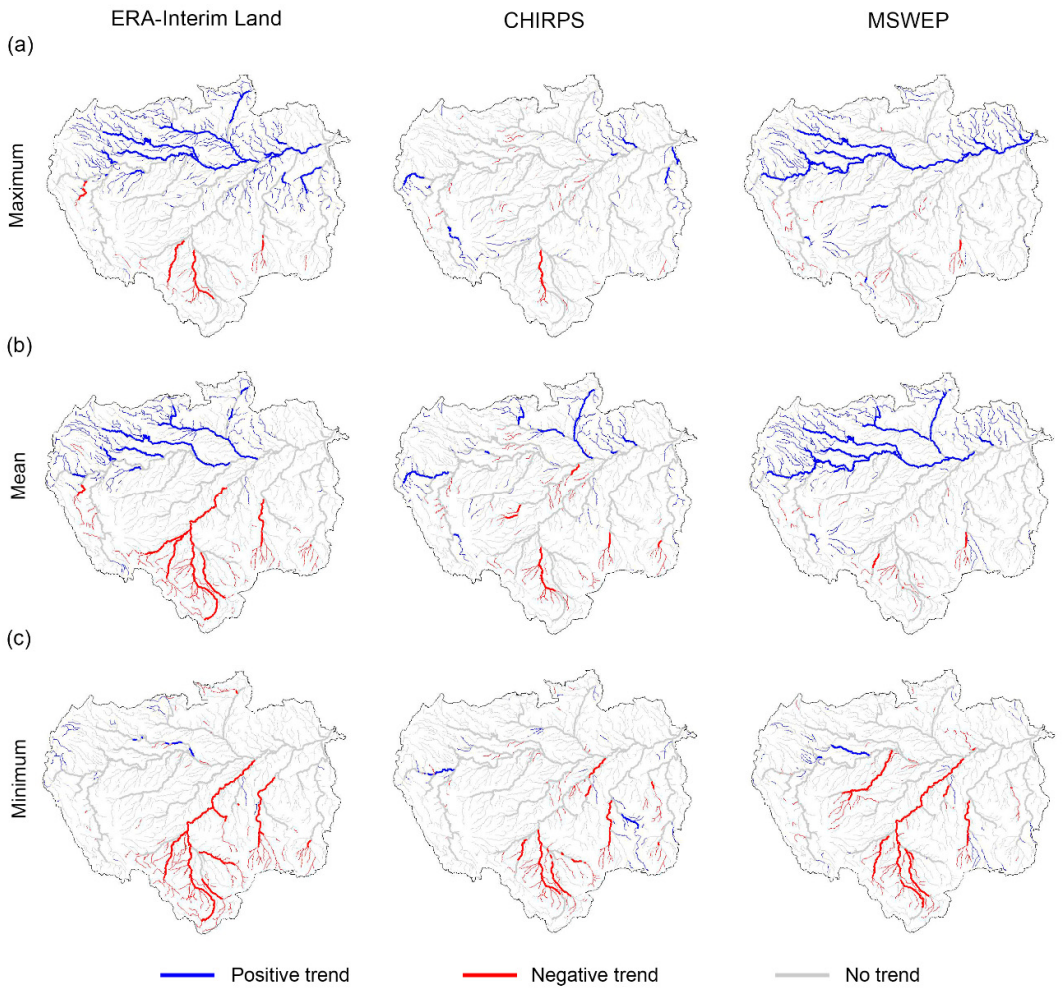

Fig. 13. Mann-Kendall trend test of (a) maximum, (b) mean and (c) minimum annual discharge for ERA Interim land, CHIRPS and MSWEP HR, from 1981 to 2010. Reaches indicate positive trend (blue), negative trend (red) and no trend (gray) considering a significance level of 0.05.

We found it remarkable that our results are compatible with other previous studies. Callède et al. (2004), who studied the evolution of discharge on the Amazon River (main stem) over almost all of the last century (1903 to 1999), found a positive trend in mean and maximum discharge, with increases of 9% and 10%, respectively. In a study based on in situ discharge data, Espinoza et al. (2009b) also show a positive trend from 1974 to 2004 for mean and maximum discharge values in the northwestern region and a negative trend in the south and southwestern region for maximum, mean and minimum values, using statistical coefficients such as Pearson, spearman rho and Kendall. Using in situ data, Molina-Carpio et al. (2017) identified a downward trend in minimum discharge for the 1985-2013 period in the southwestern region, while Lopes et al (2016), using in situ discharge data from 1975 to 2013, noticed a significant trend of more intense droughts in southern regions, similar to our results. The trends observed in the Hydrological Retrospective are in accordance with future discharge projections

in the context of global climate change (Guimberteau et al., 2013; Boisier et al., 2015; Sorribas et al, 2016). For example, Sorribas et al. (2016) projected a future increase in annual maximum discharge values in the western and north-central Amazon basin using a climate model for 2070 to 2099; the same study showed annual mean values with positive anomalies in western regions and a persistent decrease in eastern regions.

2.4. Conclusions

In this paper, we propose a Hydrological Retrospective based on regional scale hydrological and hydrodynamic modeling forced by eight daily rainfall reanalyses, from the early 20th century to present. This methodology has been developed to validate, capture and evaluate trends of extreme events over the Amazon basin in recent decades.

Validation results indicate that the use of precipitation data from climate reanalysis based only on atmospheric or ocean-atmospheric models, does not accurately represent river discharges in the Amazon basin, if the goal is to study interannual variability, floods and droughts. Hydrological Retrospective based on precipitation datasets using satellite, climate reanalysis and in situ data (ERA-Interim Land, CHIRPS and MSWEP) yielded better results, although the temporal coverage is shorter than that of "pure" climate reanalysis. Statistical coefficients vary for regions, although on average these three HR products show the most accurate performance, according to statistical metrics; this is also seen in the temporal series, when compared to observations at gauges.

In general, floods and droughts were well reported by literature at the analyzed stations, and Hydrological Retrospective results were reasonably consistent with observations. Some miss rates were detected, probably because the intensity, based on flow duration (number of days) was slight. In view of these findings, it is important to consider the potential of HR to detect events with low duration but with extreme discharge values, mainly for flood events.

Most of the floods and droughts detected by in-situ discharge were forecast well by Hydrological Retrospective, according to the forecast skill statistic. Hydrological Retrospective was also consistent with the majority of events reported by literature for three representative gauge stations. This indicates that it is possible to use Hydrological Retrospective methodology to identify past extreme events in regions where in-situ gauge availability is sparse or nonexistent, and where extreme events therefore cannot be identified. Because global precipitation datasets have improved notably in recent, this will make it possible to determine the behavior of these past events around the entire Amazon basin or other regions of the world.

As several authors have indicated, trend intensity of extreme events in some regions of the Amazon basin has been increasing in recent decades. Here we show that Hydrological Retrospective

could serve to represent an increase in the trend of maximum discharge (floods) intensity in the western and northwestern regions, rather that the entire Amazon basin, while detecting a trend toward more intense minimum discharges (droughts) in the south and central-south.

In this paper, we evaluated a single Hydrological Retrospective approach based on a regional hydrological model forced by climate reanalysis. Other approaches should also be implemented in the future, such as HR based on regional hydrological models that assimilate long records of in situ discharge or recent multiple remote sensing observations.

The HR with the best performance could be used as a practical methodology for: i) spatially and temporally distributed evaluation of past hydrological behavior, ii) creating records of floods and droughts, and iii) assessing trends of intensity of extreme events and possibly making projections in other regions of the world.

2.5. Supplementary material

The Table S1 shows a detailed information of 27 gauge discharge stations used to validation process.

Table S1. Summary of in situ discharge gauges

Gauge	Code	Name	Longitude (°)	Latitude (°)	Catchment area (km²)	River	Initial date	Final date
1	10064000	Borja	-77.548	-4.470	117,000	Marañon	02/02/1986	01/10/2015
2	10100000	Tabatinga/Nazareth	-69.933	-4.250	880,250	Solimões	01/07/1982	29/04/2015
3	13870000	Lábrea	-64.800	-7.252	220,000	Purus	05/07/1967	29/04/2015
4	14100000	Manacapuru	-60.609	-3.308	2,147,740	Solimões	29/06/1972	30/01/2015
5	14420000	Serrinha	-64.829	-0.482	279,950	Negro	30/11/1967	27/02/2015
6	14710000	Caracarai	-61.124	1.821	124,980	Branco	02/01/1967	30/08/2014
7	15275100	Rurrenabaque	-67.534	-14.445	68,900	Beni	07/08/1967	13/10/2015
8	15400000	Porto Velho	-63.920	-8.737	954,290	Madeira	11/04/1967	29/04/2015
9	15860000	Borba	-60.025	-4.897	1,324,700	Madeira	03/05/1967	30/03/2015
10	17050001	Óbidos	-55.511	-1.947	4,618,750	Amazonas	23/02/1968	30/01/2015
11	17730000	Itaituba	-55.983	-4.283	452,000	Tapajos	13/02/1968	27/02/2015
12	11400000	São Paulo de Olivença	-68.750	-3.450	1,010,000	Solimões/Amazonas	18/07/1973	17/11/2015
13	11500000	Santo Antônio do Içá	-67.933	-3.083	1,130,000	Solimões/Amazonas	14/07/1973	17/10/2011
14	12400000	Serra do Moa	-73.664	-7.447	1,030	Moa	08/05/1973	31/01/2011
15	12500000	Cruzeiro do Sul	-72.681	-7.611	37,800	Juruá	23/08/1967	31/01/2011
16	12840000	Gavião	-66.849	-4.839	164,000	Juruá	22/06/1972	28/02/2011
17	13410000	Seringal da Caridade	-68.577	-9.044	63,100	Purus	01/08/1967	31/07/2010
18	13550000	Xapuri	-68.507	-10.650	8,270	Acre	01/09/1967	31/07/2010
19	13600002	Branco	-67.800	-9.976	23,500	Acre	08/08/1967	31/07/2010
20	13650000	Floriano Peixoto	-67.368	-9.051	34,400	Acre	06/10/1968	30/04/2010
21	13710001	Valparaíso Montante	-67.400	-8.683	105,000	Purus	26/02/1975	31/07/2010
22	13750000	Seringal Fortaleza	-66.985	-7.717	154,000	Purus	15/07/1967	31/07/2010

23	13880000	Canutama	-64.384	-6.534	236,000	Purus	01/01/1973	31/07/2010
24	14680001	Fé e Esperança	-61.441	2.871	12,200	Mucajaí	07/12/1973	31/12/2007
25	15050000	Pontes e Lacerda	-59.354	-15.216	2,990	Guaporé	26/06/1971	31/12/2006
26	15250000	Guajará-Mirim	-65.348	-10.792	609,000	Mamoré	08/08/1970	31/12/2005
27	15430000	Ariquemes	-63.057	-9.932	8,140	Jamari	15/07/1970	31/12/2005

2.5.1. Mann Kendall trend test

Statistical "S" measures the tendency on information, positive values indicate an increase in concentration (positive trend) while negative values indicate a decrease in concentration (negative trend). The intensity of the trend is proportional to the magnitude of Mann-Kendall statistic S, so high values of S indicate a strongly trend.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign[x_j - x_i]$$

where *x* is the observed value in time sequential order, *n* the number of observations and sign the indicator of the function that results in values 1, 0 or -1 according the signal of $[x_j - x_i]$, the function of this indicator is as follows:

$$\text{For:} \qquad sign[x_j - x_i] = \begin{cases} 1; \text{if } x_j - x_i > 0 \\ 0; \text{if } x_j - x_i = 0 \\ -1; \text{if } x_j - x_i < 0 \end{cases}$$

In this way is calculated the Z_{MK} assuming a significance level of 5%, H_0 will be rejected (existing trend) while Z_{MK} is, in absolute value, greater than 1.96, and calculated as follow:

$$Z_{MK} = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & if \quad S > 0 \\ 0 & if \quad S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & if \quad S < 0 \end{cases}$$

2.5.2. Peirce skills score

This index is calculate as follow:

$$PSS = H - F = \frac{ad - bc}{(a+c)(b+d)}$$

where H and F mean a Hit rate and False alarm rate and were calculate based in the contingency table (Table S2).

Table S2. Contingency table for forecast accuracy quantify

Forecast	Observed	
	Detect	No detect

Detect	a	b
No detect	c	d

2.5.3. Rainfall data sets accuracy

Considering an annual mean discharge (Fig. 4) for average of ERA Interim land, CHIRPS and MSWEP HR an overestimation was achieve (19%) in the northern regions such as Serrinha mainly during dry season for northern (October to April), Caracarai and Fé e Esperança during wet season (from May to September). Besides, in central west sites such as Itaituba, São Paulo de Olivença and Santo Antônio do Içá, overestimation shows a 12%. In contrast, for the southern gauges such as Rurrenabaque, Xapuri, Pontes e Lacerda and Guajará-Mirim where discharge values were underestimated in a -23%.

Behavior at Fig. 4 could be explained by observing spatial variability (Fig. 1s) from percentage of relative error of long term (1980 – 2009) mean of total three monthly precipitation from datasets (ERA Interim Land, CHIRPS and MSWEP) interpolated for each catchment and compared in relation to HYBAM Observed Precipitation (Guimberteau et al., 2012) that is a gridded daily rainfall data set (1°x1°) produced by geostatistical interpolation from several rain gauges over Amazon basin, this data set is available in <www.ore-hybam.org>.

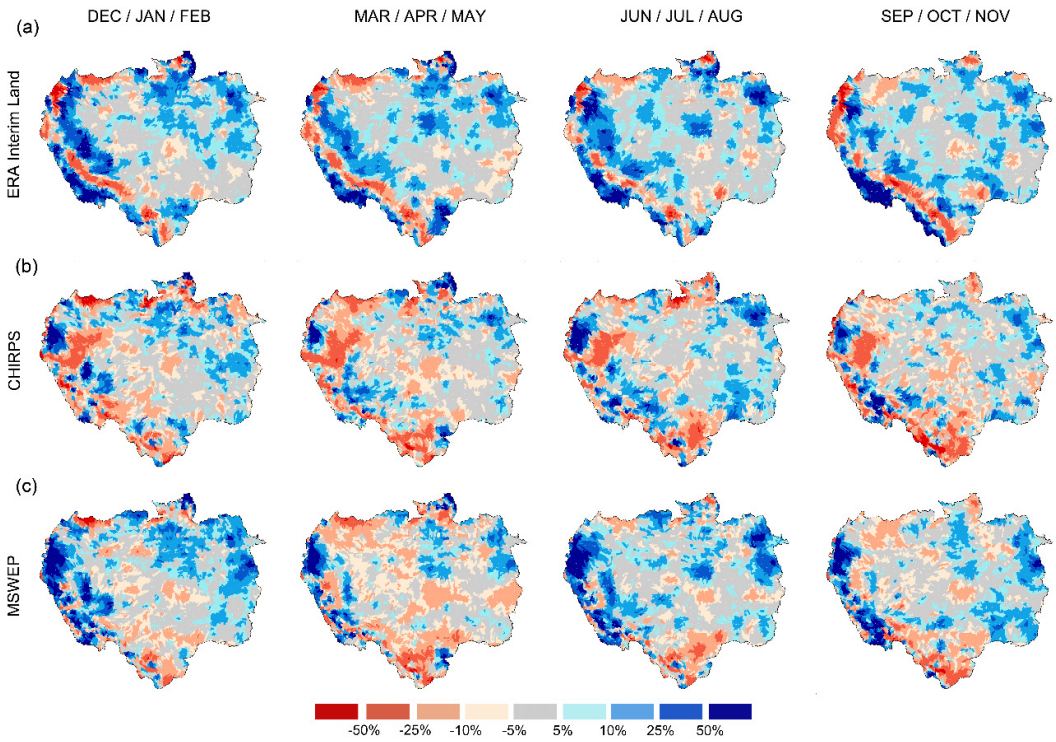

Fig. 1s. Percentage of relative error from long term mean of three monthly precipitation for (a) ERA Interim Land, (b) CHIRPS and (c) MSWEP, in relation to HYBAM Observed Precipitation dataset (1980-2009).

As can been seen in Fig. 4, south region (*Guajará–Mirim* (26) gauge) shows an underestimation of precipitation mainly from March to November, on the other hand south-east region, represented only by *Itaituba* gauge (11), display a persistent overestimation predominantly from June to November (dry season), and in the same way north-east (north hemisphere) regions shows also slight overestimation mainly from September to Abril during dry session too. In addition it is interesting to note from all datasets (notable for ERA Interim Land) a persistent overestimation in west and underestimation in east Andes regions.

Acknowledgements

The first author is grateful for a grant from the Brazilian agency CAPES and for MGB-IPH Fortran codes provided by Rodrigo Paiva. The authors are grateful for the precipitation datasets supplied by NCEP, CHG (Climate Hazards Group), ECMWF, JMA (Japan Meteorological Agency), JRC-EC (European Commission, Joint Research Centre), NOAA-CIRES and HYBAM, and the in-situ discharge data provided by ANA and HYBAM. We also thank two anonymous reviewers for their comments, which helped significantly in the improvement of this manuscript.

References

Adler, R.F., Kidd, C., Petty, G., Morissey, M., Goodman, H.M., 2001. Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP-3). Bulletin of the American Meteorological Society, 82(7), 1377-1396.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS – global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci., 17, 1161–1175.

Alho, C. J. R., Reis, R. E., Aquino, P. P. U., 2015. Amazonian freshwater habitats experiencing environmental and socioeconomic threats affecting subsistence fisheries. Ambio, 44(5), 412–425.

Aragão, L.E.O., Malhi, Y., Barbier, N., Lima, A., Shimabukuro, Y., Anderson, L., and Saatchi, S., 2008. Interactions between rainfall, deforestation and fires during recent years in the Brazilian Amazonia. Philosophical Transactions of the Royal Society of London B, Biological Sciences, 363(1498), 1779–1785.

Aragão, L.E.O.C., Malhi, Y., Roman-Cuesta, R.M., Saatchi, S., Anderson, L.O., and Shimabukuro, Y.E., 2007. Spatial patterns and fire response of recent Amazonian droughts. Geophysical Research Letters, 34(7), n/a–n/a. L07701.

Asner, G.P. and Alencar, A., 2010. Drought impacts on the amazon forest, the remote sensing perspective. New Phytologist, 187(3), 569–578.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F., 2015. Era Interim/ land, a global land surface reanalysis data set. Hydrology and Earth System Sciences, 19(1), 389–407.

Bastola, S., Misra, V., 2014. Evaluation of dynamically downscaled reanalysis precipitation data for hydrological application. Hydrological Processes, 28(4), 1989–2002.

Beck, H.E., van Dijk, A.I. J. M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., and de Roo, A., 2016. MSWEP, 3-hourly 0.25 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences Discussions, 2016, 1–38.

Beighley, R.E., Eggert, K.G., Dunne, T., He, Y., Gummadi, V., Verdin, K.L., 2009. Simulating hydrologic and hydraulic processes throughout the Amazon River Basin. Hydrological processes, 23(8), 1221–1235.

Betts, A.K., Ball, J.H., Viterbo, P., Dai, A., and Marengo, J., 2005. Hydrometeorology of the amazon in era-40. Journal of Hydrometeorology, 6(5), 764–774.

Betts, A.K., Köhler, M., and Zhang, Y., 2009. Comparison of river basin hydrometeorology in era-interim and era-40 reanalyses with observations. Journal of Geophysical Research, Atmospheres, 114(D2), n/a–n/a. D02101.

Birsan, M.-V., Molnar, P., Burlando, P., and Pfaundler, M., 2005. Streamflow trends in Switzerland. Journal of Hydrology, 314(1–4), 312 – 329.

Blacutt, L.A., Herdies, D.L., de Gonçalves, L.G.G., Vila, D.A., and Andrade, M., 2015. Precipitation comparison for the cfsr, merra, fTRMM3B42g and combined scheme datasets in Bolivia. Atmospheric Research, 163, 117 – 131. 6th Workshop of the International Precipitation Working Group.

Boisier, J.P., Ciais P., Ducharne A., and Guimberteau M., 2015. Projected strengthening of Amazonian dry season by constrained climate model simulations. Nature Climate Change, 5(7), 656-660.

Bovolo, C.I., Pereira, R., Parkin, G., Kilsby, C., Wagner, T., 2012. Fine-scale regional climate patterns in the Guianas, tropical South America, based on observations and reanalysis data. Int. J. Climatol., 32, 1665–1689.

Callède, J., Guyot, J.L., Ronchail, J., L'H^ote, Y., Niel, H., and de Oliveira, E., 2004. Evolution du débit de l'Amazone à Óbidos de 1903 à 1999 / evolution of the river amazon's discharge at Óbidos from 1903 to 1999. Hydrological Sciences Journal, 49(1), 85–97.

Carvalho, L.M.V., Jones, C., Posadas, A.N.D., Quiroz, R., Bookhagen, B., Liebmann, B., 2012. Precipitation Characteristics of the South American Monsoon System Derived from Multiple Data Sets. Science and Technology Infusion Climate Bulletin, 25(13), 4600-4620.

Castello, L., Isaac, V.J., Thapa, R., 2015. Flood pulse effects on multispecies fishery yields in the Lower Amazon. R Soc Open Sci., 2(11): 150299.

Clarke, Robin T., 2010. On the (mis)use of statistical methods in hydro-climatological research. Hydrological Sciences Journal, 55(2): 139-144, DOI: 10.1080/02626661003616819

Coe, M.T., Costa, M.H., Howard, E.A., 2008. Simulating the surface waters of the Amazon River basin impacts of new river geomorphic and flow parameterizations. Hydrological processes, 22(14), 2542–2553.

Collischonn, B., Collischonn, W., Tucci, C.E.M., 2008. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. Journal of Hydrology, 360(1-4), 207–216.

Collischonn, W., Allasia, D., da Silva, B.C., and Tucci, C.E.M., 2007. The mgb-iph model for large-scale rainfall—runoff modelling. Hydrological Sciences Journal, 52(5), 878–895.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin,X., Gleason, B.E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, ., Ross, T. F., Trigo, R.M., Wang, X.L., Woodruff, S.D., and Worley, S. J., 2011. The twentieth century reanalysis project. Quarterly Journal of the Royal Meteorological Society, 137(654), 1–28.

Cunderlik, J.M. and Burn, D.H., 2004. Linkages between regional trends in monthly maximum flows and selected climatic variables. Journal of Hydrologic Engineering, 9(4), 246–256.

de Linage, C., Famiglietti, J.S., and Randerson, J.T., 2014. Statistical prediction of terrestrial water storage changes in the amazon basin using tropical pacific and north atlantic sea surface temperature anomalies. Hydrology and Earth System Sciences, 18(6, 2089–2102.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S.B., Hersbach, H., H´olm, E.V., Isaksen, L., K°allberg, P., K¨ohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F., 2011. The era-interim reanalysis, configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597.

Dee, D.P., Balmaseda, M., Balsamo, G., Engelen, R., Simmons, A.J., and Thépaut, J.-N., 2014. Toward a consistent reanalysis of the climate system. Bulletin of the American Meteorological Society, 95(8, 1235–1248.

Dirzo, R., Raven, P.H., 2003. Global State of Biodiversity and Loss. Annual Review of Environment and Resources, 28, 137-167.

Ebert, E.E., Janowiak, J.E., Kidd, C., 2007. Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. Bulletin of the American Meteorological Society, 88(1), 47-64.

Espinoza J.C., Ronchail, J., Guyot, J.L., Cocheneau G., Filizola, N., Lavado, W., de Oliveira, E., Pombosa, R., Vauchel, P., 2009a. Spatio – Temporal rainfall variability in the Amazon Basin Countries (Brazil, Peru, Bolivia, Colombia and Ecuador. International Journal of Climatology, 29, 1574-1594.

Espinoza J.C., Guyot J.L., Ronchail J., Cocheneau G., Filizola N., Fraizy P., Labat D., de Oliveira E., Ordoñez, J.J. and Vauchel P., 2009b. Contrasting regional discharge evolutions in the Amazon Basin. Journal of Hydrology, 375, 297-311.

Espinoza, J.C., Ronchail, J., Guyot, J.L., Junquas, C., Vauchel, P., Lavado, W., Drapeau, G., and Pombosa, R., 2011. Climate variability and extreme drought in the upper Solimões River (western amazon basin, Understanding the exceptional 2010 drought. Geophysical Research Letters, 38(13), n/a–n/a. L13406.

Espinoza, J.C., Ronchail, J., Frappart, F., Lavado, W., Santini, W., and Guyot, J.L., 2013. The major floods in the Amazonas River and tributaries (western amazon basin during the 1970–2012 period, A focus on the 2012 flood. Journal of Hydrometeorology, 14(3), 1000–1008.

Espinoza, J.C., Marengo, J.A., Ronchail, J., Carpio, J.M., Flores, L.N., and Guyot, J.L., 2014. The extreme 2014 flood in south-western amazon basin, the role of tropical-subtropical south atlantic sst gradient. Environmental Research Letters, 9(12), 124007.

Espinoza, J.C., Segura, H., Ronchail, J., Drapeau, G., Gutierrez-Cori, O., 2016. Evolution of wet- and dry-day frequency in the western Amazon basin, Relationship with atmospheric circulation and impacts on vegetation. Water Resources Research, 52(11), 8546–8560.

Fan, F.M., Schwanenbergb, D., Collischonn, W., Weerts, A., 2015. Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil. Journal of Hydrology, Regional Studies, 4, 196–227.

Fan, F.M., Collischonn, W., Quiroz, K.J., Sorribas, M.V., Buarque, D.C., Siqueira, V.A., 2016. Flood forecasting on the Tocantins River using ensemble rainfall forecasts and real-time satellite rainfall estimates. Journal of Flood Risk Management, 9(3), 278–288.

Fernandes, K, Baethgen, W., Bernardes, S., DeFries, R., DeWitt, D., Goddard, L., Lavado, W., Eun Lee, D., Padoch, C., Pinedo Vasquez, M., Uriarte, M., 2011. North tropical Atlantic influence on western Amazon fire season variability. Geophys. Res. Lett. 651 38(12), L12701.

Filizola, N., Latrubesse, E.M., Fraizy, P., Souza, R., Guimarães, V., and Guyot, J.-L., 2014. Was the 2009 flood the most hazardous or the largest ever recorded in the amazon? Geomorphology, 215, 99–105. Morphological characterization and fluvial processes of large rivers at different time scales.

Frolking, S., Milliman, T., Palace, M., Wisser, D., Lammers, R., and Fahnestock, M., 2011. Tropical forest backscatter anomaly evident in SeaWinds scatterometer morning overpass data during 2005 drought in Amazonia. Remote Sensing of Environment, 115(3), 897 – 907.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J., 2015. The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. Scientific Data, 2, 150066, doi:10.1038/sdata.2015.66

Gao, L., Bernhardt, M., Schulz, K., Chen, X., Chen, Y., and Liu, M., 2016. A first evaluation of era-20cm over china. Monthly Weather Review, 144(1), 45–57.

Garreaud, R.D., Vuille, M., Compagnucci, R., Marengo, J., 2009. Present-day South American climate. Geophysical Research Letters, 33(22), doi: 10.1029/2006GL027644

Gibson, J. and for Medium Range Weather Forecasts, E. C., 1997. ECMWF Re-analysis Project Report Series, ERA description. Number v. 1. European Centre for Medium-Range Weather Forecasts.

Gloor, M., Brienen, R.J.W., Galbraith, D., Feldpausch, T.R., Schöngart, J., Guyot, J.-L., Espinoza, J.C., Lloyd, J., and Phillips, O.L. 2013. Intensification of the amazon hydrological cycle over the last two decades. Geophysical Research Letters, 40(9), 1729–1733.

Guimberteau, M., Drapeau, G., Ronchail, J., Sultan, B., Polcher, J., Martinez, J.-M., Prigent, C., Guyot, J.-L., Cochonneau, G., Espinoza, J.C., Filizola, N., Fraizy, P., Lavado, W., De Oliveira, E., Pombosa, R., Noriega, L., and Vauchel, P., 2012. Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets, Hydrol. Earth Syst. Sci., 16, 911-935, doi: 10.5194/hess-16-911-2012

Guimberteau, M., Ronchail, J., Espinoza, J.C., Lengaigne, M. Sultan, B., Polcher, J., Drapeau, G., Guyot, J-L., Ducharne, A., Ciais, P., 2013. Future changes in precipitation and impacts on extreme streamflow over Amazonian sub-basins. Environmental Research Letters, 8(1), 014035.

Gupta, H.V., Kling, H., Yilmaz, K.K., and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria, Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2), 80 – 91.

Hamed, K.H., 2008. Trend detection in hydrologic data, The mann–kendall trend test under the scaling hypothesis. Journal of Hydrology, 349(3–4), 350 – 363.

Hersbach, H., Peubey, C., Simmons, A., Poli, P., D., D., and Berrisford, P., 2013. Era-20cm, a twentieth century atmospheric model ensemble. era report series n 16. Technical Report 16, European Centre for Medium-Range Weather Forecasts. 44 p.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. Journal of Hydrometeorology, 8(1), 38-55.

Junk, W. J., 1993. Wetlands of Tropical South America, Kluwer, Dordrecht, Netherlands, 15-2, 679-739.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins,W., Deaven, D., Gandin, L., Iredell, M., Saha, D., White, G.,Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki,W., Higgins,W., Janowiak, J.,Mo, K.C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Dennis, J., 1996. The NCEP/NCAR 40-years reanalysis project. Bull. Am. Meteorol. Soc. 77, 437–472.

Karlsson, I.B., Sonnenborg, T.O., Refsgaard, J.C., Trolle, D., Børgesen, C.D., Olesen, J.E., Jeppesen, E., Jensen, K.H., 2016. Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change. Journal of Hydrology, 535, 301-317.

Karmeshu, N., 2012. Trend detection in annual temperature & precipitation using the mann kendall test – a case study to assess climate change on select states in the northeastern united states. Master's thesis submitted in partial fulfillment of requirements for master of environmental studies, Department of Earth & Environmental Science, University of Pennsylvania.

Kidd, C. and Huffman, G., 2011. Global Precipitation Measurement. Meteorological Applications 18, 334-353. doi, 10.1002/met.284

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M., 2001. The ncep–ncar 50–year reanalysis, Monthly means cd–rom and documentation. Bulletin of the American Meteorological Society, 82(2), 247–267.

Kling, H., Fuchs, M. & Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios Journal of Hydrology, 424–425, 264 – 277.

Knox, J.C.; Kundzewicz, Z.W., 1997. Extreme hydrological events, palaeo-information and climate change. Hydrological Sciences Journal, 42(5), 765-779.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K., 2015. The jra-55 reanalysis, General specifications and basic characteristics. Journal of the Meteorological Society of Japan. Ser. II, 93(1), 5–48.

Koster, R.D., Mahanama, S.P.P., Yamada, T.J., Balsamo, G., Berg, A.A., Boisserie, M., Dirmeyer, P.A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J.-H., Lee, W.-S., Li, Z., Luo, L., Malyshev, S., Merryfield, W.J., Seneviratne, S.I., Stanelle, T., van den Hurk, B.J.J.M., Vitart, F., and Wood, E.F., 2011. The second phase of the global land–atmosphere coupling experiment, Soil moisture contributions to subseasonal forecast skill. Journal of Hydrometeorology, 12(5), 805–822.

Lee, T., Kwon, H., Modarres, R., Kim, S., Chebana, F., 2016. Hydrological and Meteorological Extreme Events in Asia, Understanding, Modeling, Vulnerability, and Adaptation Measures. Advances in Meteorology, doi:10.1155/2016/2325384.

Lettenmaier, D.P., Alsdorf, D., Dozier, J., Huffman, G.J., Pan, M., Wood, E.F., 2015. Inroads of remote sensing into hydrologic science during the WRR era. Water Resources Research, 51(9), 7309–7342.

Lewis, S.L., Brando, P.M., Phillips, O.L., van der Heijden, G.M.F., and Nepstad, D., 2011. The 2010 amazon drought. Science, 331(6017), 554–554.

Lopes, A.V., Chiang, J.C.H., Thompson, S. A., and Dracup, J.A., 2016. Trend and uncertainty in spatial-temporal patterns of hydrological droughts in the Amazon basin. Geophys. Res. Lett., 43, 3307–3316.

Mann, H. B., 1945. Nonparametric Tests Against Trend Econometrica, [Wiley, Econometric Society], 13, 245-259.

Marengo, J.A., Tomasella, J., Uvo, C.R., 1998. Trends in streamflow and rainfall in tropical South America, Amazonia, eastern Brazil, and northwestern Peru. Journal of Geophysical Research, 103, 1775-1783.

Marengo, J.A., 2007. Boletim do projeto uso de cenários de mudanças climáticas regionais em estudos de vulnerabilidade e adaptação no brasil e na américa do Sul (gof-uk-cptec).

Marengo, J.A., Nobre, C.A., Tomasella, J., Oyama, M.D., de Oliveira, G.S., de Oliveira, R., Camargo, H., Alves, L.M., and Brown, I.F., 2008. The drought of Amazonia in 2005. Journal of Climate, 21(3), 495–516.

Marengo, J.A., Tomasella, J., Alves, L.M., Soares, W.R., and Rodriguez, D.A., 2011. The drought of 2010 in the context of historical droughts in the amazon region. Geophysical Research Letters, 38(12), n/a–n/a. L12703.

Marengo, J.A., Tomasella, J., Soares, W. R., Alves, L. M., and Nobre, C. A., 2012. Extreme climatic events in the amazon basin. Theoretical and Applied Climatology, 107(1), 73–85.

Marengo, J.A., Borma, L.S., Rodriguez, D.A., Pinho, P., Soares, W.R., and Alves, L.M., 2013. Recent extremes of drought and flooding in Amazonia, vulnerabilities and human adaptation. American Journal of Climate Change, 2, 87–96.

Marengo, J.A. and Espinoza, J. C. 2016. Extreme seasonal droughts and floods in Amazonia, causes, trends and impacts. International Journal of Climatology, 36(3), 1033–1050.

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., and Lane, J.E., 2009. Precipitation, measurement, remote sensing, climatology and modeling, Atmospheric Research, 94, 512–533.

Molina-Carpio, J., Espinoza, J.C., Vauchel, P., Ronchail, J., Gutierrez, B., Guyot, J.L., Noriega, L., 2017. The hydroclimatology of the upper Madeira River basin, spatio-temporal variability and trends (1967-2013). Accepted in Hydrological Sciences Journal.

Neal, J., Schumann, G., Bates, P., 2012. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resour. Res., 48, W11506, doi, 10.1029/2012WR012514.

New, M., Lister, D., Hulme, M., and Makin, I., 2002. A high-resolution data set of surface climate over global land areas. Climate Research, 21, 1–25.

Nijssen, B., O'donnell, G.M., Hamlet, A.F., Lettenmaier, D.P., 2001. Hydrologic sensitivity of global rivers to climate change. Climate change, 50(1), 143-175.

Nogués-Paegle, J., Mechoso, C.R., Fu, R., Berbery, H., Winston, C., Chao, T., Cook, K., Diaz, A., Enfield, D., Ferreira, R., Grimm, A., Kousky, V., Liebmann, B., Marengo, J., Mo, K., Neelin, D., Paegle, J., Robertson, A., Seth, A., Vera, C., Zhou, J., 2002. Progress in Pan American CLIVAR Research, Understanding the South American Monsoon. Meteorologica, 27, 3-30.

Ovando, A., Tomasella, J., Rodriguez, D.A., Martinez, J.M., Siqueira-Junior, J.L., Pinto, G.L.N., Passy, P., Vauchel, P., Noriega, L., von Randow, C., 2016. Extreme flood events in the Bolivian Amazon wetlands. Journal of Hydrology, Regional Studies, 5, 293–308.

Paiva, R.C.D., Collischonn, W., and Tucci, C.E., 2011. Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. Journal of Hydrology, 406(3–4), 170 – 181.

Paiva, R.C.D., Buarque, D.C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., and Bulhões Mendes, C.A., 2013. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resources Research, 49(3), 1226–1243.

Pereira, D.M., Szlafsztein, C.F., 2016. Ameaças e desastres naturais na Amazônia sul ocidental, análise da bacia do rio Purus / Natural hazards and disasters in south western Amazon region: analysis of Purus river basin. Ra'e Ga - O Espaço Geográfico em Análise, 35, 68–94.

Phillips, O.L., Aragão, L.E.O.C., Lewis, S.L., Fisher, J.B., Lloyd, J., L´opez-Gonz´alez, G., Malhi, Y., Monteagudo, A., Peacock, J., Quesada, C.A., van der Heijden, G., Almeida, S., Amaral, I., Arroyo, L., Aymard, G., Baker, T.R., B´anki, O., Blanc, L., Bonal, D., Brando, P., Chave, J., de Oliveira, A.C.A., Cardozo, N.D., Czimczik, C.I., Feldpausch, T.R., Freitas, M.A., Gloor, E., Higuchi, N., Jiménez, E., Lloyd, G., Meir, P., Mendoza, C., Morel, A., Neill, D. A., Nepstad, D., Pati˜no, S., Pe˜nuela, M.C., Prieto, A., Ram´ırez, F., Schwarz, M., Silva, J., Silveira, M., Thomas, A. S., Steege, H. t., Stropp, J., V´asquez, R., Zelazowski, P., Dávila, E.A., Andelman, S., Andrade, A., Chao, K.-J., Erwin, T., Di Fiore, A., C., E. H., Keeling, H., Killeen, T.J., Laurance, W.F., Cruz, A.P., Pitman, N.C.A., Vargas, P.N., Ram´ırez-Angulo, H., Rudas, A., Salamão, R., Silva, N., Terborgh, J., and Torres-Lezama, A., 2009. Drought sensitivity of the amazon rainforest. Science, 323(5919), 1344–1347.

Phipps, S J., McGregor H.V., Gergis J., Gallant A.J.E., Neukom R., Stevenson S., Ackerley D., Brown J. R., Fischer M. J., van Ommen T. D., 2013. Paleoclimate Data–Model Comparison and the Role of Climate Forcings over the Past 1500 Years. J. Climate, 26, 6915–6936.

Pohlmann, H. and Greatbatch, R.J., 2006. Discontinuities in the late 1960's in different atmospheric data products. Geophysical Research Letters, 33(22, n/a–n/a. L22803.

Pontes, P.R.M., Collischonn, W., Fan, F.M., Paiva, R.C.D., Buarque, D.C., 2015. Modelagem hidrológica e hidráulica de grande escala com propagação inercial de vazões / Hydrologic and hydraulic large-scale modeling with inertial flow routing. Brazilian Journal of Water Resources, 20(4), 888-904.

Prakash, S., Gairola, R.M., and Mitra, A.K., 2015. Comparison of large-scale global land precipitation from multisatellite and reanalysis products with gauge-based gpcc data sets. Theoretical and Applied Climatology, 121(1), 303–317.

Reichle, R. and Liu, Q., 2015. Precipitation and global land surface hydrology in the MERRALand and MERRA-2 reanalysis datasets. In EGU General Assembly Conference Abstracts, volume 17 of EGU General Assembly Conference Abstracts, page 1838.

Ronchail, J., Guyot,, J.L., Espinoza, J.C., Callède, J., Cochonneau, G., De Oliveira, E., Ordenez, J.J., Filizola, N., 2006. Impact of the Amazon tributaries on flooding in Óbidos. Climate variability and Change – Hydrological

Impacts (Procceedings of the Fifth FRIEND World Conference held at Havanna, Cuba, November 2006, IAHS Publ. 308, 220-225.

Saatchi, S., Asefi-Najafabady, S., Malhi, Y., Aragão, L., Anderson, L., Myneni, R., and Nemani, R., 2013. Persistent effects of a severe drought on amazonian forest canopy. Proceedings of the National Academy of Sciences of the United States of America, 110(2), 565–570.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins,W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R., Rutledge, G., and Goldberg, M., 2010. The ncep climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8), 1015–1057.

Saleska, S., Didan, K., Huete, A., and Da Rocha, H., 2007. Amazon forests green-up during 2005 drought. Science, 318(5850), 612.

Satyamurty, P., da Costa, C.P.W., Manzi, A.O., and Candido, L.A., 2013. A quick look at the 2012 record flood in the amazon basin. Geophysical Research Letters, 40(7), 1396–1401.

Scaife, A.A., Folland, C.K., Alexander, L.V., Moberg, A., Knight, J.R., 2008. European climate extremes and the North Atlantic Oscillation. J. Climate, 21(1), 72-83.

Schöngart, J. and Junk, W., 2007. Forecasting the flood-pulse in central amazonia by ensoindices. Journal of Hydrology, 335(1-2), 124–132.

Sena, J.A., Beser de Deus, L.A., Freitas, M.A.V., and Costa, L., 2012. Extreme events of droughts and floods in amazonia, 2005 and 2009. Water Resources Management, 26(6), 1665–1676.

Seyyedi, H., Anagnostou, E., Beighley, E., and McCollum, J., 2015. Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin. Atmospheric Research, 164-165, 37–48.

Shah, R. and Mishra, V., 2014. Evaluation of the reanalysis products for the monsoon season droughts in India. Journal of Hydrometeorology, 15(4), 1575–1591.

Sheffield, J. and Wood, E.F., 2011. Drought, Past Problems and Future Scenarios. Routledge. 192 pp.

Singh, J., Knapp, H.V., Arnold, J., Demissie, M., 2005. Hydrological Modeling of the Iroquois River Watershed Using HSPF and SWAT. Journal of the American Water Resources Association, Blackwell Publishing Ltd, 41, 343-360.

Smith, C., Lawson, N., 2012. Identifying extreme event climate thresholds for greater Manchester, UK: examining the past to prepare for the future. Meteorological Applications, 19, 26–35.

Sombroek, W., 2001. Spatial and temporal patterns of amazon rainfall. Royal Swedish Academy of Sciences, 30(7), 388–396.

Sorribas, M. V., Paiva, R.C.D., Melack, J.M., Bravo, J.M., Jones, C., Carvalho, L., Beighley, E., Forsberg, B., Costa, M.H., 2016. Projections of climate change effects on discharge and inundation in the Amazon basin. Climatic Change, 136, 555-570.

Stephenson D.B., 2000. Use of the ‘‘Odds Ratio’’ for Diagnosing Forecast Skill. American Meteorological Society, 15, 221–232.

Swierczynski, T., Ionita, M., Pino, D., 2017, Using archives of past floods to estimate future flood hazards, Eos, 98, doi, 10.1029/2017EO066221.

Wanders, N., Bierkens, M.F., Sutanudjaja, E., van Beek, R., 2014. The PCR-GLOBWB global hydrological reanalysis product. In EGU General Assembly Conference Abstracts, volume 16 of EGU General Assembly Conference Abstracts, 16, EGU2014-5369.

Werner, A.T. and Cannon, A.J., 2016. Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods. Hydrol. Earth Syst. Sci., 20, 1483–1508.

Wilks, D. S., 2011. Statistical methods in the atmospheric sciences. 3rd ed. Academic Press, 704 pp.

Xu, L., Samanta, A., Costa, M. H., Ganguly, S., Nemani, R. R., and Myneni, R. B., 2011. Widespread decline in greenness of amazonian vegetation due to the 2010 drought. Geophysical Research Letters, 38(7), n/a–n/a. L07402.

Xu, H., Xu, C., Chen, S., Chen, H., 2016. Similarity and difference of global reanalysis datasets (WFD and APHRODITE in driving lumped and distributed hydrological models in a humid region of China. Journal of Hydrology, 542, 343–356.

Yamazaki, D., Lee, H., Alsdorf, D. E., Dutra, E., Kim, H., Kanae, S., Oki, T., 2012. Analysis of the water level dynamics simulated by a global river model, A case study in the Amazon River. Water Resources Research, 48, W09508.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes, 16(9), 1807–1829. Cited By 485.

Zeng, N., Yoon, J.-H., Marengo, J. A., Subramaniam, A., Nobre, C. A., Mariotti, A., and Neelin, J. D., 2008. Causes and impacts of the 2005 amazon drought. Environmental Research Letters, 3(1), 014002.

Zhu, Q., Xuan, W., Liu, L., Xu, Y-P., 2016. Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China. Hydrological processes, 30(17), 3061–3083.

Capítulo 3

Avaliação da técnica de assimilação de dados no MGB

"Estimativa de vazões para seu uso na gestão dos recursos hídricos com modelo hidrológico semi-distribuido e assimilação de dados pelo Ensemble Kalman Filter"

Prólogo

Neste capitulo se avaliou de maneira detalhada a técnica de assimilação de dados (*data assimilation*, DA) filtro de Kalman por ensemble (*Ensemble Kalman Filter*, EnKF) no modelo hidrológico MGB para a estimativa de series de vazões e vazões de referência em comparação com métodos tradicionais de regionalização. A implementação e avaliação desta tecnica de DA no modelo MGB foi desenvolvida no trabalho de Paiva et al. (2013b). A sua adaptação para a versão do modelo MGB com propagação inercial e sua maior exploração através das diferentes análises de sensibilidade na geração dos membros do ensemble do esquema EnKF foi avaliado neste capitulo.

Para isto foi escolhida a bacia de Taquari-Antas, onde as informações foram as suficientes para desenvolver os testes tanto para a validação dos resultados de vazão, como para avaliar diferentes configurações na sensibilidade da resposta do modelo MGB aos parâmetros do EnKF e à perturbação das forçantes e das variáveis de estado do modelo.

Towards the Discharge Estimation for Water Resources Management Uses with a Semi-Distributed Model and Ensemble Kalman Filter Data Assimilation

Wongchuig, C.S., Fleischman, A., de Paiva, R.C.D., Fadel., W.A.

Paper resubmitted to the Hydrological Science Journal on May 3

Abstract

Estimating discharges is a major challenge in water resources management, and techniques such as Data Assimilation (DA) can be used to improve these estimates. This study assesses the adaptation of the Ensemble Kalman Filter DA scheme within a large scale hydrological-hydrodynamic model to improve discharge estimates. Different scenarios with assimilation and validation gauges were performed to obtain an optimal setup of assimilation parameters. These parameters were used to analyze the estimated discharge series and reference discharge values for maximum, minimum and average flows. Results showed that the perturbation of soil water storage leads to the most sensitive model responses. Furthermore, using the adequate setup, better estimates of the reference discharge values were obtained with the DA scheme, when compared to a simple regionalization method. Also leading to reduction of errors in more than double in many cases.

Key words: Water resources management; Data Assimilation; Reference Discharges; Hydrological Modelling.

3.1. Introduction

To enable society and stakeholders to develop better water resource management plans, adequate approaches to understand and predict the water-human-ecosystem interactions must be carried out (Wagener et al. 2010). In this context, adequate discharge estimation is a core issue of hydrological studies (Arnold et al. 1998), especially considering an increasing demand and competition for water resource (Postel and Wolf 2001).

In hydrological modelling, the accuracy of discharge estimation primarily depends on the uncertainties of the model structure, parametrization, the quality of hydro-meteorological forcing and the spatial and temporal availability of the observed data. During the last years, improvements in the predictions of hydrological variables were performed at both temporal and spatial scales, as in the initiatives of the Prediction in Ungauged Climates (PUC) (Merz et al. 2010) and Prediction in Ungauged Basins (PUB) (Hrachowitz et al. 2013). While the former tackles the non-stationary nature of the hydrological series, especially in systems with relevant alterations of physical characteristics (Wagener et al. 2010), PUB is characterized by the goal of understanding and predicting the hydrologic response in ungauged or poorly gauged basins. The overall aim of these initiatives is to develop and implement scientific programs to accomplish more accurate predictions (Sivapalan et al. 2003).

The estimation of hydrological responses becomes more uncertain in the context of the lack of hydrometric data, which is the current situation in many developing countries (Sivapalan et al. 2003). Simplified methodologies, based on in-situ discharge, such as regionalization, have usually been employed due to their simplicity and satisfactory application to many cases (Smakhtin 2001, Ouarda et al. 2008, Castiglioni et al. 2009, Novaes et al. 2009, Pruski et al. 2013). Regionalization methods transfer information from gauging stations to the target location usually based on the spatial proximity or physical similarities. These methods range from more common ones, such as factoring for adjacent catchment size (McCuen and Levy 2000) to more complex ones, such as artificial intelligence techniques (Beskow et al. 2016), and regionalization of weights in ensemble, multi model-based estimations (Arsenault and Brissette 2015, Razavi and Coulibaly 2016). These studies concluded that the recognition of the adequate donor station is essential for the final estimation quality. In addition, some uncertainties still remain neglected on model prediction applications (Durrans and Kirby 2004, Asadi et al. 2016, Arsenault and Brissette 2016), and must be considered in future approaches, to quantify the level of confidence in the prediction results (Reichert and Borsuk 2005, Kapangaziwiri et al. 2012).

Data Assimilation (DA) techniques have been developed to improve model predictability, especially in meteorology and oceanography, based on the fact that mathematical models do not offer a completely satisfactory description of nature (Goniadzki et al. 2011). DA may be defined as a set of techniques that combines current or past observations of the system state, with a model estimate to

generate an analysis, which is the "best unbiased" approximation of the current state of the system, taking into account its uncertainties (Reichle 2008). However, it was only in the 1990s that these techniques began to be more frequently applied in hydrology (Evensen 1994, McLaughlin 1995) and facing significant improvements since the 2000s (Walker and Houser 2005). The first applications of DA in hydrology was related to the soil moisture content (Jackson et al. 1981, Bernard et al. 1981, Prevot et al. 1984, Bruckler and Witono 1989, Ottle and Vidal-Madjar 1994, Georgakakos and Baumer 1996), while other variables such as surface temperature, snow, terrestrial water storage and streamflow, were used in more recent applications (Walker and Houser 2005). DA is primarly divided into two numerical methods, variational and sequential. The former is based on the optimal control theory (Todorov 2006); it consists of the use of control variables that allow minimizing an objective function subject to certain restrictions. This method is usually effective, although it is difficult to implement in complex models. Besides, sequential methods use a necessary probabilistic structure to quantify the uncertainties associated with model results; and they provide estimates of the entire state of the system in a sequential manner through the propagation of information (Bertino et al. 2003). Among the most popular sequential DA methods are those based on Kalman filter (KF) (Qin et al. 2008). The KF technique assumes the linearity of the system, since it explicitly considers the dynamic nature of the model and the errors of observations (Gelb 1974), which evolve over time, to produce statistically optimal estimates for linear systems.

Modifications to the KF, such as the extended Kalman filter (EKF), is presented as a KF solution for the case of dynamic systems with nonlinear behaviour, using a statistical approach of linearization at each step (Bertino et al. 2003, Sun et al. 2004). However, the EKF method has some imperfections in terms of the linear approximation for systems with highly nonlinear behaviours. Therefore, Evensen (1994) proposed a new assimilation scheme that considered the set of possible trajectories of the model's state variables for an application in oceanography, which derived from this method, called ensemble Kalman filter (EnKF). In this way, using the EnKF method, inaccuracies can be avoided due to the linearization process that affects the performance of DA, due to the strong, non-linear characteristics of a given system. EnKF-like methods, such as particle filter (PF), also use a Monte-Carlo sequential simulation (e.g. Vrugt et al., 2013; Abbaszadeh et al., 2018). Although the PF is more flexible in some assumptions such as the fact that it is not necessary that the probability distributions are Gaussian, as is with the EnKF (Hreinsson 2008). Weerts and El Serafy (2006) demonstrated in a comparative study that PF is more sensitive to the choice of model errors and observations, while EnKF is more suitable for flow prediction. Other variations of EnKF such as the retrospective ensemble Kalman filter (REnKF) or the ensemble square root filter (EnSRF), have also been used in the area of water resources. The first instance of this was idealized from the criterion of the temporal association between the state variables. Therefore, the assimilation scheme allows updating the present and past conditions (Pauwels and De

Lannoy 2006). Besides, the EnSRF method avoids the disturbance of the observations in the AD scheme (Whitaker and Hamill 2002, Clark et al. 2008).

Although not as simple as the traditional methods, and carrying higher computational costs, the extent to which DA methods may help hydrological modelling and water resources management, in the context of PUB, is still an open question. Since there are no consolidated directives on how to implement such techniques in most hydrological studies, it is necessary to develop schemes that are capable to effectively assimilate the relevant information contained in the observations (Liu and Gupta 2007); for instance, to assess their performance due to the different DA parameter setups, focusing on discharge estimates for water management. Many efforts have been described in the literature for incorporating DA techniques into rainfall-runoff hydrological models (Beck 1987, Entekhabi et al. 1994, Vrugt et al. 2005, Lee et al. 2012, Abbaris et al. 2014, Maxwell et al. 2018). Among the diverse DA techniques described above, the ensemble Kalman filter (EnKF) (Evensen 2003) has become the most popular for assimilation into land surface and large scale hydrological models (Clark et al. 2006, Vrugt and Robinson 2007, Neal et al. 2007, Paiva et al. 2013b, Revilla-Romero et al. 2016). Additionally, some assessments were made in the literature to evaluate the impact of the different EnKF's parameters, relevant in hydrological models, such as the ensemble size, localization and model structure (Paiva et al. 2013b, Thiboult and Anctil 2015, Rasmussen et al. 2015, Wongchuig et al. 2019). Moreover, water resources management aims to plan, develop and distribute water often through the use of adequate estimate reference discharges. These estimates are relevant for studies related to, for instance, the maintenance of a minimal discharge value, in order to achieve a dilution capacity of water courses (Tallaksen and Van Lanen 2004). Uncertainties in discharge estimation may lead to wrong water availability assessments, and granting water permits that are not compatible with the real capacity of the basin (Novaes 2009, Pruski et al. 2013). Although already feasible, studies that incorporate DA techniques into rainfall-runoff models to assist water resources management, such as to estimate reference discharges, are still scarce (Young 2002, Moradkhani et al. 2005a, b). The novelty of this research is the adaptation and deep assessment of a suitable configuration, based on a sensibility of DA parameters within a hydrologic-hydrodynamic model, in the context of reference discharge estimation in under-monitored basins which still remains poorly understood. Therefore, the goals of this study are:

i. To adapt a DA scheme within a large scale hydrological hydrodynamic model;

ii. To evaluate an optimal setup for DA parameters;

iii. To assess its performance to estimate reference discharges and discharge time series in a case study basin.

Section 3.2 describes the analysed study area, the hydrological model consisting of a semi-distributed hydrological-hydrodynamic model and the DA scheme and equations that are adapted. Lastly, the experimental setup is performed. Section 3.3 shows and assesses the main results of the

experiments, regarding the sensitivity analysis of different parametrizations of the DA. The discharge series and the reference discharge estimations are also evaluated. Section 3.4 widely summarizes and discusses the main issues of this research. Finally, the conclusions and perspectives are outlined in section 3.5.

3.2. Data and methods

3.2.1. Study area and hydro-meteorological data

The Taquari Antas basin (Southern Brazil) was selected as the study area, due to its relatively satisfactory availability of observed discharge data. The basin has a drainage area of around 26,000 km², located in the north-eastern portion of Rio Grande do Sul state (Figure 1), surrounded by the Uruguay, Jacuí and Caí river basins. Its outlet is in the Jacuí river; the mainstem extension is approximately 530 km (Larentis et al. 2008). The altitude ranges between ~1,000 to ~5 mamsl, and the precipitation regime shows a regular distribution throughout the year, with annual values around 1,600 mm. Finally, the time of concentration for the main sub-basins ranges from two to three days.

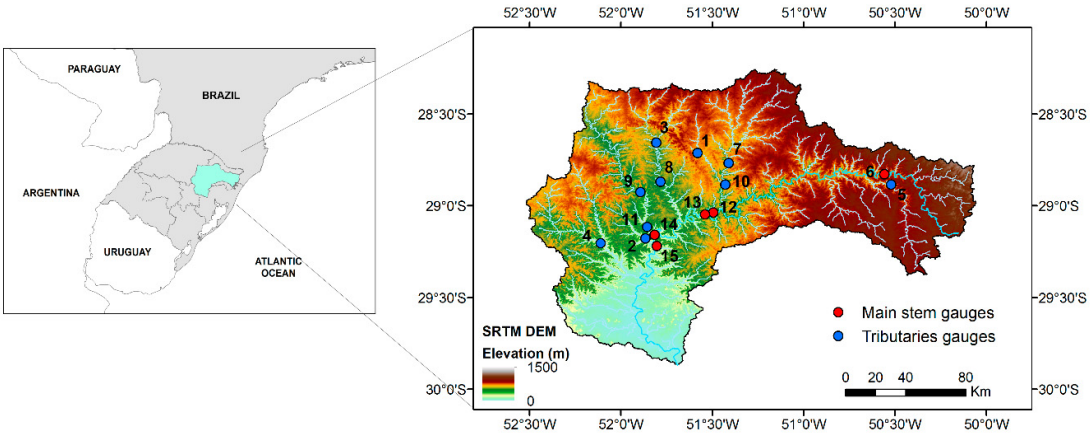

Figure 1. Location of the Taquari-Antas river basin and discharge gauges used in this study.

It is located in the "Vale do Taquari", in the downstream flat areas of the basin, where many flooding events occur due to the geomorphological characteristics. For instance, almost every year the alert level is registered at least once, by the Lajeado city Civil Defense, indicating that the Taquari river is about to overflow (Fadel 2015). Besides the usual flooding, the Taquari-Antas watershed plan detected difficulties in demand support, in the months with the highest scarcity situations, usually occurring from

November to April (SEMA 2014). The most difficult period occurs in January, when demands reach almost 80% of the Q95 reference flow (i.e. 95% exceedance). The sub-basins of Baixo Taquari and Taquari-Mirim, located in the lower Taquari-Antas basin, present higher water stress, mainly due to high demands for rice and cattle use. In the Lajeado Grande and Alto Rio das Antas regions (middle Taquari-Antas basin), water use for vegetables and livestock results in an intermediate water stress, while Alto Rio Carreiro sub-basin presents an unfavorable situation for being in the headwaters region (SEMA 2014). Along the basin, 72% of the surface water demand is used for agriculture, 23% for human consumption and five percent for industrial use.

In terms of data availability, 15 daily discharge stations from Brazilian National Water Agency (ANA) present consisting records for more than 15 years, where five of them are located in the Taquari-Antas main stem, and 10 gauges at the Forqueta, Guaporé, Carreiro, Da Prata, Turvo and Tainhas sub-basins (Figure 1). The region is also monitored by 30 rainfall stations and seven meteorological stations, provided by the ANA and the Brazilian National Institute of Meteorology (INMET). Table 1 summarizes the list of discharge stations used in this study. The data was collected of the period 1995-2010, and there was more than 75% data availability for 80% of the discharge gauges.

Table 1. Summary Brazilian Water Agency (ANA) of in situ discharge gauges

Gauge	ANA Code	Name	Longitude (°)	Latitude (°)	Catchment area (km^2)	River
1	86420000	Ponte do Prata	-51.6106	-28.6792	317	Prata
2	86700000	Ponte Jacaré	-51.9222	-29.1889	436	Arroio Jacaré
3	86480000	Passo Migliavaca	-51.8575	-28.6172	1,330	Carreiro
4	86745000	Passo Do Coimbra	-52.1897	-29.2161	791	Forqueta
5	86160000	Passo Tainhas	-50.4539	-28.8667	1,120	Tainhas
6	86100000	Passo do Gabriel	-50.4939	-28.8056	1,820	Das Antas
7	86410000	Passo Barra do Guaiaveira	-51.4244	-28.7394	2,820	Turvo
8	86500000	Passo Carreiro	-51.8333	-28.8508	1,820	Carreiro
9	86560000	Linha Colombo	-51.9531	-28.9117	2,030	Guaporé
10	86440000	Passo do Prata	-51.4456	-28.8675	3,600	Prata
11	86580000	Santa Lúcia	-51.9119	-29.1197	2,470	Guaporé
12	86460000	Monte Claro	-51.5167	-29.0333	12,100	Das Antas
13	86470000	Ponte do Rio das Antas	-51.5669	-29.0450	12,500	Das Antas
14	86510000	Muçum	-51.8681	-29.1664	16,000	Taquari
15	86720000	Encantado	-51.8542	-29.2342	19,100	Taquari

3.2.2. MGB hydrological model

The MGB (Portuguese acronym for Modelo de Grandes bacias) is a semi-distributed, rainfall-runoff model for continuous simulation of large-scale basins (Collischonn et al. 2007, Pontes et al. 2017). In the model, the basin is discretized into unit-catchments and further into hydrological response

units (HRUs). HRUs are defined by the combination of soil type and land cover classes, where vertical hydrological processes, such as canopy interception, evapotranspiration and soil infiltration, are considered to compute the generation of surface, subsurface and groundwater runoffs. Precipitation is intercepted by the vegetation up to a maximum storage capacity that depends on the estimated leaf area index for each HRU, hence evaporation from soil, vegetation and intercepted water is calculated using the Penman-Monteith equations (Shuttleworth 1993, Wigmosta et al. 1994). The main hydrological soil parameters of the model for surface runoff generation are related to the saturation excess concept, based on the variable contribution area of the ARNO model (Todini 1996). The generated runoff of each unit-catchments is routed through the linear reservoirs concept, toward the main drainage network. Henceforth, the streamflow routing within the river network is performed using a hydrodynamic module, based on the inertial approximation (Bates et al. 2010, Pontes et al. 2017). For the calibrated model, which was used as the background in the DA scheme, nine HRU and only one set of parameters (e.g. albedo, leaf area index, vegetation height, surface resistance, maximum soil water storage, among others) were used for the whole Taquari-Antas basin. This is based on the model setup from Siqueira et al. (2016); most calibration and validation processes were assessed during a period after 2010, using five gauge stations located at the main steam of the river (Passo Tainhas, Passo Barra do Guaiaveira, Santa Lúcia, Muçum and Encantado). Eleven state variables (e.g. soil water storage (W), surface reservoir volume (Vsup), sub-surface reservoir volume (Vint), groundwater reservoir volume (Vbas), unit-catchment outflow discharge (Q), among others) were considered to be updated into the EnKF scheme. Among them, only discharge was exported for further analysis, however, all state variables can be exported. The main forcings of the MGB are precipitation and climate (wind speed, relative air humidity, air temperature, insolation and atmospheric pressure). Precipitation data was obtained from ANA (period from 1995 to 2010), and long-term climatological averages from the INMET for evapotranspiration computation. The discharge dataset for assimilation and validation was also obtained from the ANA database. The hydrological model time step for this setup was based on daily information.

For more detailed information about the structure of the MGB model, see Collischonn et al. (2007) and Pontes et al. (2017).

3.2.3. Adaptation of the Ensemble Kalman Filter scheme

Compared to other sequential DA schemes, such as PF, KF or the EKF, the EnKF method is used since it enables an adequate implementation for complex systems, such as hydrological models that are dominated by strong non-linear processes (Sun et al., 2016). In the EnKF scheme, the model uncertainty is defined as a function of the spread of the model state ensemble, using a Monte Carlo simulation.

The adaptation of the EnKF DA scheme in MGB was based on the algorithm developed by Evensen (2003) and following the methodology proposed by Paiva et al. (2013b). The main goal of the

EnKF is the optimization of the model state variables, based on the model and observation errors. According to the EnKF assumptions, these errors are unbiased and follow a Gaussian distribution.

For the model, the following relation holds:

$$\boldsymbol{x}_{k+1} = M(x_k, u_k, \theta) + \boldsymbol{q}_k \quad (1)$$

where, the vector x represents model state variables, u is the model forcing, θ are the model parameters; M the model function that propagates state variables at time k with those at time (k+1); and q_k the model errors.

The observation equation is as follows:

$$\boldsymbol{y_k} = \mathbf{H}(x_k) + \boldsymbol{\varepsilon}_k \quad (2)$$

where y is a vector with observations at time k (here, observed discharge), H is the observation operator that relates model state variables x with the corresponding observation y, and ε is the observation error.

By applying the BLUE (Best Linear Unbiased Estimator) formula, the optimal correction of model states, based on the observations and all uncertainties, is as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}_e\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right) \quad (3)$$

$$\mathbf{K}_e = \mathbf{P}_e^f \mathbf{H}^T \left[\mathbf{H}\mathbf{P}_e^f \mathbf{H}^T + \mathbf{R}_e\right]^{-1} \quad (4)$$

where $\mathbf{x}^a$ and $\mathbf{x}^f$ are the analysis and forecasted (background) model state variable respectively, $\mathbf{K}_e$ the Kalman gain, $\mathbf{P}_e^f$ and $\mathbf{R}_e$are the error covariance matrices of the model and of the observation respectively.

Eq. 5 represents the matrix that contains the set of state variables of the $\mathbf{x}^f$ model in the forecasting step, where nens is the number of members of the set:

$$\mathbf{x}^f = \left(\mathbf{x}_1^f, \mathbf{x}_2^f, \mathbf{x}_3^f, \ldots \mathbf{x}_{nens}^f\right) \quad (5)$$

where, $\mathbf{x}_1^f, \mathbf{x}_2^f, \mathbf{x}_3^f, \ldots \mathrm{x}_{nens}^f$ represent the vectors that contain the model state variables for each member of the set.

In the EnKF formulation, it is assumed that the mean of the sets of state variables represents the true value, thus, it is mathematically denoted as follows:

$$\overline{\mathbf{x}^f} = \frac{\sum_{i=1}^{nens} \mathbf{x}_i^f}{nens} \quad (6)$$

where, i sub index representing each ensemble member.

However, the error associated with each i member of the set $\left(\mathbf{x}'^f_i\right)$ and the array of set anomalies ($\mathbf{A}'^f$) are defined as follows:

$$q = \mathbf{x'}_i^f = \mathbf{x}_i^f - \overline{\mathbf{x}^f} \qquad \boldsymbol{A'}^f = \left(\boldsymbol{x'}_1^f, \boldsymbol{x'}_2^f, \boldsymbol{x'}_3^f, \ldots \boldsymbol{x'}_{nens}^f\right) \tag{7}$$

In the EnKF method, the covariance matrix of errors is generated using a set of members obtained by perturbing the model forcing, state variables or parameters. This perturbation follows a spatial and temporal correlation of random errors:

$$\mathbf{P}^f \cong \mathbf{P}_e^f = \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T} \tag{8}$$

where, $\mathbf{P}_e^f$ is the covariance matrix of errors of the background (forecast); sub index *e* is related to the ensemble generation.

Several studies have been based on perturbation of only precipitation, assuming that this is the main source of error in hydrological models (Biancamaria et al. 2011, Liu et al. 2012, Paiva et al. 2013b). However, the model responses to different perturbation scenarios were assessed here. The EnKF DA scheme was evaluated in this study through the perturbation of (i) precipitation and (ii) MGB state variables. The adopted perturbation follows a log-normal distribution of errors, as proposed by Nijssen and Lettenmaier (2004):

$$V' = \frac{1+\beta}{\sqrt{E^2+1}} exp\left(\sqrt{ln[E^2+1]}S\right)V \tag{9}$$

where, V' is the perturbed variable (precipitation or model state variables), V is the non-perturbed variable, E is the variable relative error (%), β is the relative bias (here adopted as zero), and S is a random variable that follows a time evolution of model error. According to Evensen (2003):

$$S_k = \alpha S_{k-1} + \sqrt{1-\alpha^2} w_{k-1} \tag{10}$$

$$\alpha = 1 - \frac{\Delta t}{\tau_t} \tag{11}$$

where, the coefficient $\alpha \in [0\ 1\rangle$ indicates the influence of temporal correlation on model errors ($\alpha = 0$ indicates a sequence of with noise, while $\alpha = 1$ removes the stochastic forcing and represents the errors of the model with a random component that is constant over time); and τ_t is the temporal decorrelation length (time units).

Besides, w is a stochastic term that was estimated using a pseudo random field based on the two dimensional Fourier transform with mean 0 and variance 1 and a covariance function depending exponentially on a distance τ_x, which represents the length of spatial decorrelation. The parameter spatial decorrelation length was established in geographic degree units, and S was initialized with a value of zero; the consistency of the resulting simulated discharges has been already assessed in the study of Paiva et al. (2013b), through the use of different values for spatial decorrelation lengths. For detailed information, please see Evensen (2003).

State variables were also perturbed using a log-normal model considering independent perturbation, with the following being considered in the study: i) Soil water storage (W); ii) Surface linear reservoir volume (V_{SUP}); iii) Sub-surface linear reservoir volume (V_{INT}); and iv) Groundwater linear reservoir volume (V_{BAS}). For the perturbation of precipitation, precipitation relative error parameter (E), spatial decorrelation (τ_x) and temporal decorrelation (τ_t) of 50%, 1.50° and 10 days, respectively, were used as reference values. These parameters were applied by Paiva et al. (2013b), in the MGB model implementation in the Amazon basin. Perturbation scenarios for both precipitation and state variables were assessed daily, following the same time step as of the model.

Finally, for the observed discharge, an error of 10% was adopted, and the covariance matrix of the observation errors ($\mathbf{R}_e$) was estimated, following the Paiva et al. (2013b) and Clark et al. (2008) approaches:

$$\mathbf{R}_e = (\varepsilon_{obs} \cdot var_{obs})^2 \quad (12)$$

where, ε_{obs} represents the error in the observation (e.g. discharge) that can be known a priori, for example, it can be estimated by the uncertainties associated with the discharge estimation using the rating curve; var_{obs} indicates the value of the observed variable.

3.2.4. Experimental setups and evaluation metrics

The following tests were carried out to evaluate the DA scheme and discharge estimation accuracy at ungauged sites, and a flowchart is exhibited in Figure 2:

i. Sensitivity analysis of the DA scheme to (a) ensemble size, (b) discharge and logarithm of discharge as assimilation variables, (c) parameters related to precipitation errors (precipitation error and spatial and temporal decorrelation lengths), and (d) parameters related to state variables errors. For experiments (a) and (b), we adopted only the perturbation of precipitation and reference values of parameters perturbation, described in Section 2.3.

ii. Evaluation of discharge estimates by the DA method using different combinations of precipitation and state variables perturbations, in order to identify which combination leads to the best assimilation results; therefore, these tests assess the best estimation of discharge.

iii. Estimation of the reference discharges Q_{90} (90% exceedance discharge), Q_{LTA} (long-term average discharge) and $Q_{MAX\ RP=5}$ years (maximum discharge with five years of return period estimated with Gumble distribution), using the best scenario obtained in test ii and in comparison to the regionalization method, based on the potential relation between the drainage area and the reference discharge.

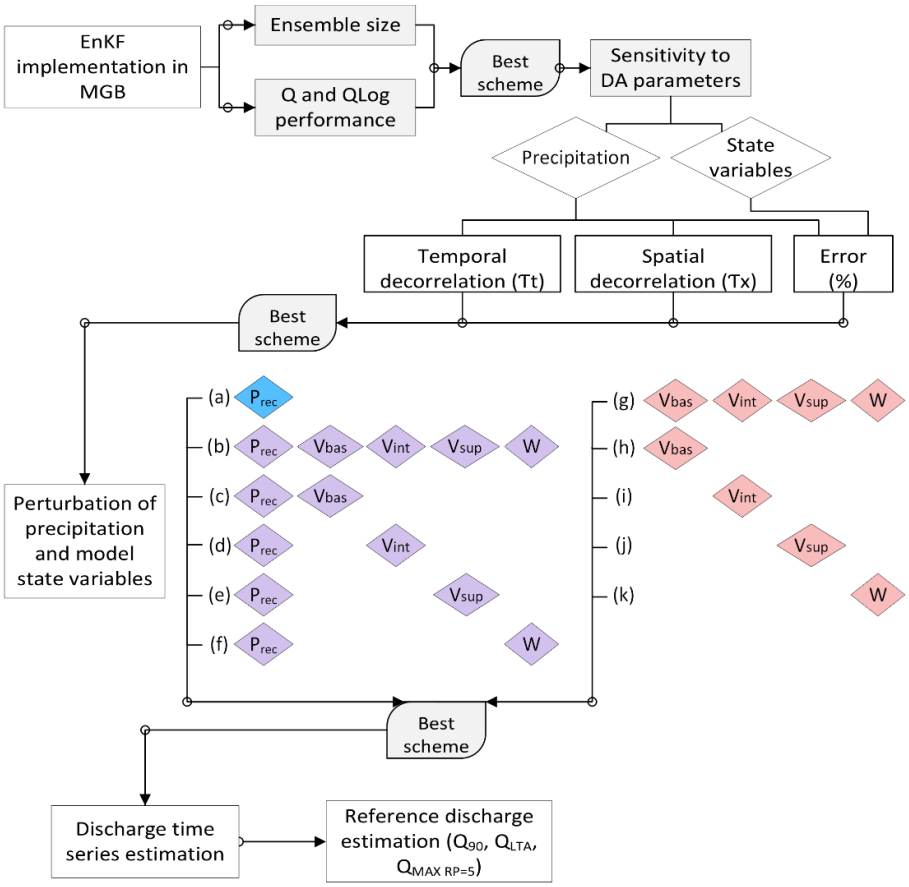

Figure 2. Flowchart describing the methodology and tests carried out in this study. The perturbed forcing and the state variables combinations are colored.

To assess the capacity for estimating discharge values in poorly monitored regions, the following tests were carried out using two sets of discharge gauges: (a) assimilation of discharge from five gauges in the Taquari-Antas main stem and validation for 10 tributary stations, and (b) assimilation of discharge from 10 tributary stations and validation in the five main stem gauges. Figure 1 shows the location of the gauges. A good assessement of adequate locations of gauges can be explored by the use of this scheme.

Evaluation of the results were based on the visual comparison of the hydrographs resulting from open-loop (i.e. deterministic simulation), EnKF simulation and observations, and on the following efficiency metrics: relative difference of root-mean-square error between open-loop simulation and the EnKF simulation in relation to observations (Δrmse), the Nash-Sutcliffe of discharges (NSE) and logarithm of discharges (NSE-Log), which focus on low flows. For the Δrmse metric (Eq. 10), our reference value is the observation (Q_{obs}), therefore, the $RMSE_{sim}$ can be estimated regarding the simulation (Q_{sim}) for both the open-loop simulation ($RMSE_{open-loop}$) and for the DA method ($RMSE_{assimilation}$).

$$RMSE_{sim} = \sqrt{\frac{1}{n}\sum_{i=1}^{nt}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2} \qquad \Delta RMSE = 100 * \frac{RMSE_{assimilation} - RMSE_{open-loop}}{RMSE_{open-loop}} \qquad (13)$$

$$NSE = 1 - \frac{\sum_{i=1}^{nt}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\sum_{i=1}^{nt}(Q_{obs}(t) - \overline{Q_{obs}})^2} \qquad NSE_{log} = 1 - \frac{\sum_{i=1}^{nt}(\log(Q_{obs}(t)) - \log(Q_{sim}(t)))^2}{\sum_{i=1}^{nt}\left(\log(Q_{obs}(t)) - \overline{\log(Q_{obs})}\right)^2} \qquad (14)$$

These evaluation metrics were assessed for both the open-loop simulation and the ensemble-mean of the EnKF simulations during a period when observed discharge was available in validation gauges. The results were averaged for the gauges. In order to compare the EnKF method with a simple regionalization method, as proposed by Tucci (2002) which is based on a power-law relation between the drainage area and reference discharge for the assimilation gauges (Eq. 15) for each validated gauge, the reference discharges, Q_{90}, Q_{LTA} and $Q_{MAX\ RP=5}$, were compared for simulated (DA and open-loop), regionalized and observed values.

$$Q_{ref} = aA^b \qquad (15)$$

where, A stands for the drainage area of the gauge station, Q_{ref} represents the discharge reference value for the regionalized site, a and b represent the coefficients of the power-law relation based on the set of gauges used for regionalization (which are the same ones used when the assimilation is performed).

The regionalization method for validated gauges was obtained using the same period as the assimilation runs, within the period of 1st January 1995 to 31st December 2010. A warm-up period of ~30 days was considered, beginning at the end of 1994; after this period, the DA updating process is initiated. The statistics metrics were considered after 1st March 1995 to avoid the influence of the initial conditions.

3.3. Results

3.3.1. Sensitivity analysis of the DA scheme

Sensitivity tests of the DA scheme were performed in order to estimate the best values for the following parameters: ensemble size for assimilation, precipitation error, spatial and temporal decorrelation, and state variables error.

The first sensitivity analysis was assessed to obtain the optimal ensemble size when only the precipitation is perturbed. Figure 3 shows the relation between Δrmse and ensemble size for the data assimilation of discharge and the logarithm of discharge, for the assimilation of observations at gauges located in, (i) the main stem and in the (ii) tributaries. Different values of performances (Δrmse) were

plotted for assimilation gauges (blue lines) and for validation gauges (red lines). For instance, for the first row and first column, the blue line represents the performance of discharge estimates by DA in gauges at the main stem, consequently, the red line stands for the performance of validated gauges at the tributaries.

Errors were generally reduced by increasing the ensemble size, which stabilized with around 100 to 200 members, and 100 members was adopted as the optimal ensemble size. A particular behaviour was noticed around an ensemble size of 10 members, which we assumed to be a spurious result due to model uncertainty or due to the stochastic behaviour of the system.

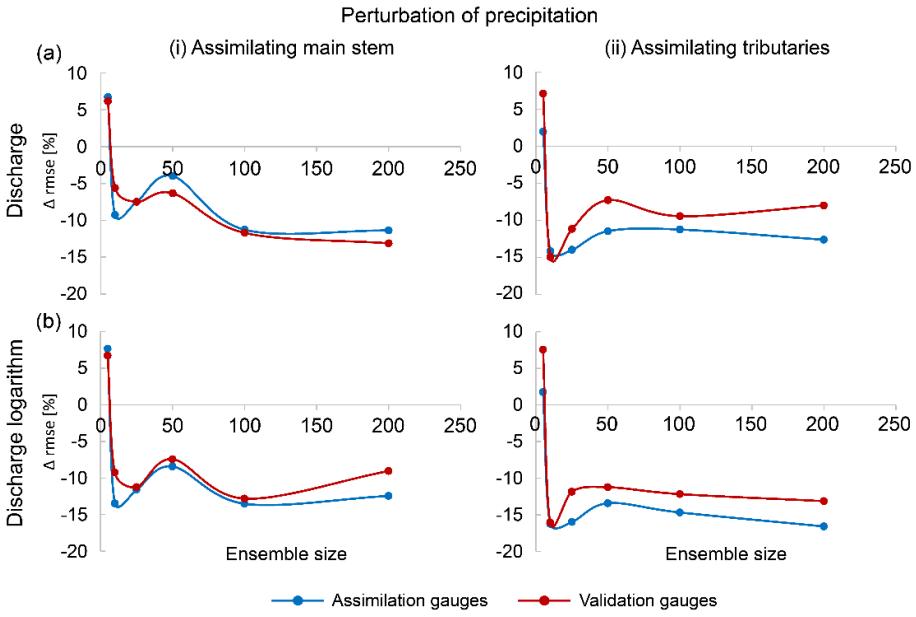

Figure 3. Sensitivity analysis of DA scheme, assessing Δrmse (%) as a function of the number of ensemble members for two scenarios, assimilating observed (a) discharge and (b) discharge logarithm in i) main stem and ii) tributaries.

To assess the impact of the perturbation of the precipitation on the ensemble simulation, Figure 4 shows the uncertainty of the model using different ensemble sizes, as compared to the assumed uncertainty of the discharge observations. The median of the uncertainty rises, while with an increase in the ensemble size, the representativeness of the ensemble spread also increases and get close to the assumed relative error of precipitation used for the perturbation (50%).

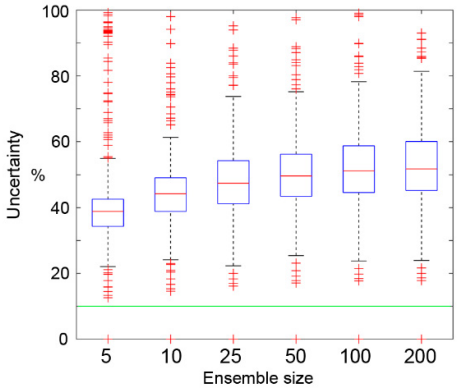

Figure 4. Boxplot of the spread of the ensemble simulation as a function of the ensemble size and assumed uncertainty of discharge observations (green line).

The second analysis (Table 2) was evaluated to quantify the advantages of the assimilation of discharge or discharge logarithm in validated gauges. Generally, the NSE and NSE-log coefficients showed similar improvements, considering the assimilation of gauges in the main stem and tributaries. On the other hand, regarding error reductions, better values resulted in simulations when assimilating the logarithm of discharge. It follows that an optimal scheme of 100 members and a discharge logarithm be considered for the following tests.

Table 2. Average of efficiency metrics NSE, NSE-Log and Δrmse for simulation discharge at validated gauges, while assimilating discharge and discharge logarithm by the gauges at main stem and tributaries.

		NSE	NSE-Log	Δrmse (%)
Open-loop		0.293	0.553	-
EnKF assimilating Main stem	**Q**	0.405	0.656	-11.706
	Qlog	0.422	0.605	**-12.811**
Open-loop		0.296	0.658	-
EnKF assimilating Tributaries	**Q**	0.408	0.779	-9.423
	Qlog	0.435	0.783	**-12.184**

* Q and Qlog represent discharge and discharge logarithm

The third analysis was performed to assess the sensitivity of the DA perturbation parameters, by testing different values of spatial decorrelation (τ_x), temporal decorrelation (τ_t) and error (E) of precipitation (Figure 5). Previous studies showed that precipitation relative error (E) has a greater impact than τ_x and τ_t (Paiva et al. 2013b). For spatial decorrelation, Δrmse improved as this value decreased. Therefore, a value of 0.0625° was adopted as the grid perturbation resolution, due to i) the exhibition of the greatest performance; ii) being approximately equal to the reach length of each unit-catchment

(~10km); and iii) the fact that performance tends to be worse for τ_x values greater than two degrees. Although the model was not sensitive to the temporal decorrelation, we chose 20 days as a relative optimal value, because this value seems to be stabilized for the assimilation of the main stem. For the precipitation error (E), Δrmse improved as E increased and stabilized at the 20% value.

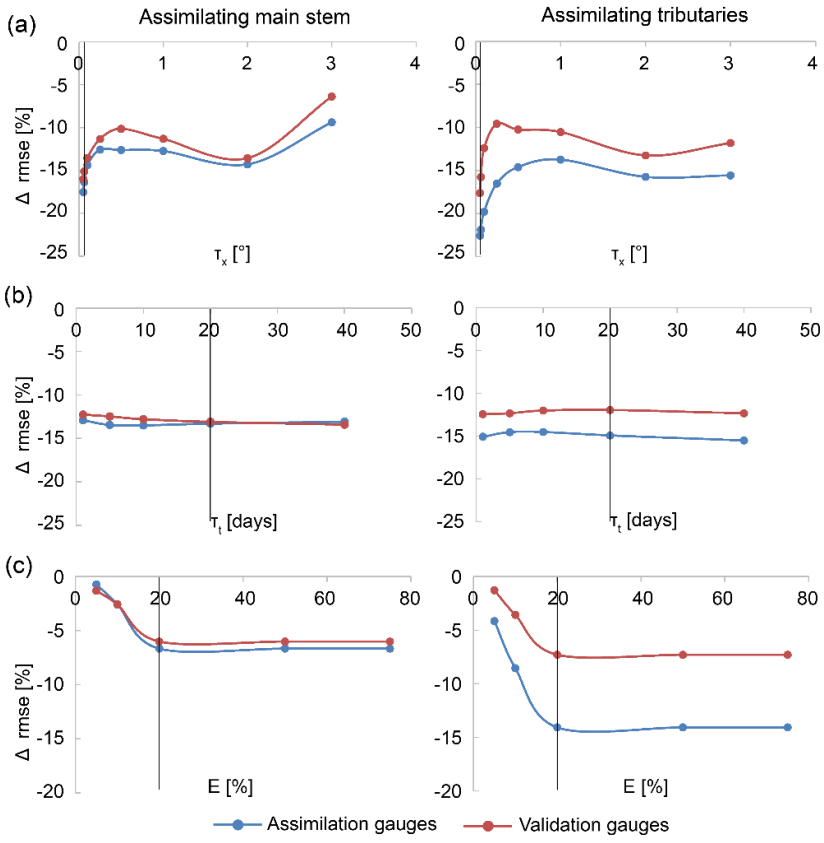

Figure 5. Sensitivity of the relative difference of root-mean-square error (Δrmse) to precipitation perturbation parameters of (a) spatial decorrelation, (b) temporal decorrelation and (c) variable error.

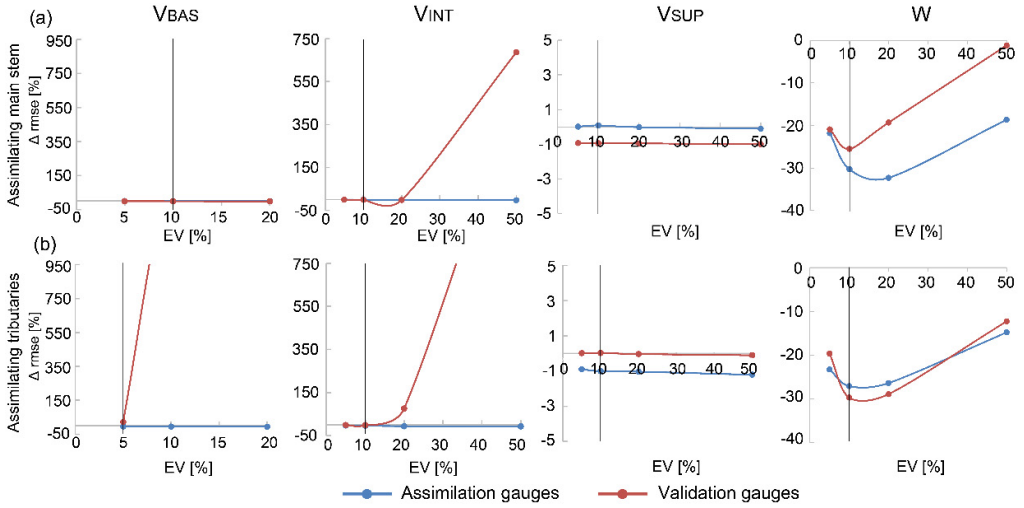

Figure 6. Sensitivity analysis for the variation of relative difference of root-mean-square error (Δrmse) to state variables (V_{BAS}, V_{INT}, V_{SUP} and W) of perturbation error, for assimilation of discharge observed by (a) main stem gauges and (b) tributaries gauges.

A further sensitivity test of the DA scheme was carried out by separately perturbing MGB state variables, assessing only the perturbation parameter of the relative error value. For state variables of perturbation (Figure 6), particular trends were observed for each of them in terms of discharge estimates by the DA in the main stem and tributaries. Considering the sensitivity of the model response as a function of the relative error parameter for different state variables, W achieved an optimal value of RMSE, where a relative error of 10% was assessed. Besides, variations on the relative error of the V_{SUP} variable induces less sensitivity to the model response, showing improvements around zero for RMSE in the validation gauges. Larger differences between assimilation and validation gauges were found mainly for the V_{INT} and V_{BAS} variables, explained by the model structure. These variables have a slower response than the V_{SUP}, thus, the assimilation at the validation gauges contain a temporal lag, which does not allow for the updating of the state variables at the time of the observed assimilation.

3.3.2. Assessment of the impact of different combined perturbations of precipitation and state variables

Different scenarios of DA were evaluated by combining the perturbation of model state variables and precipitation (Prec), considering optimum parameters obtained in the previous section: precipitation error (E) = 20%, error in state variables (EV) = 10%, spatial decorrelation (τ_x) = 0.0625° and temporal decorrelation (τ_t) = 20 days. The following 11 tests were carried out by perturbing: (a) Prec; (b) Prec, W, V_{BAS}, V_{INT} and V_{SUP}; (c) Prec and V_{BAS}; (d) Prec and V_{INT}; (e) Prec and V_{SUP}; (f) Prec and W; (g) W, V_{BAS}, V_{INT} and V_{SUP}; (h) V_{BAS}; (i) V_{INT}; (j) V_{SUP}; and (k) W. These tests are represented in Figure 2 using the colours blue (precipitation only), purple (precipitation and state variables) and red (state variables).

Figure 7 presents the results for the NSE and Δrmse metrics for the 11 tests, considering the average of validation gauges, using the two sets of gauges described in section 2.4. Scenarios (b), (f), (g) and (k) led to the best NSE values (c.a. 0.6 when assimilating main stem gauges), with larger improvements, if compared to the open-loop simulation (NSE = 0.29). It was noticed that the scenarios that achieved better performances (b, f, g and k) were related to the perturbation of soil water storage, and even the joint perturbation of precipitation and soil water storage, which may be explained by the fact that these two have a large impact in terms of runoff generation. This result is interesting and shows that water storage is a large source of errors when regarding the structure of the MGB model, likely greater than precipitation. Worst results were obtained when individual perturbation of state variables V_{BAS}, V_{INT} and V_{SUP} were performed, reaching values comparable to the open-loop simulation. Results for Δrmse metrics show the same behaviour as NSE (Figure 7b), and maximum decreased errors were around 30%. A particular behaviour was noticed for the (h) experiment (V_{BAS} variable), when examining the discharge estimates by the DA in tributaries the results for which are not showed in Figure 7 as it exceeded the graphic scale for worst results. As mentioned in the previous section, this behaviour is likely due to, according to the model structure, V_{BAS} is routed with a temporal lag, in comparison to V_{SUP} and V_{INT}; therefore, the updating of the state variables is temporally shifted. However, this could be avoided using another technique such as REnKF.

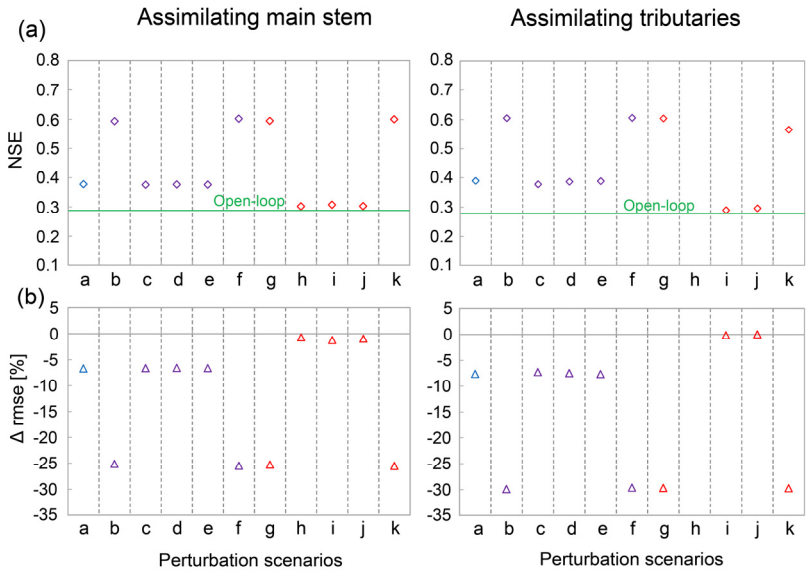

Figure 7. (a) Average NSE coefficient and (b) average Δrmse for different sets of perturbations (see text for description of each test). Tests are referred to as: a) precipitation only (blue), b-f) joint perturbation of precipitation and state variables (purple), and g-k) only model state variables (red).

3.3.3. Estimation of discharge time series with EnKF

In this section, a set of tests were carried out using the two sets of gauges (assimilating the logarithm of discharge, of the main stem or of the tributaries gauges) described in Section 2.4, adopting the assimilation scenario (b), evaluated in Section 3.2, with the perturbation of precipitation and all state variables using 100 members; E = 20%, τ_t = 20 days, τ_x = 0.0625° and EV = 10%. Table 3 shows the mean values of the NSE and Δrmse for the simulation when assimilating the logarithm of discharge. Results showed that assimilation of main stem gauges led to better performance, although both assimilation scenarios yielded an NSE between ~0.5 and ~0.6, which are better results than those obtained with assimilation in the sole context of precipitation perturbation (Table 2 of previous section).

Table 3. Average of Nash-Suttcliffe (ENS) and relative difference of root-mean-square error (Δrmse) metrics, assimilating discharge logarithm in main stem and tributaries gauges.

		Assimilating main stem gauges		Assimilating tributaries gauges	
		NSE	**Δrmse (%)**	**NSE**	**Δrmse (%)**
Assimilation Gauges	**Open-loop**	0.296	-	0.293	-
	EnKF	0.638	-30.59	0.644	-28.12
Validation gauges	**Open-loop**	0.293	-	0.296	-
	EnKF	0.593	-25.09	0.605	-29.91

The spatial distribution of error metrics is presented in Figure 8 for Δrmse, where assimilation of tributary gauges (Figure 8b) showed values below -20% (dark and light green squares), for four out of five validation gauges, while the assimilation of main stem gauges (Figure 8a) resulted in values below -20%, for five out of ten validation gauges. Isolated, headwater gauges in the eastern region of the basin obtained the worst efficiency, due to two main reasons; i) their larger number of failures of discharge records on the assimilation gauges and/or ii) due to spurious correlations, which have possibly been generated in the matrix of covariance of errors of the model (Zhu et al. 2011), because of the distant location between the two points (assimilation and validation sites) or the limited ensemble size. This last point can be assessed through the implementation of a local impact within the DA scheme, such as the local ensemble Kalman filter (LEnKF) technique.

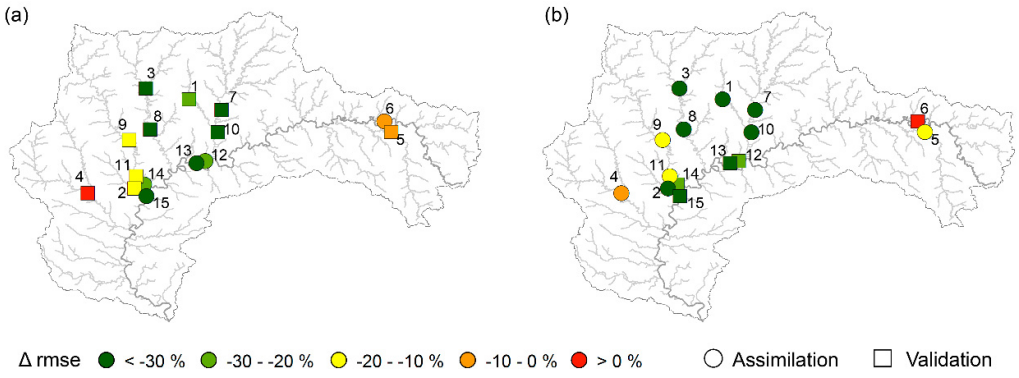

Figure 8. Spatial distribution of the relative difference of root-mean-square error (Δrmse) to (a) assimilating main stem gauges, (b) assimilating tributaries gauges.

Finally, Figure 9 presents hydrographs from observations, open-loop and EnKF estimates, for the best and worst assimilation results for validation gauges, for the two assimilation scenarios (main stem and tributaries). As expected, EnKF estimates (red line) with the highest efficiency are more similar to observations (blue line) than open-loop simulations (black line), especially for low flows. This could be explained due to the better agreement of model calibration for lower discharge values than for peak flows. Besides, even for the highest efficiency, there was a peak at the beginning of the analysis period which was not well simulated by DA and can be explained due to this event corresponding to the warm-up period. Assimilation scenarios with the lowest efficiency may even decrease performance when there are peaks in the observation, which can be explained due to a spurious correlation generated in the matrix of covariance of errors, between the distant assimilation and validation gauges likely related to a spatially variable region. Therefore, more distributed gauge observations may be necessary to achieve better efficiency in DA simulations.

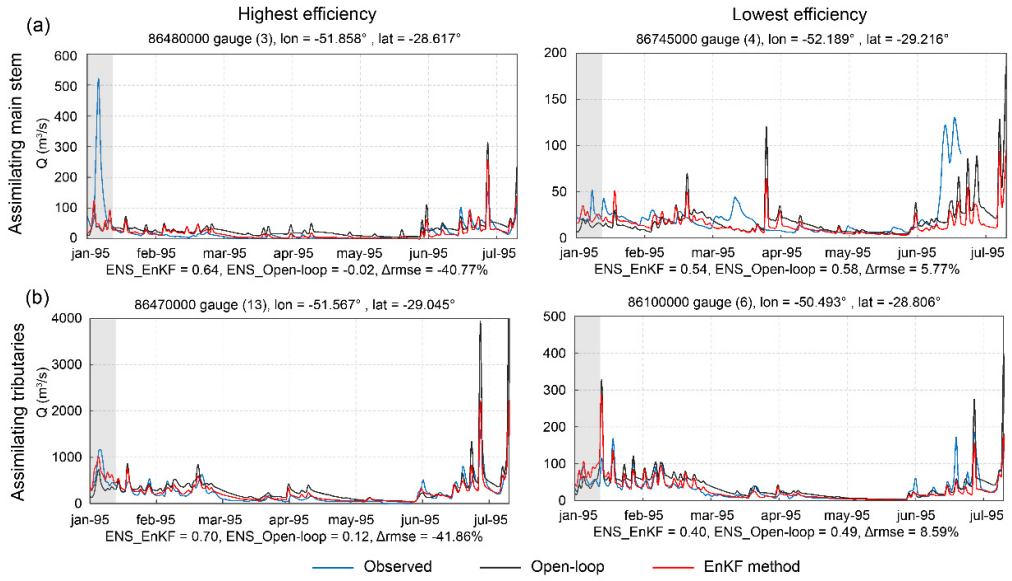

Figure 9. Hydrographs in validated gauges for observation (blue), open-loop (black) and assimilation (red), for highest and lowest efficiency, while assimilating (a) main stem gauges (validating tributaries) and (b) tributaries gauges (validating main stem) for the period from January 15 to July 15. Gray bars represent a part of the warm-up period.

It is interesting to note that both the computed time step of the model and the observations are computed at a daily time step. In the basin, however, flood waves occur and propagate extremely fast, sometimes at periods shorter than one day. This has motivated a flood forecasting system in the basin to adopt hourly time steps (Siqueira et al. 2016). This may be one of the reasons that the DA scheme did not perform well for some validated gauges, where peak flows occurred within only one-time step (see, for instance, the high peak at the end of the Figure 9b time series). Although the DA estimates at an hourly time step could be interesting to better capture the basin flood hydrograph characteristics, availability of the observed discharge at this short time step (i.e. automatic stations) is limited for the basin, and thus hindered such an analysis.

3.3.4. Estimation of reference discharges

In this section, reference discharges (Q_{90}, Q_{LTA}, $Q_{MAX\ RP=5}$) estimated with the EnKF DA method are compared with the regionalization results, for the two assimilation scenarios. Results are compared to observed reference discharges, computed from the in-situ observed data for the same period of DA simulations (see Section 2.4 final paragraph).

When assimilating gauges in main stem, Figure 10a presents the estimates with both methods for the validation gauges. Generally, values for validated gauges with assimilation (red squares) were more similar to the observed reference discharges (blue circles) than the estimates from open-loop simulations (black triangles) and regionalization methods (green diamonds). Table 4 shows with the

Δrmse metric that discharge estimates by the DA, in main stem gauges, yielded the best estimates for Q_{90} and Q_{LTA}, in comparison to the observed, simulated and regionalized reference discharges. Q_{90} increased twice in relation to regionalization, and Q_{LTA} rose more than three times. Maximum discharge ($Q_{MAX\ RP=5}$) presented slightly better results for the regionalization method.

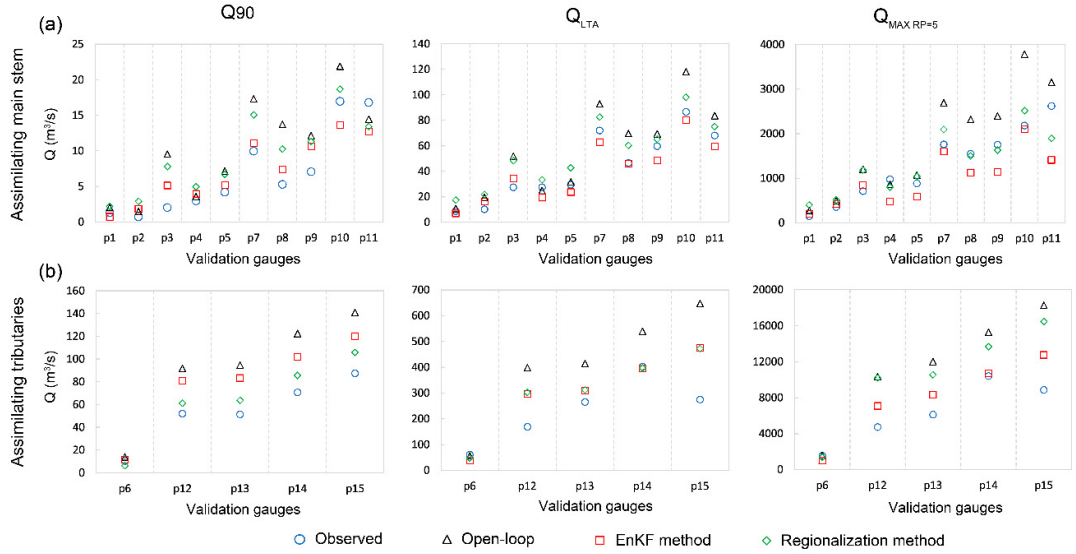

Figure 10. Reference discharged values (Q_{90}, Q_{LTA} and $Q_{MAX\ RP}$=5) for observations (blue), simulation (black), the EnKF method (red) and the regionalization method (green), while assimilating the (a) main stem gauges and (b) tributaries gauges.

In turn, the results for the estimation of reference discharges using only tributary gauges, are presented in Figure 10b and Table 4. The DA method performed better than regionalization for Q_{LTA} and $Q_{MAX\ RP=5}$. While for Q_{90}, regionalization performed significantly better in comparison with the DA and the regionalization of main stem gauges. This can be explained since a greater number of gauges (10) to develop the regression curve for regionalization implies better representativeness of the region; some results will be further discussed in the next section.

Furthermore, gains of Δrmse for EnKF, with respect to the open-loop simulations, were on average ~50% for all reference discharges when assimilating the main stem gauges, while it was more than 60% for assimilation of tributaries. For all scenarios, open-loop simulations performed worse than regionalization and EnKF ones, which may be explained as do to the relative simplified calibration scheme. Parameter sets calibrated individually, for some sub-basins, could potentially improve the open-loop simulations.

Table 4. Values of Δrmse (%) in relation to the open-loop simulations, considering EnKF and regionalization methods, and the median of validated gauges.

Using gauges on	Scheme	Q_{90}	Q_{LTA}	$Q_{MAX\ RP=5}$
Main stem	**EnKF**	**-58.88**	**-46.87**	-53.32
	Regionalization	-26.67	-14.97	**-57.27**
Tributaries	**EnKF**	-27.81	**-70.46**	**-60.34**
	Regionalization	**-71.49**	-68.49	-21.24

Values in bold show the best results

3.4. Summary and discussions

The results presented in Sections 3.1 and 3.2 are important for the comprehension of the limits and capacities of the DA scheme, for the Taquari-Antas basin. The performance of the MGB model for assimilation was evaluated for different scales of discharges (default and logarithmic scales (Section 3.1). First, the results showed that the logarithm of discharges performed slightly better, hence it was adopted for subsequent tests. The sensitivity to the ensemble size and other parameters related to precipitation and state variables errors were assessed, from which the following values were adopted: N = 100; E = 20%, τ_x = 0.0625°, τ_t = 20 days and EV = 10%, which are different from Paiva et al. (2013b) values for the Amazon river basin. The value of spatial decorrelation (0.0625°), near to unit-catchments average size (model calculation units), indicates that for the Taquari-Antas, the best results were obtained when no spatial correlation was considered for precipitation. In the Amazon application, the decorrelation length was much higher. However, in accordance to the hypothesis about the Amazon, large correlations between precipitation fields exist due to strong seasonality and less heterogeneities in the terrain, while in a mountainous and sub-tropical basin, such as the Taquari-Antas, where precipitation occurs with a high non-linear behavior throughout the year, precipitation fields present less correlation.

In Section 3.2, different configurations of the DA scheme were tested, including just the perturbation of precipitation (adopted by many authors in the literature, such as Biancamaria et al. 2011, Liu et al. 2012 and Paiva et al. 2013b), and combinations with the perturbation of model state variables, namely soil water storage and volumes of surface, sub-surface and groundwater reservoirs (W, V_{SUP}, V_{INT} and V_{BAS} variables, respectively). The tests indicated that the perturbation of all state variables, together with precipitation yielded results similar to or better than the perturbation of only precipitation, hence, used in the following steps of the study. Future studies may also assess how perturbation of model parameters affect the efficiency of DA.

The potential use of DA for improving the discharge time series estimation was assessed in Section 3.3., since one of the main uses of the distributed hydrological models is to estimate time series in ungauged regions. In this respect, results of the two scenarios of EnKF simulations (discharge estimates by DA in gauges at Taquari-Antas' main stem or from tributaries and validated with the remaining gauges) indicated relative gains, with respect to open-loop simulations (Δrmse) for validated gauges of -25 to -30%. It is important to note that better results were achieved when assimilating tributaries gauges, which can be justified by the fact that more gauges were available for assimilation use, in this case. Although the gains are evident, the high computational costs involved and the complexity required for DA implementation, considering established traditional methodologies, makes it difficult to implement. On the other hand, the tests were performed using a daily time step for the simulations, due to observational discharge data availability in the basin, even though, several flood events occur at sub-daily periods. For this reason, one important aspect for future developments of DA schemes, and its usage in discharge estimation, is the improvement of river monitoring in Brazil, increasing the number of automatic gauges throughout the country as well as improvements in the sampling frequency (hourly). That will likely help achieve better estimations in terms of basins with sub-daily response.

Regarding reference discharges (Section 3.4), tests indicated that in both scenarios of assimilation (main stem and tributaries) almost all discharge estimates showed improvements in comparison to both open-loop simulation and regionalization methods. Nevertheless, it is necessary to highlight that the results for reference discharge for low (Q_{90}) and high ($Q_{MAX\ RP=5}$) values differ when both DA and regionalization are performed. Recent studies showed the opposite results, since the DA efficiency was noticed to be greater for the lowest flows than the highest ones (Vrugt et al. 2006).

Finally, once again, the issue of simplicity in estimation with the regionalization method arises, even with associated uncertainties. One interesting possibility for the usage of DA schemes for estimation of discharges, is to use such approaches to generate a time series or reference discharges for large areas or basins (e.g., for a whole Brazilian state), which could in turn be used for water resources management studies.

3.5. Conclusions and perspectives

This research aimed to adapt and assess the EnKF DA technique into a large-scale hydrological and hydrodynamic model, to estimate reference discharges and discharge time series for water resources management.

Data assimilation techniques, such as EnKF, present some complexity in terms of its implementation, in comparison to simpler methods of discharge estimation, such as discharge

regionalization. However, proposing tools that can promote adequate discharge estimations is one of the vital challenges in water resources studies. Results presented in this article confirm previous studies, which show that model discharge series estimates may be improved with the implementation of DA schemes. Reference discharge estimates by DA also showed relative improvements when compared with a simple regionalization method, although further analysis should be assessed as described as following.

The Taquari-Antas river basin is known in Brazil for its frequent flooding events in the downstream area of Vale do Taquari region, but there are also relevant water stress situations occurring due to intense use for agriculture and livestock. Although the basin features a good number of observed discharge gauges, especially in comparison to the Brazilian context, the expansion of the hydrological monitoring network is fundamental. It could improve the use of the DA approach proposed in this study, by increasing observations at hourly time steps (instead of daily ones) that allow for better capturing the flood hydrograph characteristics (e.g., peak flow, time to peak, etc.). On the other hand, data availability in the Taquari-Antas basin allows for different test configurations of assimilation and validation gauges, throughout the basin, separating tributary stations from main stem ones.

In the Brazilian context, there is a significant need for consistent estimates of the hydrological time series and reference discharges as well as for a better spatial distribution of these estimations, for the improvement of water resources management and other hydrological applications. For instance, definition of water permits, maximum water exploitation and levels of water stress are very dependent on such reliable data. Even though this study contributes to this discussion and proposes new tools for achieving a better management scenario, related to the use of EnKF DA schemes in distributed hydrological models, to improve the quality of discharge estimates. Other similar tests could be carried out in different types of basins, in order to understand how various basin responses can be related to the improvement of discharge estimations with DA schemes.

Therefore, further research should investigate some topics such as: i) the assessment of the localization impact of the DA, considering spatial distribution of measurement stations in the river network; ii) the consideration of a lower time step (e.g., hourly) in the model simulation to assess the generating of runoff due to basin characteristics; and iii) the examination and comparison in terms of depth of other regionalization methods, since this work provides an initial insight into the latter.

Acknowledgements

The first author is grateful for a grant from the Brazilian agency CAPES, and the second author for a grant from CNPq agency. The authors are grateful to the contributions of MSc. Vinícius Siqueira for data supply, and Prof. Walter Collischonn for ideas and constructive comments. We also thank two reviewers for their comments, which helped significantly in the improvement of this manuscript.

References

Abbaris, A, H Dakhlaoui, S Thiria, and Z Bargaoui. 2014. Variational Data Assimilation with the Yao Platform for Hydrological Forecasting. Proceedings of the International Association of Hydrological Sciences 364. Copernicus GmbH: 3–8.

Abbaszadeh, P., Moradkhani, H., Yan, H. 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. Advances in Water Resources, 111: 192-204.

Arsenault, R., Brissette, F. 2015. Multi-model averaging for continuous streamflow prediction in ungauged basins. Hydrological Sciences Journal, 61(13), 2443-2454.

Arsenault, R., Brissette, F. 2016. Analysis of continuous streamflow regionalization methods within a virtual setting. Hydrological Sciences Journal, 61(15), 2680-2693.

Asadi, P., Engelke, S., Davison, A. C. 2016. Statistical Regionalization for Estimation of Extreme River Discharges. ArXiv e-prints, 1611.03219.

Bates, P. D., Horritt, M. S., Fewtrell, T. J. 2010. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. Journal of Hydrology, 387, p. 33–45.

Beck, M. B. 1987. Water quality modeling: a review of the analysis of uncertainty. Water Resources Research, 23(8), 1393-1442.

Beskow, S., de Mello, C. R., Vargas, M. M., Corrêa, L. L., Caldeira, T. L., Durães, M. F. de Aguiar, M. S. 2016. Artificial intelligence techniques coupled with seasonality measures for hydrological regionalization of Q90 under Brazilian conditions. Journal of Hydrology, 541(Part B), 1406-1419.

Bernard, R., Vauclin, M., and Vidal-Madjar, D. 1981. Possible use of active microwave remote sensing data for prediction of regional evaporation by numerical simulation of soil water movement in the unsaturated zone. Wat. Resour. Res., 17(6): 1603-1610.

Bertino, L., Evensen, G., Wackernagel, H. 2003. Sequential Data Assimilation Techniques in Oceanography. Internat. Statist. Rev. 71, no. 2, 223--241. https://projecteuclid.org/euclid.isr/1069172299

Biancamaria, S., Durand, M., Andreadis, K. M., Bates, P. D., Boone, A., Mognard, N. M., Rodriguez, E., Alsdorf, D. E., Lettenmaier, D. P., Clark, E. A. 2011. Assimilation of virtual wide swath altimetry to improve Arctic river modelling. Remote Sensing of Environment, 115, 373–381.

Bruckler, L., and Witono, H. 1989. Use of remotely sensed soil moisture content as boundary conditions in soil-atmosphere water transport modeling: 2. Estimating soil water balance. Wat. Resour. Res., 25(12): 2437-2447.

Castiglioni, S., Castellarin, A., Montanari, A. 2009. Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. Journal of Hydrology, 378, 272-280.

Clark, Martyn P.; Slater, Andrew G. 2006. Probabilistic Quantitative Precipitation Estimation in Complex Terrain. Journal of Hydrometeorology, Vol. 7, pp. 3 – 22.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., Uddstrom, M. J. 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, 31, 1309–1324.

Collischonn, W., Tucci, C. E. M. 2001. Hydrological simulation of large drainage basins (in Portuguese). RBRH: Revista Brasileira de Recursos Hídricos, 6(1), 15–35.

Collischonn, W.; Allasia, D. G.; Silva, B. C.; Tucci, C. E. M. 2007. The MGB-IPH model for large-scale rainfall-runoff modeling. Hydrological Sciences Journal, 52, 878-895.

Collischonn, B., Collischonn, W., Tucci, C. E. M. 2008. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. Journal of Hydrology, 360(1-4), 207–216.

Durrans, S.R., Kirby, J.T., 2004. Regionalization of extreme precipitation estimates for the Alabama rainfall Atlas. J. Hydrol. 295, 101–107.

Entekhabi, D. 1994. A simple model of hydrologic cycle and climate: 1. Model construct and sensitivity to landsurface boundary, Advances in Water Resources, 17, 79-91.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans (1978–2012), 99 (C5), 10143–10162. doi:10.1029/94JC00572

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367.

Fadel, A. W. 2015. Incorporação do risco de prejuízo no gerenciamento de medidas de controle de inundação. 2015. 109 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Porto Alegre, 2015.

Fan, F. M., Schwanenbergb, D., Collischonn, W., Weerts, A. 2015. Verification of inflow into hydropower reservoirs using ensemble forecasts of the TIGGE database for large scale basins in Brazil. Journal of Hydrology, Regional Studies, 4, 196–227.

Fan, F. M., Collischonn, W., Quiroz, K. J., Sorribas, M. V., Buarque, D. C., Siqueira, V. A. 2016. Flood forecasting on the Tocantins River using ensemble rainfall forecasts and real-time satellite rainfall estimates. Journal of Flood Risk Management, 9(3), 278-288.

Gelb, A. Optimal linear filtering. In Applied Optimal Estimation; MIT Press: Cambridge, MA, USA, 1974; pp. 102–155.

Georgakakos, K. P., and Baumer, O. W., (1996). Measurement and utilization of on-site soil moisture data. J. Hydrol., 184: 131-152.

Goniadzki, D.; Thibeault, M.; Lozza, H.; Quirno, M. U.; Dadamia, D. 2011. SAC-D/Aquarius Announcement of Opportunity. Assimilation of Soil Moisture Estimates into Flow-Forecasting Hydrologic Models. In: Primer Encuentro de Investigadores en Formación en Recursos Hídricos (IFRH) Instituto Nacional del Agua. Ezeiza, Argentina.

Hreinsson, Einar Ö. 2008. Assimilation of a snow covered area into a hydrologic model. Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Geography. University of Canterbury. New Zealand.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe E., Cudennec, C. 2013. A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrological Sciences Journal, 58(6), 1198-1255.

Jackson, T. J., et al., (1981). Soil moisture updating and microwave remote sensing for hydrological simulation. Hydrol. Sci. B., 26(3): 305-319.

Kapangaziwiri, E., Hughes, D. A., Wagener, T. 2012. Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa. Hydrological Sciences Journal, 57(5), 1000-1019.

Kitanidis, P. K., Bras, R. L. 1980. Real-time forecasting with a conceptual hydrologic model. 2. Applications and results. Water Resources Research, 16(6), 1034-1044.

Larentis, D. G.; Collischonn, W.; Tucci, C. E. M. 2008. Simulação da qualidade de água em grandes bacias: rio Taquari-Antas, RS. RBRH: Revista Brasileira de Recursos Hídricos, 13(3), 5-22.

Lee, H., Seo, D. J., Liu, Y., Koren, V., McKee, P., & Corby, R. (2012). Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal scale of adjustment. Hydrology and Earth System Sciences, 16(7), 2233-2251.

Liu, Y.; Gupta, H. V. 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research, v. 43, n. 7, W07401, 2007.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., Van Dijk, A. I. J. M., Van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., Restrepo, P. 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences, 16, 3863–3887.

McLaughlin, D., 1995. Recent developments in hydrologic data assimilation. Reviews of Geophysics, 33 (S2), 977–984. doi:10.1029/95RG00740

Maxwell, D. H., Jackson, B. M., & McGregor, J. (2018). Constraining the ensemble Kalman filter for improved streamflow forecasting. Journal of Hydrology, 560, 127-140.

McCuen, R. H. and Levy, B. S., 2000. Evaluation of peak discharge transposition. Journal of Hydrologic Engineering, 5(3), 278–289.

Merz, R., Parajka, J., Bloschl, G. 2010. Time stability of catchment model parameters – implications for climate impact analyses. Water Resources Research. doi: 10.1029/2010WR009505.

Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S. 2005a. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resources Research, 41(5).

Moradkhani, H., Sorooshian, S., Gupta, H., Houser, P. R. 2005b. Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources, 28, 135–147.

Nijssen, B., Lettenmaier, D. 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, Journal of Geophysical Research, 109(D2, D02103).

Neal, Jeffrey; Atkinson, Peter; Hutton, Craig. 2007. Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. Journal of Hydrology, Vol. 336, pp. 401 – 415.

Novaes, L. F., Pruski, F. F., De Queiroz, D. O., Rodriguez, R. G., Da Silva, D. D., Ramos, M. M. 2009. Modelo para a quantificação da disponibilidade hídrica: Parte 2 - Análise do comportamento do modelo para a estimativa da Q7,10 na Bacia do Paracatu. RBRH: Brazilian Journal of Water Resources, 14(1), 27-39.

Ouarda, T. B. M. J., Bâ, K. M., Diaz-Delgado, C., Cârsteanu, A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., Bobée, B. 2008. Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. Journal of Hydrology, 348, 40–58. purposes of regionalization. Journal of Hydrology, 121, 217-238.

Ottlé, C., and Vidal-Madjar, D., (1994). Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY Region. J. Hydrol., 158: 241-264.

Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., Bulhões Mendes, C.A. 2013a. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resources Research, 49(3), 1226-1243.

Paiva, R. C. D., Collischonn, W., Bonnet, M.-P., De Gonçalves, L. G. G., Calmant, S., Getirana, A., Santos Da Silva, J. 2013b. Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. Hydrology and Earth System Sciences, 17, 2929-2946.

Pauwels, V. R. N., and G. J. M. De Lannoy, 2006: Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge. J. Hydrometeor., 7, 458–477.

Pontes, P. R. M., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., Buarque, D. C., Siqueira, V. A., Jardim, P. F., Sorribas, M. V., Collischonn, W. 2017. MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS. Environmental Modelling & Software, 94, 1-20.

Postel, S. L., and Wolf, A. T. 2001. Dehydrating conflict, Foreign Policy, 9, 60–67.

Prevot, L., et al., (1984). Evaporation from a bare soil evaluated using a soil water transfer model and remotely sensed surface soil moisture data. Wat. Resour. Res., 20(2): 311-316.

Pruski, F. F., Nunes, A. A., Pruski, P. L., Rodriguez, R. G. 2013. Improved regionalization of streamflow by use of the streamflow equivalent of precipitation as an explanatory variable. Journal of Hydrology, 476, 52–71.

Qin, Changbo; Jia, Yangwen; Su, Z. (Bob); ZHOU, Zuhao; QIU, Yaqin; SUHUI, Shen. 2008. Integrating Remote Sensing Information Into a Distributed Hydrologycal Model for Improving Water Budget Predictions in Large-scale Basins through Data Assimilation. Sensors, Vol. 8, pp. 4441 – 4465.

Rasmussen, J., Madsen, H., Jensen, K. H., & Refsgaard, J. C. 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. Hydrology and Earth System Sciences, 19(7), 2999-3013.

Razavi, T., Coulibaly, P. 2016. Improving streamflow estimation in ungauged basins using a multi-modelling approach. Hydrological Sciences Journal, 61(15), 2668-2679.

Reichert, P. and Borsuk, M.E., 2005. Does high forecast uncertainty preclude effective decision support? Environmental Modelling and Software, 20(8), 991– 1001.

Reichle, R. H. 2008. Data assimilation methods in the Earth sciences. Advances in Water Resources, 31, 1411–1418.

Restrepo, P. J. 1985. A view of maximum-likelihood estimation with large conceptual hydrologic models. Applied Mathematics and Computation, 17(4), 375-403.

Revilla-Romero, B., Wanders, N., Burek, P., Salamon, P., de Roo, A., 2016. Integrating remotely sensed surface water extent into continental scale hydrology. J. Hydrol. 543, 659–670.

SEMA – SECRETARIA ESTADUAL DO AMBIENTE E DO DESENVOLVIMENTO SUSTENTÁVEL DO RIO GRANDE DO SUL. Plano da Bacia Hidrográfica do Taquari-Antas. 2014.

Shuttleworth, W.J., 1993. Evaporation. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, pp. 4.1–4.53.

Siqueira, V., Collischonn, W., Fan, F. M., Chou, S. C. 2016. Ensemble flood forecasting based on operational forecasts of the regional Eta EPS in the Taquari-Antas basin. RBRH: Brazilian Journal of Water Resources, 21(3), 587-602.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., Zehe, E. 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48(6), 857-880.

Sorribas, M. V., Paiva, R. C. D., Melack, J. M., Bravo, J. M., Jones, C., Carvalho, L., Beighley, E., Forsberg, B., Costa, M. H. 2016. Projections of climate change effects on discharge and inundation in the Amazon basin. Climatic Change, 136, 555-570.

Smakhtin, V. U. 2001. Low flow hydrology: a review. Journal of Hydrology, 240, 147-186.

Sun, C.; Walker, J.P.; Houser, P.R. A methodology for snow data assimilation in a land surface model. J. Geophys. Res. 2004, 109, D08108.

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. Hydrological Sciences Journal, 61(13), 2348-2366.

Tallaksen, L. M., Van Lanen, H. A. J. 2004. Hydrological Drought. Processes and Estimation Methods for Streamflow and Groundwater. Developments in Water Science, 48, Elsevier Science B.V., 579.

Thiboult, A., & Anctil, F. 2015. On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments. Journal of Hydrology, 529, 1147-1160.

Todini, E.: The ARNO rainfall-runoff model, J. Hydrol., 175, 339–382, 1996.

Todorov. E. Optimal Control Theory. In K. Doya, editor, Bayesian Brain: Probabilistic Approaches to Neural Coding, pages 269–298. MIT Press, 2006.

Tucci, C. (2002) Regionalização de Vazões. Primera edição. Ed. Universidades/UFRGS, Brasil. 256 pp.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., Verstraten, J. M. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41(1).

Vrugt, J.A.; Gupta, H.V.H.; Nualláin, B.; Bouten, W. 2006. Real-time data assimilation for operational ensemble streamflow forecasting. J. Hydrometeorol. 7, 548–565.Walker, J.P. and P.R. Houser, 2005. Hydrologic data assimilation. In A. Aswathanarayana (Ed.), Advances in water science methodologies (230 pp). The Netherlands, A.A.Balkema.

Vrugt, J.A. ter Braak, C.J.F., Diks, C.G.H., Schoups, G. 2013. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. Advances in Water Resources, 51: 457–478.

Wagener, T. M., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H.V., et al. 2010. The future of hydrology: An evolving science for a changing world. Water Resources Research 46: W05301.

Weerts, Albrecht H.; El Serafy, Ghada Y.H. 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research, Vol. 42, W09403, doi: 10.1029/2005WR004093.

Whitaker, J. S., and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130, 1913–1924.

Wigmosta, M.S., Vail, L.W., Lettenmaier, D.P., 1994. A distributed hydrology-vegetation model for complex terrain. Water Resour. Res. 30, 1665–1679. http://dx.doi.org/10.1029/94WR00436.

Wongchuig, C. S., Paiva, C. D. R., Siqueira, V., Collischonn, W. 2019. Hydrological Reanalysis across the 20th Century: A Case Study of the Amazon Basin. Journal of Hydrology, 570:755–773. https://doi.org/10.1016/j.jhydrol.2019.01.025.

Young, P. C. 2002. Advances in real-time flood forecasting. Philosophical Transactions of the Royal Society A, 360(1796), 1433-1450.

Zhu, J., Zheng, F., Li, X., 2011. A new localization implementation scheme for ensemble data assimilation of non-local observations. Tellus A 63, 244–255. https://doi.org/10.1111/j.1600-0870.2010.00486.x.

Capítulo 4

Reanálise Hidrológica do século 20: Estudo de caso na Bacia Amazônica

Prólogo

Resultados das Retrospectivas Hidrológicas (HR) desenvolvidas no capitulo dois, indicam que as simulações hidrológicas forçadas com as reanálises climáticas mais extensas (> 30 anos) ainda apresentam uma baixa performance para representar adequadamente a hidrologia da bacia Amazônica, mais ainda no caso de valores extremos. Por outro lado, metodologias que visam o desenvolvimento de registros mais consistentes baseados em modelos, têm utilizado técnicas como a assimilação de dados de inúmeras observações, denominadas de reanálises. Reanálises são métodos científicos voltados para o desenvolvimento de um registro retrospectivo da evolução de um determinado sistema. Estes métodos foram inicialmente desenvolvidos na área de meteorologia para o estudo do clima, desta maneira foram denominadas de reanálises climáticas (e.g. Gibson et al., 1997; Kistler et al., 2001). A técnica de assimilação de dados também tem sido usada nos últimos anos para o desenvolvimento de registros extensos e coerentes em áreas como a hidrologia (e.g. vazão, nível d'água). Denominada de reanálises hidrológicas, esta metodologia tem sido explorada por algumas iniciativas tais como o PCR-GLOBWB (Wanders et al., 2014), MERRA Land (Reichle e Liu, 2015), ERA-Interim/Land (Balsamo et al., 2015) ou o projeto Earth2Observe cujo objetivo é fornecer uma base de dados de reanálises consistentes dos recursos hídricos, muitos deles iniciando analises nos anos ~1979's. Em contrapartida, estas reanálises hidrológicas atualmente desenvolvidas, por serem produtos globais tendem a usar simplificações nos processos hidrológicos e hidráulicos (e.g. modelos de superfície global, propagação simplificada da vazão nos rios); possibilitando o aumento das incertezas nos resultados em estudos regionais.

Por outro lado, a implementação do esquema de assimilação de dados EnKF no modelo MGB apresentado no capitulo três, será usado neste capitulo para assimilar dados de séries de vazão in-situ na bacia Amazônica, onde existem postos com séries extensas de nível d'água e vazão (e.g. Manaus [1902], Lábrea [1927], Humaita [1931], Altamira [1928]). A metodologia resulta num registro extenso (século 20) de variáveis hidrológicas-hidráulicas denominado de reanálise hidrológica do século 20 (RHXX). Este RHXX se diferencia de produtos globais existentes devido à menor simplificação dos processos hidrológicos e hidráulicos (e.g. modelo hidrológico semi-distribuído, propagação hidrodinâmica nos rios). Os resultados são validados com series de vazão e nível in-situ de postos que não foram usados no processo de assimilação. Adicionalmente foram avaliados de maneira espacialmente distribuída alguns eventos hidrológicos extremos que ocorreram no último século.

Hydrological Reanalysis Across the 20th Century: A Case Study of the Amazon Basin

Wongchuig, C.S., de Paiva, R.C.D., Siqueira, V., Collischonn, W.

Paper published in Journal of Hydrology 10.1016/j.jhydrol.2019.01.025

Abstract

Adequate and accurate hydrological data is necessary to manage water resources, which are critical in developing countries where such information is limited. In recent years, global reanalysis datasets have been developed to provide this information in climatic fields and, more recently, in hydrologic fields. Nevertheless, these latest efforts have been limited to temporal coverage (~30 years) and mostly simplified hydraulic scheme routing in rivers, which can be inadequate for regional and long term scale objectives. In this article, a dataset (called hydrological reanalysis across the 20th century [HRXX] in the Amazon Basin) was developed as a case study spanning back to the year 1900 through the use of: 1) a large-scale hydrologic-hydrodynamic model (MGB) forced by a long-term climatic reanalysis of rainfall (ERA-20CM) with bias removed; and 2) a data assimilation (DA) technique coupled with a localization method (LEnKF) to use several ground observations of daily discharge within a radius of influence. Several tests were assessed to find the best bias removal method, the optimal radius of influence for the localization method and the final HRXX dataset. A total of 114 hydrological ground observations of daily information were used for assimilation and validation purposes, and several statistics indexes were employed to assess their performance. Results indicate that both bias correction and the DA with localization method greatly improved the simulations. Overestimations of the peaks in the open-loop (OL) (free run) simulation, mainly in the southern and northern regions of the Amazon Basin, were removed, and recession timing in the east-central region, as seen at the Óbidos gauge station, were corrected. An average performance of ~0.6 and ~0.7 of the Nash–Sutcliffe and Kling–Gupta indexes was reached, even when only a few of the longest ground observations were used, which can be representative of the oldest periods (since ~1930). To assess extreme events, the Pearson correlation coefficient was used for maximum and minimum annual water level anomaly values, reaching 0.6 and 0.7, respectively, at the Manaus gauge station, which is remarkable considering that the analysis covers approximately 110 years. Considering the results of this case study and the global coverage of rainfall datasets, this methodology can be transferred to other regions in order to better estimate and create a hydrological reanalysis that adequately represents the hydrologic and hydraulic spatio-temporal fields.

Keywords: Hydrological Reanalysis; Hydrological-hydrodynamic modeling; Data Assimilation; Local EnKF; Long-term dataset; Amazon Basin

4.1. Introduction

The requirements of hydrologic records have presented challenges for water management around the world, mainly in developing countries where in-situ data is temporally and spatially limited and uncertain (Sivapalan et al., 2003), which makes it difficult to characterize past hydrology in a distributed manner. Some attempts to generate large records have been made in recent years using indirect or proxy data (Jarret, 1991) that span periods beyond the range of instrumental records (Moss et al., 1988). For example, paleoflood hydrology (Denlinger et al., 2002; Webb and Jarrett, 2002) and paleohydrology (Barber and Finney, 2000; Nunnery, 2012) studies have focused on periods that usually range from the last 5000 years to 100 years (Baker et al., 2002; Benito et al., 2005). These attempts aim to reproduce the historical hydrology for periods when records were not kept; however, they still present some uncertainties, for example related to temporal scaling (Razavi et al., 2016), and they are constrained at the gauge of reconstructed sites (U.S. Army Corps of Engineers, 2003), where similar spatial restrictions are implied compared with current in-situ observations in large basins.

Moreover, in the fields of climatology and oceanography, data assimilation (DA) techniques have been developed and applied to improve forecast systems and to create a consistent and retrospective record of state-of-the-art climatology variables on a global scale through the optimal combination of models and several observations, also widely known as climatic reanalysis (e.g. Gibson et al., 1997; Kistler et al., 2001; Compo et al., 2011). In recent years, DA techniques have started to be used in the hydrological field (Sun et al., 2016), although their implementation for local (Vrugt et al., 2005; Clark et al., 2008; Xie and Zhang, 2010; Kurtz et al., 2017) and large-scale applications (Paiva, 2013b; van Dijk et al., 2014; López López et al., 2015; Liu et al., 2012a, 2012b, 2016; Andreadis et al., 2017) are still under development. Some initiatives to develop a coherent register of hydrologic variables in global coverage, usually called hydrological reanalysis or water resource reanalysis, aim to offer a database of hydrological variables, mostly from around the year 1979, such as PCR-GLOBWB (van Beek et al., 2011; Wanders et al., 2014), MERRA Land (Reichle and Liu, 2015), ERA-Interim/Land (Balsamo et al., 2015), eartH2Observe (Schellekens et al., 2017) or ERA-40 (Uppala et al., 2005; Dutra et al., 2008) since 1958. The main reason for the lack of many current hydrological reanalyses, as global products, is the use of simplifications in some procedures, such as hydraulic propagation in rivers, misrepresenting the hydraulic processes in regions with floodplain extensions, for example, the large Amazon or Niger basins, which lead to increased uncertainties primarily in the context of a regional outlook.

Recent studies have shown that a large-scale hydrologic-hydrodynamic model forced by multi-decadal precipitation datasets allows for both hydrological variability and extreme events that occurred in the last decades (e.g. Wongchuig et al., 2017; Nkiaka et al., 2017 and Lakew et al., 2017). Nevertheless, one of the limitations of such products, also known as hydrological retrospectives, is their relatively short temporal extension for the best performing datasets (Wongchuig et al., 2017), which can be extended by using long-term climatic reanalysis (Compo et al., 2011; Hersbach et al., 2013).

Moreover, a long-time series of discharges and water levels can be also combined within a DA scheme, allowing the production of an improved long-term hydrological register (since the early 1900s). From among the diverse DA techniques, the Ensemble Kalman Filter (EnKF) (Evensen, 2003) has emerged as a popular choice for assimilation into land surface and hydrological models, because it enables a simplified implementation (Moradkhani, 2008). In recent years, a modified technique called Local Ensemble Kalman Filter (LEnKF) has been developed (Patil et al., 2001; Ott et al., 2002, 2004), which implements the impact of local observations within the subset of the model state variables. One remarkable advantage is that LEnKF can avoid the spurious correlations in the model covariance matrices between the distant state variables of the model (Zhu et al., 2011), where non-zero correlations are assigned to variables that are uncorrelated because of a limited number of ensemble members (Bishop and Hodyss, 2009) or between two spatially remote locations (Hamill et al., 2001; Anderson, 2007), because no physical connection for the assimilated state variable exists. For sequential DA schemes such as EnKF, a localization method was developed based on a distant-dependent covariance localization (Houtekamer and Mitchell, 2001; Campbell et al., 2010; Zhu et al., 2011; Houtekamer, 2016); furthermore, this implementation has been demonstrated to be relevant for hydrology studies (Rasmussen et al., 2015; Munier et al., 2015; Zhang et al., 2016).

Long-term hydrological simulation requires a consistent set of forcing data, which is available in both global and long-time scales by climatic reanalysis/models. Numerous studies have demonstrated that raw output from global climate models (GCM), regional climate models (RCM) and climate reanalysis have biases mostly due to systematic model errors (Teutschbein and Seirbet, 2012; Ngai et al., 2017; Berg et al., 2003). Because of this, a bias correction scheme is required for assessing the precipitation forcing, as previous research concluded it necessary for long-term reanalysis of precipitation (Decker et al., 2012; Kim et al., 2018). On the other hand, if the hydrologic model is simulated with the precipitation forcing biased, the use of an observation within the DA scheme may result in a synthetic shift in the state variables.

Finally, the development of the “Hydrological Reanalysis Across the 20th Century” (HRXX) is proposed in this research through the use of a DA scheme of long temporal in-situ discharge records within a large-scale hydrologic-hydrodynamic model forced by long-term precipitation with bias correction. Understanding the implications of assimilating observations into large-scale hydrologic-hydrodynamic models for long temporal series is a significant exploration opportunity and may have potential advantages for water resource management (van Dijk and Renzullo, 2011). The distinctive benefits of using HRXX data are that their spatially distributed improved variables and long temporal series could be used to understand hydro climatology variability throughout the last century, to estimate projected scenarios for hydroelectrical yield companies with less uncertainty, to detect hydrological anomalies such as floods and droughts to adequately address these extreme events (Dottori et al., 2016) and to further reduce their future impacts.

In this context, the objectives of this research as a first proxy in the Amazon Basin are: 1) to develop a hydrological reanalysis methodology and product that represents hydrology characteristics across the 20th century; 2) to develop a proxy localization within a DA scheme (LEnKF); 3) to improve precipitation forcing representation through bias removal methods; and 4) to validate HRXX against discharge and water level anomalies from several in-situ gauges.

4.2. Data and methods

4.2.1. Hydrological reanalysis approach

Hydrological reanalysis (HR) can be defined as a consistent and coherent register of spatial-temporal fields of hydrologic variables. Taking as a concept their predecessor in climatology field, reanalysis usually extends over several decades, and its record is generated through the optimal combination (updating) of the model forecast with as much as possible observations used within the DA scheme. Considering a global outlook, various initiatives, mainly based on land surface models forced by climate reanalysis, have been developed to create HR datasets with the aim of improving and providing a coherent record of variables that correspond to the water cycle, such as GLDAS (Houser et al., 2002; Rodell et al., 2004), GSWP-2 (Dirmeyer et al., 2006), MERRA-Land (Reichle, 2011), ERA-Land (Balsamoet al., 2015), ERA-20CM-R (Emerton et al., 2016) and WATCH (Haddeland et al., 2011), although these products are not focused to provide a precise record of discharge or water level due to limitations in their hydraulic propagation methods. Some attempts to evaluate HR were, for instance, to assess droughts during a relatively long-term period (van Huijgevoort et al., 2013) using a large-scale-based multi-model ensemble focused on flood events (Gründemann et al., 2018); these studies call for the importance of a consistent and long temporal dataset for assessment.

To schematize and summarize the methodology of HRXX proposed in this research, Figure 1a shows the different components and processes here used, for instance the hydrological-hydrodynamic model, DA technique and bias correction as main processes to develop the HR. These processes are explained in more detail in following sections. Besides, the Figure 1b schematizes the classic sequential DA scheme and shows the time dimension of the long term DA of in-situ data proposed for this research.

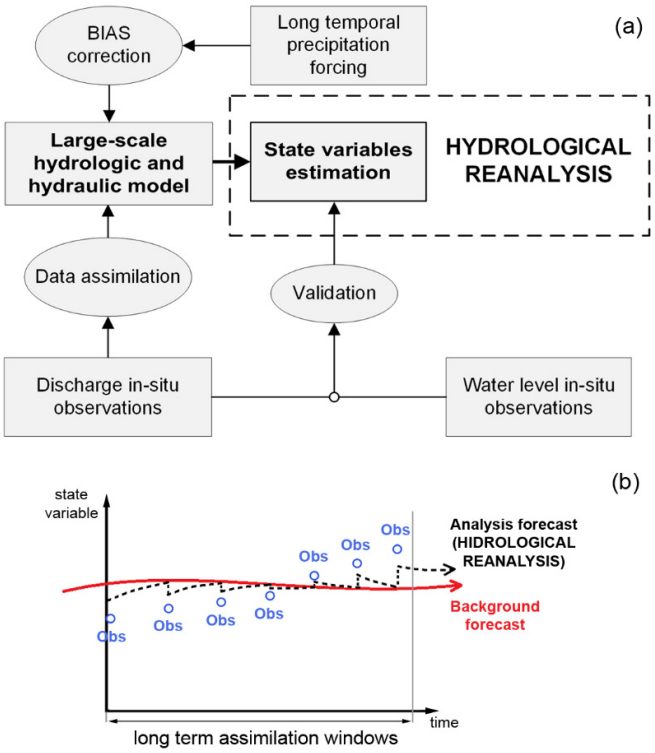

Fig. 1. (a) General chart of the hydrological reanalysis concept and (b) the time dimension of the EnKF scheme defined for this research.

4.2.2. Hydrologic-hydrodynamic model

The MGB model is a large-scale, semi-distributed hydrological model that uses physical- and conceptual-based equations to simulate the terrestrial phase of the hydrological cycle (Collischonn et al., 2007). The basin is discretized into irregular unit-catchments and further into hydrological response units (HRUs), where vertical water and energy budgets are computed individually. Each unit-catchment is also made up of a river reach that encompasses both channel and floodplain units, the latter depicted as a simple storage with ineffective flow (Pontes et al., 2017; Siqueira et al., 2018). The soil is represented as a single-layer, bucket-type model and the evapotranspiration (from soil, vegetation or open water) is calculated using the Penman-Monteith method. Surface, subsurface and groundwater runoff at each unit-catchment are produced according to water availability in the soil layer and are routed to a stream network based on a linear reservoir concept. In the most recent version, flow routing in river channels is computed using the explicit local inertial method (Bates et al., 2010; Pontes et al., 2017), which is an approximation of the full 1D Saint-Venant equations that neglect only the advective inertial term from the momentum equation (i.e., a hydrodynamic model that is able to represent backwater effects (Jacon and Cudo, 1989; Meade et al., 1991), as it happens in the Negro river at Manaus due to the stem Amazon river influence, and floodplain attenuation.

Our reasons for using the MGB model are 1) its satisfactory performance for large-scale applications (e.g., Paiva et al., 2013a; Pontes et al., 2017; Siqueira et al., 2018) and 2) the past successful attempts with the EnKF method (e.g., Paiva et al., 2013b), which demonstrate the potential of this methodology to improve spatio-temporal estimates of river discharges.

4.2.3. Long-term meteorological datasets

The 20th-century (ERA-20CM) climatic reanalysis (Hersbach et al., 2013) was chosen as the input to drive the hydrological model, mainly because the ERA-20CM can still produce statistical estimations despite assimilating no atmospheric observations (Hersbach et al., 2015), which means it is independent of synoptic observations (Gao et al., 2016) resulting in a more homogeneous performance along the temporal coverage (110 years). Further, considering previous assessments of hydrological modeling (Wongchuig et al., 2017), this dataset shows acceptable metric values (KGE and NSE) against 27 gauge observations, although BIAS still remains relatively large (average of ~20%). The ERA-20CM version used here was released in 2014 and it comprises 10-member ensemble predictions forced by the Integrated Forecast System Cy38r1, sea-surface temperature (SST) and sea ice cover (SIC) from the HadSST2 and radiative forcing CMIP5 (Titchner and Rayner 2014). The spatial resolution of the final product used herein was 0.25° and the 10 available ensemble members were assessed. The dataset is available at <https://apps.ecmwf.int/datasets/data/era20cm-daily/levtype=sfc/>, last access: 20 March 2016.

For bias correction, the HYBAM Observed Precipitation (HOP) (Guimberteau et al., 2012) dataset was used as referenced observation. HOP is a gridded daily rainfall data set (1°x1°) produced by geostatistical interpolation from 752 rain gauges across the Amazon Basin covering the period from 1980 to 2009. For more details, see Guimberteau et al., 2012 and Espinoza et al., 2016. The dataset is available at ORE-HyBam: <http://www.ore-hybam.org>, last access: 24 October 2017.

4.2.4. Bias correction

Several methodologies have been developed to remove the bias of precipitation outputs from climatic models, therefore three widely used bias adjustment methods were tested in this research. The linear scaling (LSC) approach was often assessed to remove bias of precipitation (Schmidli et al., 2006; Lenderink et al., 2007; Teutschbein and Seirbet, 2012; Eisner et al., 2012; Chen et al., 2013; Lafon et al., 2013; M'Po et al., 2016) because of its simplicity. Usually, a multiplicative correction is used for precipitation instead of an additive to avoid negative corrected values. The LSC bias removal was calculated for each unit-catchment on a monthly basis.

The disadvantage of the LSC method is that this value is used equally for extreme events and common events, and there is the possibility of overestimating corrected precipitation when: 1) the ratio of monthly values between observations and ERA is large; and 2) the daily ERA value to be corrected is larger than the corresponding observed daily value. For the LSC method to perform better, a ratio threshold was applied (LSCT), so that when ratio values exceeded this threshold, the multiplicative factor became an additive factor, and this threshold was established once several values were tested. We considered a ratio threshold greater than 1 to avoid negative values in corrected precipitation.

Because LSC is usually used to remove bias in the mean only (Lenderink et al. 2007; Teutschbein and Seirbet, 2013), quantile mapping methods were also applied, which have already been widely used in hydrological applications to correct mean and variance as well (Wood et al., 2004; Teutschbein and Seirbet, 2012, 2013; Ngai et al., 2017). The *empirical quantile mapping* (eQM) and the *parametric quantile mapping gamma distribution* (gQM) were assessed. The aim was to match the quantile of a referenced precipitation value to the observed value at the same quantile through the construction of cumulative distribution functions (CDFs) during a referenced (calibration) period. More details regarding the equations used for bias correction are described in the supplementary material.

In this study, we relied on the bias correction as a stationarity assumption, where a similar correction method is applied to both the referenced (1990-2009) and past period (before 1990).

4.2.5. Data assimilation of long-term in-situ discharge

The aim of DA techniques is the efficient use of observational information to enhance predictions of model state variables using both model and observation uncertainties (Madsen and Skotner, 2005). The EnKF method of DA, used in this research, is widely employed in hydrological applications, since it enables a simplified implementation on hydrological models where non-linear models are predominant. EnKF defines the model uncertainty as a function of the spread of the model state ensemble using a Monte Carlo approach. Implementation of the EnKF DA scheme in MGB is based on the algorithm developed by Evensen (2004) and it can be divided into two steps: forecasting and updating, as described below. For the model, the following relation holds:

$$\mathbf{x}_{k+1} = M(x_k, u_k, \theta) + q_k \quad (1)$$

where the vector **x** represents model state variables, u is the model forcing, θ is the model parameters; M is the model function that relates state variables at time k with those at time $(k+1)$; and q_k represents the model uncertainties.

The observation equation is:

$$\mathbf{y}_k = \mathbf{H}(x_k) + \varepsilon_k \quad (2)$$

where *y* is a vector with observations at time *k* (here, observed discharge), **H** a function that relates the model state variables *x* with the correspondent observation *y*, which is a matrix constructed with 1 values where there is a model prediction of the observation and 0 where there is not, and ε means the observation uncertainty.

In the stochastic formulation of EnKF, the matrix of the forecasted (or background) ensemble of model states $\mathbf{x}^f$ is represented as:

$$\mathbf{x}^f = \left(\mathbf{x}_1^f, \mathbf{x}_2^f, \dots, \mathbf{x}_{N_{ens}}^f\right) \tag{3}$$

where $\mathbf{x}_i^f$ represents each ensemble member of the model states until the total number of defined members (N_{ens}).

The model state matrix x^a is updated as the response to the new available observations, as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right) \tag{4}$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left[\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{5}$$

where the $\mathbf{x}^a$ and $\mathbf{x}^f$ are the analysis (updated) and forecasted (background) model state variable respectively, **K** is the Kalman gain, $\mathbf{P}^f$ and **R** are the covariance matrices of errors of the model (q) and observation (ε), respectively.

The particularity in the EnKF method is that the model error covariance matrix is generated by a set of members from perturbations of the model (forcing, state variables, parameters), based on a priori-known error. This error covariance matrix is estimated from the ensemble anomalies (Clark et al., 2008), as follow:

$$\begin{aligned} \mathbf{P}^f \cong \mathbf{P}_e^f &= \overline{\left(x^f - \overline{x^f}\right)\left(x^f - \overline{x^f}\right)^T} \\ \mathbf{P}^a \cong \mathbf{P}_e^a &= \overline{(x^a - \overline{x^a})(x^a - \overline{x^a})^T} \end{aligned} \tag{6}$$

where $\mathbf{P}_e^f$ and $\mathbf{P}_e^a$ are the covariance matrices of the model for the forecast and analysis respectively, and T super index means the transpose operator.

4.2.6. Experiments and case study

4.2.6.1. Amazon case study and model setup

The Amazon Basin was chosen as a proof of the concept of HRXX because it is the world's largest basin, and it drains more than 6 million km^2 and discharges ~15% of the freshwater that reaches the world's oceans, yet its ground observational network is scarce (Willmott et al., 1994; Marengo et al., 2005), and its water resources meet many human needs, such as fluvial transportation, agriculture,

fisheries and energy production. The Amazon Basin is experiencing a noticeable process of biophysical transition (Nobre et al., 2016), and hydrological extreme events such as floods and droughts have been recorded during the last years, with reported increasing of magnitude and frequency due to rainfall intensity (Marengo et al., 2013, 2016; Espinoza et al., 2018; Wang et al., 2018; Barichivich et al., 2018). Furthermore, there is an evident pattern of an upward trend of the frequency of extreme floods in the northern and on the main stem region (Callède et al., 2004; Espinoza et al., 2009; Wongchuig et al., 2017), and an upward trend of extreme drought events in southern regions (Lopes et al., 2016; Molina-Carpio et al., 2017).

Herein, we applied the same model setup and parameterization used in the South American MGB version from Siqueira et al., (2018). As such, the basin was discretized using a fixed river length of Δx = 15 km to provide a balance between computational efficiency and numeric stability with respect to the hydrodynamic routing, resulting in 12,466 unit-catchments ranging from 20 to ~2600 km^2. The model was forced with a 0.25º horizontal resolution MSWEP v1.1 precipitation (Beck et al., 2017a) and long-term monthly meteorological data from the CRU CL 2.0 dataset (New et al., 2002). In addition, the model was calibrated between 1990 and 2010 using daily river discharges and was validated using multiple data sources, including water levels (in situ and satellite altimetry), terrestrial water storage and evapotranspiration estimates. Further details about model setup and performance can be found in Siqueira et al. (2018). Datasets available at <http://www.gloh2o.org/> for MSWEP rainfall dataset, <https://crudata.uea.ac.uk/cru/data/hrg/tmc/> for CRU climatological dataset, and <http://www.snirh.gov.br/hidroweb/publico/medicoes_historicas_abas.jsf> for in-situ daily discharge.

4.2.6.2. Assimilation parameters and localization approach

In this research, assimilation parameters were assumed as known, only the precipitation was perturbed in this study, considered as the main forcing to the model (Biancamaria et al. 2011; Liu et al., 2012a; Paiva et al., 2013b). The adopted perturbation follows a log-normal distribution of errors, as proposed by Nijssen and Lettenmaier (2004):

$$P_c = \frac{1+\beta}{\sqrt{E^2+1}} exp\left(\sqrt{ln[E^2+1]}s\right) P \quad (7)$$

where P_c is the perturbed variable (precipitation or model state variables), P is the non-perturbed variable, E is relative error (%), β is relative bias (here adopted as zero), and s is a random variable that follows the time evolution of model error according to Evensen (2003):

$$S_k = \alpha S_{k-1} + \sqrt{1-\alpha^2} w_{k-1} \qquad \alpha = 1 - \frac{\Delta t}{\tau_t} \quad (8)$$

where the coefficient $\alpha \in [0\ 1\rangle$ indicates the influence of temporal correlation on the model errors (α=0 indicates a sequence of white noise, while α=1 indicates a random field of errors constant

in time, which removes the stochastic forcing); w is a stochastic term with mean 0 and variance 1 following a Gaussian distribution; and τ_t is the temporal decorrelation length (time units).

In this study a function of the discharge observation (q_{obs}) was adopted for the estimation of the matrix covariance of observation errors (σ_{obs}^2) (Clark et al., 2008).

$$\sigma_{obs}^2 = (\varepsilon_{obs} \cdot q_{obs})^2 \tag{9}$$

where ε_{obs} is the discharge error parameter that could be considered as the uncertainty on river discharge measurement, and a value of ~10% was adopted in principle, based in previous assessments (Di Baldassarre et al., 2009; Clark et al., 2008; Paiva et al., 2013b).

To afford an adequate representation of the error covariance matrix, the EnKF methodology needs a sufficient ensemble size to avoid spurious correlations between large rivers with a distant reach. In this context, to avoid this problem, Houtekamer and Mitchell (2001) proposed ignoring remote observations in the analyzed local point through the use of the "covariance localization" method. A correlation matrix was estimated using a distance-dependent correlation function (Gaspari and Cohn, 1999). Finally, the incorporation of a covariance localization matrix (Sakov and Bertino, 2011) to the equation, referred as a local ensemble Kalman filter (LEnKF) (Houtekamer and Mitchell, 2001, 2005; Ott et al., 2004; Tong, 2018), is expressed:

$$\mathbf{x}^a = \mathbf{x}^f + \boldsymbol{\rho} \times \left(\mathbf{P}^f\mathbf{H}^T\right)\left[\boldsymbol{\rho_o} \times \left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T\right) + \mathbf{R}\right]^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right) \tag{10}$$

where $\boldsymbol{\rho}$ and $\boldsymbol{\rho_o}$ are correlation matrices and × represents the Schur product.

The correlation matrices were calculated using a fifth-order function as determined by Gaspari and Cohn (1999) and suggested by Hamill et al. (2001) and Houtekamer and Mitchel (2001):

$$\boldsymbol{\rho}(r,e) = \begin{cases} 1 - \frac{1}{4}\left(\frac{e}{r}\right)^5 + \frac{1}{2}\left(\frac{e}{r}\right)^4 + \frac{5}{8}\left(\frac{e}{r}\right)^3 - \frac{5}{3}\left(\frac{e}{r}\right)^2, & 0 \le e < r; \\ \frac{1}{12}\left(\frac{e}{r}\right)^5 - \frac{1}{2}\left(\frac{e}{r}\right)^4 + \frac{5}{8}\left(\frac{e}{r}\right)^3 + \frac{5}{3}\left(\frac{e}{r}\right)^2 - 5\left(\frac{e}{r}\right) + 4 - \frac{2}{3}\left(\frac{e}{r}\right)^{-1}, & r \le e \le 2r; \\ 0, & e > 2r; \end{cases} \tag{11}$$

where e represents the distance along the river network between the observation and the analyzed river reach, and r means the radius of influence.

For this implementation, $\boldsymbol{\rho}$ is a *n* x *m* dimension, wherein each column presents the correlations at one observation location (unit-catchment) with all model state variables of all unit-catchments, and $\boldsymbol{\rho_o}$ is an *m* x *m* dimension, wherein each column presents the correlations at one observation with all other observations located in their respective unit-catchment. In this research, EnKF and LEnKF are related to Eq. 5 and 10, respectively.

4.2.6.3. Assimilation and validation datasets

Daily in-situ discharge and water level observations from 114 gauges were used in this research, divided almost equally for assimilation within the DA scheme (57) and for validation purposes (57) (Fig. 2). For assimilation and validation criteria, gauges with a coverage greater than a 30-year record and a 15-year record were chosen, respectively. These datasets were provided by the Environmental Research Observatory SO HYBAM (ORE-HyBam: <http://www.ore-hybam.org>, last access: 18 February 2017)) and the National Water Agency (ANA-Brazil) <http://www.snirh.gov.br/hidroweb/>, last access: 09 August 2017. A table with more detailed information about in-situ gauges and data quality control used in this research can be found in the supplementary material.

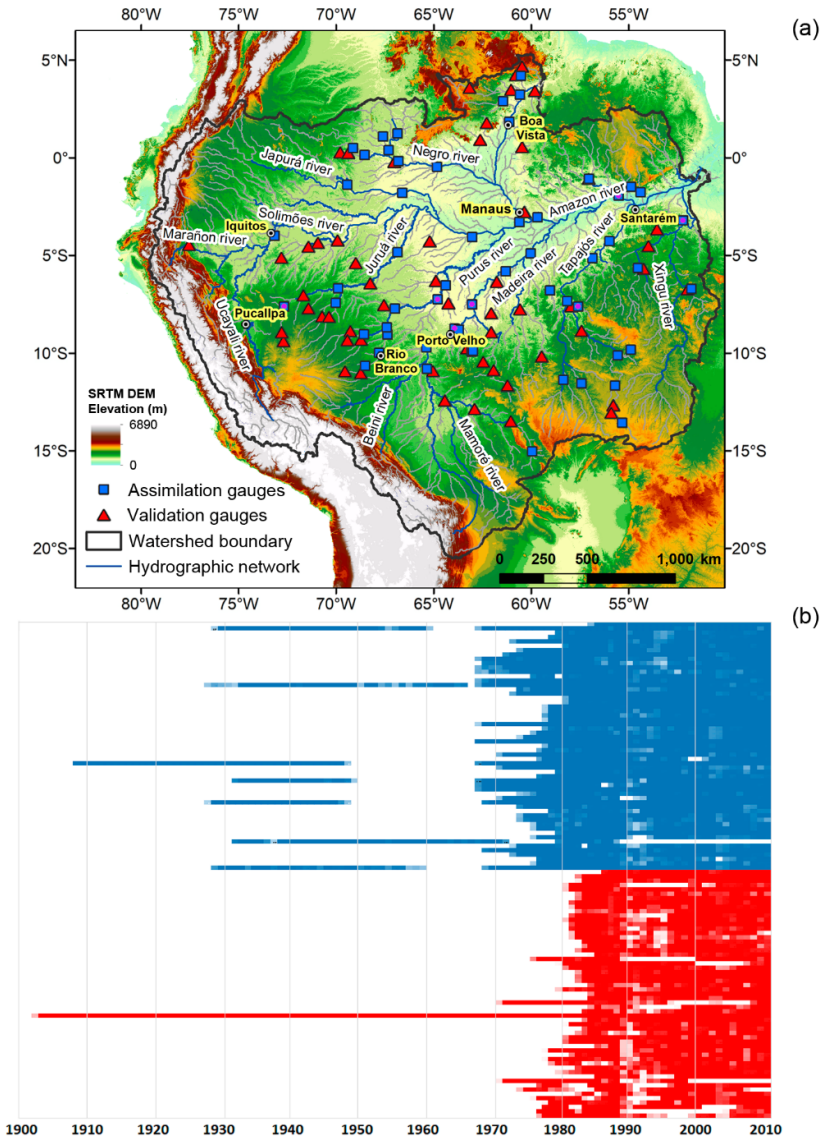

Fig. 2. (a) Amazon River Basin limits with river network, relief map from SRTM DEM, and in-situ gauges used for assimilation (blue) and validation (red); and their (b) temporal coverage. Remarked squares in magenta circles correspond to assimilation gauges with long temporal records, and main cities' names are highlighted in yellow mask with markers in black and with circles.

4.2.6.4. Experiments for Hydrological Reanalysis evaluation

The following procedures were developed to achieve the HRXX: 1) selection of the ensemble members of precipitation from the ERA-20CM climatic reanalysis, based on their performances; 2) determination of the optimal radius of influence for the case study region through the LEnKF DA scheme; 3) assessment of the bias removal methods of daily precipitation forcing in comparison with the HOP; 4) generation of the HRXX through the use of both the optimal radius of influence (within LEnKF) and the best bias corrected precipitation forcing. The statistics indexes Nash–Sutcliffe efficiency (NSE), the Kling–Gupta (KGE), the absolute value of stream flow volume error (hereafter referred to as BIAS) and the relative reduction in root-mean-square error (ΔRMSE) were used to assess performance in daily discharge and water level values produced by the HR. The NSE is a normalized statistic that determines the relative magnitude of residual variance compared with the observed data variance, while the KGE offers interesting results in model performance, because it uses the most components, such as correlation, variability and bias terms. Both NSE and KGE range between -∞ and 1 (perfect fit). The ΔRMSE compares relatively the performance (decreasing errors) of the model simulation using the DA scheme with respect to the OL (free-run) simulation.

$$NSE = 1 - \frac{\sum_{i=1}^{nt}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\sum_{i=1}^{nt}(Q_{obs}(t) - \overline{Q_{obs}})^2} \tag{12}$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \tag{13}$$

$$BIAS = \left|\frac{\sum_{i=1}^{nt}(Q_{sim}(t)) - \sum_{i=1}^{nt}(Q_{obs}(t))}{\sum_{i=1}^{nt}(Q_{obs}(t))}\right| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{nt}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2} \qquad \Delta RMSE = \frac{RMSE_{improved}{}^{*} - RMSE_{open-loop}}{RMSE_{open-loop}} \tag{15}$$

where nt is the number of observations, σ is the standard deviation and μ is the mean, and $RMSE_{improved}{}^{*}$ can be referred to as bias-correction or assimilation.

For precipitation forcing, we assessed the 10 ensemble members available from the ERA-20CM climatic reanalysis to determine the most suitable of them for following analysis. The OL simulations were carried out for each member, and all discharge gauges (113) were used for validation during the common period from 1980 to 2010. The statistical performance indexes and fewer errors with respect to the mean of the ensembles were considered when choosing the ensemble members for the following analysis. It is important to note that the evaluation of the 10 ensemble members can provide an indication of the uncertainty of the precipitation forcing, which could be taken into account for the perturbation of forcing within the DA scheme.

A general proxy was developed to estimate the optimal radius of influence for the basin in the study. We tested a range of radii between 50 and 5000 km, which was selected considering closest values of influenced area in the Amazon Basin (Emery et al., 2018). The size of 100 members was determined by previous EnKF implementation in the Amazon Basin (Paiva et al., 2013b) and for other hydrological studies (Huang et al., 2008; Tong et al., 2012); nevertheless, this could result in a huge computational demand because of: 1) the long simulation/assimilation window (110 years); 2) the large spatial discretization of the study area; and 3) several radius of influence scenarios to be assessed. Therefore, for this specific experiment only, the ensemble size of 25 members was used to estimate the optimal radius of influence. The validation was carried out using a set of 56 in-situ discharge gauges described in section 2.6.3, during the period from 1980 to 2010.

To assess the bias removal methods of ERA-20CM precipitation, a referenced period from 1990 to 2009 and a validated period from 1980 to 1989 were considered. These periods were chosen because observation data availability (1980–2009) from the HOP dataset, described in section 2.3. The validation performance was assessed for both precipitation and discharge. For discharge, we carried out two periods of validation: 1) from 1980 to 1989 using all discharge gauges (113); and 2) assessing the performance of bias removal methods for past eras during the common period from 1931 to 1948, we also used seven discharge gauges with long records. For precipitation validation, all unit-catchments were assessed from 1980 to 1989.

For the HRXX, two scenarios of assimilation were considered: 1) using all assimilation gauges (57), validated against 56 gauges; and 2) using only seven gauges with the longest time records, validated against the 106 remaining gauges. The second scenario was assessed because we consider validation necessary when just a few gauges (seven) are used for assimilation purposes, which happened the most during the period from 1908 to 1970 of the simulation. Finally, also assessed was the performance of the anomalies in water levels at the longest data record in the Amazon Basin, located in Manaus (which was not used for assimilation purposes due to Manaus records are controlled by Solimões river level variability generating a hysteresis effect in the rating curve). It was carried out from 1903 to 2010.

The methodology developed in this research has the objective of estimating a long and consistent spatio-temporal register of hydrological variables, mainly discharge and water level, through: 1) hydrologic-hydrodynamic modeling; 2) a DA method; 3) a simple localization approach; and 4) bias correction of precipitation forcing. To summarize these procedures, Figure 3 and Table 1 show the general framework of the methodology and experiments that were applied in this research.

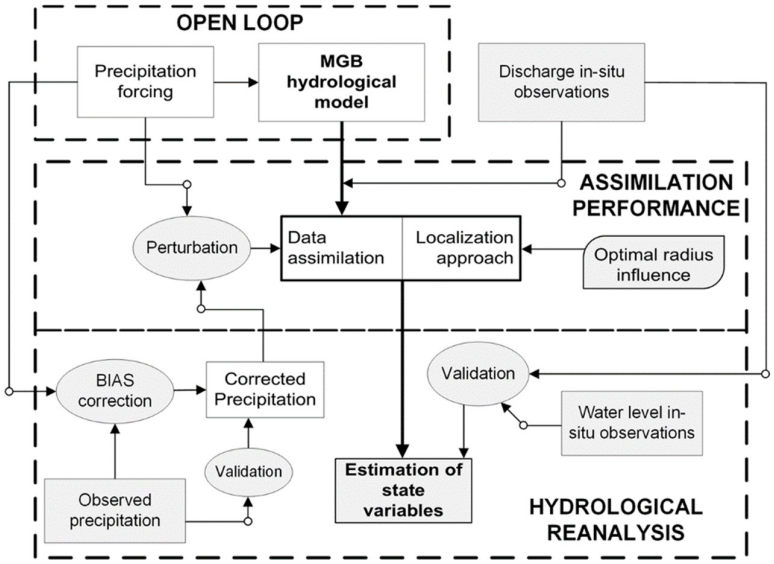

Fig. 3. General scheme describing the methodology carried out in this study.

Table 1. Description of the setup for each experiment assessed in this research.

Experiments	DA scheme	Localization	Bias removal	Assimilation gauges	Validation			Goal
					Variable *	Period	Gauges	
1. OL	No	No	No	None	D	1980–2010	113/56/106	To determine the ensemble of ERA-20CM/ Background
2. OL bias corrected	No	No	Yes	None	D	1980–1989 / 1931–1948	113/7	To determine an adequate bias removal method
						1980-2010	56/106	
3. Precipitation bias corrected					P	1980–1989	All unit-catchments	
4. EnKF	Yes	No	No	All (57)	D	1980–2010	56	Background of localization scheme
5. LEnKF	Yes	Yes	No	All (57)	D	1980–2010	56	To determine the optimal radius of influence for Amazon Basin
6. HRXX based on only seven gauges	Yes	Yes	Yes	Longest (7)	D	1980–2010	106	To fairly compare the performance when only a few stations are assimilated, especially before 1980
7. HRXX	Yes	Yes	Yes	All (57)	D	1980–2010	56	Final product
				Longest (7)			106	
				All (57)	WA	1903–2010	1	

* D = Discharge, P = Precipitation, WA = Water level anomalies

4.3. Results and discussions

4.3.1. Open-loop ensemble members' performance

The OL simulation, also known as free-run simulation, was considered as the background for following tests. As we expected, the OL simulations using the 10 ensemble members of ERA-20CM climatic reanalysis showed medium performance in the Amazon Basin according to the statistical

indexes. This performance was already noted in previous studies (Wongchuig et al., 2017). The average performances for the median of 10 members of the ensemble were 0.61, 0.40 and 20.6% for KGE, NSE and BIAS, respectively.

There was no clear significant difference among the members' performance for daily discharge, according to a student's t-test at 95% confidence level. Therefore, for the following analyses, the members with less RMSE in the daily precipitation with respect to the mean of the 10 ensemble members was chosen. Nevertheless, the uncertainty of the 10 ensemble members can also be considered, through their ensemble standard deviation, for the perturbation of the precipitation within the DA scheme.

4.3.2. Assessment of the localization approach

A general approach to estimating the optimal value for the radius of influence across a river routing network is presented in this section through the use of the polynomic equation developed by Gaspari and Cohn (1999). The localization experiments were carried out considering a range of radius values closest to previous studies (Emery et al., 2018). The spatial distribution of the correlation values from the 57 assimilation gauges, considering several radii of influence, is shown in Figure 4.

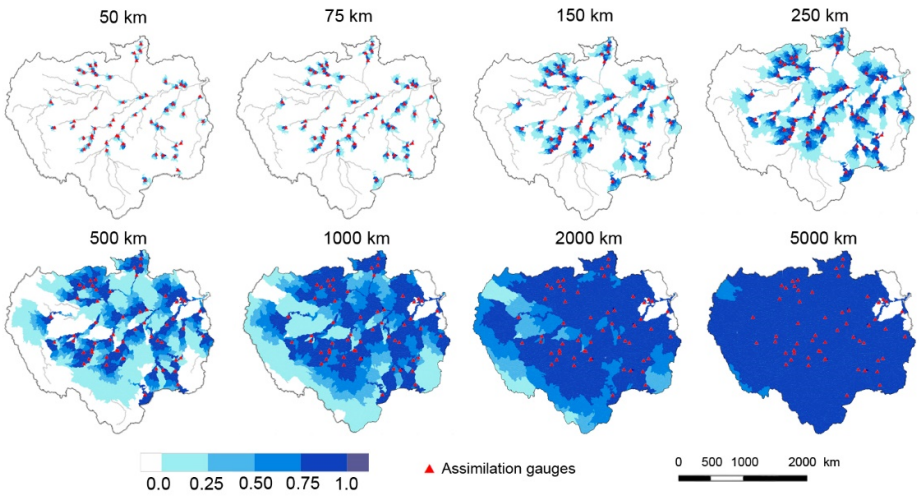

Fig. 4. Map of correlation values according to radius influence, considering all assimilation gauges.

For each simulation, statistical indexes were applied to find the optimal radius. Figure 5 shows the boxplots of performance considering the validation and assimilation gauges. The LEnKF simulations showed more clearly the sensible approach of the use of the localization technique in assimilation (Fig. 5b) than in validation (Fig. 5a) gauges, according to the statistical indexes. When the localization technique was applied (LEnKF), the performance increased directly proportional with the radius of influence until a certain value.

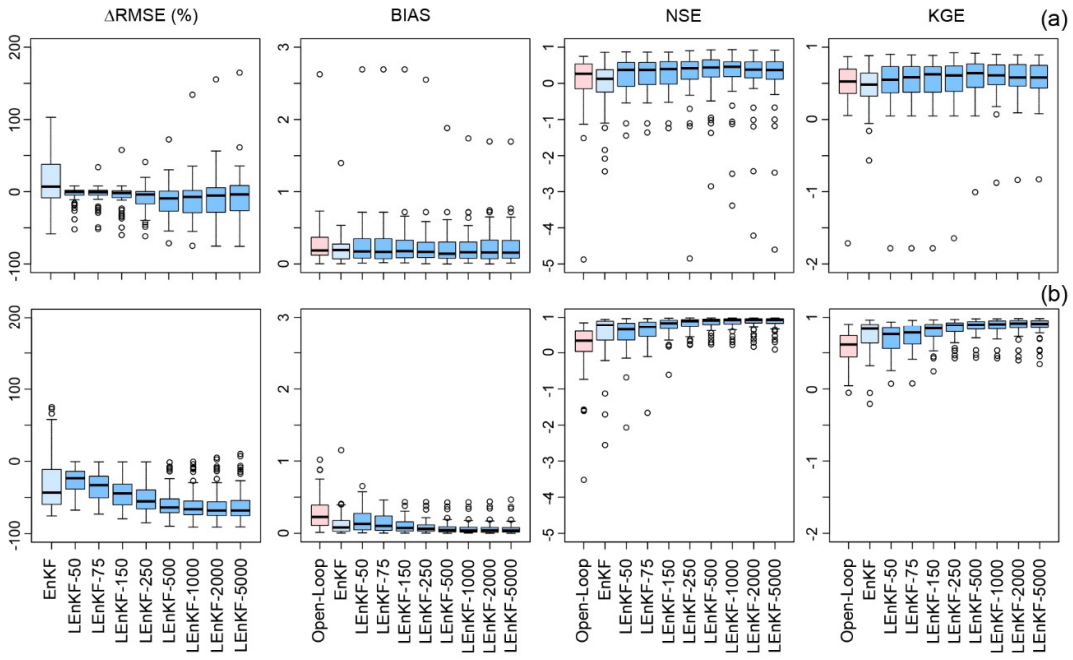

Fig. 5. Boxplots of RMSE, BIAS, NSE and KGE index considering the open-loop (red), the assimilation without localization (sky blue) and with localization (blue) for several radii of influence in (a) validation and (b) assimilation gauges.

Figure 6 shows a general increase in performance when different radii of influence are applied within the localization method. In addition, better results were found when validation is performed in gauges with large contribution areas (> 25,000 km^2). It is interesting to note that for the ΔRMSE metric, EnKF showed an increase in errors compared with the OL when all the validated gauges were used, which can be explained due to the spurious correlations between the remote locations of assimilation gauges and the updated unit-catchments. The opposite behavior was observed when validation gauges with large contribution areas were analyzed, because these are usually closest to the assimilation gauges used: those with long series and therefore located in biggest contribution areas.

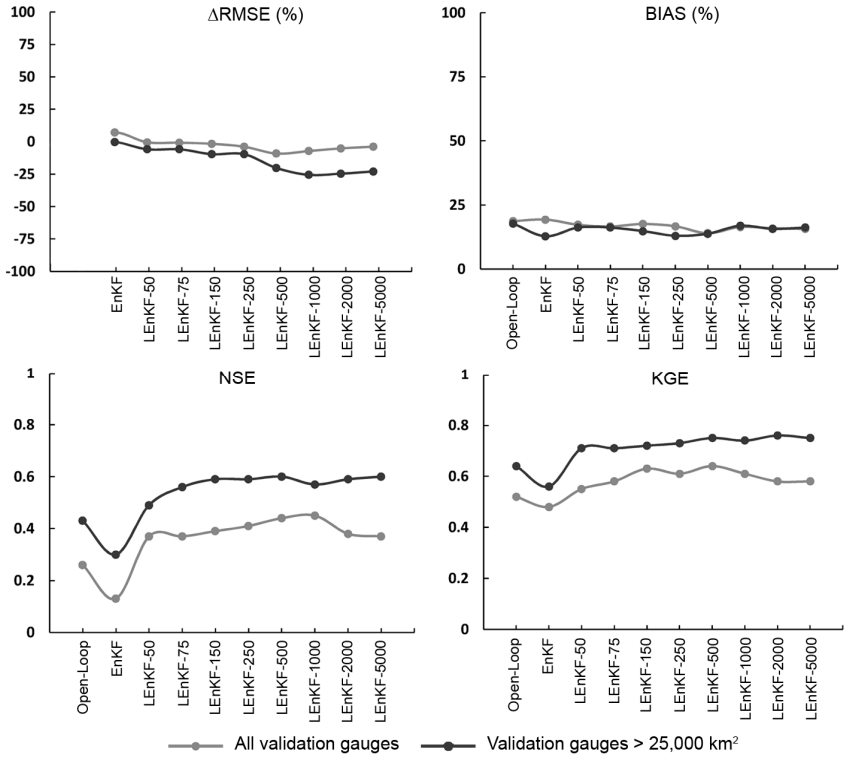

Fig. 6. Median of the performance for open loop, EnKF and LEnKF for different radii of influence in validation gauges.

In general, the localization technique greatly improved DA performance, which was already observed by Paiva et al. (2013b) for large basins such as the Amazon Basin. In general, the optimal values were found between 500 km and 2,000 km, which fits close to values found by Emery et al., (2018). Consequently, the radius of influence of 1000 km was chosen because of its relatively good performance and greater spatial coverage (Fig. 4), ~92% of which represents ~5.4 10^6 km^2. The following simulations were assessed using 100 ensemble members.

4.3.3. Validation of data assimilation scheme, localization and bias correction

In this section, we present the performance of the EnKF DA scheme in validated gauges when a size of 100 ensemble members and an optimal radius of influence of 1000 km, as established in previous section, were used (LEnKF). Figure 7 shows improvements in performance for all statistical indexes for the assimilation and validation gauges when the localization technique is used. For the

median of validation gauges, the ΔRMSE index improved from -1% to -7% (Fig. 7a). Considering the comparison between the OL (red) and LEnKF (blue), the statistics were consistent for the BIAS, NSE and KGE, improving from 19% to 17%, from 0.28 to 0.47 and from 0.56 to 0.62, respectively. Improvements were found for the NSE index for ~18% of the validation gauges, in which values less than 0.50 went to greater than 0.75; and approximately 24% of all gauges with values greater than 0.25 of BIAS were reduced to 19%.

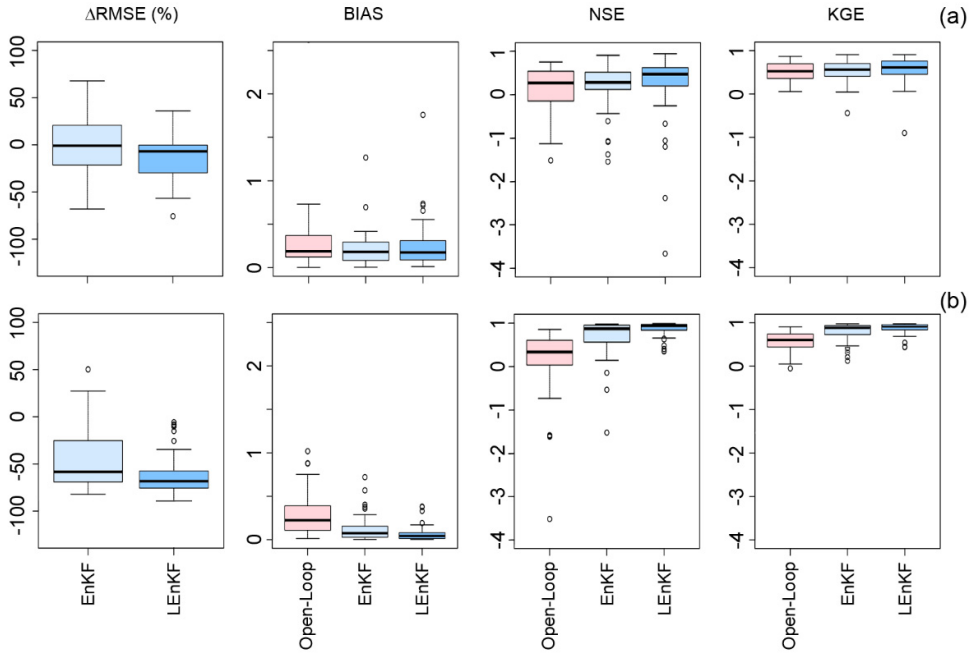

Fig. 7. Boxplots of RMSE, BIAS, NSE and KGE index considering the open loop (red), EnKF (sky blue) and LEnKF (blue) for optimal radii of influence in (a) validation and (b) assimilation gauges.

The performance of bias correction methods was assessed against precipitation and discharge records during the validation period (1980–1989), and for the longest in-situ records of discharge gauges during the oldest period (1931–1948). Figure 8 shows the analysis of mean monthly precipitation during the validation period, where HOP, ERA-20CM raw and ERA-20CM bias removed were plotted. For the northwest region, the precipitation was underestimated almost all year for ERA-20CM reanalysis, and mainly during the rainy season in the north region. For the western area, precipitation was underestimated during the dry period and overestimated during the wet period. For the central and southern regions, the precipitation was underestimated mainly during the dry season, and on average, this behavior was also seen for the whole Amazon Basin. The bias correction methods improved monthly precipitation in all regions; we can also see the LSC method (green lines) overestimating values during some dry months, which is a problem for this specific method and was described in section 2.4. Overall, the LSC threshold (LSCT) method has the best performance to correct precipitation on a monthly scale,

considering the Pearson correlation coefficient, which improved from 0.97 for the ERA-20CM raw to 0.99 for the LSCT.

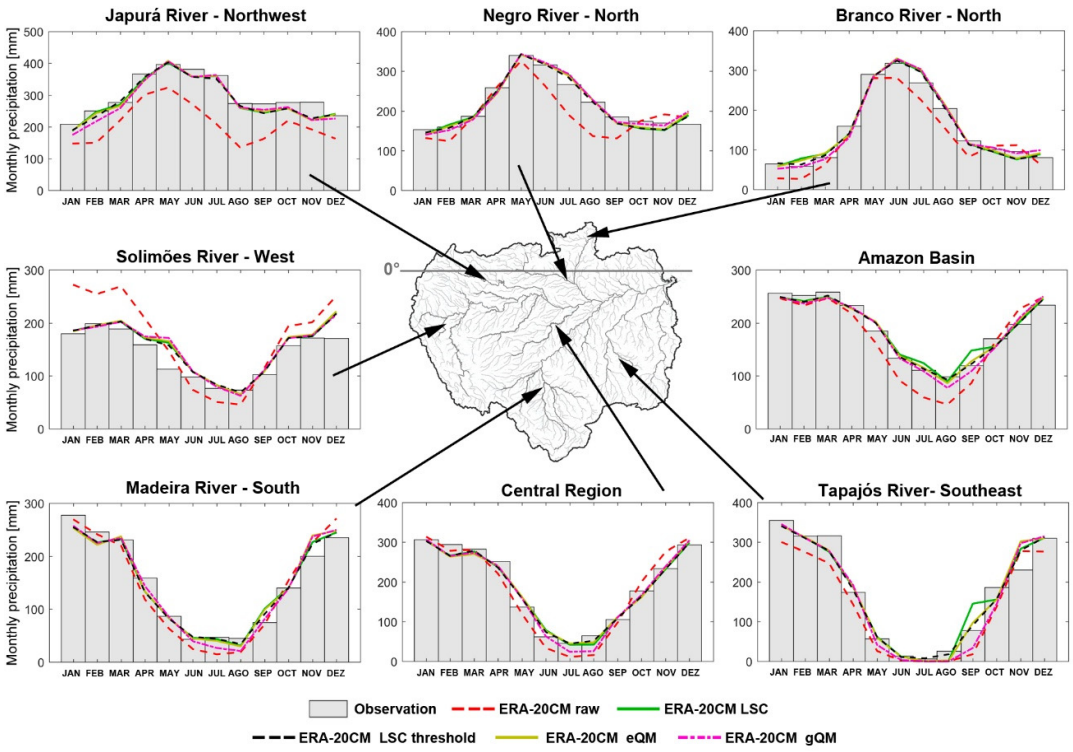

Fig. 8. Mean monthly precipitation during validation period (1980-1989) for HOP (gray bars), ERA-20CM raw and ERA-20CM with bias removal methods in different regions and for the entire Amazon Basin.

Figure 9 shows the bias values for the ERA-20CM precipitation forcing spatially distributed and aggregated in tri-monthly means, before and after bias removal. The ERA-20CM shows peculiar behavior in the extreme rainfall regions (hotspots), as related by Wongchuig et al. (2017), where the MSWEP and ERA Interim Land rainfall datasets also presented the particular characteristic of overestimation in the eastern flank of the Andes while showed underestimation in most hotspots, similarly to Espinoza et al. (2015) using the TRMM-PR dataset. The bias removal method LSCT (Fig. 9b), which has best performance, was effective in constraining bias between the ranges of -5% and 5% for ~80% of the Amazon Basin's area during the span of the referenced and validation periods.

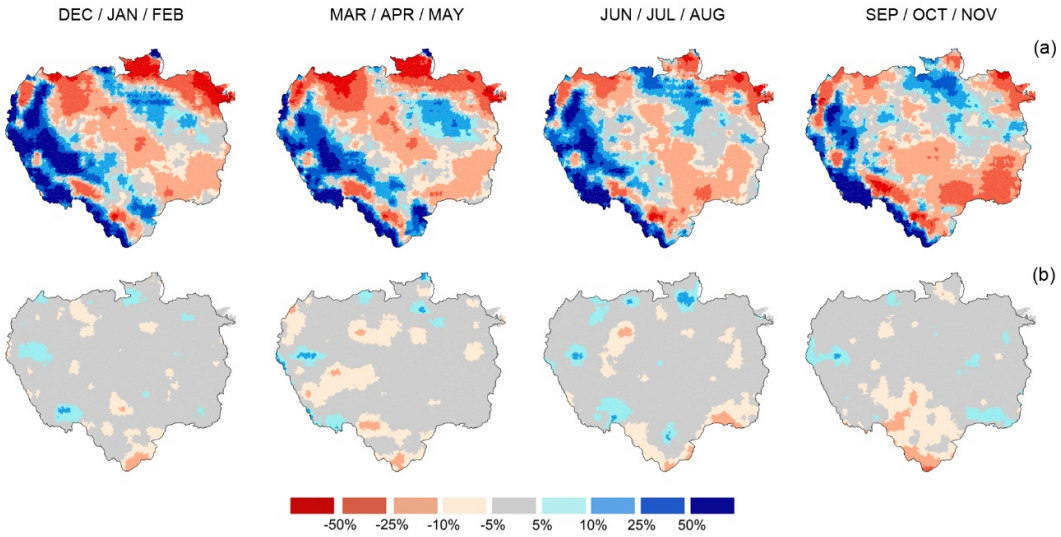

Fig. 9. Percentage of bias from long-term monthly mean precipitation for (a) ERA-20CM raw data and (b) corrected ERA-20CM version, in relation to the HOP dataset (1980–2009).

The daily observed discharge values were also used to assess the bias removal performance. Figure 10 and Table 2 show the statistical index values of daily discharge for the OL simulations before (gray) and after (sky blue) each bias removal method. We can observe that the best improvement on bias was done by the LSCT method during the validation period, which was reduced from 20.1% to 14.3% considering the average of all discharge gauge (113). There was also an improvement in the NSE and KGE indexes, using the LSCT method, from 0.32 to 0.44 and from 0.55 to 0.63, respectively. The OL LSC shows the worst performance (Table 3), even worse than the OL ERA raw during the validation period.

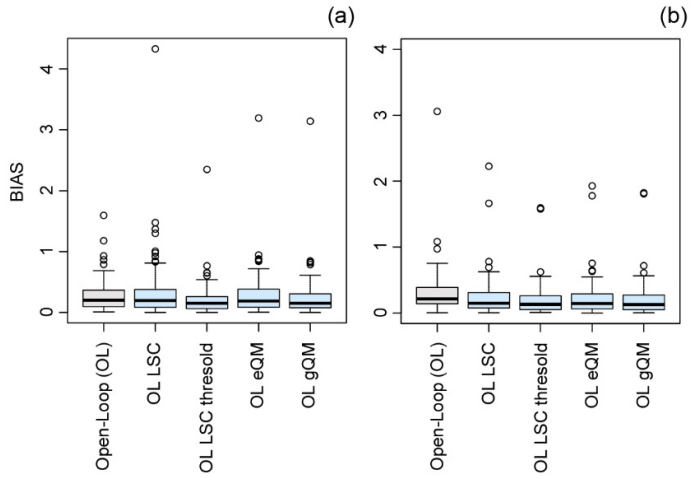

Fig. 10. Boxplot of BIAS discharge performance using corrected precipitation by different methods of bias correction (sky blue), for (a) validation period 1980–1989 and (b) referenced period 1990–2009.

Table 2. Comparison of median statistic index for bias correction methods for precipitation during validation (1980–1989) and referenced period (1990–2009).

Statistic index	Validation period			Referenced period		
	BIAS (%)	**NSE**	**KGE**	**BIAS (%)**	**NSE**	**KGE**
OL ERA raw	20.1	0.32	0.55	21.4	0.31	0.56
OL LSC	19.9	0.22	0.53	14.7	0.38	0.59
OL LSC threshold (LSCT)	**14.9**	**0.44**	**0.63**	**13.1**	**0.48**	**0.64**
OL eQM	18.7	0.21	0.54	14.1	0.33	0.60
OL gQM	15.5	0.31	0.61	12.6	0.39	0.62

* Values in bold represent the best results.

The hydrographs of the OL simulation and the OL using precipitation with bias removal are shown in Figure 11. The three best and the three worst gauges, considering the average bias of all methods, were chosen to be plotted during five years for the validation and referenced periods. For instance, during the validation period (Fig. 11a and b), the bias removal methods performed better at the São Felipe (northern region), Valparaiso and Ipixuna (south-western region) stations in relation to OL ERA raw (dashed red line); this improvement is clearly indicated by the reduction of peaks during the rainy season, while OL ERA raw overestimates discharge values. This correction is evidently noted at the Ipixuna station (Fig. 11a, where bias was reduced from 68% to 7% during the validation period. Further, the Ji-Parana and Boa Esperança (south-central and eastern regions respectively) gauges (Fig. 11b) performed the worst for bias. It is interesting to note that at the Altamira (eastern region) and Boa Esperança gauges, there is a peak on discharge (green line) at the end of 1986 when the LSC method was used, which explains the decreasing performance when this method was assessed (Table 2). This problem of overestimation was mentioned in section 2.4, which justifies the application of the threshold on the ratio and what results in a better general performance.

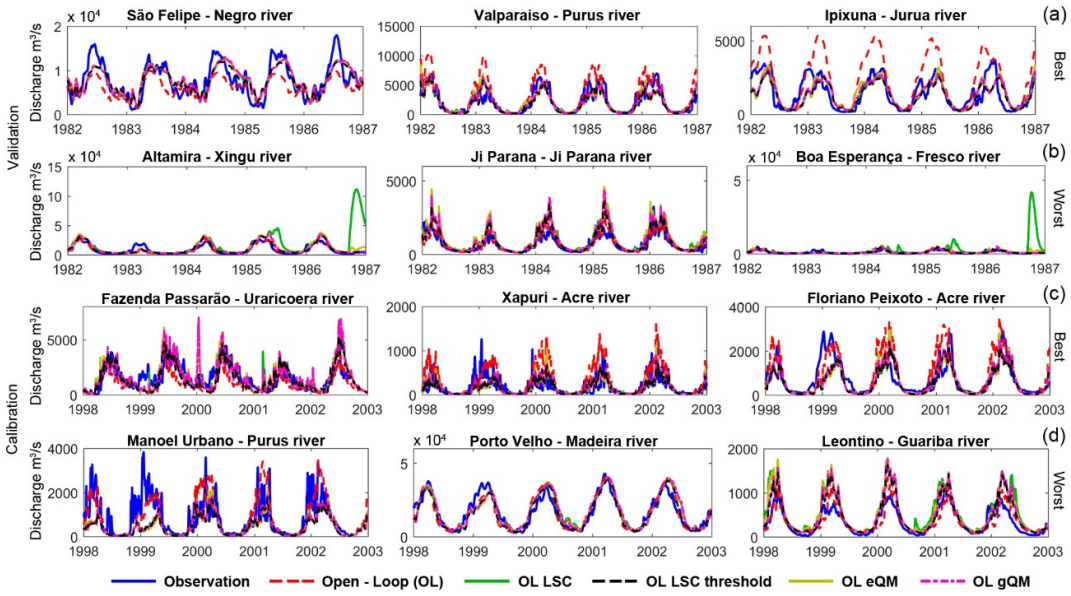

Fig. 11. Hydrographs during five year of the validation period (1980–1989) and the referenced period (1990–2009) for best (a and c) and worst (b and d) gauge performances considering bias mean efficiency of all bias correction methods in relation to OL simulation.

For the referenced period (Fig.11c and d), the bias of the three stations plotted were improved by all bias correction methods, reducing the overestimation on discharge by OL ERA raw in almost all years, for instance, in the Fazenda Passarão gauge (northern region), where bias was reduced from 34% to 1%. On the other hand, for the worst gauge stations, discharge with bias correction was not much different in comparison with OL ERA raw, as shown mainly at the Porto Velho and Leontino gauges.

The hydrographs for the seven gauges with longest records are shown in Figure 12 and geographical location of each station can be seen at the Figure 1s in supplementary material. The assessment of bias removal methods in these gauges illustrates their performances during a distant temporal era from the referenced period. Improvements can be clearly seen at the Labrea, Cruzeiro do Sul and Óbidos gauges, where the peaks were removed for the two first-mentioned gauges and enhancements during the recession period for Óbidos. During the common period from 1931 to 1948, the bias improvement for the average of the seven gauges went from 18% to 9%, and achieved the best enhancement at Óbidos, from 14% to 3.4%, which is generally remarkable because the referenced period is temporarily distant. These results indicate that the precipitation input dataset has mainly a systematic bias, as described in section 2.3, which was acceptably removed by the bias removal methods that mesh with their non-stationary assumptions.

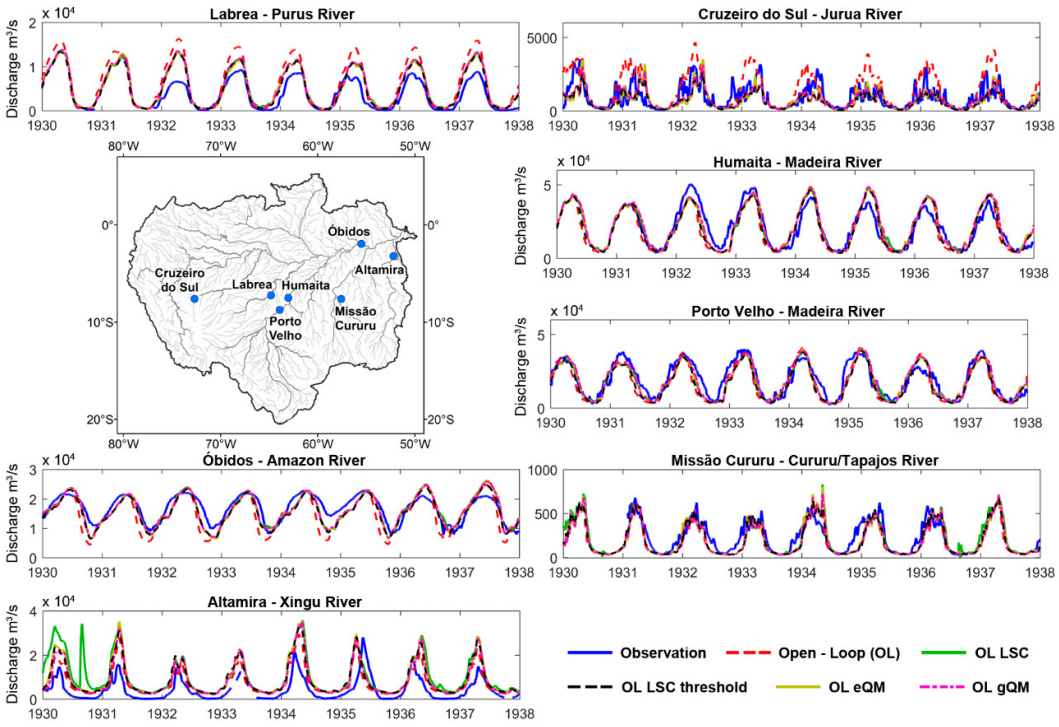

Fig. 12. Hydrographs over eight years within a common period from 1931 to 1948 for the seven gauges with the longest records.

Due to statistical performance, we have chosen the OL LSCT to assess the further experiments for the HR development, which means that the final spatio-temporal analysis is going to take the best setup of the LEnKF scheme, when and where observation records were available within an optimal radius of influence, in a synergy with the improved bias removed precipitation forcing during the years 1900 to 2010.

Figure 13 in section 3.4 displays the hydrographs for the assimilation and validation gauges at some stations during a period when observation records are partially available. The bias of the OL (red line) simulations is shown clearly at the Serrinha, Óbidos, Labrea and Cruzeiro do Sul gauges, while discharge values were improved by LEnKF (gray and dashed black lines) when observations were available, which is clearly seen since 1968. It is interesting to see how this improvement spread to the validation gauges, most clearly at Fazenda Borangaba and Santos Dumont. By contrast, discharge improvements were not observed well before 1968 for LEnKF without bias removal (gray line). The LEnKF scheme was used to improve state variables in spatial and temporal fields using a local assimilation, nonetheless this improvement is done when and where observations are available within an area of influence, which means that state variables remain biased for several unit-catchments before ~1970 due to: 1) biased precipitation ERA-20CM forcing; 2) limited temporal availability from in-situ

observation records to be assimilated; 3) spatial restriction of state variables updated by localization technique. To avoid these constraints and in order to assess the best dataset in as great an area and as much temporal coverage as possible, the bias correction was performed to the precipitation forcing from 1900. Even though we considered the stationarity assumption of bias correction methods, the climatic variability should be represented by in-situ observations used in the DA scheme. Afterwards the corrected precipitation and the LEnKF scheme will be used to develop the HRXX dataset.

4.3.4. HRXX and multi-decadal behavior

After validating the performance of the LEnKF DA scheme and bias correction of the precipitation forcing, HRXX in the Amazon Basin was developed considering the conditions for the optimal radius of influence and best bias correction method. HRXX was forced with the bias corrected precipitation since 1900, and the DA scheme of in-situ discharge records was effectively done by data availability since 1908, ~1928 and mostly since the end of the 1960s.

Figure 13 illustrates the hydrographs for the assimilation and validation gauges at some stations when observation records are partially available to denote the improvements by both the DA scheme and bias correction. For Serrinha and Jusante da Cachoeira do Caju (Fig. 13a), the assimilated observations were available from late 1967, when the improvements to LEnKF (gray and dashed black lines) compared with the OL (red line) are evidenced by a reduction of peaks in the assimilated gauge, which spread to the validation gauge appreciably since the middle of 1982, when observations are available. Similar behavior occurs at Labrea with Fazenda Bocanegra and at Cruzeiro do Sul with Santos Dumont, where improvements by DA spread when observations were available since 1968. In all these cases, we highlight the improvements by bias correction when observations were not available (before 1968), thus HRXX (dashed black lines) shows the correction of overestimated peaks by OL, which was not possible to do by LEnKF without bias removal (gray line). Improvements were also important at the Óbidos gauge, where a rapid recession was removed by a DA scheme when observations were available since early 1968 and by bias removal before that (Fig. 13b).

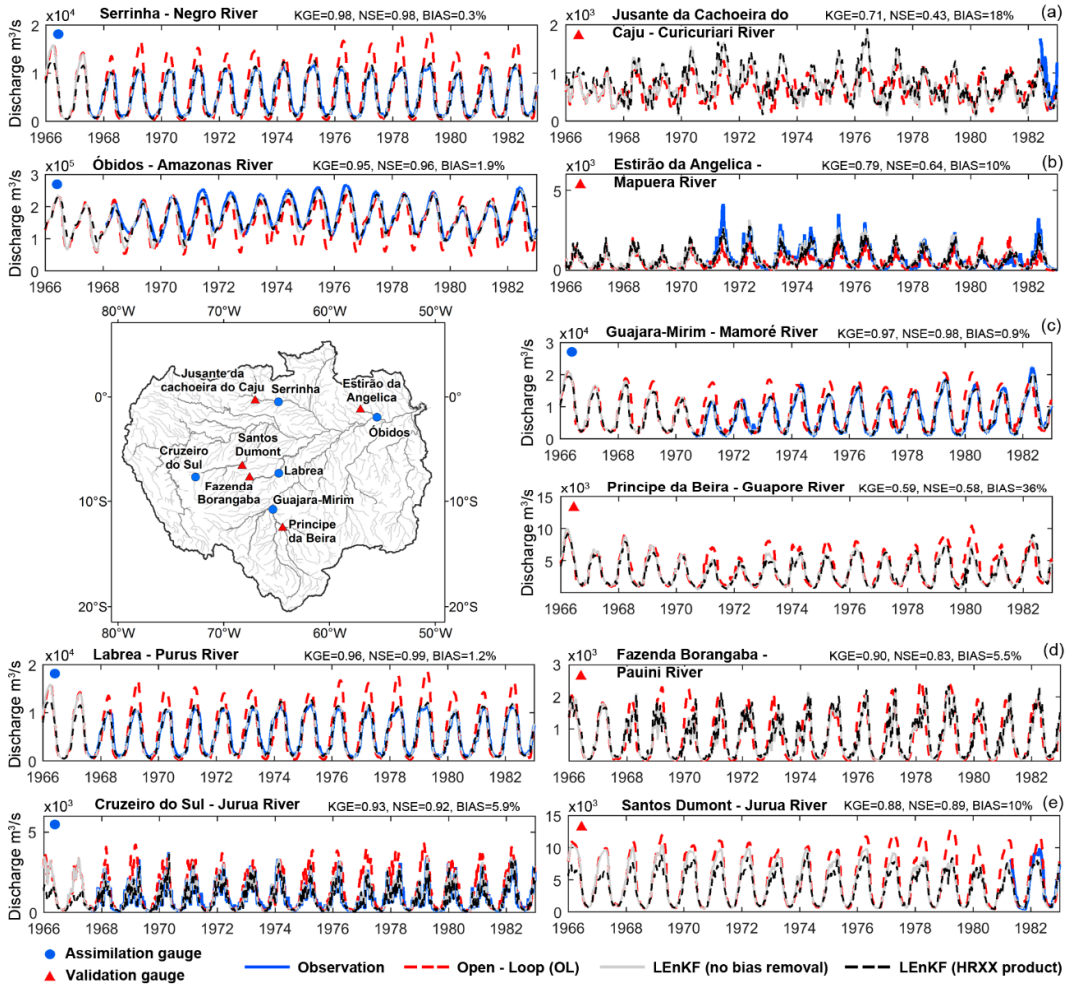

Fig. 13. Hydrographs in assimilation (blue circles) and validation (red triangles) gauges for in-situ observed discharge (blue lines), OL discharge (red lines), LEnKF without bias removal (gray lines) and HRXX (dash black lines), over similar periods where observed record is partial available; and performance values are showed for the HRXX.

The HRXX performance is shown in Figure 14, in which the statistical indexes are spatially assessed. The KGE and NSE indexes showed values above 0.6 for 63% and 34% of validation gauges, respectively, which represent an improvement of 20% and 18% in the number of validation gauges with respect to the OL. For the validation, 46% have values of BIAS index below 10%, indicating an improvement in the number of gauges of 25% in comparison with the OL. The results of ΔRMSE indicate a reduction of RMSE with respect to the OL of more than 25% for almost 34% of validation gauges.

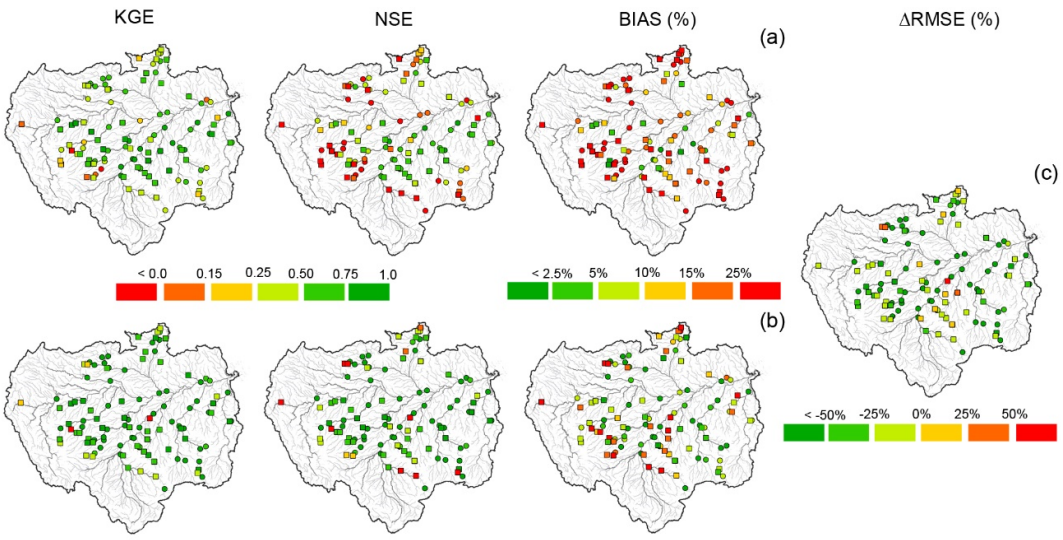

Fig. 14. Performance for daily discharge of KGE, NSE and BIAS indexes for the (a) OL and (b) HRXX product and (c) improvements in RMSE at assimilation (circles) and validation (squares) gauges.

Finally, Table 3 summarizes the performance of the statistical indexes for the different scenarios/experiments assessed in this research. From OL to the HRXX final product, we can see a consistent improvement in performance, in which values could be considered as satisfactory for HRXX, even for the experiment in which only the seven gauges with the longest records were used in the DA reaching an average of ~10% of BIAS, reductions of RMSE of ~14% and satisfactory values of 0.51 and 0.66 for NSE and KGE, respectively, which could be considered in some ways as representative of the oldest periods (before 1970).

Table 3. Summary of comparison of statistical indexes on discharge for different simulations scenarios for the average of the validation gauges.

Statistical index	Assimilation gauge stations	Validation gauge stations	All validation gauges				Validation gauges > 25,000 km²			
			ΔRMSE (%)	BIAS (%)	NSE	KGE	ΔRMSE (%)	BIAS (%)	NSE	KGE
OL	**57**	56	-	18.68	0.26	0.52	-	17.84	0.43	0.64
OL + bias correction			-4.2	14.16	0.36	0.59	-2.9	11.29	0.45	0.64
EnKF			-1.0	18.16	0.28	0.57	-10.2	13.32	0.42	0.67
LEnKF			-6.8	17.32	0.47	0.62	**-26.9**	16.52	0.58	0.75
HRXX (LEnKF + bias correction)			**-14.8**	**10.55**	**0.52**	**0.65**	**-23.0**	**10.05**	**0.67**	**0.78**
OL	7 longest records	106	-	22.31	0.31	0.55	-	16.26	0.39	0.62
OL + bias correction			-9.0	11.57	0.44	0.63	-13.8	10.43	0.53	0.73
HRXX (LEnKF + bias correction)			**-13.9**	**9.94**	**0.51**	**0.66**	**-18.6**	**9.80**	**0.58**	**0.76**

* Values in bold represent better results.

Because one objective in this research was to represent acceptable values since the beginning of 1900 and our validation common period was mostly done since 1980, the Manaus water level record since 1903 was also used to validate HRXX in this research. As described in section 2.6.4, discharge data by Manaus gauge was not used for assimilation purposes, therefore they are independent. Figure 15 shows the water level anomalies for two periods, from 1901 to 1930 (Fig. 15a) and from 1978 to 2009 (Fig. 15b) where the observations, OL and HRXX were plotted. It is clearly seen in the reduction of peaks from OL to HRXX, and the enhancements from the rapid recession during the dry period. The sky blue vertical lines, in the years 1908 and 1928, indicate when observations were available at Porto Velho (Madeira River) and at Óbidos (Amazon River), of which the latter is located downstream of the Manaus gauge and therefore influenced by the DA scheme. The extreme drought in 1926 (Marengo et al. 2016) was not represented by HRXX because discharge observations by Óbidos, used in DA scheme, just started in 1928. Besides, the 1979-81, 1982-83, and 1995 drought events as well as the 1989, 1999 and 2009 flood events were better represented. The performance of the NSE and KGE indexes improved from 0.44 to 0.86 and from 0.72 to 0.87, respectively, during the period from 1903 to 2010. HRXX shows a good performance even when observations were not available due to improvements by the bias removal of precipitation forcing, which is further enhanced by LEnKF when observations are accessible.

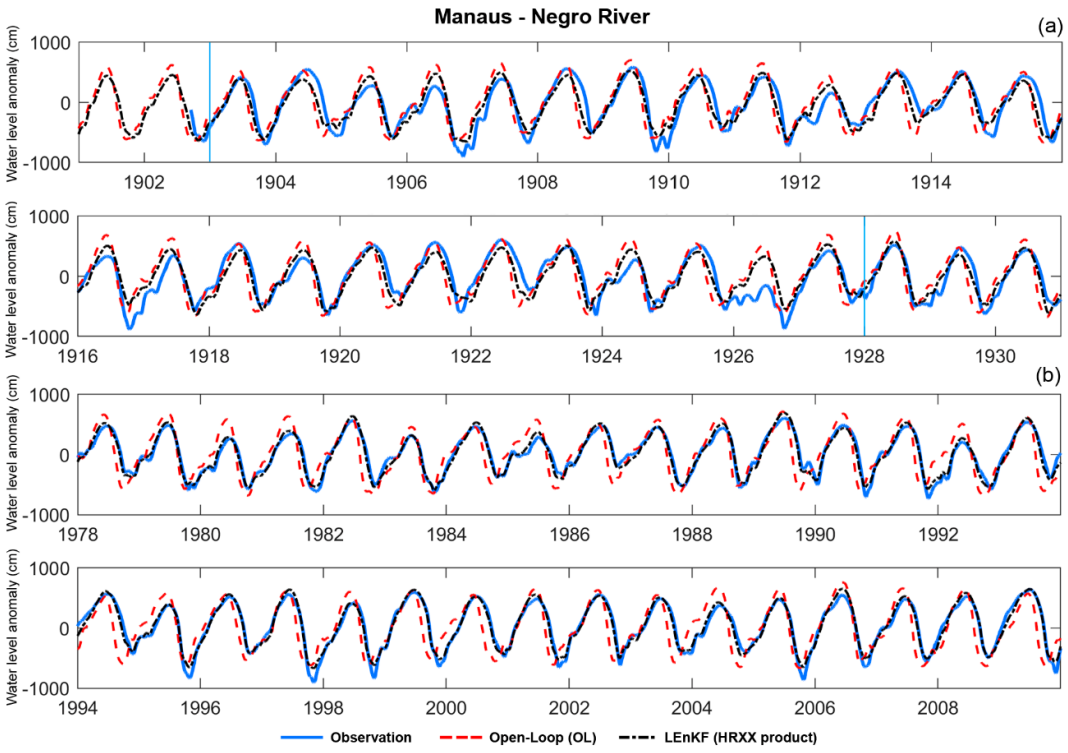

Fig. 15. Water level anomaly comparison of in-situ observations (blue), OL (red) and the HRXX product (black) at the Manaus gauge (a) from 1901 to 1930 and (b) from 1978 to 2009.

As several studies have shown, extreme events in the Amazon Basin have become more frequent and intense during recent decades. These events, such as floods and droughts, are important because of their economic and social impacts. Some of these extreme events that occurred in the past were assessed in this research, corresponding to the two oldest and two most recent floods and droughts cited in Marengo et al. (2016), which were based mostly on the impact on the water level at the Manaus gauge. The spatial impact of these drought and flood events are plotted in Figures 16a and 17a, respectively; the intensity of the event in days was estimated based on a simple approach described in Wongchuig et al. (2017), where the number of days above or below certain discharge or water level thresholds were counted to determine the impact of the flood or drought events, respectively. The comparison between observation and simulation (HRXX) was based on daily water level values at the Manaus gauge. In addition to that, it is important to highlight that several flood after 2009 such as 2012 and 2014 years (Marengo et al., 2016; Barichivich et al., 2018) were not registered by the HRXX due to our simulation window.

Considering the drought events, the years 1906, 1926, 2005 and 2010 were chosen for this analysis; the two oldest drought years were mainly registered at the Manaus gauge (Jenkins et al., 2009; Marengo et al., 2016), being the drought of 1926 perhaps one of the worst on record (Williams et al., 2005). Moreover, the 2005 and 2010 events were also registered at the Óbidos gauge and reported in the literature (Espinoza et al., 2011; Marengo et al., 2011) through observation records in the Madeira and Solimões rivers in the southern and western regions, as shown in Figure 16a. Figure 16b shows the water level anomalies of minimum annual values for observations, OL and HRXX, where three drought events (1906, 2005 and 2010) are well-represented in the vertical red lines for certain drought years. For the common period between observation and simulation, the Pearson correlation coefficient for the water level anomalies of minimum annual values increased from 0.38 for the OL to 0.72 for HRXX. In spite of relative well agreement for correlation between HRXX and observations, drought events seem to be underestimated by HRXX, mainly before 1928, when observations were not available to be assimilated.

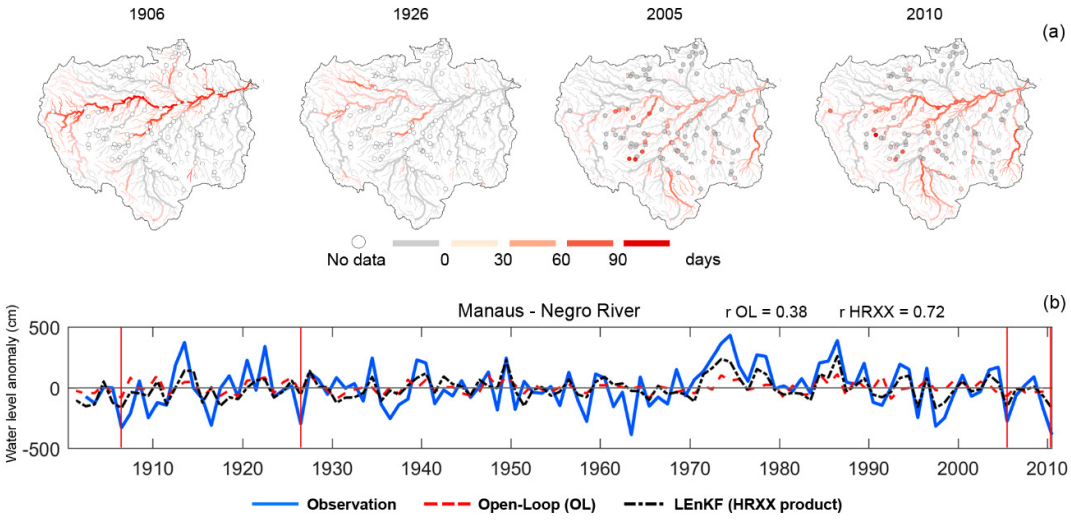

Fig. 16. (a) Maps of the intensity on days of some droughts for HRXX in each reach river and Observed data (circles), and (b) water level anomalies for minimum annual observation (blue), OL (red) and HRXX (black).

For floods, the events in 1909, 1976, 1999 and 2009 were selected. The 1909 event was detected only at the Manaus gauge, while the last three events were also detected at the Óbidos gauge (Marengo et al., 2016). The intensity of the 1909 flood event was represented in the Negro River close to the confluence with the Amazon River (Fig. 17a), which was less intense in comparison with the 1976, 1999 and 2009 events. These events were also reported in the literature through observation records at Manaus (Satyamurty et al., 2013), Madeira and Purus rivers, most notably during 2009 event (Marengo et al. 2012; Pereira and Szlafsztein, 2016) as shown in Figure 17a. For these three recent flood events, Figure 17b shows a clear concordance between observation and HRXX; the Pearson correlation coefficient for the water level anomalies of maximum annual values increased from 0.18 for the OL to 0.60 for HRXX, which indicates a significant improvement in the performance of detecting maximum events according this methodology. Finally, considering the anomalies of the mean annual values of water level at Manaus, it increased from 0.22 for OL to 0.69 for HRXX.

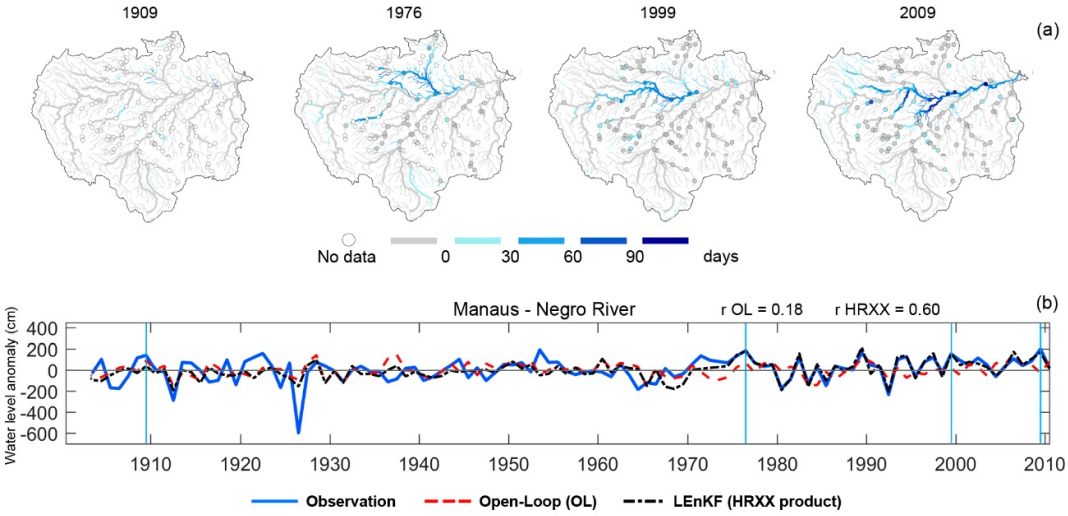

Fig. 17. (a) Maps of the intensity on days of some floods for HRXX in each reach river and Observed data (circles) and (b) water level anomalies for maximum annual observation (blue), OL (red) and HRXX (black).

4.4. Conclusions and perspectives

This research aimed to develop, in a retrospective manner, a consistent register of hydrological variables through the use of a large-scale hydrologic-hydrodynamic model, DA scheme and bias removal method.

LEnKF was assessed in this research using a primary approximation of a general radius of influence. The local implementation showed improvements with respect to the EnKF scheme in the Amazon Basin, even using a relatively large size of ensemble members, which demonstrates that local consideration regarding the correlation between assimilation and updated points that are separated by a far distance or physically disconnected on a river network should be taken into account for large-scale basin studies. The optimal radius of influence was reached at 1000 km, which when used within the LEnKF scheme for improved discharge representation, increasing the NSE and KGE performance by ~70% and ~9%, respectively, with respect to the EnKF scheme.

Four bias removal techniques were assessed because 1) climatic reanalysis is usually biased and 2) a DA scheme only updates state variables when and where observations are available. The improvement in absolute values of bias during the period of validation was assessed for monthly precipitation and for daily discharge for the validation period (1980-1989) and during the oldest period (1931-1948). The linear scaling method using a threshold achieved the best performance among other methods tested in this study. The bias values for the ERA-20CM raw dataset range spatially from -50% to 50%, which was finally constrained by bias removal from -5% to 5% for ~80% of the study area. The

improvements on BIAS for daily discharge values was ~40% during the validation period, and achievements of 0.45 and 0.64 for NSE and KGE respectively

HRXX proposed herein is based in the large-scale MGB model forced with the bias removed precipitation and coupled in synergy with a LEnKF scheme. Its performance was assessed for daily discharge observations at 56 validation gauges and for water level anomalies at the Manaus gauge (Negro River) during the longest records in Amazon Basin. Historic observed flood and drought events were also assessed for HRXX. Results indicate that improvements were clearly achieved by HRXX, which takes advantage of both a DA scheme and a bias removal method, in which statistical performance could be considered as satisfactory to represent adequate discharge and water levels across the 20th century.

In this paper, we evaluated the HR methodology to generate a consistent record of hydrological variables on a large scale by the use of long temporal records of in-situ discharge and climatic precipitation reanalysis datasets that are globally available. This new dataset is the first approach of long term data assimilation of in-situ data in the Amazon Basin. Despite the fact that our methodology presents some simplifications, results show adequately for large scale. The HRXX could be useful for the spatial and temporal hydrological characterization (e.g. climate change assessment) for several past decades, once this basin is considered an important large biome with a considerable control on the carbon cycle and global climate.

Other approaches should also be implemented in the future, such as complex approximations to estimate the radius of influence in the LEnKF scheme considering spatial and temporal variations for each unit-catchment, or better bias removal methods that consider non-stationary series. Despite this, HRXX developed in this study could be used as a methodology in other large basins around the world: 1) to evaluate spatially and temporally distributed hydrological variability across the last century; 2) to create consistent records of hydrological extreme events, such as floods and droughts; and 3) to make availability a complementary dataset in regions with sparse hydro-meteorological network.

4.5. Supplementary material

In-situ datasets and quality control

The Table 1s and 2s show a detailed information of 114 gauge stations used for assimilation and validation proposes respectively. Besides, Figure 1s shows the geographical distribution of all gauges used in this research for assimilation (Fig. 1sa) and validation (Fig. 1sb) purposes.

Table 1s. Summary of in situ assimilation discharge gauges.

Gauge	Institution Code	Name	Longitude (°)	Latitude (°)	Catchment area (km^2)	River	Data period
1	10075000	Tamshiyacu	-73.1617	-4.0031	750,000	Amazonas (Peru)	1984 - 2010
2	12500000	Cruzeiro Do Sul	-72.6667	-7.6167	37,800	Juruá	1928 - 2010
3	12550000	Eirunepe-Montante	-69.9000	-6.6833	77,300	Juruá	1979 - 2010
4	12680000	Envira	-70.0225	-7.4281	49,700	Tarauaca	1978 - 2010
5	12840000	Gavião	-66.8506	-4.8392	164,000	Jurua	1972 - 2010
6	12845000	Vila Bittencourt	-69.4167	-1.4000	206,000	Caqueta/Japura	1980 - 2010
7	12850000	Acanaui	-66.6000	-1.8167	249,000	Caqueta/Japura	1973 - 2010
8	13150000	Itapeua	-63.0278	-4.0578	1,780,000	Solimões	1971 - 2010
9	13410000	Seringal da Caridade	-68.5683	-9.0350	63,100	Purus	1967 - 2010
10	13550000	Xapuri	-68.5067	-10.6497	8,270	Acre	1968 - 2010
11	13600002	Rio Branco	-67.8000	-9.9758	23,500	Acre	1967 - 2010
12	13650000	Floriano Peixoto	-67.3681	-9.0506	34,400	Acre	1967 - 2010
13	13710001	Valparaiso	-67.4000	-8.6833	105,000	Purus	1975 - 2010
14	13750000	Seringal Fortaleza	-66.9847	-7.7172	154,000	Purus	1967 - 2010
15	13870000	Labrea	-64.8000	-7.2522	226,000	Purus	1927 - 2010
16	13880000	Canutama	-64.3839	-6.5342	236,000	Purus	1973 - 2010
17	14100000	Manacapuru	-60.6094	-3.3083	2,200,000	Solimões	1972 - 2010
18	14110000	Cucui	-66.8525	1.2153	74,300	Negro	1980 - 2010
19	14230000	Missão Icana	-67.5947	1.0744	23,600	Icana	1980 - 2010
20	14250000	São Felipe	-67.3128	0.3717	124,000	Negro	1977 - 2010
21	14260000	Uaracu	-69.1281	0.4769	40,200	Uaupés	1977 - 2010
22	14280001	Taraqua	-68.5386	0.1303	44,300	Uaupés	1977 - 2010
23	14330000	Curicuriari	-66.8022	-0.2006	194,000	Negro	1977 - 2010
24	14420000	Serrinha	-64.8289	-0.4819	293,000	Negro	1967 - 2010
25	14515000	Fazenda Passarão	-60.5711	3.2078	50,200	Uraricoera	1977 - 2010
26	14550000	Maloca do Contao	-60.5281	4.1675	5,780	Cotingo	1975 - 2010
27	14680001	Fe e Esperança	-61.4406	2.8708	12,200	Mucajai	1973 - 2010
28	14710000	Caracarai	-61.1236	1.8214	126,000	Branco	1967 - 2010
29	15030000	Jatuarana	-59.6783	-3.0517	2,930,000	Amazonas	1977 - 2010
30	15120001	Vila Bela da Santis.Trindade	-59.9492	-15.0081	22,500	Guaporé	1976 - 2010
31	15250000	Guajara-Mirim	-65.3478	-10.7925	609,000	Mamoré	1970 - 2010
32	15320002	Abuna	-65.3647	-9.7031	921,000	Madeira	1976 - 2010
33	15400000	Porto Velho	-63.9203	-8.7367	976,000	Madeira	1908 - 2010
34	15430000	Ariquemes	-63.0570	-9.9317	8,140	Jamari	1970 - 2010
35	15431000	Fazenda Rio Branco	-62.9878	-9.8872	988	Branco	1979 - 2010
36	15550000	Santa Isabel	-63.7106	-8.7986	12,700	Candeias	1976 - 2010
37	15630000	Humaita	-63.0203	-7.5078	1,090,000	Madeira	1931 - 2010
38	15700000	Manicore	-61.3019	-5.8167	1,150,000	Madeira	1967 - 2010
39	15860000	Fazenda Vista Alegre	-60.0253	-4.8972	1,310,000	Madeira	1967 - 2010
40	15910000	Sucunduri	-59.0422	-6.7950	12,700	Sucunduri	1973 - 2010
41	16650000	Cach Da Porteira-Conj 1	-57.0469	-1.0875	77,600	Trombetas	1970 - 2010
42	17050001	Óbidos	-55.5111	-1.9472	4,670,000	Amazonas	1927 - 2010
43	17090000	Boca Do Inferno	-54.8733	-1.5000	19,800	Curuá	1973 - 2010

44	17093000	Fontanilhas	-58.3428	-11.3583	55,900	Juruena	1978 - 2010
45	17120000	Porto dos Gauchos	-57.4228	-11.5364	37,100	Arinos	1973 - 2010
46	17200000	Porto Roncador	-55.3336	-13.5569	10,800	Teles Pires	1973 - 2010
47	17280000	Cachoeirão	-55.7017	-11.6531	34,600	Teles Pires	1975 - 2010
48	17340000	Indeco	-55.5700	-10.1125	52,200	Teles Pires	1975 - 2010
49	17350000	Cachimbo	-54.8858	-9.8172	1,010	Braço Sul	1979 - 2010
50	17430000	Barra do São Manuel Jusante	-58.1553	-7.3397	333,000	Tapajós	1975 - 2010
51	17450000	Missão Cururu	-57.5833	-7.6167	7,345	Cururu / Tapajós	1931 - 1988
52	17650000	Jatoba	-56.8539	-5.1525	387,000	Tapajós	1972 - 2010
53	17730000	Itaituba	-55.9833	-4.2833	458,000	Tapajós	1968 - 2010
54	18200000	Arapari	-54.3958	-1.7758	12,400	Maicuru	1972 - 2010
55	18500000	Boa Esperança	-51.7828	-6.7192	42,400	Fresco	1976 - 2010
56	18650000	Cajueiro	-54.5211	-5.6539	35,600	Curuá (Xingu)	1975 - 2010
57	18850000	Altamira	-52.2106	-3.2122	448,000	Xingu	1928 - 2010

Table 2s. Summary of in situ validation discharge and water level gauges.

Gauge	Institution Code	Name	Longitude (°)	Latitude (°)	Catchment area (km^2)	River	Data period
1	10064000	Borja	-77.5480	-4.4700	115,000	Marañon	1986 - 2010
2	10100000	Tabatinga	-69.9330	-4.2500	874,000	Solimões	1982 - 2010
3	10200000	Palmeiras do Javari	-72.8000	-5.1333	16,500	Javari	1982 - 2010
4	10300000	Santa Maria	-71.4167	-4.5667	25,200	Curuca	1980 - 2010
5	10500000	Estirão do Repouso	-70.9333	-4.3667	61,400	Javari	1980 - 2010
6	12100000	Colocação Caxias	-69.0000	-5.4000	10,700	Jutaí	1982 - 2010
7	12360000	Foz do Breu	-72.7025	-9.4017	7,690	Juruá	1982 - 2010
8	12370000	Taumaturgo	-72.7833	-8.9333	16,100	Juruá	1981 - 2010
9	12520000	Ipixuna	-71.6842	-7.0508	56,100	Juruá	1981 - 2010
10	12530000	Fazenda Paranacre	-71.4167	-7.7167	2,070	Gregorio	1982 - 2010
11	12600001	Tarauaca-Jusante	-70.7169	-8.1356	15,600	Tarauaca	1981 - 2010
12	12650000	Feijo	-70.3683	-8.1519	17,000	Envira	1980 - 2010
13	12700000	Santos Dumont	-68.2439	-6.4417	144,000	Juruá	1981 - 2010
14	12880000	Estirão da Santa Cruz	-65.2017	-4.2922	13,700	Tefe	1981 - 2010
15	13180000	Manoel Urbano	-69.2681	-8.8842	32,800	Purus	1981 - 2009
16	13300000	Seringal São Jose	-68.7167	-9.3167	11,200	Iaco	1981 - 2010
17	13405000	Seringal Guarany	-69.4167	-9.3500	6,110	Caeté	1983 - 2010
18	13450000	Assis Brasil	-69.5500	-10.9333	3,760	Acre	1983 - 2010
19	13470000	Brasileia	-68.7450	-11.0178	7,020	Acre	1982 - 2010
20	13740000	Fazenda Borangaba	-67.5500	-7.5500	23,300	Pauini	1984 - 2010
21	13885000	Cristo	-64.2433	-7.4644	7,030	Mucuim	1975 - 2010
22	13886000	Bacaba	-64.8839	-6.3172	37,800	Cuniua	1979 - 2010
23	14300000	Pari Cachoeira	-69.7850	0.2494	1,970	Tiquié	1980 - 2010
24	14310000	Cunuri	-69.3781	0.2094	4,250	Tiquié	1982 - 2010

25	14350000	Jusante da Cachoeira do Caju	-67.0092	-0.2461	13,100	Curicuriari	1982 - 2010
26	14440000	Posto Ajuricaba	-62.6219	0.8842	16,900	Demeni	1982 - 2010
27	14488000	Uaicas	-63.1692	3.5497	16,100	Uraricoera	1984 - 2010
28	14495000	Fazenda Cajupiranga	-61.0367	3.4381	36,900	Uraricoera	1979 - 2010
29	14526000	Bom Fim	-59.8156	3.3783	9,860	Tacutu	1984 - 2010
30	14530000	Vila Surumu	-60.7939	4.1961	2,280	Surumu	1984 - 2010
31	14540000	Fazenda Bandeira Branca	-60.4706	4.6306	3,210	Cotingo	1970 - 2010
32	14750000	Missão Catrimani	-62.2833	1.7500	6,180	Catrimani	1984 - 2010
33	14845000	Fazenda São Jose	-60.4661	0.5178	5,680	Jauaperi	1983 - 2010
34	15130000	Pimenteiras	-61.0453	-13.4819	54,400	Guaporé	1983 - 2010
35	15150000	Pedras Negras	-62.8992	-12.8514	110,000	Guaporé	1980 - 2010
36	15200000	Principe da Beira	-64.4253	-12.4267	341,000	Guaporé	1983 - 2010
37	15248010	Boca do Pompeu-Jus.(Ceron)	-65.0433	-10.9175	4,390	Ouro Preto	1982 - 2010
38	15432000	Mineração Ponte Massanga	-63.2872	-9.7603	852	Massangana	1980 - 2010
39	15558000	Pimenta Bueno	-61.1922	-11.6836	10,100	Apedia	1980 - 2010
40	15559000	Sitio Bela Vista	-61.2150	-11.6525	16,100	Jiparaná	1984 - 2010
41	15560000	Ji-Parana	-61.9355	-10.8736	32,800	Jiparaná	1977 - 2010
42	15565000	Jaru	-62.4656	-10.4458	3,960	Jaru	1981 - 2010
43	15580000	Tabajara	-62.0539	-8.9333	60,200	Jiparaná	1977 - 2010
44	15650000	Maloca Tenharim	-62.0503	-7.9578	3,630	Marmelos	1978 - 2010
45	15670000	Nova Esperança	-61.7667	-6.3592	26,300	Marmelos	1982 - 2010
46	15750000	Humboldt	-59.4644	-10.1675	15,200	Aripuanã	1979 - 2010
47	15795000	Leontino	-60.5667	-7.7667	16,200	Guariba	1985 - 2010
48	16500000	Estirão da Angelica	-57.0567	-1.1011	25,800	Mapuera	1970 - 1999
49	17210000	Teles Pires	-55.7917	-12.6742	13,900	Teles Pires	1976 - 2010
50	17230000	Lucas do Rio Verde	-55.9039	-13.0511	5,435	Verde (Tapajós)	1973 - 2007
51	17410000	Santa Rosa	-57.4203	-8.8597	131,000	Teles Pires	1982 - 2010
52	17420000	Três Marias	-57.9503	-7.6147	138,000	Teles Pires	1975 - 2010
53	18250000	Uruara	-53.5542	-3.6772	2,960	Para do Uruará	1977 - 2010
54	18460000	Boa Sorte	-51.9833	-6.7500	210,000	Xingu	1977 - 2010
55	18600000	Laranjeiras	-54.2394	-5.6997	58,700	Iriri	1976 - 2010
56	18700000	Pedra do O	-54.0008	-4.5417	122,000	Iriri	1976 - 2010
57	14990000	Manaus*	-60.0269	-3.1366	712,000	Negro	1902 - 2010

* Correspond to water level information for validation purposes

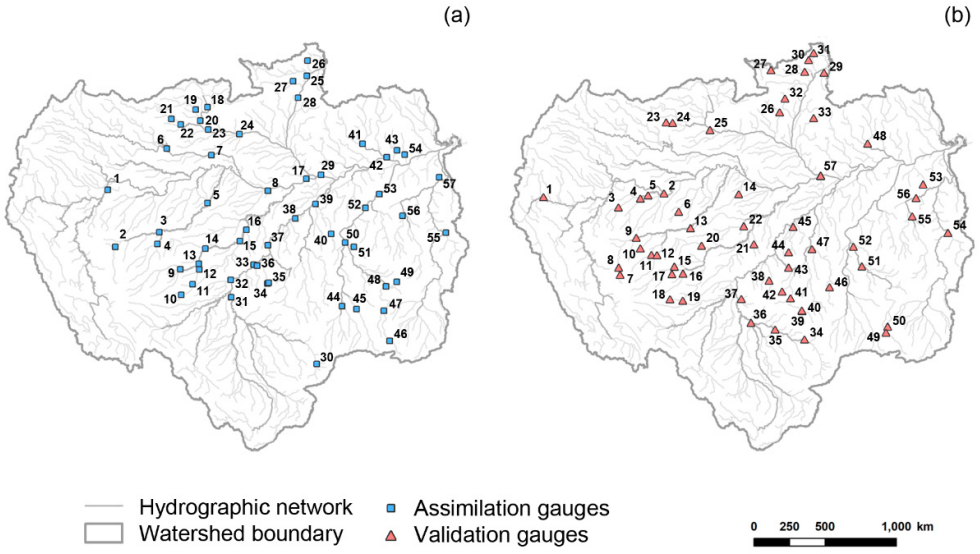

Fig. 1s. Map of localization of gauges for (a) assimilation and (b) validation gauges.

For selection and quality control of data from gauge stations, the following criteria were considered:

1. For 137 total downloaded gauges from HYBAM and ANA, just those with more than ~1000 km^2 and with record greater than 15 and 30 years were chosen for validation and assimilation purposes respectively.
2. Outliers at few gauges were removed from the series through graphical analysis (e.g. Labrea, Fazenda Passarão, Fazenda Bandeira Branca).
3. To assess the gauges with longest level series, the temporal record of discharge was extended for when just level was available. This extension was estimated for Cruzeiro do Sul, Humaita, Porto Velho and Missão Cururu gauges. Therefore, due to the rating curves were extended for periods when the river geometry probably changed, the discharge error parameter for these gauges was assumed as the greatest value estimated by Getirana and Peters-Lidard (2013) of 15%, a greater value than the a priori assumption (10%).

Bias correction methods

In this section the formulations for the bias correction methods used in this research are presented. For the *linear scaling* (LSC) and the *linear scaling threshold* (LSCT) approach, following equations were used:

$$P_{cor,mini,d} = P_{ERA,mini,d} \boldsymbol{x} \frac{\left(\overline{P_{obs,mini,ref}}\right)_m}{\left(\overline{P_{ERA,mini,ref}}\right)_m} \quad (1)$$

if $\frac{\left(\overline{P_{obs,mini,ref}}\right)_m}{\left(\overline{P_{ERA,mini,ref}}\right)_m}$ > threshold; $P_{cor,mini,d} = P_{ERA,mini,d} + \left(\frac{\left(\overline{P_{obs,mini,ref}}\right)_m - \left(\overline{P_{ERA,mini,ref}}\right)_m}{n_{days,m}}\right)$ (2)

where $P_{cor,mini,d}$ is the daily corrected precipitation, calculated by the product of $P_{ERA,mini,d}$ which is a reference daily precipitation and the ratio between the monthly mean of the observed precipitation $\left(\overline{P_{obs,mini,ref}}\right)_m$ and the reanalysis precipitation $\left(\overline{P_{ERA,mini,ref}}\right)_m$ within the reference period.

For the quantil mapping, considering the *parametric quantile mapping gamma distribution* (gQM), the Gamma probability distribution with α and β as shape and scale parameters respectively (Eq. 3) was assumed as suitable for both observed and ERA daily precipitation, such distribution has been demonstrated to be appropriate for precipitation (Piani et al. 2010).

$$f_N(x|\alpha,\beta) = x^{\alpha-1} \cdot \frac{1}{\beta^{\alpha} \cdot \Gamma(\alpha)} \cdot e^{\frac{-x}{\beta}}; \quad x \geq 0; \ \alpha,\beta \geq 0 \quad (3)$$

To perform the bias correction for each mini-catchment by eQM and gQM, each daily ERA precipitation amount ($P_{ERA,mini,d}$) within month m is searched in the CDF of the ERA for the same month, afterward this cumulative probability value is located in the CDF of the observations. Then this amount is considered as the corrected precipitation for correspond day ($P_{cor,mini,d}$).

$$P_{cor,mini,d} = F^{-1}_{obs,mini,ref,m}\left(F_{ERA,mini,ref,m}\left(P_{ERA,mini,d}\right)\right) \quad (4)$$

where $F^{-1}_{obs,mini,ref,m}$ correspond to the inverse of the CDF for observations and consequently $F_{ERA,mini,ref,m}$ is the CDF of ERA, for each month forthe referenced period.

The Figure 2s shows the threshold assessments to determinate an adequate value for the LSC method considering the RMSE index on daily discharge values during a validation period (1980-1989). It is clearly seen an improvement when threshold values trend to 1, which was finally considered in the LSCT method for bias correction.

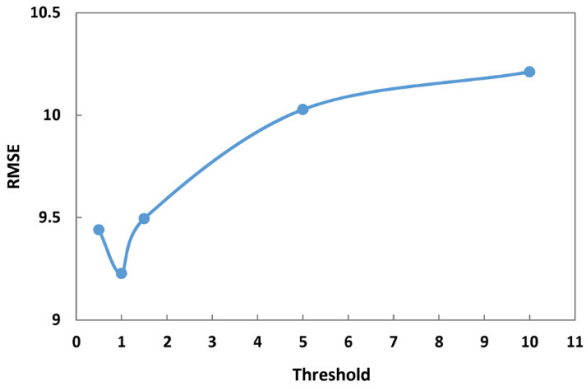

Fig. 2s. Performance on daily discharge (RMSE) considering different values of thresholds for the LSCT bias correction method.

Open loop for 10 ensembles member's performance

The Fig. 3s shows the boxplot of statistical indexes for each ensemble member against all (113) in-situ discharge gauges. It is noticed that all the members have a similar performance during the validation period (1980-2010).

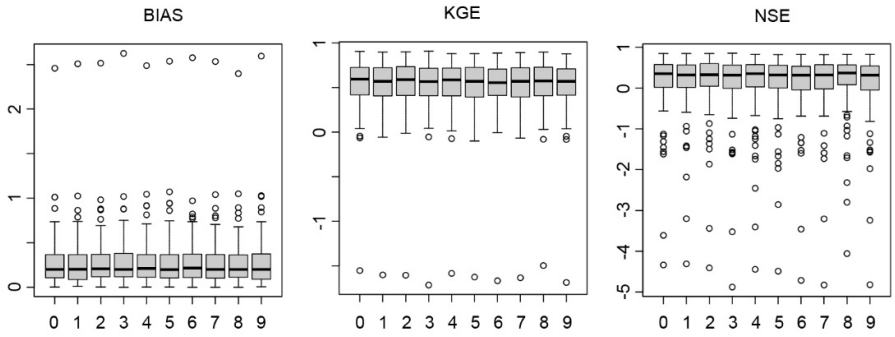

Fig. 3s. Performance on BIAS, KGE and NSE index for Open Loop simulation of each member considering all gauges.

The median of the performance for each ensemble member is showed in the Table 3s where all members have similar values, and could be considered that results are within a range of uncertainty. Therefore, a student's t-test is necessary to verify the existence of a significant difference between them.

Table 3s. Median of performance for each ensemble member of the ERA 20-CM reanalysis.

ERA 20-CM member	BIAS (%)	KGE	NSE
0	19.91	0.59	0.35
1	20.28	0.57	0.32
2	20.68	0.59	0.33
3	19.98	0.57	0.32
4	21.09	0.59	0.35
5	19.56	0.57	0.33
6	21.43	0.55	0.32
7	19.94	0.57	0.32
8	19.89	0.57	0.37
9	20.08	0.57	0.32

Localization performance

In this section detailed values of the performance for each radius of influence used in the LEnKF scheme is showed in the Table 4s. It is noticed that localization method improves discharge performance proportionally to the radius of influence until certain values which reach between 500 to 2000 km.

Table 4s. Median of performance for Open Loop, EnKF and LEnKF for different radius of influence, in validation gauges using 25 members of ensemble.

		All validation gauges				Gauges > 25,000 km^2			
	Area of DA influence (%)	ΔRMSE (%)	BIAS (%)	NSE	KGE	ΔRMSE (%)	BIAS (%)	NSE	KGE
Open Loop	-	-	18.7	0.26	0.52	-	17.8	0.43	0.64
EnKF	-	7.1	19.3	0.13	0.48	-0.2	**12.8**	0.30	0.56
LEnKF - 50	8.2	-0.5	17.3	0.37	0.55	-5.9	16.2	0.49	0.71
LEnKF - 75	12.7	-0.8	16.6	0.37	0.58	-5.9	16.2	0.56	0.71
LEnKF - 150	26.0	-1.7	17.6	0.39	0.63	-9.7	**14.8**	0.59	0.72
LEnKF - 250	40.7	-3.9	16.7	**0.41**	**0.61**	-9.7	**13.0**	0.59	0.73
LEnKF - 500	68.3	**-9.2**	**14.0**	**0.44**	**0.64**	**-20.3**	**13.8**	**0.60**	**0.75**
LEnKF - 1000	91.7	**-7.3**	**16.4**	**0.45**	**0.61**	**-25.6**	16.9	0.57	0.74
LEnKF - 2000	95.9	**-5.2**	**15.9**	0.38	0.58	**-24.7**	15.7	**0.59**	**0.76**
LEnKF - 5000	95.9	-3.9	**15.7**	0.37	0.58	-23.0	16.2	**0.60**	**0.75**

* Values in bold represent the best results

Acknowledgments

The first author is grateful for a grant from the Brazilian agency CAPES and for EnKF codes provided by Evensen. The authors are grateful for the precipitation datasets supplied by ECMWF and the in-situ discharge and water level data provided by HYBAM and ANA institutions. We also thank two anonymous reviewers for their comments, which helped significantly in the improvement of this manuscript. The developed HRXX data for the Amazon Basin that support the findings of this study are available from the corresponding author upon request.

References

Anderson, J.L. 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. Physica D: Nonlinear Phenomena, 230, 99–111.

Andreadis, K.M., Das, N., Stampoulis, D., Ines, A., Fisher, J.B. Granger, S., Kawara, J., Han, E., Behrangi, A. 2017. The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation. PLoS ONE 12(5): e0176506.https://doi.org/10.1371/journal.pone.0176506

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F. 2015. Era Interim/ land: a global land surface reanalysis data set. Hydrology and Earth System Sciences, 19(1):389–407.

Barichivich, J., Gloor, E., Peylin, P., Brienen, R.J.W., Schönegart, J., Espinoza, J.C., Pattnayak, K.C. 2018. Recent intensification of Amazon flooding extremes driven by strengthened Walker circulation, Science Advances, 4, doi:10.1126/sciadv.aat8785.

Bates, P.D., Horritt, M.S., and Fewtrell, T.J. 2010. A simple inertial formulation of the shallow water equations for efficienttwo-dimensional flood inundation modelling, Journal of Hydrology, 387, 33-45, 10.1016/j.jhydrol.2010.03.027

Baker, V.R., Webb, R.H., House, P.K., 2002. The scientific and societal value of paleoflood hydrology. In: House, P.K., Webb, R.H., Baker, V.R., Levish, D.R. (Eds.), Ancient Floods, Modern Hazards: Principles and Applications of Paleoflood Hydrology. Water Science and Application Series, 5, 127–146.

Barber, V.A., Finney, B.P. 2000. Late quaternary paleoclimatic reconstructions for interior Alaska based on paleolake-level data and hydrologic models. Journal of Paleolimnology. 24(1): 29-41.

Di Baldassarre, G. and Montanari, A. 2009. Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913-921, https://doi.org/10.5194/hess-13-913-2009

Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., and de Roo, A. 2017. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589-615, https://doi.org/10.5194/hess-21-589-2017

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., and Wood, E.F. 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201-6217, https://doi.org/10.5194/hess-21-6201-2017

van Beek, L.P.H., Wada, Y., and Bierkens, M.F.P. 2011. Global monthly water stress: I. Water balance and water availability, Water Resour. Res., 47, W07517, doi:10.1029/2010WR009791

Benito, G., Thorndycraft, V.R. 2005. Palaeoflood hydrology and its role in applied hydrological sciences. Journal of Hydrology, 313: 3–15.

Berg, A.A., Famiglietti, J.S., Walker, J.P. and Houser, P.R. 2003. Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, J. Geophys. Res., 108, 4490, doi: 10.1029/2002JD003334, D16.

Biancamaria S., Durand, M., Andreadis, K.M., Bates, P.D., Boone, A., Mognard, N.M., Rodríguez, E., Alsdorf, D.E., Lettenmaier, D.P., and Clark, E.A. 2011. Assimilation of Virtual Wide Swath Altimetry to Improve Arctic River Modeling, Remote Sens. Environ., 115(2), 373-381, doi: 10.1016/j.rse.2010.09.008

Bishop, C.H. and Hodyss, D. 2009. Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. Tellus A, 61: 84-96. doi:10.1111/j.1600-0870.2008.00371.x

Callède, J., Guyot, J.L., Ronchail, J.L'H., ote, Y., Niel, H., de Oliveira, E., 2004. Evolution du débit de l'Amazone à Óbidos de 1903 à 1999 / evolution of the river amazon's discharge at Óbidos from 1903 to 1999. Hydrol. Sci. J. 49 (1), 85–97.

Campbell, W.F., Bishop, C.H., Hodyss, D. 2010. Vertical covariance localization for satellite radiances in ensemble Kalman filters. Mon. Wea. Rev., 138, 282–290, doi:10.1175/2009MWR3017.1.

Chen 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America

Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J. 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources, v. 31, 1309–1324.

Collischonn, W., Allasia, D., da Silva, B.C., Tucci, C.E.M., 2007. The mgb-iph model for large-scale rainfall—runoff modelling. Hydrol. Sci. J. 52 (5), 878–895.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, ., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D., and Worley, S.J. 2011. The twentieth century reanalysis project. Quarterly Journal of the Royal Meteorological Society, 137(654):1–28.

Decker, M., Brunke, M. A.,Wang, Z.,Sakaguchi, K., Zeng, X. B., Bosilovich, M. G. 2012: Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. J. Climate, 25, 1916–1944, doi:10.1175/JCLI-D-11-00004.1.

Denlinger, R.P., O'Connell, D.R.H., House, P.K., 2002. Robust determination of stage and discharge: an example from an extreme flood on the Verde River, Arizona. In: House, P.K., Webb, R.H., Baker, V.R., Levish, D.R. (Eds.), Ancient Floods, Modern Hazards: Principles and Applications of Paleoflood Hydrology, Water Science and Application Series, vol. 5, pp. 127–146.

van Dijk, A.I.J.M., Renzullo, L.J. 2011. Water resource monitoring systems and the role of satellite observations. Hydrology and Earth System Sciences 15(1), 39–55.

van Dijk, A.I.J.M., Renzullo, L.J., Wada, Y., Tregoning, P. 2014. A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. Hydrol. Earth Syst. Sci., 18: 2955–2973.

Dirmeyer, P.A., Gao, X., Zhao, M., Guo, Z., Oki, T. and Hanasaki, N. 2006. GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface. Bull. Amer. Meteor. Soc., 87, 1381–1398, https://doi.org/10.1175/BAMS-87-10-1381

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L. and Hirpa, F.A. 2016. Development and evaluation of a framework for global flood hazard mapping, Adv. Water Resour., 94, 87–102, doi:10.1016/j.advwatres.2016.05.002.

Dutra, E., P. Viterbo, and P.M.A. Miranda. 2008. ERA-40 reanalysis hydrological applications in the characterization of regional drought, Geophys. Res. Lett., 35, L19402, doi: 10.1029/2008GL035381.

Emerton, R., Cloke, H.L., Stephens, E.M., Zsoter, E., Woolnough, S.J., Pappenberger, F. 2017. Complex picture for likelihood of ENSO-driven flood hazard. Nature communications, 8(14796), http://dx.doi.org/10.1038/ncomms14796

Emery, C.M., Paris, A., Biancamaria, S. Boone, A., Calmant, S., Garambois, P.-A., da Silva J. S. 2018. Large scale hydrological model river storage and discharge correction using a satellite altimetry-based discharge product. Hydrol. Earth Syst. Sci., 2017, 1-54. https://doi.org/10.5194/hess-2017-516

Espinoza, J.C., Guyot, J.L., Ronchail, J., Cocheneau, G., Filizola, N., Fraizy, P., Labat, D., de Oliveira, E., Ordoñez, J.J., Vauchel, P., 2009. Contrasting regional discharge evolutions in the Amazon Basin. J. Hydrol. 375, 297–311.

Espinoza, J.C., Ronchail, J., Guyot, J. L., Junquas, C., Vauchel, P., Lavado, W., Drapeau, G., Pombosa, R. 2011. Climate variability and extreme drought in the upper Solimões River (western Amazon Basin): understanding the exceptional 2010 drought. Geophys. Res. Lett. 38(13): L13406.

Espinoza, J.C., Chavez, S., Ronchail, J., Junquas, C., Takahashi, K., Lavado, W. 2015. Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity and relations with large-scale atmospheric circulation. Water Resources Res. 51, doi:10.1002/2014WR016273.

Espinoza, J.C., Segura, H., Ronchail, J., Drapeau, G., Gutierrez-Cori, O. 2016. Evolution of wet- and dry-day frequency in the western Amazon basin, Relationship with atmospheric circulation and impacts on vegetation. Water Resour. Res. 52 (11), 8546–8560.

Espinoza, J.C., Ronchail, J., Marengo, J.A., Segura, H. 2018. Contrasting North-South changes in Amazon wet-day and dry-day frequency and related atmospheric features (1981-2017). Climate Dynamics. doi: 10.1007/s00382-018-4462-2

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367.

Evensen, G. 2004. Sampling strategies and square root analysis schemes for the EnKF. Ocean Dynamics 54: 539–560.

Fleischmann, A., Paiva, R.C.D., Collischonn, W. 2019. Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison. Journal of Hydrology X, 3(2019): 100027.

Gao, L., Bernhardt, M., Schulz, K., Chen, X., Chen, Y., Liu, M. 2016. A First Evaluation of ERA-20CM over China. Mon. Weather Rev. 144 (1), 45–57.

Gaspari, G. and Cohn, S.E. 1999. Construction of correlation functions in two and three dimensions. Q.J.R. Meteorol. Soc., 125: 723-757. doi:10.1002/qj.49712555417

Getirana, A. C. V.; Peters-Lidard, C. 2013. Estimating water discharge from large radar altimetry datasets. Hydrol. Earth Syst. Sci., 17, 923–933.

Gibson, J. and for Medium Range Weather Forecasts, E. C. 1997. ECMWF Re-analysis Project Report Series: ERA description. Number v. 1. European Centre for Medium-Range Weather Forecasts.

Gründemann, G.J., Werner, M., and Veldkamp, T.I.E. 2018. The potential of global re-analysis datasets in identifying flood events in Southern Africa, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-164, in review, 2018.

Guimberteau, M., Drapeau, G., Ronchail, J., Sultan, B., Polcher, J., Martinez, J.-M., Prigent, C., Guyot, J.-L., Cochonneau, G., Espinoza, J.C., Filizola, N., Fraizy, P., Lavado, W., De Oliveira, E., Pombosa, R., Noriega, L., and Vauchel, P. 2012. Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets, Hydrol. Earth Syst. Sci., 16, 911-935, doi:10.5194/hess-16-911-2012

Haddeland, I., Clark, D.B., Franssen, W., Ludwig, F., Voß, F., Arnell, N.W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S.N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G.P. and Yeh, P. 2011. Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results. J. Hydrometeor., 12, 869–884, https://doi.org/10.1175/2011JHM1324.1

Hamill, T.M., Whitaker, J.S., Snyder, C. 2001. Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. Mon. Wea. Rev., 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Hersbach, H., Peubey, C., Simmons, A., Poli, P.D.D., Berrisford, P., 2013. Era-20cm, a Twentieth Century Atmospheric Model Ensemble. Era report Series n 16. Technical Report 16. European Centre for Medium-Range Weather Forecasts, p. 44.

Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P. and Dee, D. 2015. ERA-20CM: a twentieth-century atmospheric model ensemble. Q.J.R. Meteorol. Soc, 141: 2350-2375. doi:10.1002/qj.2528

Houtekamer, P.L. and Mitchell, 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Wea. Rev., 129, 123–137.

Houtekamer, P.L. and Mitchell, H.L. 2005. Ensemble Kalman filtering. Q.J.R. Meteorol. Soc., 131: 3269-3289. doi:10.1256/qj.05.135

Houtekamer, P.L. and Zhang, F. 2016. Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. Mon. Wea. Rev., 144, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1

Houser, P., Rodell, M. 2002. GLDAS: An Important Contribution to CEOP. GEWEX Newsletter, May 2002, 2,8,9,16

Huang, C., Li, X., Lu, L., Gu, J. 2008. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sensing of Environment, 112:888–900.

van Huijgevoort, M.H., Hazenberg, P., van Lanen, H.A., Teuling, A.J., Clark, D.B., Folwell, S., Gosling, S.N., Hanasaki, N., Heinke, J., Koirala, S., Stacke, T., Voss, F., Sheffield, J. and Uijlenhoet, R. 2013. Global Multimodel Analysis of Drought in Runoff for the Second Half of the Twentieth Century. J. Hydrometeor., 14, 1535–1552, https://doi.org/10.1175/JHM-D-12-0186.1

Jacon, G.; Cudo, K.J. Curva-chave: análise e traçado. Brasília: DNAEE, 1989. 273 p.

Jarrett, R.D., 1991, Paleohydrology and its value in estimating floods and droughts, in Paulson, R.W., Chase, E.B., Roberts, R.S., and Moody, D.W., Compilers, National Water Summary 1988-89--Hydrologic Events and Floods and Droughts: U.S. Geological Survey Water-Supply Paper 2375, 105-116.

Kim, D.-I., Kwon, H.-H., and Han, D. 2018. Exploring the Long-Term Reanalysis of Precipitation and the Contribution of Bias Correction to the Reduction of Uncertainty over South Korea: A Composite Gamma-Pareto Distribution Approach to the Bias Correction, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-36

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M. 2001. The ncep–ncar 50–year reanalysis: Monthly means cd–rom and documentation. Bulletin of the American Meteorological Society, 82(2):247–267.

Kurtz, W., Lapin, A., Schilling, O.S., Tang, Q., Schiller, E., Braun, T., Jumkeler, D., Vereecken, H., Sudicky, E., Kropf, P., Franssen, H-J.H., Brunner, P. 2017. Integrating hydrological modelling, data assimilation and cloud computing for real-time management of water resources. Environmental Modelling & Software, 93: 418-435.

Lakew, H.B.; Moges, S.A.; Asfaw, D.H. 2017. Hydrological Evaluation of Satellite and Reanalysis Precipitation Products in the Upper Blue Nile Basin: A Case Study of Gilgel Abbay. Hydrology 2017, 4, 39.

Lenderink, G., Buishand, A., Van Deursen, W. 2007. Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. Hydrology and Earth System Sciences Discussions, European Geosciences Union, 11: p. 1145-1159.

Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P. 2012a Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16, 3863–3887, doi:10.5194/hess-16-3863-2012

Liu S, Shao Y, Yang C, Lin Z, Li M. 2012b. Improved regional hydrologic modelling by assimilation of streamflow data intoaregional hydrologic model. Environ Model Softw. 31:141-149. https://doi.org/10.1016/j.envsoft.2011.12.005

Liu, Y., Wang, W., Hu, Y., Cui, Wei. 2016. Improving the Distributed Hydrological Model Performance in Upper Huai River Basin: Using Streamflow Observations to Update the Basin States via the Ensemble Kalman Filter. Advances in Meteorology, Article ID 4921616.2016: 14p. http://dx.doi.org/10.1155/2016/4921616

Lopes, A.V., Chiang, J.C.H., Thompson, S.A., Dracup, J.A. 2016. Trend and uncertainty in spatial-temporal patterns of hydrological droughts in the Amazon basin. Geophys. Res. Lett. 43, 3307–3316.

López López, P., Wanders, N., Schellekens, J., Renzullo, L.J., Sutanudjaja, E.H. Bierkens, F.P. 2015. Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations. Hydrol. Earth Syst. Sci. Discuss., 12: 10559–10601.

Madsen, H. and Skotner, C., 2005. Adaptive state updating in real-time river flow forecasting - A combined filtering and error forecasting procedure. Journal of Hydrology, 308: 302–312. doi:10.1016/j.jhydrol.2004.10.030

Marengo, J.A. 2005. Characteristics and spatio-temporal variability of the Amazon River Basin Water Budget. Climate Dynamics, 24: 11-22.

Marengo, J.A., Tomasella, J., Alves, L.M., Soares, W., Rodriguez, D.A. 2011. The drought of 2010 in the context of historical droughts in the Amazon region. Geophys. Res. Lett. 38: 1–5.

Marengo, J.A., Tomasella, J., Soares, W.R., Alves, L.M., Nobre, C.A., 2012. Extreme climatic events in the amazon basin: Climatological and hydrological context of recent floods. Theoret. Appl. Climatol. 107 (1), 73–85.

Marengo, J.A., Borma, L.S., Rodriguez, D.A., Pinho, P., Soares, W.R., Alves, L.M., 2013. Recent extremes of drought and flooding in Amazonia, vulnerabilities and human adaptation. Am. J. Clim. Change 2, 87–96.

Marengo, J.A. and Espinoza, J.C. 2016. Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts. International Journal of Climatology, 36(3):1033–1050.

Meade, R.H., Rayol, J.M., Da Conceicão, S.C., Natividade, J.R.G. 1991. Backwater effects in the Amazon River basin of Brazil. Environmental Geology and Water Sciences, 18(2), 105-114.

Molina-Carpio, J., Espinoza, J.C., Vauchel, P., Ronchail, J., Gutierrez, B., Guyot, J.L., Noriega, L., 2017. The hydroclimatology of the upper Madeira River basin: spatio-temporal variability and trends (1967–2013). Hydrol. Sci. J. http://dx.doi.org/10.1080/02626667.2016.1267861.

Moradkhani, H. 2008. Hydrologic Remote Sensing and Land Surface Data Assimilation. Sensors, 8, 2986-3004.

Moss M.E., Hindall, S.M., Lins, H.F., Eberle, M. 1988. Water Resources in the Twenty-first Century: A Study of the Implications of Climate Uncertainty. Department of the Interior, U.S. Geological Survey, p. 17.

Munier, S., Polebistki, A., Brown, C., Belaud, G. and Lettenmaier, D.P. 2015. SWOT data assimilation for operational reservoir management on the upper Niger River Basin, Water Resour. Res., 51, 554–575, doi: 10.1002/2014WR016157.

New, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface climate over global land areas. Clim. Res. 21, 1–25.

Ngai, S.T., Tangang, F., Juneng, L. 2017. Bias correction of global and regional simulated daily precipitation and surface mean temperature over Southeast Asia using quantile mapping method. Global and Planetary Change, 149, 79–90.

Nijssen, B., Lettenmaier, D. 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, Journal of Geophysical Research, 109(D2, D02103).

Nkiaka, E., Nawaz, N.R., Lovett, J.C. 2017. Evaluating Global Reanalysis Datasets as Input for Hydrological Modelling in the Sudano-Sahel Region. Hydrology 2017, 4, 13.

Nobre, C.A., Sampaio, G., Borma, L.S., Castilla-rubio, J.C., Silva, J.S., Cardoso, M. 2016. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. PNAS 113(39):10759-10768. https ://doi.org/10.1073/pnas.16055 16113

Nunnery, A.; Baker, P. A.; Coe, M. T.; Fritz, S. C.; Rigsby, C. A. 2011. A quantitative history of precipitation and hydrologic variability for the last 45 ka: Lake Titicaca, Salar de Coipasa and Salar de Uyuni, Peru and Bolivia. In AGUGeneral AssemblyConference Abstracts #PP41B-1769

Nunnery, A. 2012. A Reconstruction of Precipitation and Hydrologic Variability on the Peruvian and Bolivian Altiplano During the Late. Thesis for the degree of Doctor of Philosophy in Earth and Ocean Sciences in the Graduate School of Duke University.

Ott, E., Hunt, B.R., Szunyogh, I., Corazza, M., Kalnay, E., Patil, D.J., and Yorke, J. E.2002. Exploiting low-dimensionality of the atmospheric dynamics for efficient ensemble Kalman Filtering, http://arXiv:physics/0203058

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., and Yorke, J. 2004. A local ensemble Kalman filter for atmospheric data assimilation, Tellus, 56A, 415–428

Paiva, R.C.D., Buarque, D.C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., Bulhões Mendes, C.A. 2013a. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resources Research, 49(3): 1226-1243.

Paiva, R.C.D., Collischonn, W., Bonnet, M.-P., De Gonçalves, L.G.G., Calmant, S., Getirana, A., Santos Da Silva, J. 2013b. Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. Hydrology and Earth System Sciences, 17: 2929-2946.

Patil, D.J.S., Hunt, B.R., Kalnay, E., Yorke, J.A., and Ott, E. 2001. Local Low Dimensionality of Atmospheric Dynamics, Phys. Rev. Lett., 86, 5878–5881

Pereira, D.M., Szlafsztein, C.F. 2016. Ameaças e desastres naturais na Amazônia sul ocidental, análise da bacia do rio Purus / Natural hazards and disasters in south western Amazon region: analysis of Purus river basin. Ra'e Ga – O Espaço Geográfico em Análise 35, 68–94.

Piani, C., Haerter, J.O., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. Theor. Appl. Climatol. 99 (1), 187–192. http://dx.doi.org/10.1007/s00704-009-0134-9.

Pontes, P.R.M., Fan, F.M., Fleischmann, A.S., de Paiva, R.C.D., Buarque, D.C., Siqueira, V.A., Jardim, P.F., Sorribas, M.V., and Collischonn, W. 2017. MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systemscoupled with open source GIS, Environmental Modelling & Software, 94, 1-20, 10.1016/j.envsoft.2017.03.029

Rasmussen, J., Madsen, H., Jensen, K.H., and Refsgaard, J.C. 2015. Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance, Hydrol. Earth Syst. Sci., 19, 2999-3013, https://doi.org/10.5194/hess-19-2999-2015

Razavi, S., Elshorbagy, A., Wheater, H., Sauchyn, D. 2016. Time scale effect and uncertainty in reconstruction of paleohydrology. Hydrol. Process.,30(13):1099-1085.

Reichle, R.H., Koster, R.D., Lannoy, G.J.M.D., Forman, B.A., Liu, Q., Mahanama, S.P.P., and Touré, A. 2011. Assessment and enhancement of Merra land surface hydrology estimates. Journal of Climate, 24(24):6322–6338.

Reichle, R. and Liu, Q. 2015. Precipitation and global land surface hydrology in the MERRALand and MERRA-2 reanalysis datasets. In EGU General Assembly Conference Abstracts, volume 17 of EGU General Assembly Conference Abstracts, page 1838.

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, 2004. The Global Land Data Assimilation System. Bull. Amer. Meteor. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381

Sakov, P. and Bertino, L. 2011. Relation between two common localisation methods for the EnKF.Comput Geosci, 15: 225. https://doi.org/10.1007/s10596-010-9202-6

Satyamurty, P., da Costa, C.P.W., Manzi, A.O., Candido, L.A., 2013. A quick look at the 2012 record flood in the amazon basin. Geophys. Res. Lett. 40 (7), 1396–1401.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G.P. 2017. A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth Syst. Sci. Data, 9, 389-413, https://doi.org/10.5194/essd-9-389-2017

Schmidli, J., Frei, C.,Vidal, P.L. 2006. Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods, Int. J. Climatol., 26, 679–689.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E. 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal, 48(6), 857-880.

Siqueira, V.A., Paiva, R.C.D., Fleischmann, A.S., Fan, F.M., Ruhoff, A.L., Pontes, P.R.M., Paris, A., Calmant, S., and Collischonn, W. 2018. Toward continental hydrologic–hydrodynamic modeling in South America, Hydrol. Earth Syst. Sci., 22, 4815-4842.

Sun, L., Seidou, O., Nistor, I., Liu, K. 2016. Review of the Kalman type hydrological data assimilation. Hydrological Sciences Journal, 61(13): 2348-2366.

Teutschbein, C., Seibert, J. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology, 456–457: 12-29.

Teutschbein, C., Seibert, J. 2013. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? Hydrology and Earth System Sciences, 17(12):5061-5077.

Titchner, H.A., and Rayner, N.A. 2014. The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2:1. Sea ice concentrations. J. Geophys. Res. Atmos., 119, 2864–2889, doi:10.1002/2013JD020316.

Tong, J-X., Hu, B.X., Yang, J-Z. 2012. Using an Ensemble Kalman Filter Method to Calibrate Parameters and Update Soluble Chemical Transfer from Soil to Surface Runoff. Transp Porous Med., 91:133–152.

Tong, X.T. 2018. Performance Analysis of Local Ensemble Kalman Filter. Journal of Nonlinear Sci., 1–46. https://doi.org/10.1007/s00332-018-9453-2

Uppala, S.M., KÅllberg, P.W., Simmons, A.J., Andrae, U. , Bechtold, V.D., Fiorino, M. , Gibson, J.K., Haseler, J. , Hernandez, A. , Kelly, G.A., Li, X. , Onogi, K. , Saarinen, S. , Sokka, N. , Allan, R.P., Andersson, E. , Arpe, K. , Balmaseda, M.A., Beljaars, A.C., Berg, L.V., Bidlot, J. , Bormann, N., Caires, S. , Chevallier, F. , Dethof, A. , Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L. , Janssen, P.A., Jenne, R., Mcnally, A.P., Mahfouf, J., Morcrette, J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. 2005. The ERA-40 re-analysis. Q.J.R. Meteorol. Soc., 131: 2961-3012. doi:10.1256/qj.04.176

U.S. Army Corps of Engineers. 2003. Application of Paleohydrology to Corps Flood Frequency Analysis. Report Document RD-47.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41(1).

Wanders, N., Bierkens, M.F., Sutanudjaja, E., van Beek, R., 2014. The PCR-GLOBWB global hydrological reanalysis product. In: EGU General Assembly Conference Abstracts, volume 16 of EGU General Assembly Conference Abstracts, 16, EGU2014-5369.

Wang, X-Y., Li, X., Zhu, J., Tanajura, C.A.S. 2018. The strengthening of Amazonian precipitation during the wet season driven by tropical sea surface temperature forcing. Environ. Res. Lett. 13 094015. doi.org/10.1088/1748-9326/aadbb9

Webb, R.H., Jarrett, R.D., 2002. One-dimensional estimation techniques for discharges of paleofloods and historical floods. In: House, P.K., Weeb, R.H., Baker, V.R., Levish, D.R. (Eds.), Ancient Floods, Modern Hazards: Principles and Applications of Paleoflood Hydrology. Water Resources Monograph, vol. 5. AGU, Washington, DC, pp. 111–125.

Willmott, C.J., Robeson, S.M., Feddema, J,J. 1994. Estimating Continental and Terrestrial Precipitation Averages from Rain-Gauge Networks. International Journal of Climatology, 14: 403-414.

Wongchuig, S.C., Paiva, R.C.D., Espinoza, J.C., Collischonn, W. 2017. Multi-decadal Hydrological Retrospective: Case study of Amazon floods and droughts. Journal of Hydrology, 549: 667-684.https://doi.org/10.1016/j.jhydrol.2017.04.019

Wood, A., Leung, L.R., Sridhar, V., Lettenmaier, D.P. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate outputs. Clim. Chang. 62: 189–216. http://dx.doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Xie, X., Zhang, D. 2010. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Advances in Water Resources, 33: 678–690.

Zhang, D., Madsen, H., Ridler, M.E., Kidmose, J., Jensen, K.H., and Refsgaard, J.C. 2016. Multivariate hydrological data assimilation of soil moisture and groundwater head, Hydrol. Earth Syst. Sci., 20, 4341-4357, https://doi.org/10.5194/hess-20-4341-2016

Zhu, J., Zheng, F. and Li, X. 2011. A new localization implementation scheme for ensemble data assimilation of non-local observations. Tellus A, 63: 244-255. doi:10.1111/j.1600-0870.2010.00486.x

Capítulo 5

Reanálise hidrológica na era dos satélites: Estudo de caso na bacia do rio Purus

Prólogo

A partir das últimas décadas encontram-se disponíveis uma grande quantidade de informação de produtos de sensoriamento remoto, as quais seriam úteis para propor a implementação de uma reanálise hidrológica na era dos satélites. Neste capítulo é apresentado o último artigo da tese, e que foi desenvolvido em base ao uso de dados sintéticos de sensoriamento remoto da futura missão SWOT no esquema de assimilação de dados e modelagem hidrológica. A missão SWOT, em princípio, será capaz de fornecer informação principalmente do nível d'água com maior resolução espacial, temporal e acurácia que missões atuais, maiores detalhes desta missão são descritos no Anexo B. Desta maneira os objetivos deste capitulo foram: i) a implementação e avaliação da assimilação das informações do futuro satélite SWOT no modelo MGB; ii) avaliar a sensibilidade do esquema de assimilação de dados a diferentes cenários de incertezas de parâmetros hidrológicos e hidráulicos; iii) implementar as bases técnicas para a assimilação de múltiplas observações no modelo hidrológico MGB.

A técnica de DA de múltiplas informações prevê como perspectiva futura deste capítulo o aproveitamento conjunto de uma grande quantidade de informação de sensoriamento remoto atualmente disponível (e.g. altimetria, gravimetria, etc.). Desta maneira, espera-se uma melhor representação física das variáveis de estado do modelo.

Assimilation of SWOT data into a large-scale hydrologic-hydrodynamic model: Case study of Purus basin

Wongchuig, C.S., de Paiva, R.C.D., Biancamaria, S., Collischonn, W.

Abstract

Global estimates of river dynamics are necessary to manage water resources, mainly in developing countries where insitu observation is limited. Remote sensors such as nadir altimeters can complement ground data, however current altimeters miss a large number of continental surface water bodies. What is expected to be surpassed by the future Surface Water and Ocean Topography (SWOT) mission, since it would be able to observe the seasonality of large portion of rivers and lakes globally. In this research the synthetic data of the SWOT mission was assimilated through the ensemble Kalman filter (EnKF) scheme in a large scale hydrologic and hydrodynamic model (MGB). A tropical basin (Purus) was chosen to perform the so-called "twin experiment" and the forcing and parameters of the model setup were corrupted to achieve the uncertainties of global hydrological models (GHMs). Where the SWOT community will mainly assess the SWOT data in real scenarios. The SWOT-like observations of discharge (Q), water surface elevation (WSE) and flooded water extent (FEW) were used to recover the original model. The results indicate that considering the expected errors of the SWOT observations, DA of these variables has a large impact in the improvements of hydrologic and hydrodynamic simulations in global and continental scales what implies a promissory usefulness by SWOT scientific community. SWOT-like discharge can be able to improve simulations by using hydrological models in global scale in approximately ~40% for the reduction of errors of daily discharge. Moreover, when the multiple-observation DA methodology was implemented, the results were even more promising. Considering these results, it is interesting to think about the use of the methodologies developed in this research, using multiple current remote sensing products such as satellite altimetry, terrestrial water storage, extension of flooded surfaces, among others.

Keywords: Continental modeling; Data assimilation; Observing system simulation experiment; Surface Water and Ocean Topography; Tropical basin

5.1. Introduction

The global estimates of river dynamics, including, discharge (Q), water surface elevation (WSE) and flooded water extent (FWE) still remains as a challenge in hydrologic field. They are important for applications such as hydrologic forecasting, flood risk, land use change scenarios, food production, water supply and quality among others (Wood et al. 2011). In addition, most basins in the world remains poorly gauged (Hrachowitz et al., 2013; Loukas and Vasiliades, 2014), especially in numerous developing countries where ground network is characterized for their sparse and irregular distribution. Usually, discharge of a specific gauging station are estimated by converting measured water level into flow rate by means of an existing stage-discharge relation, or rating-curve. The construction of stage-discharge relationships is based on several assumptions therefore, introducing uncertainties, in which errors in discharge measurements can range approximately from 6% to 20% (Pellettier, 1987; Schmidt, 2002; Herschy, 2002; Clark et al., 2008). In addition, it has been noticed a global decrease in measurements networks during last decades (Alsdorf et al., 2003, 2007; Silvapalan 2003), and their availability for scientific purposes could be limited due to institutional or international policies, restricting the exchange of data (Alsdorf et al., 2007). In addition, monitoring the dynamics in the extension of water, storage and flow in areas that are not controlled by defined channel networks (e.g. braided rivers, wetlands, etc.) are not measured by traditional insitu gauges.

Hydrological models could provide these estimates by simulating the water continental cycle at large scales (e.g. global and continental). These models have been improved in the last years in conceptual and physical structure and in computational efficiency, providing capabilities to simulate hydrologic and hydraulic processes for continental to global domains (Gao et al., 2010; Paiva et al., 2013a; Yamazaki et al., 2013; Sood and Smakhtin, 2015), however being locally relevant or representative (Wood et al., 2011; Bierkens et al., 2015). Nevertheless, there are still many uncertainties that should be kept in mind such as hydro-meteorological forcing, model structure, digital elevation model, calibration parameters and data availability and accessibility of which the models are sensitives (Döll et al., 2008; Thielen et al., 2010; Kauffeldt et al. 2016; Siqueira et al., 2018). Usually global hydrological models (GHM) provide estimates of discharge with errors ranging from 52% to 103% (relative root-mean-square errors RRMSE) in South America based on estimatives for HTESSEL, LISFLOOD and WaterGAP. Besides, Fleischmann et al. (2019) assessed a synthetic evaluation of water level for hydrological models at the global scale, which shows uncertainties of ~7.4m and ~1.3m for total and anomalies respectively.

In this context, hydro-meteorological observations are fundamental to help improving these hydrological models (Alsdorf et al., 2007; López López et al., 2017). Ground observations (e.g. discharge or water level in river channels) are traditionally used in water resources to calibrate and validate hydrological models (e.g. Gupta et al., 1998; Madsen, 2000; Wanders et al., 2014). By contrast, considering the limitation of current ground network measurements described above, remote sensing

has been used during recent years to complement it because i) greater spatial coverage (Engman, 1995; Tang et al., 2009; van Dijk and Renzullo, 2011; Wanders et al. 2014) and ii) accessibility from different sensors with a wide range of spatiotemporal, radiometric and spectral resolutions. Hence, remote sensing monitoring are suitable for the use for global and continental hydrological applications (Smith, 1997; Melesse et al. 2007; López López et al. 2017) and/or in large and remote regions, such as the Amazon basin (Koblinsky et al., 1993; Hess et al., 2015) or the Congo basin (Alsdorf et al., 2016; Becker et al., 2018). This also leads to further improvements in the representation of the hydrological physical processes (Yu, 2015).

Concurrently, data assimilation (DA) techniques has been developed during last years in climatic and oceanographic fields, to provide an effective way to combine information from models and measurements into an optimum analysis or estimation (Reichle, 2008). Many hydrological models and DA schemes has been widely used during last years to improve the estimations of their state variables. By using remote sensing observations such as soil moisture (Massari et al. 2015; Baguis and Roulin, 2017; Crow and Ryu, 2009), terrestrial water storage change (Tapley et al., 2004; Khaki et al., 2018), flooded water extent (Lai et al., 2014), and synthetic observations (Andreadis et al., 2007; García-Pintado et al., 2013; Fisher et al., 2018; Oubanas et al., 2018a).

At this point, numerous nadir radar altimeters missions have been also used in hydrologic studies by the assimilation of water surface elevations (WSE) (Paiva et al., 2013; Michailovsky et al., 2013) or discharge generated by rate curves altimetry based (Emery et al., 2018). These missions were initially meant to observe the circulation of the ocean and sea level change (Fu and Ubelmann, 2013), such as TOPEX/Poseidon (1992), followed by Jason-1 (2001), ENVISAT (2002) and more recently Jason-2 (2008) and Jason-3 (2016). They have an accuracy to detect water level ranging from 0.28m to 1.2m (O'Loughlin et al., 2016). Radar altimeters measure along the satellite's ground track missing a large number of continental surface water bodies (e.g. small lakes and reservoirs, most rivers, etc.) (Fu and Ubelmann, 2013). Henceforth, the spatial and temporal resolution of the present altimetry record is not sufficient for addressing and for reveal the spatial features of flow wave propagation and flood dynamics for medium to large rivers (David et al., 2011). Based on the development of wide-swath altimetry, a new satellite mission based on SAR (synthetic aperture radar) was recommended in 2007 to NASA (The National Aeronautics and Space Administration) by the NRC (National Research Council), under the name of SWOT (Surface Water and Ocean Topography) (Durand et al. 2010). This mission in collaboration developed by NASA, CNES (Centre National d'Etudes Spatiales), CSA/ASC (Canadian Space Agency/Agence Spatiale Canadienne), and UKSA (United-Kindom Space Agency), is scheduled to be launched in 2021. The SWOT mission is planned to make high-resolution measurement of the elevation of land surface water and ocean topography. SWOT is designed, in principle, to be able to observe the seasonality of large portion of river and lakes globally (Biancamaria et al., 2015), which will help to understand the surface water dynamics (Bates et al., 2014). SWOT will provide global

estimates of WSE and flooded water extent (FWE) with expected accuracy of less than 10cm and 15 % for 1 km^2 areas, i.e. 100m (width) x 10km (long) (Rodriguez 2015; Biancamaria et al., 2016; Durand et al., 2016). In addition, several efforts have been made for estimate discharge by using various algorithms (Durand et al., 2016), however due to the orbit cycle (~21 days) and algorithm uncertainties which can reach, in better cases, 35% for RRMSE, discharge from SWOT will not be a replacement of in-situ discharge stations (Biancamaria et al., 2016). Nevertheless, the assimilation of SWOT measurements into hydrologic models can be the alternative to improve the accuracy of such estimates and enhance the spatio temporal coverage.

Previous studies (Biancamaria et al., 2011; Pedinotti et al., 2014; Durand et al., 2016; Oubanas et al., 2018a, b; Tuozzolo et al., 2019) have been concentrated on development of DA techniques to ingest SWOT synthetic data (hereafter called SWOT-like observations) or observations from AirSWOT (an airborne variant of the SWOT instrument) into hydrodynamic models. They showed the benefits to improve hydraulic variables; nevertheless, proof of concept experiments were developed by these researches in specific test cases and with simplified representation of SWOT errors. For instance, most of them were applied for only single reaches of rivers, or even when a complex river network is used, such as for the Congo basins (Revel et al., 2019), unrealistic corrupted scenarios were assessed in the contex of the Observing System Simulation Experiment (OSSE). In addition, to achieve SWOT synthetic observations from water surfaces, approximations were mostly made by using a SWOT spatio-temporal sampling scheme, obtaining the swath coverage by deriving the prior known ground track, and adding the level of noise expected from the SWOT mission (Yoon et al., 2012, 2013; Willson et al., 2015; Paiva et al., 2015). However, these simplifications can lead to misrepresent most realistic errors. Besides, the SWOT Science Simulator (Desroches et al., 2018) is currently available, which is expected to reproduce the planned performance (spatial and temporal sampling and observational accuracy) of the KaRIn instrument on-board the SWOT satellite.

These opportunities call our attention due to the importance to explore and anticipate the use of the SWOT information in large-scale and in more complex hydrological systems before launch (Bates et al., 2014). Therefore, many applications in real scenarios will be addresses by SWOT community on continental and global scales (Yamazaki et al., 2011; Neal et al., 2012; Schumann et al., 2013, 2014; Bates et al., 2014; Biekens et al., 2015). In addition, the benefit of assimilating the different SWOT observations for enhancing estimates from continental to global hydrological models is not yet explored.

The purpose of this paper is to assess the integration of DA from SWOT-like observations into a large-scale hydrologic-hydrodynamic model. Therefore, several experiments are delineated to unveil scientific questions such as: How the assimilation of different SWOT-like observations can improve global and continental hydrological models? How sensible is the DA scheme to the uncertainties of the parameters of the model? Is the multi-variables assimilation necessary to further improve large-scale

hydrological models? Therefore, i) we implement and assess the DA of SWOT-like observations into a hydrologic-hydrodynamic large-scale model; ii) we assess the sensibility of the model parameters to different corrupted scenarios and iii) finally we present the multi-variables assimilation as an improved DA scheme.

Section 5.2 describes the general framework and the design of the SWOT Observing System Simulation Experiment, where a brief description of the SWOT mission, the study area, the semi-distributed hydrological-hydrodynamic model and the DA scheme used in this research are presented. In the section 5.3, the characterization of hydrological errors of the model are delineated. Section 5.4 shows the extraction of the SWOT-like observations and in the section 5.5 the experimental setup is performed. Section 5.6 shows the main results of the experiments, considering the goal of achieving uncertainty from global models, the sensitivity of the model to different corrupted scenarios and the improvements on performance by using SWOT information at global scales. Finally, the conclusions and perspectives are outlined in section 5.7.

5.2. Design of the SWOT Observing System Simulation Experiment

This study is designed in the context of data assimilation framework using an Observing System Simulation Experiment (OSSE) (Biancamaria et al., 2016), also known as twin experiment. The OSSE can be explained as a modeling experiment used to evaluate the impact of new observing systems when actual observational data is not available. One of their applicacions is the assessement of future satellite observations (Masutani et al., 2010; Biancamaria et al., 2016), as the case of the SWOT mission. Therefore, a hydrologic and hydrodynamic MGB model was used to compute the “truth” and the “corrupted” versions, in the context of the OSSE. The corrupted version of the model was notably set up to represent the uncertainties of global models. The pseudo observations (SWOT-like observations) are extracted with the SWOT simulator described in section 5.4. Therefore, SWOT-like observations are used into the Ensemble Kalman Filter (EnKF) DA scheme to update the state variables of the corrupted version of the model. The EnKF has been widely used in hydrological assessements and by the SWOT scientific community (e.g. Andreadis et al., 2007; Durand et al., 2008; Biancamaria et al., 2011; Yoon et al., 2012; Andreadis and Schumann 2014, Munnier et al., 2015; Revel et al., 2019). The simulation period for the orbits of the SWOT was defined from 2011 to 2012 due to the flood reported in 2012 (Marengo and Espinoza, 2016) that would allow for the testing of more SWOT samples.

5.2.1. The SWOT mission

The major payload technology carry on SWOT satellite is a Ka-band Radar Interferometer (KaRIN), which is a wide swath radar interferometer. These ground swaths will cover 50 km separated by a gap of 20 km (Alsdorf et al., 2007; Biancamaria et al., 2016). To obtain synthetic SWOT data, previous studies (e.g. Yoon, et al., 2013; Munier, et al., 2015; Paiva, et al., 2015) have tested a simple method where spatio-temporal sampling was calculated through SWOT pre-defined orbits and swath tracks. Besides, JPL and CNES institutions have developed the SWOT Science Simulators. One of them, the High Resolution (HR) SWOT simulator has some limitations to be used in large scale applications, due to its notable input demands such as a high resolution Digital Elevation Model (DEM) (10 meters), only applicable to a study domain of ~100km x ~100km. Therefore, the Large Scale L2 HR Pixel Cloud Simulator (LSPCS), which is a large scale version of the SWOT simulator, was used in this research. It considers a white noise increases along the swath (look angle), although some errors measurements were not taken into account i.e. interferogram computation, layover, etc. The main advantage of this simplified version are the less restrictive inputs requirement (DEM and vector maps of water mask) in comparison with the SWOT High Resolution simulator. In addition, it allows a multi temporal simulation with relative less computational requirements. Detailed information about LSPCS can be found at Desroches et al (2018).

5.2.2. Case study

5.2.2.1. The Purus basin

The Purus basin has an extent of ~370,000 km^2 (Fig. 1) and is one of the major tributaries of the Solimões river. The Purus River is characterized for the high concentration of supended sediments, mainly composed by silt and clay that comes from the Andes (McClain and Naiman 2008), this characteristic influences in the geomorphology of the main river (Latrubesse 2008). Due to its meandering nature, it has a large river length of ~3,380 km (Goulding et al., 2003). The high interconnectivity between the main river and floodplains, the combination of lakes, unflooded and flooded forests, floating-meadows and the abundance in nutrients, makes this a highly productive and diverse region, defining a large number of habitats with fish diversity (Goulding et al., 2003), what characterizes it as a region with large environmental importance.

The mean annual discharge is estimed in 9,500 m^3s^{-1} at Beaba (Cariuacanga) gauge station (near to their confluence with the Solimões River). The main direction (south-west to northeast) of the Purus River has a channel pattern of tortuous meandering (Latrubesse 2008). Therefore, the whole Purus basin has been chosen as a case study due to offers a good observational sampling from remote sensing in the tropical region. Purus basin can be considered as a large basin with a more complex river network than previous assessements of SWOT satellite, which mostly used only single reach rivers.

. In the Purus river basin, the SWOT observation frequency was calculed for each river reach of 10 km. Therefore, 97% of the river reaches, larger than 50m, will be observed by the satellite and the number of observations ranges from zero to three for each 21-day cycle.

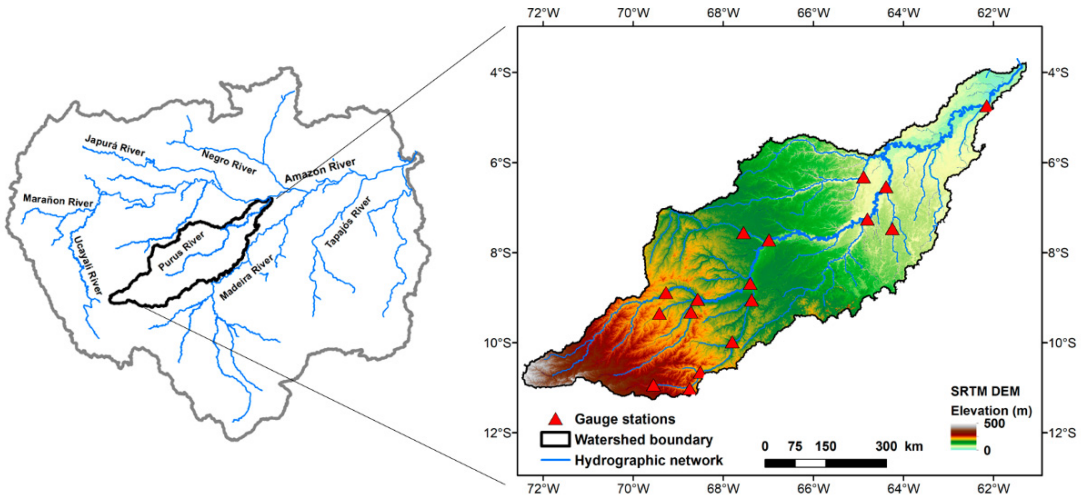

Figure 1. Map of Purus Basin, in-situ gauges' location and digital elevation model

5.2.2.2. The hydrologic-hydrodynamic model

In this research we used the large-scale hydrologic-hydrodynamic MGB model. The MGB model was selected due to (i) its satisfactory performance for large-scale applications (e.g., Paiva et al., 2013a; Pontes et al., 2017; Siqueira et al., 2018) and (ii) past successful data assimilation applications (e.g., Paiva et al., 2013b; Wongchuig et al., 2019), which demonstrate the potential of this methodology to improve spatio-temporal estimates of hydraulic variables.

The MGB is a semi-distributed hydrological model which uses physical and conceptual based equations to simulate the continental phase of the hydrological cycle (Collischonn et al., 2007). The watershed is discretized into irregular unit-catchments and further into hydrological response units (HRUs), where vertical water and energy budget are computed individually. For the hydrodynamic component, each unit-catchment is composed by a river reach that includes both channel and floodplain units, the latter depicted as a simple storage with ineffective flow (Pontes et al., 2017; Siqueira et al., 2018). The surface, sub-surface and groundwater runoff produced at each unit-catchment into the hydrologic module are routed to the stream network based on linear reservoir concept. The main hydrological soil parameters of the model are related with the saturation excess concept based on the variable contributing area concept of the ARNO model (Todini, 1996). The flow routing in river channels is computed using the explicit local inertial method (Bates et al., 2010; Pontes et al., 2017), which is an approximation of the full 1D Saint-Venant equations and is able to represent backwater

effects and floodplain attenuation. The main hydraulic state variables which will be related with SWOT-like observations are discharge, water level and flooded area extension.

The model setup used here over the Purus basin is an application of Pontes et al. (2017), which was discretized into 1613 unit-catchments and river reaches of ~10 km. The model was calibrated using MSWEP v 1.2 precipitation datasets (Beck et al., 2016), with spatial resolution of 0.25° x 0.25° and daily time step for a period spanning 15 years (2000- 2015) and monthly meteorological data obtained from the CRU CL 2.0 dataset (New et al. 2002). The model parameters related to soil water budget were calibrated using discharge data from in situ stream gauges (17 stations) and a single set of soil parameter for all basin has been established. The model was validated against discharge data from stream gauge stations during the last 5 years of simulation by using NSE and KGE indexes.

5.2.2.3. Data assimilation scheme

Among data assimilation techniques, the Ensemble Kalman Filter (EnKF) (Evensen, 1994; Burgers et al., 1998) has become popular due to its relative simple conceptual formulation and computational implementation (Evensen, 2003). The EnKF method has been widely used by the SWOT scientific community such as Andreadis et al. (2007), Durand et al. (2008), Biancamaria et al. (2011); Yoon et al. (2012), Andreadis and Schumann (2014), Munnier et al. (2015), Revel et al. (2019), among others since it enables a simplified implementation on hydrological models where non-linear behaviour are predominant. It uses a Monte Carlo formulation based on the ideas of the Kalman filter (KF). For example, in the Kalman filter (KF) methodology, a single state estimate is used to produce an analysis; in contrast, the main idea in the EnKF is that an ensemble, which is a statistical sample of state estimates, is used to calculate an analysis state for each member. From a forecast ensemble the sample mean and error covariances are calculated. During the analysis stage these are then used to calculate a single Kalman gain that is used to assimilate the observations and produce an analysis state ensemble.

The implementation of the EnKF DA scheme in MGB was based on the algorithm developed by Evensen (2004). The main goal of EnKF is the optimization of the model state variables ($\mathbf{x}^a$) estimate based on model and observation uncertainties. Eleven state variables were considered to be updated into the EnKF scheme (e.g. soil water storage, surface reservoir volume, sub-surface reservoir volume, groundwater reservoir volume, catchment outflow discharge, water level, flooded water extent, among others).

The DA scheme is divided into two steps, forecasting and updating, which are described below:

During the stochastic formulation of EnKF DA, the matrix of the forecasted ensemble of model states $\mathbf{x}^f$, also called background, is represented as:

$$\mathbf{x}^f = \left(\mathbf{x}_1^f, \mathbf{x}_2^f, \dots, \mathbf{x}_{N_{ens}}^f\right) \quad (1)$$

where $\mathbf{x}_i^f$ represents each ensemble member of the model states until the total number of defined members (N_{ens}).

In the updating formulation the $\mathbf{x}^a$ is estimated by the Kalman filter as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}_e\left(\boldsymbol{y} - \mathbf{H}\mathbf{x}^f\right) \quad (2)$$

$$\mathbf{K}_e = \mathbf{P}_e^f \mathbf{H}^T \left[\mathbf{H}\mathbf{P}_e^f \mathbf{H}^T + \mathbf{R}_e\right]^{-1} \quad (3)$$

where $\mathbf{x}^a$ and $\mathbf{x}^f$ are the estimated and forecasted model state variable respectively, $\mathbf{K}$ the Kalman Gain, $\mathbf{P}^f$ and $\mathbf{R}$ covariance matrices of model (q) and observation (ε) uncertainties, respectively, $\mathbf{H}$ is a function that relates the model state variables $\mathbf{x}^f$ with the correspondent observation $\mathbf{y}$.

The covariance matrix of the state variables is generated by a set of members from perturbations of the model (forcing, state variables, parameters), based on a known error:

$$\mathbf{P}^f \cong \mathbf{P}_e^f = \frac{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^T}{N_{ens} - 1} = \frac{\mathbf{x'}^f \mathbf{x'}^{f^T}}{N_{ens} - 1} \quad (4)$$

where the ensemble mean $\overline{x^f}$ denotes the best guess and x'^f represents the ensemble of perturbations.

$$\mathbf{y} = \left(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{N_{ens}}\right) \quad (5)$$

$$\mathbf{E} = \left(\mathbf{y}_1 - \bar{y}, \mathbf{y}_2 - \bar{y}, \dots, \mathbf{y}_{N_{ens}} - \bar{y}\right) \quad (6)$$

where y and E indicate the ensemble matrix of perturbed observations and perturbations respectively.

Therefore, the observation error covariance matrix is calculated:

$$\mathbf{R} = \frac{\mathbf{E}\mathbf{E}^T}{N_{ens} - 1} \quad (7)$$

Finally resulting the updating formulation as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{x'}^f \mathbf{x'}^{f^T} \mathbf{H}^T \left(\mathbf{H}\mathbf{x'}^f \mathbf{x'}^{f^T} \mathbf{H}^T + \mathbf{E}\mathbf{E}^T\right)^{-1} \left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right) \quad (8)$$

Finally, the covariance matrix of the observation errors ($\mathbf{R}_e$) was estimated following the Clark et al. (2008) approach:

$$\mathbf{R}_e = (\varepsilon_{obs} \cdot var_{obs})^2 \quad (9)$$

where ε_{obs} represents the error in the observation that can be known a priori and var_{obs} indicates the value of the observed variable.

Several improvements to the EnKF scheme were made, for instance the Localization Analysis (LA) was proposed by Houtekamer and Mitchell (2001) who use the “covariance localization” method

to limit the updating of state variable into a spatially area from the observation. This is important to avoid spurious correlations between two points mainly because i) the use of limited number of ensemble members, mostly to avoid large computational effort and ii) large distant or physically disconnected points, which could have significant impact in large scale regions. Detailed information is described in supplementray material. To provide improvements on hydrologic-hydrodynamic simulations by the use of SWOT- like observations, a DA scheme from multi-variable was also implemented. This technique was recently assessed in hydrology once several remote sensing data are currently available; in addition, several authors have realized benefits due to the better representation of more hydrological-hydrodynamic processes (Khaki et al., 2019).

In the Kalman Filter assumptions, the errors of the model should be estimated for the computation of the analysis or updated state variables. In this context, given the difficulty in knowing hydrological model errors, in the Ensemble Kalman Filter it is common to estimate these errors from the perturbation of the model's forcing such as precipitation for the ensemble generation. Since is assumed that rainfall is the main source of error in hydrological models (Biancamaria et al. 2011; Liu et al., 2012; Paiva et al., 2013b). Therefore, the EnKF DA scheme was evaluated in this study through perturbation of precipitation. The adopted perturbation follows a log-normal distribution of errors, as proposed by Nijssen and Lettenmaier (2004):

$$P_c = \frac{1+\beta}{\sqrt{E^2+1}} exp\left(\sqrt{ln[E^2+1]}s\right) P \tag{10}$$

where P_c is perturbed variable (precipitation or model state variables), P non-perturbed variable, E variable relative error (%), β relative bias (here adopted as 0.0), and s is a random variable that follows a time evolution of model error according Evensen (2003):

$$\begin{aligned} S_k &= \alpha S_{k-1} + \sqrt{1-\alpha^2} w_{k-1} \\ \alpha &= 1 - \frac{\Delta t}{\tau_t} \end{aligned} \tag{11}$$

The coefficient $\alpha \in [0\ 1\rangle$ indicates the influence of temporal correlation on the model errors (α=0 indicates a sequence of white noise, while α=1 indicates a random field of errors constant in time, which removes the stochastic forcing); w is a stochastic term with mean 0 and variance 1; and τ_t is the temporal decorrelation length (time units).

5.3. Designing Hydrological model errors

In this section, we conducted the considerations for the corruption of the "truth" model based on the perturbation of the precipitation forcing and several parameters of the model. Therefore, to generate the open-loop simulation, errors in global and continental models were adopted to represent uncertainties of simulations on these scales. The estimative of the ensemble of uncertainty on discharge

of three GHMs, also assessed by Siqueira et al. (2018), was used as reference. Considering that the sources of hydrological model uncertainty for global and continental scales are primarily represented by precipitation forcing, river bathymetry, floodplain topography and water balance parameters, these were considering to corrupt the model. Therefore, a literature review about errors of these parameters in global and continental scales was conducted, which were used to corrupt the "truth" model to achieve similar results of uncertainty of GHMs. In addition, the same metrics of errors assessed by Durand et al. (2016) were used in this research. For RRMSE, the ensemble of the models (EGHMs) shows a median of 73% for daily discharge, that we will try to achieve in our model with all errors combined. The assessed models were HTESSEL, LISFLOOD, WATERGAPand the ensemble of these global models (EGHMs), The MGB results from South America model (Siqueira et al. 2018) was also used for comparison.

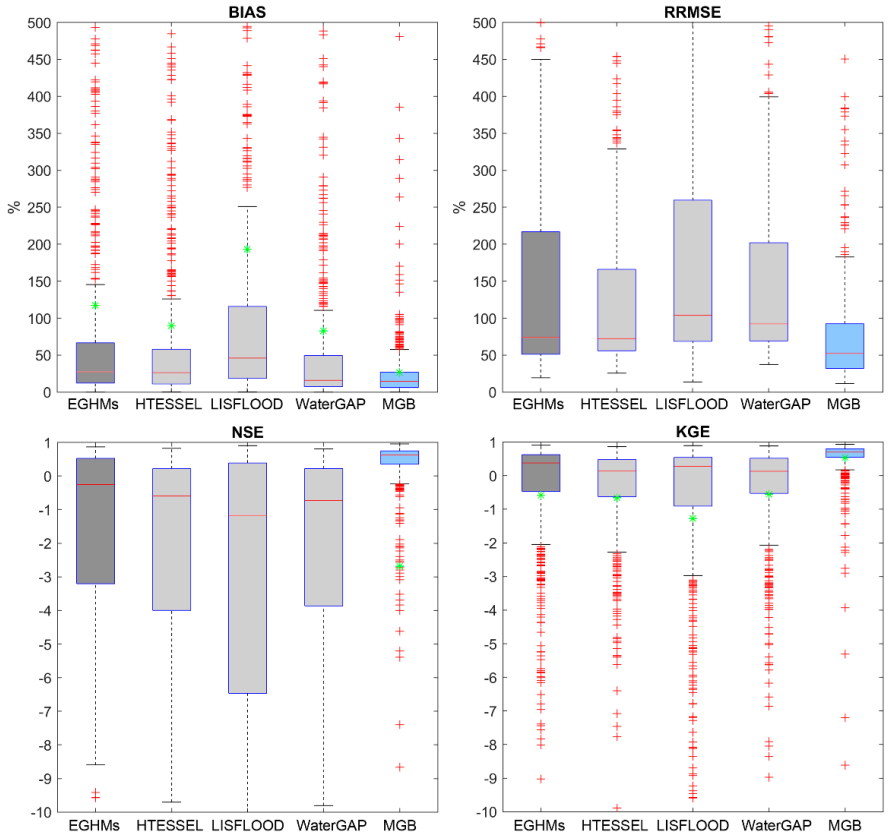

Figure 2. Boxplots of the performance of daily discharge for global and continental scale hydrological models, showing the ensemble of global models (EGHMs), HTESSEL, LISFLOOD and WaterGAP models and the MGB model.

The Figure 2 shows that the MGB discharge simulations in continental scale has better performance among the assessed GHMs (Fig. 2). Hence, for our study we will focus on the ensemble of them that represents the best estimator, and the specific RRMSE index was taken into account in this

section to reach uncertainties of discharge for global models; this index reaches an approximately value of 73% for the median of RRMSE.

5.3.1. Rainfall

Precipitation forcing is the major source of hydrological model uncertainty (Tapiador et al., 2012; Weiland et al., 2015). Among precipitation datasets based on insitu gauges, satellite and reanalysis are typically applied at the global scale (Saha et al., 2011); however, there are many uncertainties due to dependency on data availability in time and space that can produce spurious trends and artificial variability (Tapiador et al. 2012; Essou and Brissette, 2017). The uncertainty of 22-precipitation datasets was assessed by Beck et al. (2017), which could give us an order of magnitude of the associated bias that achieves ~17% for global datasets. Another more extensive assessment of 30 global rainfall datasets (Sun et al. 2017) results for South America in values of systematic errors and random errors that ranges between 20% to 40% and 60% to 80% respectively. Therefore, the values for systematic and random errors of ~25% and ~70% respectively were assumed in this research as representative for South America. These magnitudes were used to generate the corrupted scenarios of rainfall based on a lognormal distribution of errors (Nijssen and Lettenmaier, 2004), and a multiplicative error as recommended by Tian et al. (2013) to avoid negative values.

5.3.2. Hydrological parameters

Most of continental and global hydrological models use parameters that are chosen based on a priori datasets without extensive calibration against in situ discharge data (Siqueira et al., 2018, Beck et al., 2017). The uncertainty of model parameters was represented by uniformly random selection of values in the range found by Siqueira et al. (2018) in the calibration of MGB model for different South America catchments. The range of these parameters are: Wm = 50–1500 mm, b = 0.02-1.5, Kbas = 0.01-3.0 mm day^{-1}, Kint = 0.1-50 mm day^{-1}, XL = 0.1-0.67, Cs = 5-35, Ci = 20-200 and Cb = 800-6000 h.

5.3.3. Floodplain bathymetry

Since the river bathymetry based on the DEM is central to represent adequately flooded water extent in hydrodynamic models (Bates et al., 2014), the hypsometric curve was also perturbed for each catchment, using a simple bias additive error to the floodplain based on the approximation of the DEM errors. The assumed absolute bias is one of the main source of errors in global DEMs (Yamazaki, et al. 2017) which is around of ~10m.

5.3.4. Hydraulic parameters

Large-scale hydrological models usually represent river bathymetry as rectangular cross sections width (w) and depth (d). These parameters could be estimated using hydraulic geometry relations as developed by Andreadis et al. (2013) in a global dataset of river geometry. Uncertain river

width and depth from these estimates can be modelled as showed in Table 1, where β and ε are regional bias and reach random errors. The relations proposed by Andreadis et al. (2013) and Moody and Troutman (2002) present estimates ranging from approximately 1/3 to 3 at 95% confidence interval for width and 1/2.5 to 2.5 at 95% for depth which means that uncertainty could be modelled as a log-normally distributed random noise. With a multiplicative standard deviation or shape parameter (σ*) of approximately 1.73 and 1.58, which transformed to a normal σ become in 0.55 and 0.46 respectively. In addition, bias and noise show the same order of magnitude, therefore a simplification could be assumed by breaking errors as β_w ~logN (1, 1.22^2), ε_w ~logN (1, 1.22^2) (relative error of 39% for bias and noise) and β_d ~logN (1, 1.12^2), ε_d ~logN (1, 1.12^2) (relative error of 33% for bias and noise), considering total errors modelled as $\varepsilon_T = \sqrt{\beta^2 + \varepsilon^2}$ similar to proposed by Durand et al. (2016), where ε_T represents total errors.

For the Manning's coefficient, the values for natural streams (main channels and floodplain) were used as a proxy of uncertainty (Gichamo et al., 2012; Anees et al., 2017), which ranges from 0.02 to 0.16. Therefore, considering a mean of the Manning's value of ~0.056 for the main channels and floodplain, at 95% confidence interval, we can corrupt it with a multiplicative noise (ε_n) that follows a lognormal distribution with a standard deviation of ~1.69, which represents approximately a relative error close to ~50%.

Table 1. Summary of the variables of the model and corruption setup

	Truth Model	Corrupted Model	Errors	Reference
Rainfall	MSWEP	Adopted a multiplicative error perturbation	β_p ~logN with relative error of ~25% ε_p ~logN with relative error of ~70%	Nijssen and Lettenmaier (2004), Sun et al. (2017)
Hydrological parameters	Calibrated	Uniform random selection within the parameters' interval	-	Siqueira et al. (2018)
Floodplain Topo (DEM)	SRTM Bare earth	Perturb the hypsometric curve using a regional bias which follows a Gaussian normal distribution. $flp' = flp + \beta_{flp}$	β_{flp} ~N with error of ~10m	Farr et al. (2007); Yamazaki et al. (2017)
Hydraulic parameters (channel width and depth and Manning)	Geomorphologic equations (Paiva et al., 2013a)	$w' = w * \beta_w * \varepsilon_w$ $d' = d * \beta_d * \varepsilon_d$ $n' = n * \varepsilon_n$	β_w and ε_w ~logN (1, 1.12^2) or with relative error of 39% β_b and ε_b~logN (1, 1.22^2) or with relative error of 33% ε_n ~logN (1, 1.69^2) or with relative error of ~50%	Andreadis et al. (2013) Moody and Troutman (2002) Anees et al. (2017)

5.4. SWOT-like observations

The SWOT-like observations into the LSPCS were generated in two steps: i) the estimation of the orbits for the study area, and ii) generation of the pixel clouds by using a multitemporal simulation. Considering that the information provided by the LSPCS SWOT simulator, among others, are primarily WSE errors and not real observations, the best setup of the hydrological model, also called the "best effort" is assumed to be true. The vectorized water masks of each day, provided by the truth version of the MGB, were used as an input for the LSPCS, which are intersected with the two swaths of the predefined SWOT orbits, therefore, the pixel cloud information is generated within the respectives water masks. The hydrodynamic scheme included into MGB model was used to recover the true variables such as *WSE*, discharge (*Q*) and *FWE* values. Thus, the SWOT-like observations were estimated as the "truth" model plus a measurement error. The SWOT-like observation is obtained as follows:

$$y = y^t \cdot \varepsilon \tag{12}$$

Where y represents the pseudo-observation (error-containing observation), $\mathbf{y^t}$ the true observation (error-free observation) and ε the error in observations.

For the *WSE*, the mean of the pixel cloud of SWOT *WSE* errors within each catchment were used to perturb the truth *WSE*, and the standard deviation of these errors was assumed as the uncertainty of the *WSE* observation.

The *Q* and *FWE* pseudo-observations were estimated by adding a multiplicative noise, which is assumed to follow a Gaussian-distributed with zero mean (Xue et al., 2007). It means that ε (in equation 12) is calculated as $(1 + \alpha r)$ where α represents specific percentage factor and r a Gaussian-distributed random number with zero mean and standard deviation of 1 (Xue et al., 2007). The specific percentage factor α add the uncertainty to the dispersion (standard deviation) of the observations. Because SWOT simulator only provides errors on *WSE*, the true *Q* and *FWE* were extracted at the same time and space of the *WSE*. The uncertainties were assumed by the estimates of the global algorithms uncertainty (Durand et al., 2016) and the SWOT scientific measurement requirements (SWOT, 2017) respectively.

In the EnKF scheme, the innovation matrix $[\mathbf{y} + \epsilon_n - \mathbf{H}\mathbf{x}_n^f]$ is estimated by the difference between the ensemble matrix of perturbed pseudo-observation and the ensemble matrix of state variables ($\mathbf{x}_n^f$) mapped to the measurement space by the $\mathbf{H}$ operator. While ϵ_n represents an ensemble matrix random noise which follows a Gaussian distribution with mean equal to zero (Evensen, 2003) and taking also into account the uncertainties of the observations. Henceforth in OSSE SWOT experiments, the standard deviation of the pixel cloud to create the ensemble matrix of perturbed pseudo-observation was used for water level assimilation. For *Q* and *FWE*, the uncertainties were assumed from algorithm's error estimation and scientific requirements respectively as showed in Table 2.

The estimated standard deviation of *WSE* errors provided by LSPCS ($\sigma_{\varepsilon\text{LSPCS}}$) are still large because of the high pixel resolution. However, errors should decrease exponentially with pixel aggregation (Andreadis et al., 2007, Bates et al. 2014). Therefore, they were aggregated (averaged) considering the number of pixel clouds within one km^2 through the standard error estimation (Altman et al., 2005) (Eq. 13).

$$\varepsilon^{LSPCS} = \frac{\sigma_{\varepsilon LSPCS}}{\sqrt{N}} \tag{13}$$

where $\sigma_{\varepsilon LSPCS}$ is the standard deviation of the sampled pixel clouds and N is the sampled size represented by the number of pixel clouds.

The multiplying parameter is considered to be systematic error and stochastic error in discharge, what means that a bias (ε_{bias}) and a random noise (ε_{noise^t}) were added to the true of discharge. Based in the equation 14 and following a multiplicative perturbation, the error (ε) could be represented as ($\varepsilon_{bias} * \varepsilon_{noise^t}$) for lognormal distribution (Table 2). To define the uncertainty of discharge estimation by SWOT, the discharge errors estimations of Durand et al. (2016) were used in this research. Considering that discharge errors were expressed in terms of the relative residuals of discharge (RR) which is defined as:

$$RR_t = \frac{Q^{corrupted}{}_t - Q^{true}{}_t}{Q^{true}{}_t} \tag{14}$$

Based on the equations 12 and 14, the corrupted variable is expressed in terms of RR as following:

$$Q^{corrupted} = Q^{true} \cdot (1 + RR) \tag{15}$$

The $(1 + \text{RR})$ is assumed as a Gaussian-distributed random number, therefore the mean (MRR) and the standard deviation (SDRR) of the discharge relative residual (RR) (Durand et al., 2016) could be used in the way that:

$$mean(1 + RR) = 1 + MRR \tag{16}$$

$$std(1 + RR) = SDRR \tag{17}$$

These values were already estimated by Baratelli et al. (2018) from Durand et al. (2016) who considered values around -18.8% and 15% as the median of MRR and SDRR respectively for all river algorithms.

Lastly, for *FWE* perturbation only a random noise ($\varepsilon_{noise^t}^{FWE}$) was added to the true value. For this purpose, the value of 15% of uncertainty established in the SWOT scientific requirements document (SWOT, 2017) was used.

Table 2. Summarize of the perturbation setup for SWOT-like observations

SWOT-like observation	Uncertainty (assimilation / corruption)		Corrupted model by measurement errors (Gaussian distribution)	# of realizations
WSE	ε^{LSPCS}	min ~0.10 m for 1km^2	$WSE^{corrupted} = WSE^{true} + \mu_{\varepsilon\text{LSPCS}}$	One
Q	Algorithm's uncertainties, (Durand et al., 2016; Baratelli et al., 2018).	MRR and SDRR -18.8% and 15%	$Q^{corrupted} = Q^{true} * \varepsilon_{bias}^{Q\ c} * \varepsilon_{noise^t}^{Q}$	10
FWE	Mission performance and error budget reference (SWOT, 2017).	σ_{FWE} 15 %	$FWE^{corrupted} = FWE^{true} * \varepsilon_{noise^t}^{FWE}$	10

Therefore, in this study true variables (Q and FWE) are assumed to be corrupted at all-time steps by the random errors of the error terms ($\varepsilon_{noise^t}^{Q}$ and $\varepsilon_{noise^t}^{FWE}$), while the systematic error (ε_{bias}^{Q}) for Q is constant in time and spatially correlated between contiguous catchments within the river network to ensure continuity.

$$\varepsilon_{bias}^{Q} \sim logN\left(1, e^{MRR^2}\right) \quad \varepsilon_{noise^t}^{Q} \sim logN\left(1, e^{SDRR^2}\right) \tag{18}$$

$$\varepsilon_{noise^t}^{FWE} \sim logN\left(1, e^{\sigma_{FWE}{}^2}\right) \tag{19}$$

The continuity assumption for discharge is an important issue for some algorithms tested by Durand et al. (2016) (e.g. AMHG and MetroMan). To ensure that, three contiguous reaches within the river network were considered as "full spatial correlated". To generate the random numbers by a multivariate normal distribution, the following equation was used here, which is based in the statement that the linear combination of two independent random variables having a normal distribution also has a normal distribution:

$$\varepsilon_{bias}^{Q\ c} = \rho \cdot \varepsilon_{bias}^{Q\ 1} + \sqrt{1-\rho^2} \cdot \varepsilon_{bias}^{Q\ 2} \tag{20}$$

where $\varepsilon_{bias}^{Q\ 1}$ and $\varepsilon_{bias}^{Q\ 2}$ are independent realizations and ρ is the correlation matrix of one and zeros, where one represents full correlation.

To estimate the multiplicative errors, the adopted perturbation follows a log-normal distribution of errors (Nijssen and Lettenmaier, 2004), described in Eq. 10.

5.5. Experiments setting

Table 3 summarizes the experiments defined in this research, where different scenarios are assumed to assess the improvements in the states variables of the model using DA scheme and complementary localization and multi-source techniques.

Table 3. Summary of the experiments and main characteristics and setup

Experiment	Assimilated SWOT-like observation *	Product	Realizations
1	-	To achieve the uncertainties of GHMs by corrupted parameters of the model	10
2	WSE, FWE, Q and MV	DA sensivity based on individual and full-corrupted scenarios	10
3	WSE', FWE' and Q'	Improving by assimilating the anomaly of the SWOT-like observations	10

* WSE=Water Surface Elevation, FWE=Flooded Water Extent, Q=Discharge, MV=Multi-variable, " ' " represents the variable's anomaly

In addition, a localization experiment was implemented to estimate the optimum radius of influence in the Purus Basin and to assess the improvements by using this technique. Nevertheless, due to the computational demand, it was only assessed by assimilating WSE. Results are showed in the suplementary material section.

The corruption of the model was performed to achieve similar performances in discharge estimates by GHMs, based on the uncertainties of some important forcing and parameters of the model such as rainfall, floodplain topography, hydrologic and hydraulic. For a better representation and reduction of uncertainties of this experiment, 10 realizations were performed for the corruption of each forcing and parameters.

The following experiment was stablished to assess the sensibility of the corrupted models responses by assimilating different SWOT-like observations. As the Table 4 shows, individual corrupted models were assessed and the full-corrupted model that corresponds to the closest model version to the EGHM performance evaluated in previous sections. The four assessments were done by assimilating each SWOT-like observation. A reference setup by using an ensemble size of 100 members, 10 realizations and no localization method were assumed within the DA technique. An additional experiment was assessed by using the multi observation method in the EnFK technique, considering that the assimilation of combined observations at the same time could be more physical representative by

improving of state variables of the model. The goal of the last experiment is to improve the state variables of the model by assimilating the anomalies of the SWOT-like observations, henceforth bias on the observations would be avoided.

Table 4. Summary of the perturbed variables taken into account for the corrupted models regarding the experiments.

Experiment	SWOT-like obs / Corrupted model	Rainfall	Hydrologic parameters	Floodplain parameters	Hydraulic parameters		
					Depth	Width	Manning
2	**WSE / FWE / Q**	**x**					
			x				
				x			
					x	**x**	**x**
		x	**x**	**x**	**x**	**x**	**x**
	WSE + FWE + Q	**x**	**x**	**x**	**x**	**x**	**x**
3	WSE' / FWE' / Q'	x	x	x	x	x	x

* + means the multi-variable assimilation approach, " ' " represents the variable's anomaly

The performance of the simulations has been evaluated by the difference in the relative root mean square error (ΔRRMSE), calculated for each updated state variable i.e. water level, discharge and flooded water extent. The root mean square error is calculated follows:

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{S_i - T_i}{T_i}\right)^2}{N}} \tag{21}$$

$$\Delta RRMSE = \frac{RRMSE_{assimilation} - RRMSE_{open-loop}}{RRMSE_{open-loop}} \tag{22}$$

where N is the number of days of the windows simulations, S represents the simulation (i.e. Open Loop, DA EnKF), T means the true model. The $\Delta RRMSE$ compares relatively the performance (decreasing of errors) of the model simulation using DA scheme respect with to the open-loop (free run simulation).

The same metrics by Durand et al. (2016) based on the relative residuals were used in this research:

$$RRMSE^2 = MRR^2 + SDRR^2 \tag{23}$$

5.6. Results and discussions

5.6.1. Simulation of model uncertainty

Figure 3 shows the RRMSE statistic of discharge for the corrupted model when each perturbed input (forcing and parameters) is assessed, as many authors described (Clark et al., 2008), the rainfall is

the most sensible parameter in our evaluation. Besides, the full-corrupted model achieves a median of RRMSE of 64%, even if it does not achieve to values of estimates RRMSE of the ensemble of GHM (~73%), we assume as a close value that could be representative for GHM and therefore for the objectives of this research.

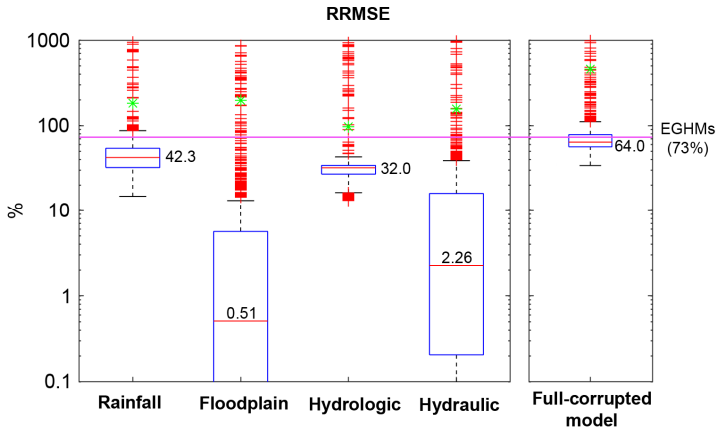

Figure 3. Boxplot in semi-log scale of RRMSE of discharge for the corrupted model due to each independent corrupted parameter, and for the full-corrupted one. Magenta line represents RRMSE value for the ensemble of GHMs. Green dots represent the mean values.

For the corrupted model when only the floodplain was considered, the errors were lower than other parameters, this could be explained due to most part of the basin is not inundated during the year, being this sensitive only in the lower basin; similar behavior occurs when hydraulic parameters were corrupted.

Besides, Figures 4a and 4b show the RMSE for WSE and RRMSE FWE, where larger errors are observed for rainfall and hydraulic corrupted parameters; these can be explained due to the strong influence of hydraulic parameters to estimate both WSE and FWE. As comparison, Fleischmann et al. (2019) found RMSE values of WSE of ~7.4m for the median, at global scale. This higher value can be explained due to the large range of slopes on their study area for global scale (500m of spatial resolution) that results in abrupt changes in riverbed elevation, therefore, results in greater uncertainties.

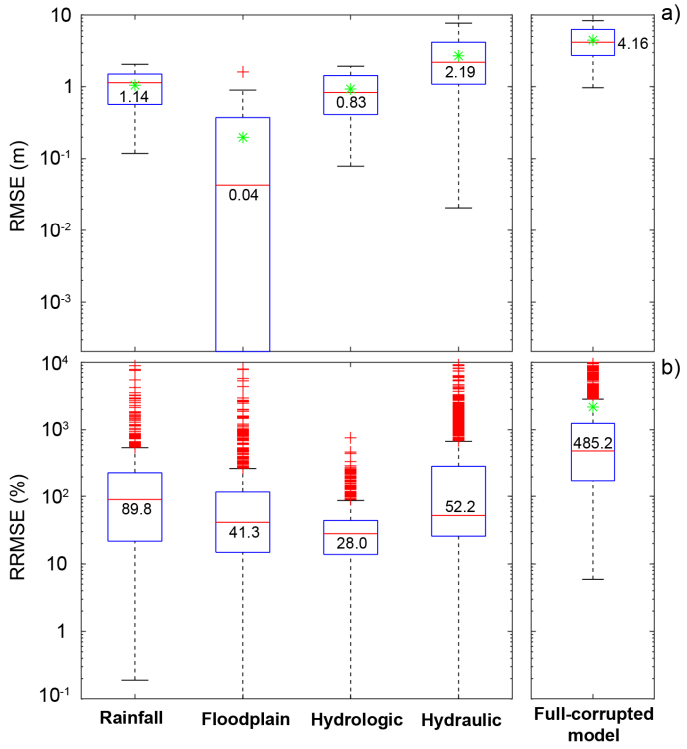

Figure 4. Boxplot in semi-log scale of a) RMSE of water surface elevation and b) RRMSE of flooded water extent for the corrupted model due to each independent corrupted parameter, and for the full-corrupted one. Green dots represent the mean values.

The spatial influence of floodplain and hydraulic errors in the estimates of WFE is showed in Fig. 5a, b. Where higher RRMSE values are mainly located in the lower part of the basin that is characterized to have flattened areas and holds a large part of the Purus' floodplains at the Piagaçu-Purus Sustainable Development Reserve covering approximately 8,342.45 km^2 of which around 40% represents periodically flooded areas (Albernaz and Venticinque, 2003).

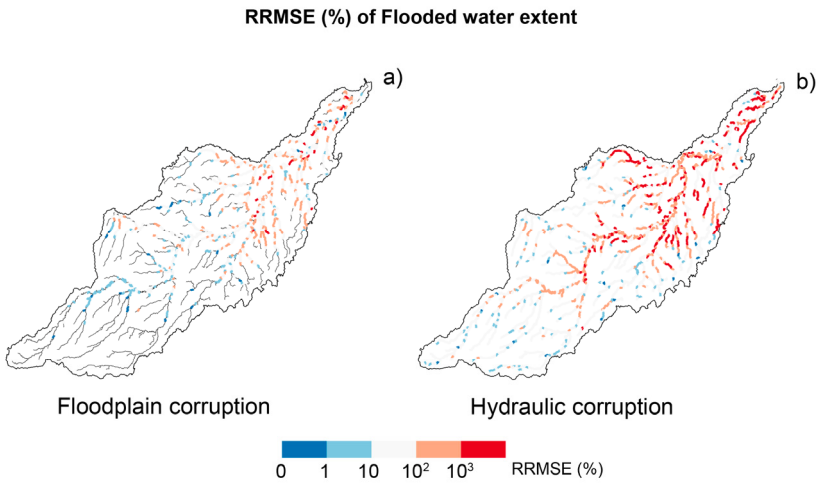

Figure 5. Spatial distribution of the RRMSE of WSE when the (a) floodplain and (b) hydraulic parameters are perturbed.

5.6.2. Sensitivity of model to corrupted parameters

The localization technique was assessed showing variable performances for different radii of influences, in general terms, an ensemble size of 100 members and a radius of influence of 500 km performed better considering the reduction of the RRMSE index. Detailed results can be found in the supplementary material.

Fig. 6 shows the reduction of RRMSE for WSE, FWE and Q of different version the corrupted models, when independent SWOT-like observation were assimilated. In a general overview, the assimilation of WSE improved all other variables but with lower improvements for the FWE when the floodplain is corrupted; this could be explained due to the strong relation between the floodplain and water level, depicted in the hypsometric curve that was perturbed. In addition, for the median, any improvement was realized for Q when hydraulic parameters were corrupted due to the direct relation of width, depth and Manning coefficient on discharge.

The assimilation of FWE showed improvements in all the corrupted scenarios only for the same updated state variable (FWE), besides the worst improvement was realized for improve Q, especially when hydraulic parameters were corrupted. The assimilation of Q improved all corrupted scenarios for all the updated state variables (WSE, FWE and Q), although worst improvements were showed when hydraulic parameters were corrupted. In a first overview of these results, state variable improvements related to SWOT-like DA scheme seems coherent regarding the corrupted scenarios, that can show us a general idea of the impact of future SWOT data in GHMs considering independent uncertainties.

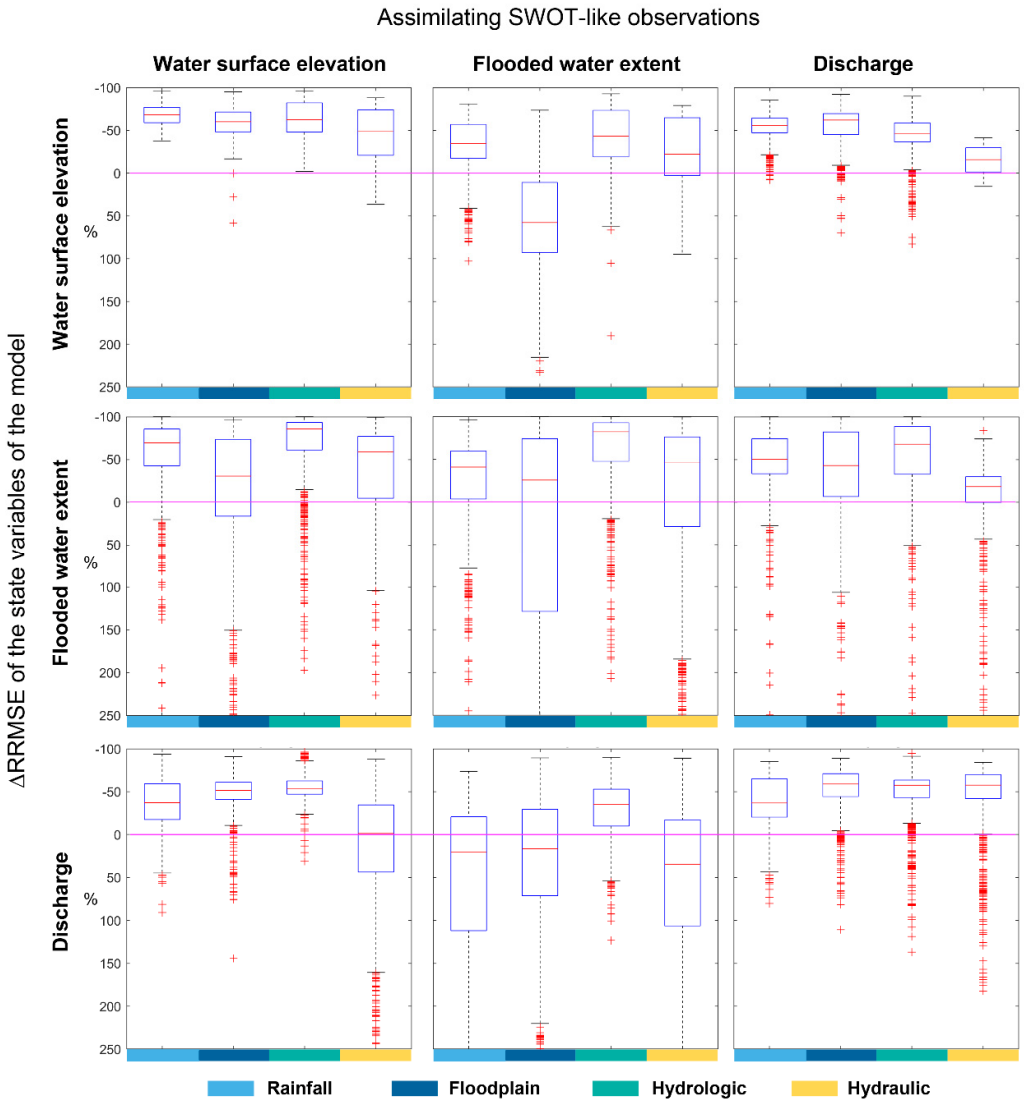

Figure 6. Boxplot of reduction of RRMSE for WSE, FWE and Q for different corrupted model's versions by assimilating correspondent SWOT-like observations.

1. The assimilation of WSE clearly improves all variables in corrupted scenarios. However, it shows worse improvements in Q when the model has corrupted hydraulic parameters, what can be explained due to the strong relation of Q and WSE estimates.
2. When FWE was assimilated, FWE was the unique variable with improvements in all scenarios. Deterioration of WSE estimates were produced when the floodplain was corrupted. For Q the aggravation is showed for all scenarios except for the corruption of hydrologic parameters, likely due to the poor relationship with the hydrodynamic processes.
3. Finally, the assimilation of Q improves all variables and scenarios. Lower improvements are showed for WSE and FWE when hydraulic parameters are corrupted, due to this is the best proxy for the state of the catchment.

5.6.3. Improvements on performance at continental and global scale

The improvements in RRMSE when the full-corrupted model was assessed is shown in Fig. 7. In general, the better performances in the state variables of the model were constrained to the DA of correspondent SWOT-like observations Besides, some aggravation on performance in Q was ocurred when WSE and FWE were assimilated, explained due to the strong sensibility of discharge to hydraulic parameters, of whose uncertainties were added to the others. These results demonstrate that SWOT-like observations coupled in a large-scale hydrological model with a EnKF method, could have the ability to recover the corrupted model in most cases even when this has large errors as showed by GHMs. In addition, the multi-variable (MV) DA was also performed and showed in Fig. 7b, where interesting results can be discussed due to any decreasing of performance was depicted, however all the improvements seem to be strong related with the correspond worst improvement when independent SWOT-like observations were assimilated. For instance, the improvements of performance on WSE due to MV assimilation appears to be close to results when Q is assimilated (the worst improvement).

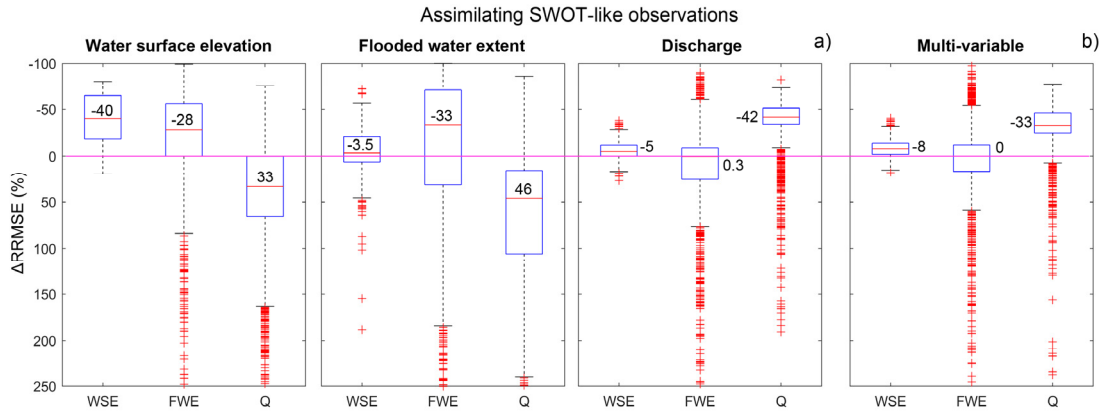

Figure 7. Boxplot of reduction of RRMSE index for WSE, FWE and Q for the full-corrupted model version by assimilating correspondent SWOT-like observations and MV approach.

As showed in Fig. 7, the MV assimilation was performed, improving the efficiency in all the assessed state variables of the model without decreasing of performance of any state variable as occurs when independent single state variables are assimilated (Table 5). It is important to highlight that the reduction of errors on the simulated discharge can be compared with another assessment (Wongchuig et al., 2019), who assimilated in-situ observations for the whole Amazon basin. Those results reached a reduction of error that ranges from ~15% to ~20% for a period of ~100 years. In contrast, the SWOT assessment in the Purus basin reached reduction of errors in discharge of more than the double (~ -41%).

This can be explained due to larger uncertainties were represented for our model setup, therefore the reduction of errors is likely greater, even when uncertainties of SWOT-like discharge (~24%) are greater than in-situ discharge observations (~15%).

Table 5. Values of median of improvements in RRMSE (%) for all catchments when are assimilating SWOT-like observations individually and all together (MV).

	SWOT-like observations			
	WSE	**FWE**	**Q**	**MV**
WSE	-39.96	-3.50	-5.18	-7.95
FWE	-28.17	-33.31	**0.32**	0.00
Q	**33.19**	**45.98**	-41.64	-32.76

* Values in bold indicate decreasing in performance

Finally, Fig. 8 shows the spatial distribution of variations in RRMSE, where blue color means greater improvements (reduction of errors).It is clearly that for the median, the Q is the most challenging state variable to be improved due to the high sensibility to the hydraulic parameters and floodplain topography. The improvements in Q when WSE and FWE were assimilated, were only realized in the lower part of the main river where the impact of corruption of hydraulic variables are relatively less sensible than the upper part of the basin. It is also quite clear that the use of multi-variable DA could overcome these issues, even when high uncertainties were considered.

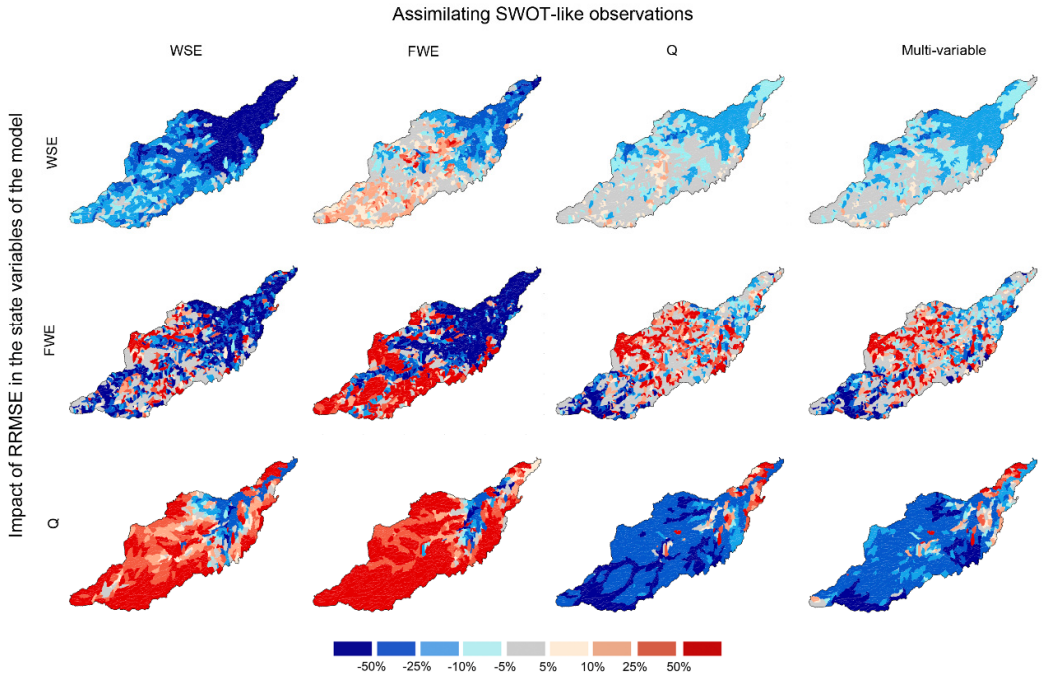

Figure 8. Maps of spatial distribution of the reduction on RRMSE by assimilating absolute values of SWOT-like observations.

Therefore, under the assumed considerations in this research, the SWOT data will perform adequately for global and continental scales, being this remote sensing information useful for hydrological and hydrodynamic assessments, even more so when multi-variable DA is performed.

The hydrographs and water level series were also showed in Fig. 9 and 10 respectively. The reduction of uncertainties is clearly showed by DA scheme (gray envelope) during the presence of SWOT-like observations from October 2011 to October 2012. For Q time series the SWOT-like observations are showed as an envelope of the realizations of the perturbations of the true, were the true is always within this envelope. Therefore, the number of realizations used here could be considered as representative.

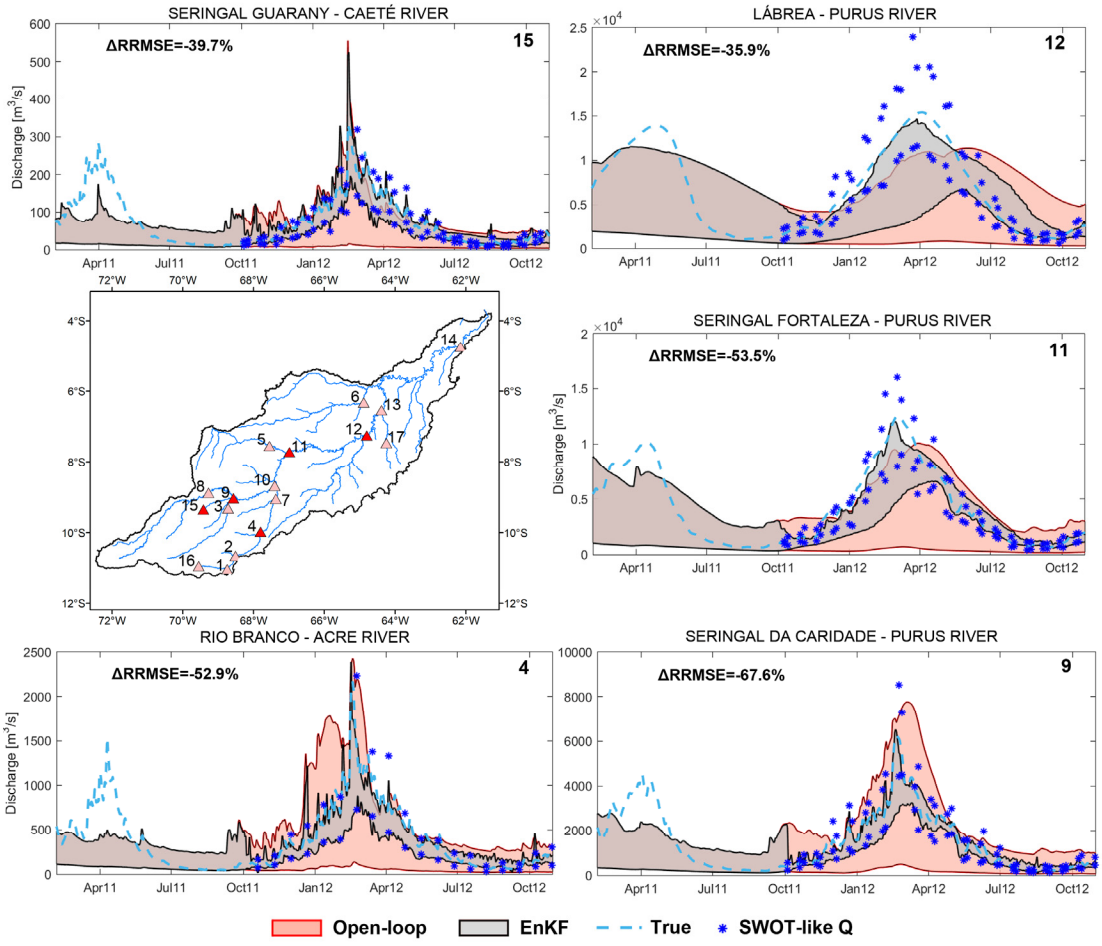

Figure 9. Hydrographs for a few sites at main stem and tributaries.

In the case of the time series of water level (Fig. 9), the SWOT-like observations are showed as just one realization and are close to the true due to high accuracy. The uncertainties were also reduced by DA scheme (gray envelope) from the Open-loop (red envelope). Some noise was showed by the DA series mainly in tributary rivers where less accuracy was likely found.

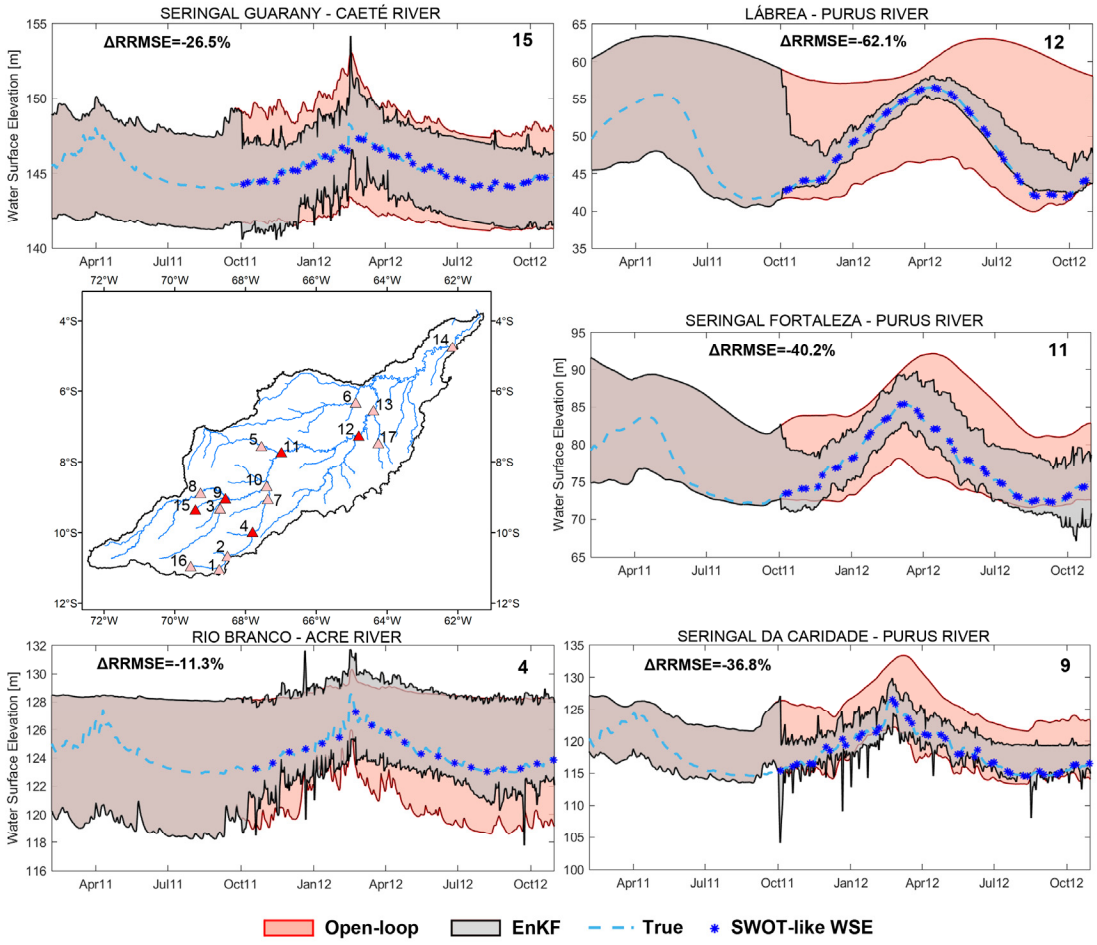

Figure 10. Time series of water level for a few sites at main stem and tributaries.

5.6.4. Performance by assimilating anomalies of SWOT-like observations

The improvements in RRMSE by assimilating anomalies were performed. The full-corrupted model was used for this experiment. Fig. 11a shows that best performances in the state variables were achieved by assimilating correspondent anomalies of SWOT-like observations as occurs in the experiment in section 6.3. For the assimilation of WSE', the performance results in worse results than when WSE was assimilated what could be explained due to absolute values (WSE) are larger accurate than the anomalies (WSE') which finally are influenced by the mean of the Open-loop simulations. Besides the performance by assimilating FEW' and Q' clearly presents improvements in the other estate

variables when compared assessments using absolute values (FEW and Q), that is explained due to the removing of the bias presented on the absolute values, primarily in Q where values are ~18%. Finally, the multi-variable (MV) DA of anomalies was also performed and showed in Fig. 11b, where similar results were achieved by assimilating absolute values. It is also interesting that the performance was improved for all state variables in a mean of ~8% which is also motivating for improvements in hydrological models in global and continental scale.

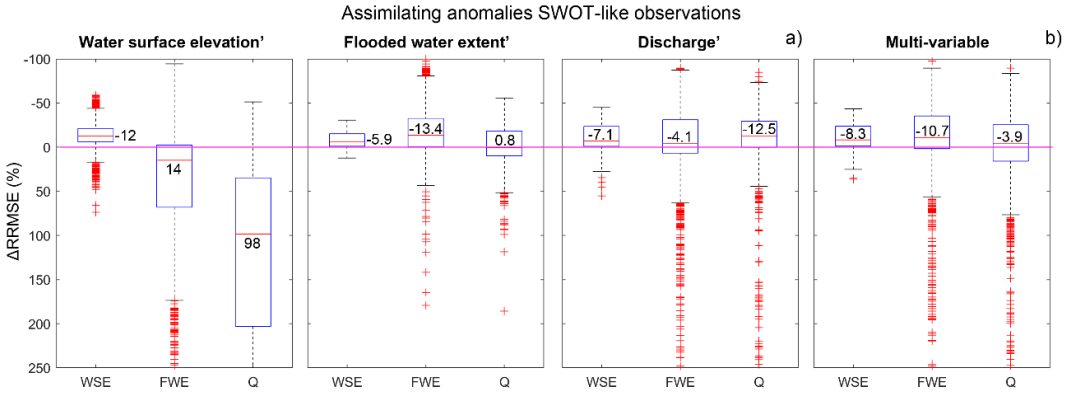

Figure 11. Boxplot of reduction of RRMSE index for WSE, FWE and Q for the full-corrupted model version by assimilating correspondent anomalies of SWOT-like observations and MV approach.

The Fig. 12 shows the spatial distribution of improvements in RRMSE by assimilating anomalies, evidently the assimilation of WSE' decrease the performance of FWE and Q due to reasons explained lines above, therefore, this setup is not recommended to be used for GHMs. Besides, the assimilation of FWE' and Q' improved all state variables although slightly lower than the assimilation of absolute values. Henceforth the assimilation of anomalies can improve simulations due to the removal of bias in SWOT-like observations, although if bias between observations and open-loop are large, once this last represents uncertainties of GHMs, the improvements are constrained to this limitation.

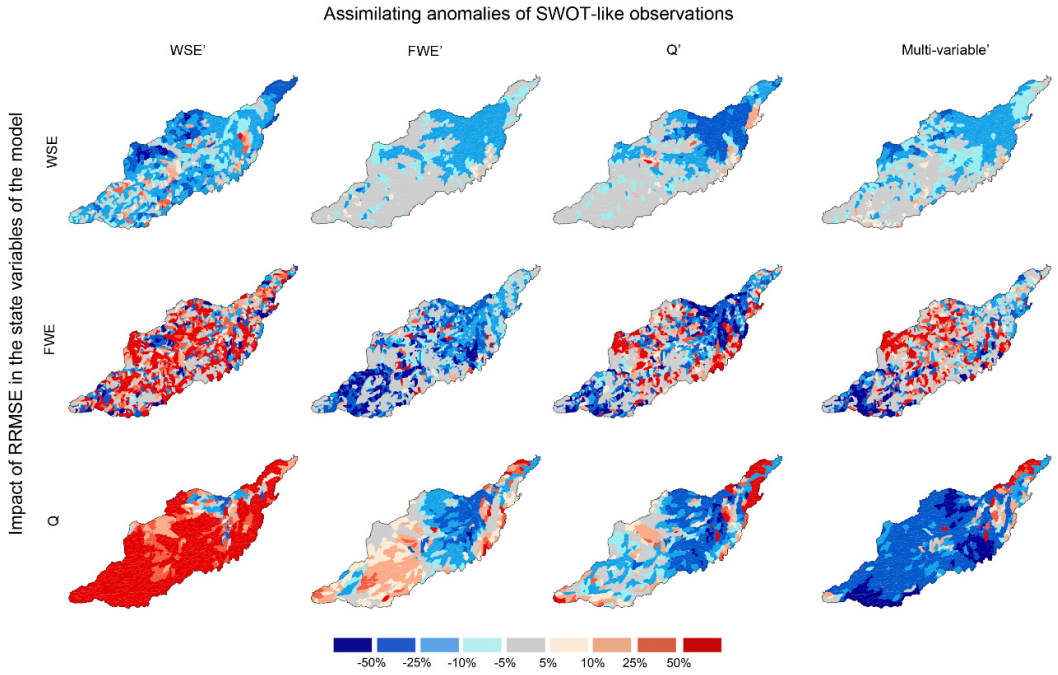

Figure 12. Maps of spatial distribution of the reduction on RRMSE by assimilating anomalies of SWOT-like observations.

5.7. Conclusions

This research aimed to implement and assess the DA of SWOT-like observations within a large-scale hydrological model that emulates the uncertainties of GHMs. For this, the use of the large-scale SWOT pixel cloud simulator and a large-scale model in a region with complex drainage network as a case of study was performed.

The values of uncertainties of the GHMs were achieved by the corruption of a few set of forcing and parameters of the model using the most representative values of errors found in the literature. The precipitation and the hydrologic parameters were found to be those that most influence the errors for the estimation of discharge in the model.

The sensitivity analysis of the improvements on the state variables of the model regarding individual error scenarios showed that the hydraulic parameters and the floodplain topography were the worst scenarios for the performance in Q when WSE and FWE were assimilated; which depict the strong influence of these parameters in Q simulation. Besides, the only aggravation on WSE was figured out when the FWE was assimilated in the scenario of errors of floodplain topography that also relates the large relation among WSE, FWE and floodplain due to the hypsometric curve used in our hydraulic simulations.

When the full-corrupted model is assessed, which means an emulated GHM, improvements in some state variables of the model were not achieved by the assimilation of individual SWOT-like observations. In addition, the improvements in Q by assimilating SWOT-like discharge in global scale were larger than other assessments by using in-situ discharge observations in regional scale. Henceforth, SWOT products of discharge can be able to improve simulations by using hydrological models in global and continental scale in approximately ~40%. Besides, the improvements of all state variables were only reached by the use of the multi-variable DA scheme even when the uncorrelated observations were assumed as a simplification.

The improvements by the using of anomalies of SWOT-like observations were assessed. Experiments by assimilating WSE' depicted that results were worse than where WSE was assimilated, due to the large influence of the bias between the open-loop and SWOT-like observations. Henceforth, is recommendable the assimilation of absolute values of WSE over its anomalies, once the first already has higher accuracy. Besides, the assimilation of FEW' and Q' improved all state variables although these were not greater than when absolute values were assimilated, for the cases of the correspondent state variable (e.g. improvements on Q by assimilating SWOT-like Q). These mean that even when bias in observations were removed, the bias between open-loop and SWOT-like observations has also large influence. Lastly, a smart combination of assimilation of SWOT-like observations should be considered regarding the uncertainties of GHMs.

Finally, other approaches should be implemented in the future when SWOT real data is available such as further explore multi-variable DA technique by the use of less simplified assumptions. However, even under large uncertainties by GHMs, SWOT data shows that it could have a large impact in the improvements of hydrologic and hydrodynamic simulations in global and continental scales; what implies a promissory usefulness by SWOT scientific community.

5.8. Supplementary material

Implementation and performance for localization technique

To implement LA, a correlation matrix was used considering a distant-dependent correlation function proposed by Gaspari and Cohn (1999). The incorporation of the covariance localization matrix (Sakov and Bertino, 2011) to the traditional EnKF (equation 9), and sometime referred as Local Ensemble Kalman Filter (Houtekamer and Mitchell, 2001, 2005; Ott et al., 2004; Tong, 2018):

$$\mathbf{x}^a = \mathbf{x}^f + \boldsymbol{\rho} \times \left(\mathbf{x'}^f \mathbf{x'}^{fT} \mathbf{H}^T\right) \left[\boldsymbol{\rho_o} \times \left(\mathbf{H}\mathbf{x'}^f \mathbf{x'}^{fT} \mathbf{H}^T\right) + \mathbf{E}\mathbf{E}^T\right]^{-1} \left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right) \quad (9)$$

where ρ and ρ_o are correlation matrices and x represents the Schur product. The correlation matrices were calculated using a fifth-order function (Gaspari and Cohn, 1999).

$$\boldsymbol{\rho}(r,e) \begin{cases} 1-\frac{1}{4}\left(\frac{e}{r}\right)^5+\frac{1}{2}\left(\frac{e}{r}\right)^4+\frac{5}{8}\left(\frac{e}{r}\right)^3-\frac{5}{3}\left(\frac{e}{r}\right)^2, & 0 \le e \le r; \\ \frac{1}{12}\left(\frac{e}{r}\right)^5-\frac{1}{2}\left(\frac{e}{r}\right)^4+\frac{5}{8}\left(\frac{e}{r}\right)^3+\frac{5}{3}\left(\frac{e}{r}\right)^2-5\left(\frac{e}{r}\right)+4-\frac{2}{3}\left(\frac{e}{r}\right)^{-1}, & r \le e \le 2r; \\ 0, & e > 2r; \end{cases}$$

where e represents the distance between the observation point and the updated site, r means the distance parameter related to a radius of influence.

In this section, the performance for each radius of influence and ensemble size used in the LEnKF scheme is showed in the Figures 1s, 2s and 3s. It is noticed that assimilating SWOT-like observations of WSE, the localization method improves the performance of WSE, FWE and Q proportionally to the radius of influence until certain values which approximately reachs 500 km. These improvements are clearly greater for smallest ensemble size.

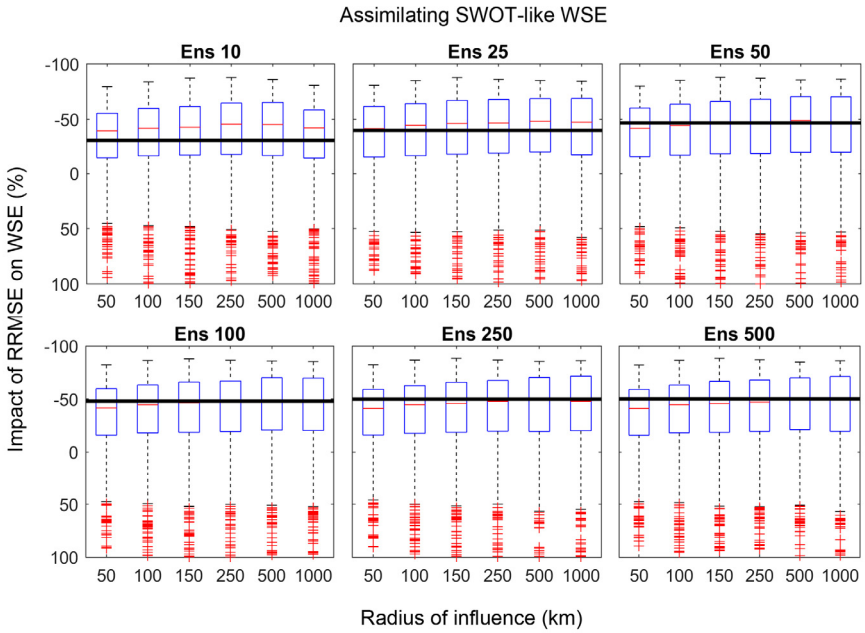

Fig. 1s. Boxplot of reduction on RRMSE of WSE by assimilating SWOT-like WSE using different ensemble sizes and radii of influence, compared with the median of the not localization scheme (black line).

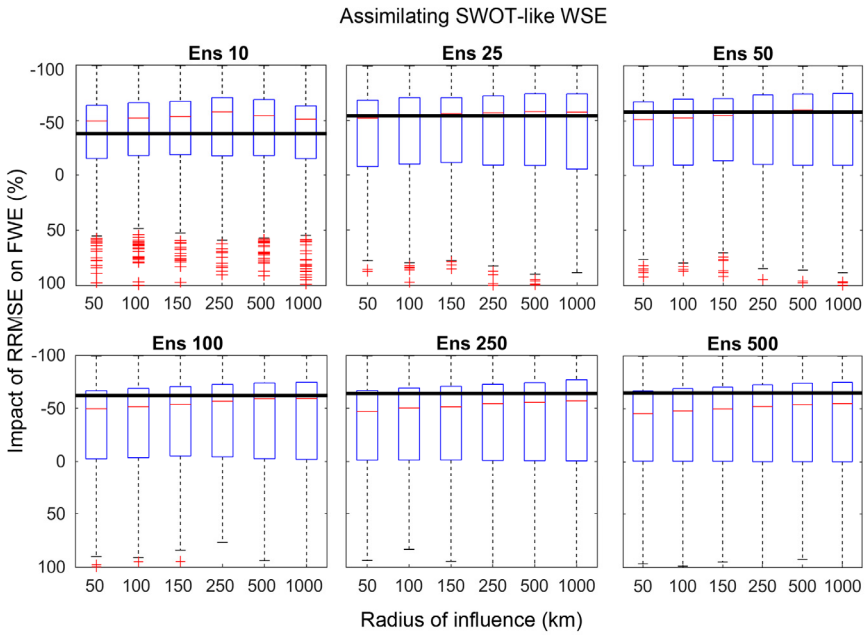

Fig. 2s. Boxplot of reduction on RRMSE of FWE by assimilating SWOT-like WSE using different ensemble sizes and radii of influence, compared with the median of the not localization scheme (black line).

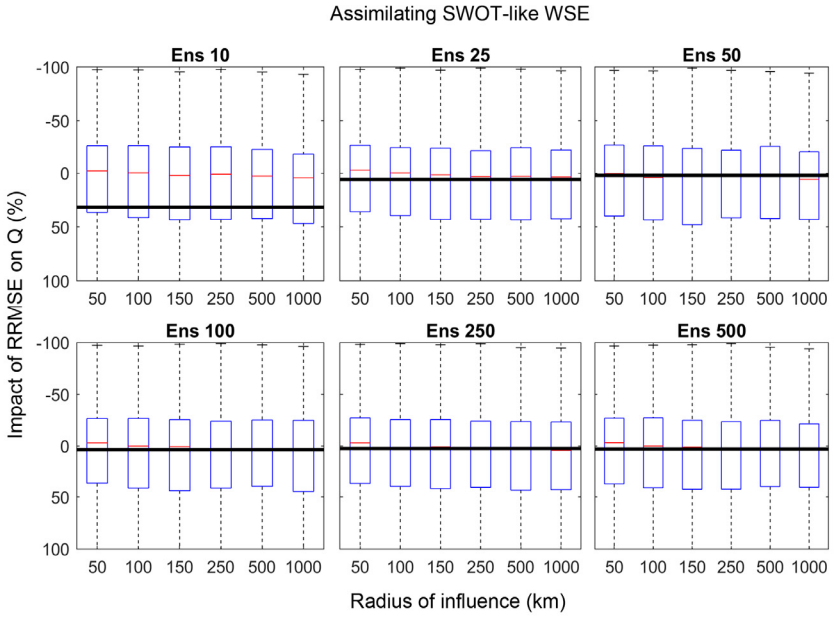

Fig. 3s. Boxplot of reduction on RRMSE of Q by assimilating SWOT-like WSE using different ensemble sizes and radii of influence, compared with the median of the not localization scheme (black line).

Acknowledgments

The first author is grateful for a grant from the Brazilian agency CAPES and for the facilities provided during their internship doctorate by the French institutions LEGOS (Laboratoire d'Etudes en Géophysique et Océanographie Spatiales) and GET (Laboratoire Géosciences Environnement Toulouse) laboratories from the OMP (L'Observatoire Midi-Pyrénées).

References

Adler, R.F., Kidd, C., Petty, G., Morissey, M., Goodman, H.M., 2001. Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP-3). Bulletin of the American Meteorological Society, 82(7), 1377-1396.

Albernaz, A.L.K.M. and Venticinque, E.M. 2003. Reserva de Desenvolvimento Sustentável Piagaçu-Purus: Características e Limites Geográficos In Piagaçu-Purus: Bases Científicas para Criação de uma Reserva de Desenvolvimento Sustentável (C.P. de Deus, R. Da Silveira & L.H.R. Py-Daniel, eds). IDSM, Manaus, p.3-12.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS – global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci., 17, 1161–1175.

Alho, C. J. R., Reis, R. E., Aquino, P. P. U., 2015. Amazonian freshwater habitats experiencing environmental and socioeconomic threats affecting subsistence fisheries. Ambio, 44(5), 412–425. http://doi.org/10.1007/s13280-014-0610-z

Altman, D. G., and Bland, J. M. 2005. Standard deviations and standard errors. BMJ, 331(7521), 903. https://doi.org/10.1136/bmj.331.7521.903

Alsdorf, D.E., Rodríguez, E. and Lettenmaier, D.P., 2007. Measuring surface water from space. Rev. Geophys., 45, RG2002, doi:10.1029/2006RG000197.

Andreadis, K. M. and Lettenmaier, D. P. 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model, Adv. Water Resour., 29, 872–886.

Andreadis, K.M., Clark, E.A., Lettenmaier, D.P., Alsdorf D.E. 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. Geophys Res Lett 34:L10403

Andreadis, K.M., Schumann, G.J.-P., Pavelsky, T., 2013. A simple global river bankfull width and depth database. Water Resources Research 49, 7164e7168.

Andreadis, K.M., Schumann, G.J.P. 2014. Estimating the impact of satellite observations on the predictability of large-scale hydraulic models. Adv Water Resour 73:44–54. doi:10.1016/j.advwatres.2014.06.006

Anees, M.T., Abdullah, K., Nordin, M.N.M., Ab Rahman, N.N.N., Syakir, M.I. and Kadir, M.O.A., 2017. One-and Two-Dimensional Hydrological Modelling and Their Uncertainties. In Flood Risk Management. IntechOpen.

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., et al. (2015). Accelerating advances in continental domain hydrologic. Water Resources Research, 51:10,078–10,091.

Baguis, P. and Roulin, E. 2017. Soil Moisture Data Assimilation in a Hydrological Model: A Case Study in Belgium Using Large-Scale Satellite Data. Remote Sens. 2017, 10, 820; doi:10.3390/rs9080820.

Bates, P.D., Horritt, M.S., Fewtrell, T.J., 2010. A simple inertial formulation of the shallow water equations for efficient two dimensional flood inundation modelling. Journal of Hydrology 387, 33e45.

Bates, P.D., Neal, J.C., Alsdorf, D., Schumann, G.J.-P. 2014. Observing Global Surface Water Flood Dynamics. Surveys in Geophysics, 35(3): 839-852.

Beck, H.E., van Dijk, A.I. J. M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., and de Roo, A., 2016. MSWEP, 3-hourly 0.25 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences Discussions, 2016, 1–38.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J. 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrol. Earth Syst. Sci., 21, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017.

Becker, M., Papa, F. Frappart, F., Alsdorf, D., Calmant, S., Da Silva, J. S., Prigent, C., Seyler, F. 2018. Satellite-based estimates of surface water dynamics in the Congo River Basin. International Journal of Applied Earth Observation and Geoinformation, 66: 196-209.

Bennett, A. F., (1992.) Inverse methods in physical oceanography. Cambridge University Press, 346 pp.

Bernard, R., Vauclin, M., and Vidal-Madjar, D., (1981). Possible use of active microwave remote sensing data for prediction of regional evaporation by numerical simulation of soil water movement in the unsaturated zone. Wat. Resour. Res., 17(6): 1603-1610.

Bertino, L., Evensen, G., Wackernagel, H. Sequential Data Assimilation Techniques in Oceanography. Internat. Statist. Rev. 71 (2003), no. 2, 223--241. https://projecteuclid.org/euclid.isr/1069172299

Biancamaria, S; Andreadis, K., Durand, M., Clark, E., Rodriguez, E., Mognard, N., Alsdorf, D., Lettenmaier, D.P., Oudin, Y. 2010. Preliminary characterization of SWOT hydrology error budget and global capabilities. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., 3(1), 6–19, doi:10.1109/JSTARS.2009.2034614.

Biancamaria, S.; Durand, M.; Andreadis, K. M.; Bates, P. D.; Boone, A.; Mognard, N. M.; Rodriguez, E.; Alsdorf, D. E.; Lettenmaier, D. P.; Clark, E. A. 2011. Assimilation of virtual wide swath altimetry to improve Arctic river modelling. Remote Sensing of Environment, 115, 373–381.

Biancamaria, S., D. P. Lettenmaier, and T. M. Pavelsky, 2016. The swot mission and capabilities for hydrology. Surveys in Geophysics, 1–31, doi:10.1007/s10712-015-9346-y.

Bierkens MFP, Bell VA, Burek P, Chaney N, Condon LE, David CH, de Roo A, Do¨ll P, Drost N, Famiglietti JS, Flo¨rke M, Gochis DJ, Houser P, Hut R, Keune J, Kollet S, Maxwell RM, Reager JT, Samaniego L, Sudicky E, Sutanudjaja EH, van de Giesen N, Winsemius H, Wood EF. 2015. Hyper-resolution global hydrological modelling: what is next? Everywhere and locally relevant. Hydrol Process 29(2):310–320. doi:10.1002/hyp.10391

Bruckler, L., and Witono, H. (1989). Use of remotely sensed soil moisture content as boundary conditions in soil-atmosphere water transport modeling: 2. Estimating soil water balance. Wat. Resour. Res., 25(12): 2437-2447.

Burgers, G., van Leeuwen, P. J., Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter, Monthly Weather Review, 126, 1719–1724.

Brenner, A.C., DiMarzio, J.P., Zwally, H.J., 2007. Precision and Accuracy of Satellite Radar and Laser Altimeter Data Over the Continental Ice Sheets. IEEE Transactions on Geoscience and Remote Sensing, 45(2), 321-331.

Brooks, R. L., 1982. Lake elevation from satellite radar altimetry from a validation area in Canada. Reporte, Geosci. Res.

Charney, J. G., Halem, M., and Jastrow, R., (1969). Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci. 26: 1160-1163.

Clark, Martyn P.; Slater, Andrew G. 2006. Probabilistic Quantitative Precipitation Estimation in Complex Terrain. Journal of Hydrometeorology, Vol. 7, pp. 3 – 22.

Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J., 2008. Hydrological data assimilation with the Ensemble Kalman Filter: use of streamflow observations to update states in a distributed hydrological model. Adv. Water Resour. 31, 1309–1324.

Costa, S.M.A., Matos, A.C.O.C., Blitzkow, D., 2011. Validation of the land water storage from gravity recovery and climate experiment (GRACE) with gauge data in the amazon basin. Boletim de Ciências Geodésicas, 18 (2), 262-281.

Crow, W.T. and Ryu, D. 2009. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. Hydrol. Earth Syst. Sci., 13, 1–16.

Daley, R., (1991). Atmospheric data analysis. Cambridge University Press, 460 pp.

David, C.H., Maidment, D.R., Niu, G.-Y., Yang, Z.-L., Habets, F. and Eijkhout. V. 2011. River Network Routing on the NHDPlus Dataset, J. Hydrometeorol., 12(5), 913–934.

Desroches, D., Pottier, C., Blumstein, D., Biancamaria, S., Poughon, V., Fjortoft, R. Large Scale Pixel Cloud Simulator and Hydrology Toolbox. SWOT Science Team Meeting, Montreal, Canada 23 June 2018.

Döll, P., Berkhoff, K., Bormann, H., Fohrer, N., Gerten, D., Hagemann, S. and Krol. M. 2008. Advances and visions in large-scale hydrological modelling: findings from the 11th Workshop on Large-Scale Hydrological Modelling. Adv. Geosci., 18: 51–61.

Durand, M.T., Andreadis, K.M., Alsdorf, D.E., Lettenmaier, D.P., Moller, D., Wilson, M. 2008. Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. Geophys Res Lett 35:L20401. doi:10.1029/2008GL034150

Emery, C. M., Paris, A., Biancamaria, S., Boone, A., Calmant, S., Garambois, P.-A., and Santos da Silva, J. 2018. Large-scale hydrological model river storage and discharge correction using a satellite altimetry-based discharge product, Hydrol. Earth Syst. Sci., 22, 2135-2162, https://doi.org/10.5194/hess-22-2135-2018.

Engman, E.T. 1995. Recent advances in remote sensing in hydrology. Reviews of Geophysics, Supplement, 967-975.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans (1978–2012), 99 (C5), 10143–10162. doi:10.1029/94JC00572

Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. 2007. The shuttle radar topography mission. Rev. Geophys.,45, RG2004, doi:10.1029/2005RG000183.

Fisher, C. K., Pan, M., and Wood, E. F.: Spatiotemporal Assimilation/Interpolation of Discharge Records through Inverse Streamflow Routing, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-109, in review, 2018.

Gao, H., Tang, Q., Shi,X., Zhu, C., Bohn,T. J., Su,F., Sheffield,J., Pan,M., Lettenmaier,D.P and Wood,E.F., 2010. Water Budget Record from Variable Infiltration Capacity (VIC) Model. In Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records. pp. 120-173.

García-Pintado, J., Neal, J.C., Mason, D.C., Dance, S., Bates, P.D. 2013. Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling. J. Hydrol., 495:252–266. doi:10.1016/j.jhydrol.2013.03.050

Gaspari, G., & Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. Quarterly Journal Royal Meteorological Society, 125(554), 723−757.

Gelb, A. Optimal linear filtering. In Applied Optimal Estimation; MIT Press: Cambridge, MA, USA, 1974; pp. 102–155.

Georgakakos, K. P., and Baumer, O. W. 1996. Measurement and utilization of on-site soil moisture data. J. Hydrol., 184: 131-152.

Gichamo T, Popescu I, Jonoski A, Solomatine D. 2012. River cross-section extraction from the ASTER global DEM for flood modeling. Environmental Modelling & Software, 31:37–46. DOI: 10.1016/j.envsoft.2011.12.003

Goulding, M., Barthem, R., Ferreira, E.J.G. 2003. The Smithsonian Atlas of the Amazon. Smithsonian Institution Press, Washington D.C. 256p.

Hess, L.L., J.M. Melack, A.G. Affonso, C. Barbosa, M. Gastil-Buhl, and E.M.L.M. Novo, 2015, Wetlands of the Lowland Amazon Basin: Extent, Vegetative Cover, and Dual-season Inundated Area as Mapped with JERS-1 Synthetic Aperture Radar, Wetlands 35: 745-756. doi:10.1007/s13157-015-0666-y

Herschy, W. 2002. The uncertainty in a current meter measurement, Flow Meas. Instrum., 13, 281–284.

Koblinsky, C. J., Clarke, R. T., Brenner, A. C., and Frey, H. 1993. Measurement of river level variations with satellite altimetry, Wat. Resour. Res., 29(6), 1839-1848

Houtekamer, P.L. and Mitchell, H.L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129, 123–137.

Houtekamer, P.L. and Mitchell, H.L. 2005. Ensemble Kalman filtering. Quarterly Journal of the Royal Meteorological Society,131, 3269–3289.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe E., Cudennec, C. 2013. A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrological Sciences Journal, 58(6), 1198-1255.

Hreinsson, Einar Ö. 2008. Assimilation of a snow covered area into a hydrologic model. Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Geography. University of Canterbury. New Zealand.

Jackson, T. J., et al., (1981). Soil moisture updating and microwave remote sensing for hydrological simulation. Hydrol. Sci. B., 26(3): 305-319.

Kalman, R.E. A new approach to linear filtering and prediction problems. J. Basic Eng. 1960, 82, 35–45.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., Thielen, J. 2016. Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. Environmental Modelling & Software, 75: 68-76.

Khaki, M., Forootan, E., Kuhn, M., Awange, J., Longuevergne, L., Wada, Y. 2018. Efficient basin scale filtering of GRACE satellite products. Remote Sensing of Environment, (204):76–93.

Khaki, M., Hoteit, I., Kuhn, M., Forootan, E., Awange, J. 2019. Assessing data assimilation frameworks for using multi-mission satellite products in a hydrological context. Science of The Total Environment, (647): 1031-1043.

Lai, X., Liang, Q., Yesou, H., Daillet, S., 2014. Variational assimilation of remotely sensed flood extents using a 2-D flood model. Hydrol. Earth Syst. Sci. 18, 4325–4339. http://dx.doi.org/10.5194/hess-18-4325-2014.

Latrubesse, E.M. 2008. Patterns of anabranching channels: The ultimate end-member adjustment of mega rivers. Geomorphology, 101: 130–145.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., Van Dijk, A. I. J. M., Van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., Restrepo, P. 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences, 16, 3863–3887.

Loukas, A. and Vasiliades, L. 2014. Streamflow simulation methods for ungauged and poorly gauged watersheds. Nat. Hazards Earth Syst. Sci., 14:1641–1661.

López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., and Bierkens, M. F. P. 2017. Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. Hydrol. Earth Syst. Sci., 21, 3125-3144, https://doi.org/10.5194/hess-21-3125-2017, 201

Massari, C., Brocca, L., Tarpanelli, A., Moramarco, T. 2015. Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe? Remote Sens. 2015, 7, 11403-11433; doi:10.3390/rs70911403.

Masutani, M., Woollen, J.S., Lord, S.J., Emmitt, G.D., Kleespies, T.J., Wood, S.A., Greco, S., Sun, H., Terry, J., Kapoor, V., Treadon, R., Campana, K.A. 2010. Observing system simulation experiments at the National Centers for Environmental Prediction, J. Geophys. Res., 115, D07101.

McClain, M.E., Naiman, R.J. 2008. Andean Influences on the Biogeochemistry and Ecology of the Amazon River. Bioscience, 58: 325.

McLaughlin, D. 1995. Recent developments in hydrologic data assimilation. Reviews of Geophysics, 33 (S2), 977–984. doi:10.1029/95RG00740

Melesse, A.M., Weng, Q., Thenkabail, P.S., and Senay, G.B. 2007. Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. Sensors, 7: 3209-3241.

Munier, S., Polebistki, A., Brown, C., Belaud, G., Lettenmaier, D.P. 2015. SWOT data assimilation for operational reservoir management on the upper Niger River Basin. Water Resour Res. doi:10.1002/2014WR016157

Neal, Jeffrey; Atkinson, Peter; Hutton, Craig. 2007. Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. Journal of Hydrology, Vol. 336, pp. 401 – 415.

Neal, J.C., Schumann, G.J.P., Bates, P.D. 2012. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resour Res 48:W11506. doi:10.1029/2012WR012514.

Nijssen, B., Lettenmaier, D. 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, Journal of Geophysical Research, 109, n. D2, D02103.

O'Loughlin, F. E., Neal, J., Yamazaki, D., and Bates, P. D. 2016. ICESat-derived inland water surface spot heights. Water Resour. Res., 52, 3276–3284, doi:10.1002/2015WR018237.

Ottlé, C., and Vidal-Madjar, D., (1994). Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY Region. J. Hydrol., 158: 241-264.

Oubanas, H. Gejadze, I., Malaterre, P-O., Durand, M., Wei, R., Frasson, R.M.P., Domeneghetti, A. 2018a. Discharge estimation in ungauged basins through variational data assimilation: The potential of the SWOT mission. Water Resources Research, https://doi.org/10.1002/2017WR021735

Oubanas, H. Gejadze, I., Malaterre, P-O., Mercier, F. 2018b. River discharge estimation from synthetic SWOT-type observations using variational data assimilation and the full Saint-Venant hydraulic model. Journal of Hydrology, v. 559, p. 638-647.

Paiva, R.C.D., Buarque, D.C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., Bulhões Mendes, C.A., 2013a. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resour. Res. 49 (3), 1226–1243.

Paiva, R.C.D., Collischonn, W., Bonnet, M.-P., De Gonçalves, L.G.G., Calmant, S., Getirana, A., Santos Da Silva, J., 2013b. Assimilating in situ and radar altimetry data into a large-scale hydrologic-hydrodynamic model for streamflow forecast in the Amazon. Hydrol. Earth Syst. Sci. 17, 2929–2946.

Paiva, R.C.D., Durand, M.T., Hossain, F. 2015. Spatiotemporal interpolation of discharge across a river network by using synthetic SWOT satellite data. Water Resour Res. doi:10.1002/2014WR015618

Papa, F., Pringent, C., Aires., F., Jimenez, C., Rossow, W.B., Matthew, E., 2010. Interannual variability of surface water extent at the global scale, 1993-2004. J. Geophys. Res., 115, D12.

Papa, F., Frappart, F., Malbeteau, Y., Shamsudduha, M., Vuruputur, V., Sekhar, M., Ramillien, G., Prigent, C., Aires, F., Pandey, R.K., Bala, S., Calmant, S., 2015. Satellite-derived surface and sub-surface water storage in the Ganges–Brahmaputra River Basin. Journal of Hydrology: Regional Studies, 4(Part A), 15-35.

Parrens, M., Al-Bitar, A., Frappart, F., Paiva, R.C.D., Wongchuig, C.S., Papa, F., Yamazaki, D., Kerr, Y. 2019. High resolution mapping of inundation area in the Amazon basin from a combination of L-band passive microwave, optical and radar datasets. Int. J. Appl. Earth. Obs. Geoinformation, 81:58–71. https://doi.org/10.1016/j.jag.2019.04.011

Pauwels, V. R. N., and G. J. M. De Lannoy, 2006: Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge. J. Hydrometeor., 7, 458–477.

Pedinotti, V., Boone, A., Ricci, S., Biancamaria, S., Mognard, N.M. 2014. Assimilation of satellite data to optimize large-scale hydrological model parameters: a case study for the SWOT mission. Hydrol Earth Syst Sci 18(11):4485–4507. doi:10.5194/hess-18-4485-2014

Pelletier, M. P. 1987. Uncertainties in the determination of river discharge: a literature review, Can. J. Civil Eng., 15, 834–850.

Pontes, P. R. M., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., Buarque, D. C., Siqueira, V. A., Jardim, P. F., Sorribas, M. V., Collischonn, W. 2017. MGB-IPH model for hydrological and hydraulic simulation of large floodplain river systems coupled with open source GIS. Environmental Modelling & Software, 94, 1-20.

Prevot, L., et al., (1984). Evaporation from a bare soil evaluated using a soil water transfer model and remotely sensed surface soil moisture data. Wat. Resour. Res., 20(2): 311-316.

Qin, Changbo; Jia, Yangwen; Su, Z. (Bob); ZHOU, Zuhao; QIU, Yaqin; SUHUI, Shen. 2008. Integrating Remote Sensing Information Into a Distributed Hydrologycal Model for Improving Water Budget Predictions in Large-scale Basins through Data Assimilation. Sensors, Vol. 8, pp. 4441 – 4465.

Reichle, R. H. 2008. Data assimilation methods in the Earth sciences. Advances in Water Resources, 31, 1411–1418, https://doi.org/10.1016/j.advwatres.2008.01.001.

Revel, M., Ikeshima, D., Yamazaki, D., Kanae, S.A 2019. Physically Based Empirical Localization Method for Assimilating Synthetic SWOT Observations of a Continental-Scale River: A Case Study in the Congo Basin. Water, 11, 829.

Rollett, A. D. and Manohar, P. (2005). The Monte Carlo Method. In Continuum Scale Simulation of Engineering Materials (eds D. Raabe, F. Roters, F. Barlat and L. Chen). doi:10.1002/3527603786.ch4

Scanlon, B. R., Zhang, Z. Save, H., Wiese, D. N., Landerer, F. W., Long, D., Longuevergne, L., Chen, J. 2016. Global evaluation of new GRACE mascon products for hydrologic applications, Water Resour. Res., 52, doi:10.1002/2016WR019494.

Schmidt, A. R.: Analysis of stage-discharge relations for open channel flow and their associated uncertainties, Ph.D. Thesis, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering, p.349, 2002.

Schumann G.J.P., Neal J.C., Voisin N., Andreadis K.M., Pappenberger F., Phanthuwongpakdee N., Hall A.C., Bates P.D. 2013. A first large-scale flood inundation forecasting model. Water Resour Res 49(10):6248–6257. doi:10.1002/wrcr.20521

Schumann G.J.P., Bates P.D., Neal J.C., Andreadis K.M. 2014. Fight floods on a global scale. Nature 507(7491):169

Siqueira, V.A., Paiva, R.C.D., Fleischmann, A.S., Fan, F.M., Ruhoff, A.L., Pontes, P.R.M., Paris, A., Calmant, S., and Collischonn, W. 2018. Toward continental hydrologic–hydrodynamic modeling in South America, Hydrol. Earth Syst. Sci., 22, 4815-4842.

Sood, A. and Smakhtin, V., 2015. Global hydrological models: a review. Hydrol. Sci. J. 60(4), 549e565. http://dx.doi.org/10.1080/02626667.2014.950580.

Sulistioadi, Y. B., Tseng, K.-H., Shum, C. K., Hidayat, H., Sumaryono, M., Suhardiman, A., Setiawan, F., and Sunarso, S. 2015. Satellite radar altimetry for monitoring small rivers and lakes in Indonesia, Hydrol. Earth Syst. Sci., 19, 341-359, https://doi.org/10.5194/hess-19-341-2015.

Sun, C.; Walker, J.P.; Houser, P.R. A methodology for snow data assimilation in a land surface model. J. Geophys. Res. 2004, 109, D08108.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.-L. 2017. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev. Geophys., 56 (2017), pp. 79-107, 10.1002/2017RG000574

SWOT. 2017. SWOT Project: MISSION PERFORMANCE AND ERROR BUDGET

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., and de Castro, M. 2012. Global precipitation measurement: Methods, datasets and applications, Atmos. Res., 104–105, 70–97.

Tapley B.D., Bettadpur S., Watkins M., Reigber C. 2004. The gravity recovery and climate experiment: mission overview and early results. Geophys. Res. Lett., 31, p. L09607, 10.1029/2004GL019920

Thielen, J., Pappenberger, F., Salamon, P., Bogner, K., Burek, P., de Roo, A., 2010. The State of the Art of Flood Forecasting-Hydrological Ensemble Prediction Systems, p. 145.

Todini, E.: The ARNO rainfall-runoff model, J. Hydrol., 175, 339–382, 1996.

Todorov. E. Optimal Control Theory. In K. Doya, editor, Bayesian Brain: Probabilistic Approaches to Neural Coding, pages 269–298. MIT Press, 2006.

Tuozzolo, S., Lind, G., Overstreet, B., Mangano, J., Fonstad, M., Hagemann, M., et al. 2019. Estimating river discharge with swath altimetry: A proof of concept using AirSWOT observations. Geophysical Research Letters, 46. https://doi.org/10.1029/2018GL080771

Vrugt, J.A.; Robinson, B.A. 2007. Treatment of uncertainty using Ensemble methods: Comparison of secuential data assimilation and Bayesian model averaging. Water Resources Research, Vol. 43, W01411, doi: 10.1029/2005WR004838.

Walker, J.P. and P.R. Houser, 2005. Hydrologic data assimilation. In A. Aswathanarayana (Ed.), Advances in water science methodologies (230 pp). The Netherlands, A.A.Balkema.

Wanders, N., Bierkens, M.F.P., de Jong, S.M., de Roo, A., Karssenberg, D. 2014. The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. Water Resour. Res. 50, 6874–6891. http://dx.doi.org/10.1002/2013WR014639.

Weerts, Albrecht H.; El Serafy, Ghada Y.H. 2006. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research, Vol. 42, W09403, doi: 10.1029/2005WR004093.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130, 1913–1924.

Wongchuig, C.S., de Paiva, R.C.D., Siqueira, V., Collischonn, W. 2019. Hydrological Reanalysis Across the 20th Century: A Case Study of the Amazon Basin. Journal of Hydrology, (570):755-773. doi.org/10.1016/j.jhydrol.2019.01.025

Wood EF, Roundy JK, Troy TJ, van Beek LPH, Bierkens MFP, Blyth E, de Roo A, Doell P, Ek M, Famiglietti J, Gochis D, van de Giesen N, Houser P, Jaffe P, Kollet S, Lehner B, Lettenmaier DP, Peters-Lidard C, Sivapalan M, Sheffield J, Wade A, Whitehead P. 2011. Hyperresolution global land surface modeling: meeting a grand

challenge for monitoring Earth's terrestrial water. Water Resources Research 47: W05301. DOI: 10.1029/2010WR010090.

Yamazaki, D., Alsdorf, D., Kanae, S., Oki, T., Andreadis, K. 2011. Toward global-scale data assimilation using SWOT: Requirements for global hydrodynamics models. IGARSS proceeding, Proceedings of Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, pp.3031-3034 in Vancouver, Canada, 24-29 July, 2011

Yamazaki, D., De Almeida, G. A. M., and Bates, P. D. 2013. Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector based river network map, Water Resour. Res., 49, 7221–7235, https://doi.org/10.1002/wrcr.20552.

Yamazaki D., D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J.C. Neal, C.C. Sampson, S. Kanae & P.D. Bates. 2017. A high accuracy map of global terrain elevations Geophysical Research Letters, vol.44, pp.5844-5853, 2017 doi: 10.1002/2017GL072874

Yoon, Y., Durand, C. J. Merry, Clark, E.A., Andreadis, K., Alsdorf, D. 2012. Estimating river bathymetry from data assimilation of synthetic SWOT measurements. Journal of Hydrology, v. 464–465, p. 363–375.

Yoon, Y., Durand, M., Merry, C. J. and Rodrıguez, E. 2013. Improving temporal coverage of the SWOT mission using spatiotemporal kriging, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., 6(3), 1719–1729. doi:10.1109/JSTARS.2013.2257697.

Yu, Z. 2015. Modeling and Prediction. College of Hydrology and Water Resources, Hohai University, Nanjing, China; and University of Nevada Las Vegas, Las Vegas, NV, USA

Capítulo 6

Conclusões gerais e perspectivas futuras

Neste capítulo final são apresentadas as conclusões gerais que bordam de maneira integral os cinco capítulos da tese. Adicionalmente uma seção de perspectivas de trabalhos futuros é adicionado como guia para a continuação e aprimoramento das metodologias e resultados apresentados nos capítulos anteriores.

6.1. Conclusões gerais

Esta pesquisa procura preencher diversas lacunas no conhecimento, porem várias perguntas científicas foram abordadas a continuação. Estas são desenvolvidas ao longo da tese desde o capitulo 2 até o capitulo 5, principalmente mediante o uso da modelagem hidrológica e hidrodinâmica, técnicas de assimilação de dados e o uso de informações como reanálises climáticas de precipitação, extensas informações de vazão in-situ e de sensoriamento remoto. Algumas técnicas complementares foram implementadas para melhorar o desempenho das simulações, tais como métodos de remoção de viés da precipitação, técnicas de localização e assimilação de múltiplas observações no esquema de EnKF.

Bases de dados globais de precipitação diária são adequados para representar importantes processos como a variabilidade hidrológica ou eventos extremos na bacia Amazônica?

Quão sensível é a resposta do modelo MGB aos parâmetros do esquema de DA EnKF e a perturbação de diferentes forçantes e variáveis de estado?

A modelagem hidrológica-hidrodinâmica de grande escala conjuntamente com reanálises climáticas e técnicas de assimilação de dados de longas séries de observações de vazão in-situ, serão capazes de criar um registro o suficientemente adequado para a representação da variabilidade hidrologia e de eventos extremos que aconteceram nos últimos 100 anos?

Informação de sensoriamento remoto do futuro satélite SWOT será realmente adequado para melhorar as simulações dos modelos hidrológicos em nível global e continental?

A DA de múltiplas variáveis é capaz de representar melhor os processos hidrológicos, desta maneira melhorar ainda mais as simulações?

Desta maneira, as conclusões gerais desta pesquisa são apresentadas a continuação.

6.1.1. Retrospectiva hidrológica

A simulação hidrológica-hidrodinâmica foi desenvolvida na bacia Amazônica, mediante o uso de oito bases de dados globais de precipitação como forçantes do modelo. Os resultados foram validados

mediante o uso de 27 estações de vazão in-situ no período de 1981 até 2010. Foi avaliada a variabilidade hidrológica, e a representatividade dos eventos hidrológicos extremos e da sua tendência através da análise das series simuladas de vazão. Esta análise é denominada de retrospectiva hidrológica.

As validações usando vários índices estatísticos indicaram que apenas três das oito bases de dados globais obtiveram uma performance adequada (ERA-Interim Land, CHIRPS e MSWEP), e que estas bases estão principalmente relacionadas com o uso de conjunto de modelos climáticos, dados in-situ e derivados de satélite, assim também estas bases de dados estão disponíveis principalmente nos últimos 30 anos. Desta maneira estas foram selecionadas para outras avaliações relacionadas com a capacidade de representar eventos hidrológicos extremos.

Mediante o uso de tabelas de contingência os três produtos foram capazes de maneira geral de capturar eventos como secas e cheias comparadas com as observações e reportadas na literatura. Desta maneira, é possível utilizar estas retrospectivas hidrológicas como uma primeira tentativa para representar de maneira espacialmente distribuída a variabilidade hidrológica numa serie temporalmente extensa. Isto nos permite por exemplo identificar eventos hidrológicos extremos que aconteceram no passado em locais onde observações in-situ são muito espalhadas ou não existem. Adicionalmente, a tendência temporal nos eventos extremos foi avaliada indicando um aumento de tendência nas cheias na região oeste e noroeste enquanto um aumento na intensidade das secas na região sul e centro sul. Estes resultados coincidem com resultados de diversos autores mediante o uso de series de observações in-situ.

Outras aproximações como técnicas de assimilação de dados de extensas séries de vazão in-situ poderiam ser implementadas para melhorar a representação da hidrologia para as reanálises mais extensas (~100 anos), as que mostraram uma inadequada performance. Embora, esta primeira aproximação pode ser implementada em outras regiões do mundo considerando: i) a disponibilidade global das forçantes de precipitação, ii) as incertezas e limitações destas forçantes e do modelo hidrológico.

6.1.2. Sensibilidade do MGB no esquema de EnKF

Técnicas de assimilação de dados pelo esquema de EnKF foram implementadas e avaliadas para a estimativa de séries temporais de vazão e vazões de referência em locais com poucas observações in-situ na bacia do rio Taquari Antas, no sul do Brasil. Diversos esquemas e testes de sensibilidade do modelo MGB aos diferentes parâmetros da DA e a perturbação da precipitação e as variáveis de estado foram avaliados. As estimativas das vazões de referência foram comparadas com metodologias simplificadas como a regionalização de vazões.

A partir de um esquema ótimo dos parâmetros da DA e da perturbação da precipitação e do armazenamento d'água no solo (variáveis que geram uma reposta mais sensível por parte do MGB), os resultados indicam que, de maneira geral, o método de EnKF é adequado na geração de series temporais de vazão na bacia do Taquari Antas. Além, se mostra adequado para estimar vazões de referência, apesar de sua relativa maior complexidade quando comparado com metodologias simplificadas. Por outro lado, certa diminuição da performance na estimativa da vazão foi evidenciada em algumas regiões o que provavelmente é devido a: i) alta variabilidade espacial de algumas regiões o que leva a extrapolar correlações espúrias em regiões pela metodologia de DA, e/ou ii) a pouca representatividade dos processos pelo MGB da rápida resposta em algumas regiões da bacia como as cabeceiras.

Nesse sentido é recomendado que em regiões com alta variabilidade espacial sejam implementadas técnicas que restringem espacialmente o impacto da assimilação; por outro lado dependendo dos objetivos e escala do estudo, outros processos hidrológicos devem ser considerados assim como o cálculo em intervalos de tempo menor (e.g. horário).

6.1.3. Reanálise hidrológica de longo período e na era dos satélites

A partir dos resultados e aprendizados dos capítulos 2 e 3, foi proposto o desenvolvimento do método denominado de reanálise hidrológica, como equivalente ao método aplicado em áreas da climatologia denominado de reanálise climatológica. Desta maneira, mediante o uso da modelagem hidrológica-hidrodinâmica, extensas forçantes de precipitação e métodos de assimilação de dados de extensos registros de vazão in-situ na bacia Amazônica, foi desenvolvido o produto denominado de reanálise hidrológica do século 20 (HRXX). Técnicas de remoção de viés da precipitação assim como de localização no esquema de DA EnKF (LEnKF) foram adicionalmente implementadas na metodologia.

Os resultados indicam que a consideração da restrição espacial do impacto da assimilação das observações in-situ através da técnica LEnKF consegue melhoras da performance em relação a técnica tradicional de EnKF, mostrando que esta restrição espacial devido a alta variabilidade espacial da bacia Amazônica e/ou devido a não conectividade física entre regiões muito distantes (e.g. bacias localizadas em cabeceiras diferentes), é muito importante e deve ser considerada. O incremento na performance do LEnKF num raio de influência "ótimo" previamente testado de 1000 km atinge valores de 70% e 9% para os índices de NSE e KGE, em relação com o EnKF.

É conhecido que o incremento da performance mediante o uso das técnicas de DA acontecerá espacial e temporalmente na presença das observações. Porém, considerações de remoção de viés foram assumidas devido a: i) reanálises climáticas de precipitação possuirem grandes valores de erro sistemático, ii) melhora das simulações nos períodos em que observações não estavam disponíveis. Das

metodologias testadas, a que teve melhor performance conseguiu remover o viés da precipitação num intervalo original de -50% até 50% para -5% até 5% em aproximadamente 80% da bacia Amazônica quando consideramos valores mensais.

A validação das simulações foi realizada mediante 56 postos in-situ de vazão e uma estação de nível. Esta validação considerou diferentes etapas, partindo de algumas implementações até todas em conjunto, a que é considerada como o produto final (HRXX). A HRXX utiliza os esquemas de localização (LEnKF) e de remoção de viés da chuva. A HRXX atinge a melhor performance tanto para representar a variabilidade hidrológica como uma correlação aceitável em relação a valores máximos e mínimos anuais nos últimos 100 anos. Esta base de dados está disponível de livre acesso e pode ser utilizada para diferentes fins científicos considerando as limitações e incertezas associadas. Apesar das simplificações assumidas, esta metodologia pode ser replicada e inclusive aprimorada para sua aplicabilidade em outras áreas onde poucas, mas extensos registros de informações de vazão ou de nível d'água in-situ estão disponíveis.

Grandes bacias nos países em desenvolvimento apresentam limitações em relação a observações in-situ devido principalmente a sua baixa densidade, desta maneira informações complementares como as provenientes de sensoriamento remoto podem ser adequadas para complementar observações in-situ devido a sua alta abrangência espacial. Embora a aplicabilidade de sensores atuais na observação de corpos de agua continental ainda apresentam certas restrições para representar adequadamente a dinâmica da agua em muitas regiões. Desta maneira o método de DA EnKF foi avaliado no modelo MGB mediante o uso de informações sintéticas da futura missão SWOT a qual está especialmente projetada para observar uma grande porcentagem de corpos d'água a nível global, sendo os principais produtos o nível d'água, extensão de áreas inundadas e vazão. Adicionalmente a técnica de DA de múltiplas observações foi implementada.

A avaliação teve como objetivo emular as incertezas de modelos hidrológicos globais (GHMs) mediante a corrupção do MGB, a qual será a escala de estudo (e.g. global, continental) pelos principais usuários dos produtos derivados do SWOT. Assim, foi usado um programa de software (Large Scale Pixel Cloud Simulator) que simula a orbita e os possíveis erros que o SWOT terá uma vez entre na fase de operação. O estudo foi implementado na bacia do rio Purus utilizando a metodologia do "twin experiment".

Valores das incertezas de modelos globais foram atingidos mediante a corrupção de forçantes e variáveis de estado do MGB. Resultados indicam que no Purus o MGB é mais sensível a precipitação e aos parâmetros hidrológicos do solo.

Devido as grandes incertezas dos modelos globais, o esquema de assimilação de dados conseguiu recuperar as variáveis de estado na maioria dos casos, mas reduziu a performance em outros casos. Embora quando a técnica de múltiplas observações é avaliada, considerando os valores da

mediana, em nenhum caso foi reduzida a performance. Isto indica uma clara vantagem no uso de múltiplas observações e que resulta coerente devido a que estas conseguem atualizar uma maior quantidade de processos do sistema ao mesmo tempo, um campo que está sendo utilizado recentemente também, por exemplo, no processo de calibração de modelos hidrológicos.

Esta primeira aproximação da avaliação de dados sintéticos do SWOT em grande escala através de um modelo distribuído resulta em melhoras de performance promissórias uma vez que dados reais sejam implementados em escalas globais e continentais.

6.2. Perspectivas futuras

Os resultados e conclusões da tese indicam que as metodologias implementadas são capazes de melhorar a capacidade de simulação do modelo hidrológico em grande escala, a pesar de algumas simplificações serem assumidas.

A partir disso, nesta seção são apresentadas algumas perspectivas futuras que permitirão melhorar a representação física dos processos hidrológicos e, portanto, a capacidade de simulação da modelagem hidrológica e hidrodinâmica em grande escala mediante o MGB.

i. Eventualmente, quando são usadas séries extensas de precipitação, é recomendado o uso de metodologias de remoção de viés que considere a não estacionareidade das series de precipitação, a pesar da boa representatividade de nossas simulações no desenvolvimento do produto HRXX e de que o uso da assimilação de observações in-situ já insere nossa análise a não estacionareidade do clima.
ii. O método de localização espacial no esquema LEnKF pode ainda ser melhorado mediante a desconsideração de regiões com regime de vazão altamente diferente (e.g. observação no rio principal e assimilação num rio tributário), de outro modo isto pode levar a expandir valores espúrios na DA.
iii. Observações de sensoriamento remoto normalmente possuem erros sistemáticos como por exemplo dados de altimetria derivada de satélite devido a diferentes sistemas de referência ou dados de vazão que proverá a futura missão SWOT estimados por diversos algoritmos. Conseguiram-se incrementos na performance assimilando os valores absolutos destas informações, embora esta pode ser melhorada mediante a assimilação das anomalias ou anomalias padrão, o que removeria os erros sistemáticos das observações.
iv. Técnicas de DA de múltiplas observações se mostram como provisórias devido a representatividade de maior quantidade de processos na dinâmica hidrológica. Desta forma inúmeros produtos provenientes de sensores remotos disponíveis nas últimas décadas podem ser

usados para sua implementação em nível regional e continental para a melhora das simulações através de modelos hidrológicos.

ANEXOS

ANEXO A

Equacionamento da técnica de Kalman Filter por Ensemble

A expressão. a.1 representa a matriz que contém o ensemble das variáveis de estado do modelo $\mathbf{x}^f$ na etapa do forecasting com dimensões ***nsta***×***nens***, onde nens é o número de membros do ensemble:

$$\mathbf{x}^f = \left(\mathbf{x}_1^{\mathrm{f}}, \mathbf{x}_2^f, \mathbf{x}_3^f, \ldots \mathbf{x}_{nens}^f\right) \qquad (a.1)$$

onde $\mathbf{x}_1^f, \mathbf{x}_2^f, \mathbf{x}_3^f, \ldots \mathbf{x}_{nens}^f$ representam os vetores que contem as variáveis de estado do modelo para cada membro do ensemble.

Como descrito linhas acima, na formulação do EnKF é assumido que a média do ensemble das variáveis de estado representa é a verdade, desta maneira é denotado matematicamente como:

$$\overline{\mathbf{x}^f} = \frac{\sum_{i=1}^{nens} \mathbf{x}_i^f}{nens} \qquad (a.2)$$

Porém, o erro associado com cada membro i do ensemble ($\mathbf{x'}_i^f$) e a matriz de anomalias do ensemble das variáveis de estado ($\mathbf{A'}^f$) são definidas:

$$\mathbf{x'}_i^f = \mathbf{x}_i^f - \overline{\mathbf{x}^f} \qquad \boldsymbol{A'}^f = \left(\boldsymbol{x'}_1^f, \boldsymbol{x'}_2^f, \boldsymbol{x'}_3^f, \ldots \boldsymbol{x'}_{nens}^f\right) \qquad (a.3)$$

As variáveis de estado ($\mathbf{x}^f$) consideradas a ser atualizadas no modelo MGB são: i) armazenamento d'água no solo (W), volume do reservatório superficial (Vsup), volume do reservatório sub-superficial (Vint), volume do reservatório subterrâneo (Vbas), escoamento gerado em cada minibacia (Q_{CEL}), água interceptada (SI), vazão (Q), nível d'água (Y), extensão da área inundada (Area) e volume d'água armazenada (Vol).

Então, a matriz de covariância dos erros do modelo ($\mathbf{P}_e^f$) pode ser estimada diretamente a partir da matriz de anomalias do ensemble:

$$\mathbf{P}_e^f = \frac{\mathbf{A'}^f \mathbf{A'}^{f^T}}{nens - 1} \qquad (a.4)$$

No esquema do EnKF proposto por Evensen (1994) e reafirmado por Burgers et al. (1998), se assume que as observações têm um comportamento aleatório. Desta maneira as observações são perturbadas adicionando um erro pseudo-aleatório que segue uma distribuição gaussiana com uma covariância igual a $\mathbf{R}_e$. Desta maneira são definidos nens vetores (i=1,2,3,...nens) de observações perturbadas, em que cada vetor y_i possui dimensões *nobs*×1. Assim, estes vetores de observações perturbadas podem ser alocados numa matriz **D**, de dimensões *nobs*×*nens*:

$$\mathbf{y}_i = \mathbf{y} + \varepsilon_i \quad \varepsilon_i \sim N(0, R_e) \qquad \boldsymbol{D} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \ldots \boldsymbol{y}_{nens}) \qquad (a.5)$$

Semelhante com a matriz de perturbações das variáveis de estado, o ensemble das perturbações das observações com média igual a zero, é armazenada na matriz **E**:

$$\mathbf{E} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \dots \boldsymbol{\varepsilon}_{nens}) \qquad (a.6)$$

Assim a matriz de covariância de erros das observações do ensemble é definido como R_e:

$$\mathbf{R}_e = \frac{\mathbf{E}\mathbf{E}^T}{nens - 1} \qquad (a.7)$$

Baseado nas Eq. 1.8 e 1.9, para o caso do EnKF a atualização das variáveis de estado é feita sobre cada membro do ensemble, e a definição do ganho de Kalman é semelhante:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_e\left(\boldsymbol{y}_i - \mathbf{H}\mathbf{x}_i^f\right) \qquad \boldsymbol{K}_e = \boldsymbol{P}_e^f \boldsymbol{H}^T \left[\boldsymbol{H}\boldsymbol{P}_e^f \boldsymbol{H}^T + \boldsymbol{R}_e\right]^{-1} \qquad (a.8)$$

Finalmente o esquema do EnKF apresenta a seguinte equação geral:

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}_e\left(\bar{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}^f}\right) \qquad (a.9)$$

Por outra parte, os erros das observações são parametrizados seguindo o esquema de Clark et al. (2008), o qual foi adotado para a estimativa da matriz de covariância dos erros das observações (**R**).

$$\sigma_{obs}^2 = (\varepsilon_{obs} \cdot var_{obs})^2 \qquad (a.10)$$

onde $\boldsymbol{\varepsilon_{obs}}$ representa o erro ou incerteza na observação que pode ser conhecida a priori, por exemplo para a vazão podem ser mesuradas a incertezas na estimativa pela curva chave, $\boldsymbol{var_{obs}}$ indica o valor da variável observada.

Geração dos elementos do ensemble no EnKF

O esquema de EnKF pela natureza estocástica da técnica, requer a implementação da perturbação das forçantes, parâmetros ou variáveis de estado do modelo, embora dentre eles, a forçante da precipitação é considerada como a principal fonte de incerteza nos modelos hidrológicos (Biancamaria et al. 2011; Liu et al., 2012a; Paiva et al., 2013b). Desta maneira, para o processo de assimilação de dados no desenvolvimento das reanálises hidrológicas, será avaliada a técnica de EnKF através da perturbação da precipitação. Segundo o trabalho de Clark et al. (2008) a precipitação pode ser perturbada através de um esquema multiplicativo seguindo um multiplicador φ_p.

$$P' = P\varphi_p \qquad (a.11)$$

Este multiplicador é expresso como:

$$\varphi_p = \left(1 - \varepsilon_p\right) + 2u_p\varepsilon_p \qquad (a.12)$$

onde u_p é um numero aleatorio uniforme, ou seja varia entre 0 e 1, desta maneira φ_p é uma realização que vem de uma distribuição uniforme que varia entre $1 - \varepsilon_p$ e $1 + \varepsilon_p$. Desta maneira por exemplo se o erro ε_p é de 20%, a distribuição do multiplicador varia de 0.8 até 1.2 mas de maneira aleatoria.

Continuando essa lógica, a perturbação adotada ao longo desta tese também será multiplicativa mas segue uma distribuição log-normal dos erros, como proposto por Nijssen e Lettenmaier (2004) e que foi adotado em outros trabalhos (Andreadis e Lettenmeier, 2006; Paiva et al., 2013). A distribuição log-normal dos erros foi adotada principalmente para evitar os valores negativos na precipitação perturbada, e o esquema apresenta o seguinte equacionamento:

$$P' = \frac{1+\beta}{\sqrt{E^2+1}} exp\left(\sqrt{ln[E^2+1]}s\right) P \qquad (a.13)$$

em que P' indica a precipitação perturbada, P a precipitação não perturbada, E o erro relativo da precipitação (%), β o viés relativo, e s indica um número aleatório ou campo de perturbações o qual segue uma correlação espacial e temporal.

O valor de w é um campo de números pseudoaleatórios espacialmente correlacionados por meio do algoritmo baseado na transformada de Fourier bidimensional apresentada por Evensen (2003) com média 0, variância 1 e uma função de covariância que diminui espacialmente e^{-1} vezes o valor da distancia τ_x que é denominado de comprimento da decorrelacao espacial.

Enquanto a evolução temporal dos erros, este segue a seguinte dependência (Evensen, 2003):

$$S_k = \alpha S_{k-1} + \sqrt{1-\alpha^2} w_{k-1} \qquad (a.14)$$

$$\alpha = 1 - \frac{\Delta t}{\tau_t} \qquad (a.15)$$

em que α ϵ [0 1› é um coeficiente que indica a influência da correlação temporal na estocasticidade do modelo de erros, α=0 indica uma sequência de ruído branco (white noise) enquanto α=1 remove a forçante estocástica e representa os erros do modelo com um componente aleatório que é constante no tempo, w é o termo estocástico ou ruído "pseudoaleatório" com média 0 e variância 1 e espacialmente correlacionado, tanto s como w são vetores com dimensão *ncat*×1, sendo ncat o número de mini bacias, finalmente τ_t é o tempo de decorrelação e é usado para determinar o coeficiente α, enquanto Δt é a magnitude do passo de tempo do modelo (Clark et al., 2008).

Técnica de localização

Ao longo deste subcapitulo serão apresentados alguns métodos associados a técnica de DA para melhorar sua performance e representatividade física, isto conduzirá a redução de incertezas das simulações no modelo hidrológico.

Muitos autores têm identificado que técnicas de DA podem gerar resultados contrários aos esperados em algumas situações, como no caso da aplicação em regiões muito grandes ou quando se assumem poucos elementos do ensemble no caso do método do EnKF, desta maneira o ganho de Kalman é afetado pela amostragem e o erro da analise pode exceder os erros do background ou forecast (Houtekamer e Zhang, 2016; Zhang et al., 2016). Desta maneira, algumas modificações têm sido implementadas na versão tradicional do EnKF (Patil et al., 2001; Ott et al., 2002, 2004) que é conhecido em alguns casos de filtro de Kalman por ensemble local (*local ensemble Kalman filter*, LEnKF). Estas modificações adicionam o impacto local das observações dentro do ensemble de variáveis de estado do modelo, cuja vantagem se encontra principalmente em evitar correlações espúrias na matriz de erros do modelo (Zhu et al., 2011). Devido ao custo computacional que implica o cálculo das matrizes com um número grande de elementos do ensemble (~>100), tem sido identificado que as aplicações que usam um número limitado de elementos do ensemble atribuem de maneira espúria correlações diferentes que zero para variáveis que não estão correlacionadas (Bishop e Hodyss, 2009; Campbell et al. 2009) ou entre pontos que se encontram distantes ou fisicamente desconectados (Hamill et al., 2001; Anderson, 2007).

No caso de esquemas sequencias de DA como o EnKF, o método de localização foi desenvolvido com base em uma matriz de covariância dependente da distância (Houtekamer e Mitchell, 2001; Campbell et al., 2009; Zhu et al., 2011; Houtekamer, 2016). Desta maneira esta técnica tem demonstrado relevância na sua implementação em estudos hidrológicos (Rasmussen et al., 2015; Munier et al., 2015; Zhang et al., 2016).

O equacionamento geral da localização avaliado nesta tese está baseado na implementação da matriz de correlações na Eq. a.16 como desenvolvido por Sakov e Bertino (2011).

$$\mathbf{A}^a = \mathbf{A}^f + \boldsymbol{\rho} \times \left(\mathbf{A}'^f \mathbf{A}'^{f^T} \mathbf{H}^T\right) \left[\boldsymbol{\rho}_o \times \left(\mathbf{H}\mathbf{A}'^f \mathbf{A}'^{f^T} \mathbf{H}^T\right) + \mathbf{E}\mathbf{E}^T\right]^{-1} \mathbf{D}' \qquad (a.16)$$

onde $\boldsymbol{\rho}$ e $\boldsymbol{\rho}_o$ sao as matrizes de correlacao, e $\times$ representa o operador do produto Hadamar ou produto Schur onde cada elemento da matriz é multiplicado pelo valor correspondente do elemento da outra matriz das mesmas dimensões. As matrizes $\boldsymbol{\rho}$ e $\boldsymbol{\rho}_o$ (também denominada versão compacta de $\boldsymbol{\rho}$) tem dimensões de *nsta*×*nobs* e *nobs*×*nobs* respectivamente; enquanto D' representa a matriz de innovacao.

A matriz de correlações foi realizada seguindo uma equação polinomial de quinto grau que segue uma dependência espacial considerando um raio de influencia pré-definido. Este esquema foi proposto por Gaspari e Cohn (1999) e avaliada por Hamill et al. (2001) e Houtekamer e Mitchel (2001). O esquema geral para a determinação da matriz de correlação (Eq. a.17) segue a seguinte sequencia: i) são determinadas as mini bacias da cabeceira (Fig. a. 1a) e logo são gerados ***ncat*** vetores com as numerações correspondentes de cada mini bacia; ii) logo é gerada uma matriz de dimensão *ncat*×*ncat* que corresponde com a distância entre cada mini bacia com as outras mini bacias através da rede de drenagem; iii) finalmente é gerada a matriz **rho** de dimensões *ncat*×*ncat* correspondente com os valores da correlação espacial que varia entre 0 e 1. Esta matriz se caracteriza por possuir valores unitários na diagonal principal, valores <1 para as mini-bacias dentro da área de influência e valores 0 para mini bacias fora da área de influência ou que não tem conexão física com a mini bacia de referência (e.g. afluente vizinho) (Eq. a.17). Quando existe uma observação numa mini bacia especifica (Fig. a.1b) esta gera uma área de influência ao longo da rede de drenagem por médio do raio estabelecido. O equacionamento da função de correlação em relação ao raio de influência e a distância entre as observações com as outras mini-bacias é apresentado na seção 4.2.6.2.

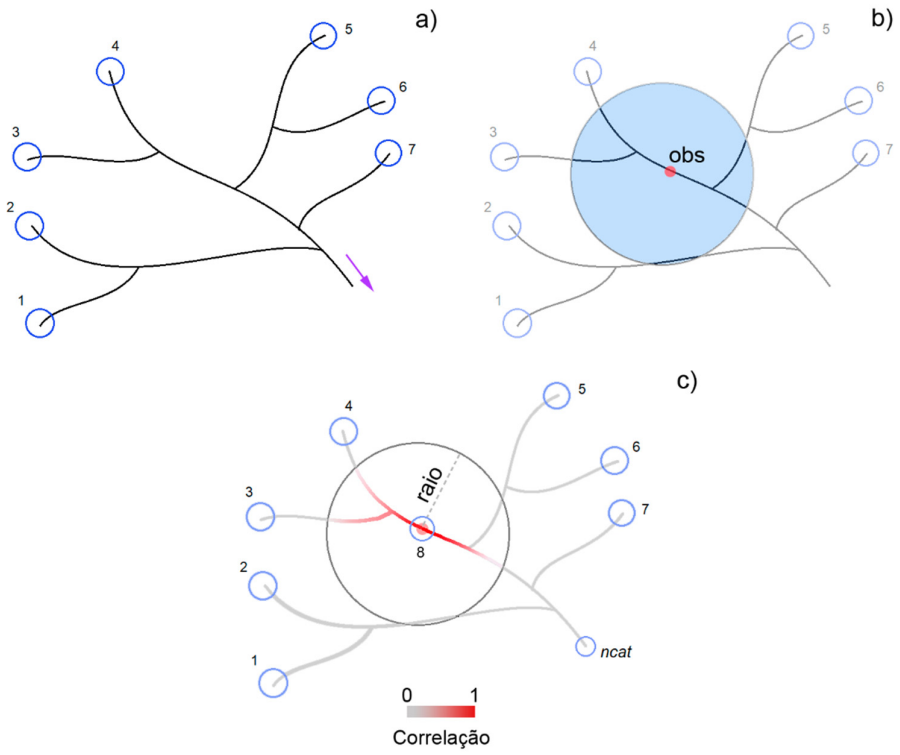

Figura a.1. Esquema de localização na rede de drenagem.

$$\mathbf{rho} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & [0,1[& 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & [0,1[& 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & [0,1[& 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & [0,1[& 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & [0,1[& 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & [0,1[& 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & [0,1[& 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & [0,1[& 0 \\ [0,1[& [0,1[& [0,1[& [0,1[& [0,1[& [0,1[& [0,1[& [0,1[& \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{\boldsymbol{ncat \times ncat}} \qquad (a.17)$$

Técnica de assimilação de múltiplas observações no EnKF

Modelos conceituais ou semi-conceituais apresentam incertezas na representação dos processos físicos, pela simplificação destes processos ou pela definição dos seus parâmetros. Estes parâmetros podem esconder e/ou compensar erros ou a verdadeira representatividade dos processos físicos do ciclo hidrológico. Este fenômeno é denominado de equifinalidade, ou seja, se chegam aos mesmos resultados ótimos com diferentes conjuntos de parâmetros.

Neste sentido a implementação da DA de multi-observações requereu um arranjo diferente nas matrizes $\mathbf{HA'}^f$, $\mathbf{E}$ e $\mathbf{D'}$ apresentadas na Eq. a.18. Assumindo $\boldsymbol{n}$ tipo de observações, as matrizes terão dimensões de $nobsT \times nens$, $nobsT \times nobsT$ e $nobsT \times nens$ respectivamente, e o arranjo é da seguinte maneira:

$$nobsT^k = \sum_{i=1}^{n} nobs_i^k \qquad (a.18)$$

$$\mathbf{HA'}^f = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{\boldsymbol{nobsT \times nens}}$$

$$\mathbf{E} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & & & & 0 & 0 & 0 \\ 0 & & & & 0 & 0 & 0 \\ 0 & & & & 0 & 0 & 0 \\ 0 & & & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & & \end{bmatrix}_{\boldsymbol{nobsT \times nobsT}} \qquad (a.19)$$

$$\mathbf{D'} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{\boldsymbol{nobsT \times nens}}$$

onde $nobs_i$ representa o numero de observações de dado tipo de observações num instante k. Na Eq. a.19, os valões assumidos como zero indicam que as observações não estão correlacionadas e as cores indicam uma fonte especifica de observações.

ANEXO B

Informações de sensoriamento remoto

a. Altimetria por satélite

A altimetria por satélite toma vantagem das propriedades refletivas dos corpos d'água, desta maneira um altímetro por nadir abordo de um satélite mede o tempo de viagem bidirecional do pulso desde que é emitido pela antena, refletido pela superfície e capturado de novo pelo sensor (AVISO, 1996). Inicialmente os altímetros de radar foram concebidos para serem usados em grandes superfícies relativamente regulares como os oceanos, tais como o TOPEX/Poseidon (1992-2005), seguido pelo Jason-1 (2001-2013) e Jason-2 (2008-presente) e mais recentemente o Jason-3 (2016-presente), ENVISAT (2002-2012) e ICESat (2003-2010). Isto se apresentou como uma limitação para seu uso em continentes devido à maior irregularidade da superfície afetada pela topografia ou a vegetação, embora os primeiros esforços na altimetria por satélite através dos radares em superfícies liquidas continentais como grandes reservatórios foram testadas por Brooks (1982) assim como em grandes lagos na América do Norte. Segundo León et al. (2008) estes tipos de estudos foram desenvolvidos inicialmente com o lançamento do satélite Seasat no ano 1978, seguidamente pelo Geosat entre os anos 1986 e 1989, e finalmente o Topex/Poseidon entre 1992 e 2001, produtos que também foram avaliados em grandes lagos da África (Mercier et al., 2002). Muitos sensores de altimetria por nadir que medem o nível da superfície d'água também tem sido bastante usados em estudos hidrológicos (Paiva et al., 2013a, Sulistioadi et al., 2015).

Adicionalmente, e de maneira paralela aos sensores por radar, tem se explorado a altimetria mediante a tecnologia laser, originalmente desenvolvidas para a medição das elevações nas superfícies de gelo continental (e.g. ICESat). Segundo Brenner et al. (2007) os altímetros por radar e laser possuem tanto vantagens como desvantagens, enquanto o altímetro por radar possui imprecisões na penetração de alguns tipos de superfícies como a vegetação, este pode atravessar diferentes condições atmosféricas como por exemplo as nuvens. Por outra parte, os altímetros laser podem mensurar com melhor precisão, embora possuem dificuldades de atravessar as nuvens.

Dentro do campo da altimetria aplicado à hidrologia, tem se usado o termo "estação virtual" que é considerado como o local em que se dá a intersecção entre uma superfície liquida monitorada (rios, lagos, reservatórios, entre outros) com o traço da leitura do satélite em que é possível criar uma serie temporal dos dados de altimetria. Nos últimos anos diferentes trabalhos foram abordados mediante o uso de informação de altimetria e suas aplicações em hidrologia (Leon et al., 2006; Santos da Silva, 2008; Roux et al., 2008), de maneira paralela foi desenvolvida uma grande base de dados de altimetria por radar em corpos de agua continentais (e.g. rios largos, lagos, reservatórios) pelo grupo de pesquisa LEGOS (Laboratoire d'Etudes en Géophysique et Océanographie).

Informações das missões Jason-1, Jason-2 e ENVISAT já processadas para a bacia Amazônica, encontram-se disponíveis no site <http://hydroweb.theia-land.fr/>, último acesso 07 de agosto de 2018,

adicionalmente algumas outras missões com informação disponível na bacia Amazônica são apresentadas na Tabela b.1.

Tabela. b.1. Resumo de alguns produtos de altimetria por satélite disponíveis.

Missão	Instituição	Referência	Acurácia (cm)	Cobertura temporal	Footprint (m)	Distância entre traços (km)	Ciclo (dias)	Delay na disponibilidade
TOPEX/ Poseidon	NASA/CNES	Frappart et al. (2006)	35	1992/08 - 2005/12	~600	315	9.9	Dados reprocessados
Jason-1	NASA/CNES	Jarihani et al. (2013)	107	2001/12 - 2013/07	~300	315	9.9	Dados reprocessados
Jason-2	NASA/CNES	Jarihani et al. (2013)	28	2008/06 - presente	~300	315	9.9	~3 a 5 hrs, ~24 hrs e 1 a 2 meses
Jason-3	NASA/CNES	-		2016/01 - presente	~300	315	9.9	
ENVISAT	ESA/CNES	Frappart et al. (2006)	28	2002/03 - 2012/06	~400	80	35	Dados reprocessados
SARAL/ Altika	ISRO/CNES	Schwatke et al. (2015)	11	2013/02 - presente	~173	80	35	~3 a 5 hrs, ~48 hrs e 2 meses
Sentinel-3 A/B/C/D	ESA	-	-	2016 (A) - presente	~300	52 (A+B)	27	-
ICESat	NASA	Urban et al. (2008)	10	2003/02 - 2010/08	~70	30	91	Dados reprocessados
ICESat-2	NASA	-	5-20	2018 - presente	~17	-	91	-

Adaptado de O'Loughlin et al. (2016) e McCabe et al. (2017)

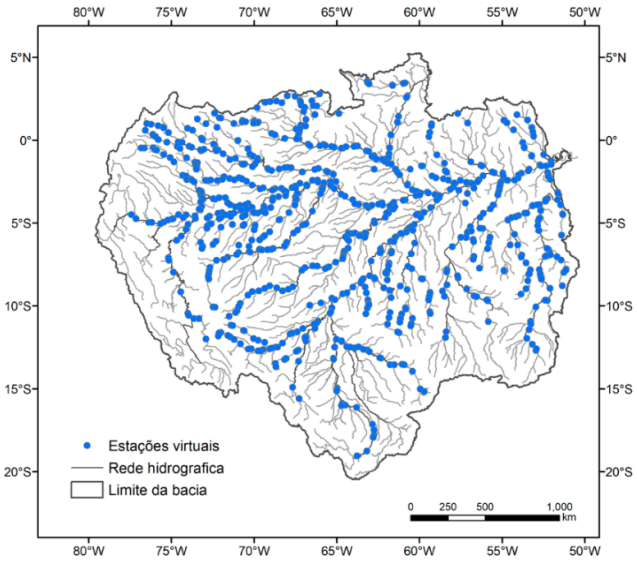

Figura. b.1. Mapa de localização das estações virtuais dos satélites Jason-1, Jason-2 e ENVISAT na bacia Amazônica (fonte hydroweb.theia-land).

b. Gravimetria

A missão GRACE (Gravity Recovery and Climate Experiment) foi desenvolvida pela NASA conjuntamente com a agencia espacial alemã German Aerospace Center (DLR), com a finalidade de mapear as variações temporais do campo gravitacional da terra. Os resultados permitem detectar indiretamente a variação no armazenamento d'água na superfície terrestre em ciclos de 30 dias aproximadamente, considerando que o deslocamento d'água nos continentes é uma das principais causas

de flutuações de massa na terra (Rodell and Famiglietti, 2002; Strassberg et al., 2009). A missão compreende dois satélites separados aproximadamente 220 km (Rodell and Famiglietti, 2002), desta maneira a variação no campo gravitacional altera a distância entre eles, e os sensores são capazes de detectar essas variações com a sensibilidade de um micrometro (Strassberg et al., 2009). Devido que outros elementos podem causar uma alteração no campo gravitacional, tais como flutuações atmosféricas, mudanças na topografia, efeitos de maré e movimentos de gelo ou neve, a informação detectada pelo GRACE prevê uma serie de correções para desprezar estes outros efeitos.

A missão GRACE tem contribuído com a compreensão das variações nos campos espaço-temporais do armazenamento d'água continental (Costa et al., 2011; Tapley et al., 2004). Por outro lado, diversos estudos indicaram que as estimativas da variação no armazenamento d'água pelo GRACE são adequadas quando comparadas com medições in-situ e modelos hidrológicos em escala regional (Ramillien et al., 2008), além disso outros estudos também mostraram que os produtos do GRACE podem ser usados na modelagem hidrológica (Lettenmaier e Famiglietti, 2006; Vaz de Almeida et al., 2009; Güntner, 2008; Paiva et al., 2013a).

Informaçoes disponíveis para a assimilação dos dados do GRACE no modelo hidrológico MGB podem ser atribuídas a média dos produtos processados pelos centros CSR (University of Texas / Center for Space Research), o GFZ (GeoForschungsZentrum Potsdam) e o JPL (Jet Propulsion Laboratory) (disponível em <https://grace.jpl.nasa.gov/data/get-data/>). Estes dados (Tabela b.2) possuem uma resolução temporal de aproximadamente 1 mês e resolução espacial de ~1° x ~1°.

Tabela. b.2. Resumo das características das missões GRACE.

Missão	Instituição	Referência	Cobertura temporal	Resolução espacial	Ciclo (dias)
GRACE	NASA/DLR	Tapley et al. (2004)	2002/04 - 2017/05	1°x1° (~111 km no equador)	~30
GRACE-FO	NASA/DLR	GRACE-FO (2018)	2018/05 - presente	-	~30

Os dados da solução mascon (mass concentration pelo acrônimo em inglês) (Tabela b.3) seriam usados nestas avaliações, estes têm a vantagem em comparação com os produtos anteriores (spherical harmonics) devido (i) redução do ruído na amplitude na transição oceano continente; (ii) uso de restrições geofísicas com postprocessamento menos empírico, o que facilita para o usuário menos experiente nessa área (Scalon et al., 2016). A Figura b.2 mostra o produto que seria usado, o GRACE processado pelo JPL na versão RL06v1.

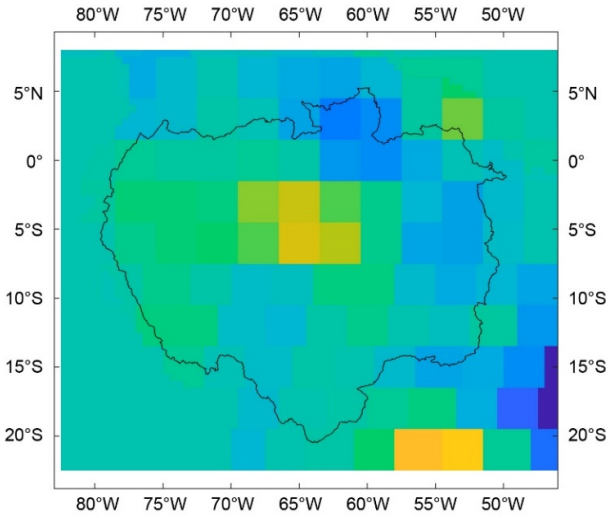

Figura. b.2. Mapa da distribuição espacial do produto GRACE mascon JPL RL06v1 na bacia Amazônica.

Tabela. b.3. Resumo das soluções GRACE mascon disponíveis.

Versão mascon	Instituição	Referência	Cobertura temporal	Resolução espacial	Enlace
Release 06 RL06v1	JPL	Wiese et al. (2018)	2002/04 - 2017/05	1° x 1°	https://podaac.jpl.nasa.gov/dataset/
Release 05 RL05v1	CSR	Save et al. (2016)	2002/04 - 2017/05	1° x 1°	http://www2.csr.utexas.edu/grace/
v02.4	GSFC	Luthcke et al. (2013)	2003/01 - 2016/07	3° x 3°	https://neptune.gsfc.nasa.gov/gngphys/

c. Áreas inundáveis

Produtos de áreas inundáveis (Papa et al., 2010) tem sido usado previamente (Paiva et al., 2013a) para a validação dentro do modelo hidrológico MGB, embora ainda não para processos de assimilação de dados. Deste modo alguns possíveis produtos são apresentados a seguir.

As estimativas de áreas inundáveis por sensoriamento remoto têm sido especialmente valoradas em regiões pouco monitoradas e com planícies de inundação sujeitas a eventos como cheias (Thomas et al., 2011). Diferentes estudos têm demonstrado que a informação de satélite (e.g. Landsat, IKONOS, MODIS) através de imagens no espectro visível/infravermelho são capazes de prover adequados registros de áreas inundadas (Wang, 2004; Brakenridge and Anderson, 2006), como por exemplo o GSWO (Global Surface Water Occurence) (Pekel et al., 2016), um dos mais recentes produtos globais de agua superficial baseado em imagens do Landsat entre o período de 1984 a 2015. Embora os produtos baseados em imagens no espectro visível/infravermelho apresentam deficiências em regiões cobertas por florestas e nuvens (Smith, 1997; Aires et al., 2017). Para superar estes inconvenientes em regiões com grandes extensões de floresta (e.g. Amazonas), produtos de sensores ativos no espectro do mico

ondas (e.g. SAR) tem sido usado em diferentes estudos (Townsend, 2002; Horrit et al., 2001; Brivio et al., 2002) para detectar áreas inundadas nessas condições, a partir da medição do índice de retro espalhamento (IR). Dentre eles os sensores RADARSAT-1, Sentinel 1 (Pham et al., 2016), JERS-1 (Saatchi et al., 2000; Rosenqvist et al., 2002; Hess et al., 2013), AlosPALSAR (Rosenqvist et al., 2004), este último avaliado tanto na região Amazónica como no Congo (Arnesen et al., 2013; Lee et al., 2014). Por outro lado, não tem sido uma tarefa fácil estabelecer uma metodologia que defina um limiar adequado do IR para determinar as extensões de áreas inundadas e que automatize os processos de classificação das imagens SAR (Brisco et al., 2009; White et al., 2014; Li e Wang, 2015; Westerhoff et al., 2013), o que dificultaria o pré-processamento das áreas inundáveis a partir destes produtos para os fines de sua assimilação no MGB.

Por outra parte, tem sido desenvolvido nos últimos anos a base de dados GIEMS (Global Inundation Extent from Multi-Satellites) que prove de informação global de média mensal das extensões de agua superficial entre os períodos de 1993 a 2007 através da combinação de observações de satélites no espectro do visível, infravermelho próximo e micro-ondas passivo/ativo (Prigent et al., 2007, 2012; Papa et al., 2010) e que tem sido testado em aplicações regionais para estimar as variações do armazenamento d'água superficial e sub-superficial nas bacias dos rios Ganges e Brahmaputra (Papa et al., 2015). Este produto foi aprimorado mediante técnicas de downscaling utilizando principalmente mapas de probabilidade de inundações a partir de o SRTM, gerando produtos em resoluções de ~500m (Fluet-Chouinard et al., 2015) e ~90m (Aires et al., 2017) denominados GIEMS-D15 e GIEMS-D3 respectivamente no período 1993-2007. Este último foi comparado na Amazônia com produtos SAR (Hess et al., 2013) e o GSWO obtendo resultados bastante coerentes com a dinâmica de inundações apresentada nestes produtos. Alguns produtos disponíveis na bacia Amazônica são resumidos na Tabela b.4.

Tabela. b.4. Resumo das características de alguns produtos de área inundável.

Produto	Referência	Cobertura temporal	Cobertura espacial	Resolução espacial	Resolução temporal
GIEMS-D3	Aires et al. (2017)	1993 - 2007	Global	~25 km	~30 dias
SWAF-HR	Parrens et al. (2019)	2010 - 2016	Amazonas	~1 km	~3 dias

d. Informações sintéticas do satélite SWOT

Missões atuais de altimetria por satélite ou produtos de extensões e variabilidade de áreas inundáveis ainda não proveem de uma resolução adequada para um completo entendimento da dinâmica d'água continental (Pedinotti et al. 2014). Para suprir estas necessidades, a missão SWOT (Surface Water and Ocean Topography) está sendo desenvolvida pela colaboração entre as agências espaciais estadunidense NASA, francesa Centre National d'Etudes Spatiales (CNES), e recentemente pelas agências canadense (CSA/ASC) e inglesa (UKSA). Esta missão tem previsão de lançamento para o ano 2021, e está especificamente planejada para prover de informação em alta resolução espacial e temporal,

principalmente, mapas de níveis de superfície d'água (maiores a 250m x 250 m), declividade da superfície d'água e estimativas de vazões (para rios mais largos que 100 m) (Durand et al., 2014; Prigent et al., 2015). Estas melhoras serão possíveis devido às capacidades do radar interferométrico na banda Ka (KaRIN), composto por duas antenas separadas 10m e que observarão duas faixas terrestres (swath) de 60km a cada lado do nadir do satélite, separados por um espaço (gap) de 20km (Rodriguez, 2012). Por outra parte, as medições do interferómetro por radar são sensíveis à estabilidade do satélite, para abordar este problema o SWOT possui um altímetro convencional para calibração e validação das informações do interferômetro (Fu and Ubelmann, 2014). Maiores detalhes técnicos podem ser encontrados em Durand et al. (2010) e Fu and Rodriguez (2004). Aplicações preliminares usando dados sintéticos do SWOT no esquema de assimilação de dados têm sido estudadas durante os últimos anos, por exemplo para melhorar parâmetros do modelo como o coeficiente de rugosidade de Manning (Pedinotti et al., 2014), com fines de manejo de reservatórios (Munier et al., 2015), estimativas da batimetria do rio (Yoon et al., 2012), ou aprimoramentos em modelos hidrológicos (Biancamaria, et al, 2011).

Neta pesquisa, as informações a serem usadas no esquema de assimilação de dados dentro do modelo hidrológico MGB serão dados sintéticos proporcionados pelo Large Scale SWOT Pixel Cloud Simulator. Diferentes cenários poderão ser avaliados como a assimilação destas informações em rios com planícies, nas cabeceiras ou no rio principal, ou a avaliação da melhora das variáveis de estado devido a assimilação de diferentes observações ou todas em conjunto.

Tabela. b.5. Resumo das características da missão SWOT.

Missão	Instituição	Referência	Cobertura temporal	Resolução espacial	Ciclo (dias)
SWOT	NASA/CNES	Biancamaria et al. (2015)	2021 - ~2024	~50 - 100 m	20.86