THÈSE DE DOCTORAT DE L'UNIVERSITÉ
PIERRE ET MARIE CURIE

Spécialité : **Hydrogéologie**

École Doctorale de Géosciences, Ressources Naturelles et Environnement

présentée par

Ana SCHNEIDER

pour obtenir le grade de :

Docteur de l'Université Pierre et Marie Curie

Estimation of the base flow time constant for global scale applications

---

Soutenue le 22 Juin 2017 devant le jury composé de :

M^{me} Agnès DUCHARNE	Directrice de thèse
M^{me} Anne JOST	Encadrante
M. Pierre RIBSTEIN	Examinateur
M^{me} Shannon STERLING	Rapporteur
M. Aaron BOONE	Rapporteur
M^{me} Pauline ROUSSEAU-GUEUTIN	Examinatrice
M. Jan POLCHER	Examinateur

UMR 7619 - METIS
4 place Jussieu, Case 105
75252 - Paris - CEDEX 05

UPMC
École Doctorale de Géosciences, Ressources Naturelles et Environnement
Campus Jussieu
Paris - France

To Amin and Sarah.

REMERCIEMENTS

Mes remerciements vont tout d'abord à Agnès Ducharne et Anne Jost pour m'avoir acceptée en tant que doctorante pendant ces trois ans et demi et pour la confiance qu'elles m'ont accordée tout au long de cette thèse. J'aimerais remercier spécialement Agnès pour m'avoir choisie parmi d'autres candidats (alors que je me trouvais encore au Brésil) et pour m'avoir encore soutenue par la prolongation de mon contrat, après la naissance de ma petite princesse Sarah. J'ai eu la chance d'avoir été encadrée par Agnès et Anne. Vous étiez toujours disponibles, vous avez lu et corrigé toutes mes publications (mêmes les petits résumés de la Journée des Doctorants), vous étiez là chaque fois que j'ai eu besoin de discuter avec vous. Je ne vous remercie pas seulement pour le côté scientifique, mais aussi humain : vous avez accepté ma condition de maman avec sérénité et toujours avec un grand sourire, c'est pourquoi je vous remercie énormément.

Je remercie les membres du jury pour avoir accepté d'évaluer mon travail, pour leurs expertise et conseils.

Je remercie le KIC Climat (Knowledge and Innovation Community) de me soutenir financièrement durant les trois ans de cette thèse et de me donner l'opportunité d'étendre mon réseau de contacts avec d'autres scientifiques, non seulement en Europe, mais aussi au Canada (où j'ai passé deux mois et pu connaître Tom Gleeson). Je remercie l'Agence Nationale de la Recherche et le programme LEFE/INSU du CNRS pour le financement de ce projet. Je remercie aussi l'IDRIS (Institut du Développement et des Ressources en Informatique Scientifique) pour les simulations avec le modèle ORCHIDEE.

Merci à Tom Gleeson, avec qui j'ai eu beaucoup de discussions sur la densité de drainage et la constante de temps de l'écoulement de base. Il m'a montré qu'on peut faire de la recherche d'une manière décontractée, sans arrêter de penser aux questions qui nous motivent : pourquoi ça ? Pourquoi pas d'une autre manière ?

Merci à tous les chercheurs, doctorants et ingénieurs qui m'ont aidée au cours de cette thèse. Un grand merci à Philippe Marty, qui m'a accueillie au troisième étage, qui m'a aidé à déchiffrer mes erreurs en Fortran et qui fait un très bon barbecue. Merci pour avoir accueilli ma petite famille (merci à Marianne aussi !) et pour avoir partagé de très bons moments avec nous. Merci à Matthieu Guimberteau pour avoir été toujours disponible, pour les discussions, pour me donner des idées quand j'étais bloquée et pour m'avoir appris à utiliser ferret. Merci à Trung Nguyen Quang pour sa gentillesse et pour les discussions toujours productives. Merci aussi à Jan Polcher d'avoir partagé ses scripts qui nous ont énormément aidées et pour avoir répondu à nos questions toujours en un instant. Merci à Ardalan Tootchi pour les discussions enrichissantes sur l'hydrologie, le climat et le développement des bébés. Merci aussi pour dire que mon bébé est "cute". Merci aussi à Marie Silvestre qui était toujours disponible quand ArcGIS a décidé de ne pas marcher. Merci à Raphaël Fauches et Nejla Aissa Grouz qui m'ont accueilli au quatrième étage et m'ont fait me sentir chez moi. Merci pour de m'avoir écoutée, pour le partage, pour l'envoi de liens inutiles, pour les croissants le matin et m'avoir tenu compagnie les week-ends. Merci pour l'amitié que vous m'avez accordée. Vous êtes top !

Merci à l'équipe du laboratoire METIS qui m'a accueillie pendant ma thèse : Nadine Bost (toujours avec un sourire, merci !), Valérie Girard (toujours patiente pour m'expliquer les démarches administratives), Jean-Robert Gowe, Pierre Ribstein, Jean-Marie Mouchel, Damien Jougnot, Christophe Bonnet, Ludovic Oudin, Fayçal Rejiba, Danièle Valdes-Lao (merci de la part de Sarah aussi !), Thomas Verbeke, Bénédicte Sabatier, Antsiva Ramarson, Bahar Salvati, Florence Habets, Aurélien Baro, Justine Tirogo, Tatiana Dmitrieva, Marine Dangeard, Sylvain Théry, Vincent Thieu, Cyril Schamper, Alain Tabbagh, Julien Thiesson, Roger Guérin, Feras Abdulsamad... et tous ceux qui font partie de l'équipe mais ne sont listés ici.

Un grand merci à ma famille : mon père qui m'a donné le goût pour la science et qui m'a toujours soutenue dans mon odyssée. Mes frères Márcio et Carlos, qui ont été présents même de loin. En dernier, mais pas le moindre, je voudrais remercier Amin pour être mon ami, mon mari et pour m'avoir donné le meilleur cadeau de ma vie : ma petite princesse Sarah, qui a été ma motivation pour continuer et finir cette thèse. Je ne peux pas oublier de dire un grand merci aussi pour tous les soirs où tu l'aurait été gardée lorsque j'ai dû rester tard pour rédiger ce manuscrit. Sans ton aide, la thèse aurait été trop dure.

RÉSUMÉ

Estimation de la constante de temps du débit de base pour applications à l'échelle globale

Le débit de base est la décharge naturelle des aquifères vers les rivières (Younger, 2009), ou elle peut être définie comme la partie du flux d'écoulement qui est submergée par la décharge des eaux souterraines (Carter *et al.*, 2015). Le débit de base peut être la seule source d'eau pour les rivières, étant un élément important dans la gestion de l'eau, et une variable clé dans les modèles de surface (LSM). La quantification du débit de base dépend de l'estimation, car il n'existe actuellement aucune technique pour le mesurer directement. Les LSM à grande échelle estiment le débit de base en utilisant deux approches principales : en utilisant une solution de l'équation de diffusivité ou à partir d'un réservoir linéaire unidimensionnel, dans lequel le débit de base est estimé par le rapport du volume du réservoir et d'une constante de temps de débit de base (τ) qui représente un temps de résidence. L'équation de diffusivité n'est pas linéaire, ce qui entraîne plusieurs méthodes basées sur différentes hypothèses pour arriver à une solution. Cependant, l'équation de diffusivité peut être résolue numériquement, dans laquelle le débit de base dépend des données hydrogéologiques. Les solutions numériques exigent des ordinateurs puissants, une demande qui augmente avec des données à grande échelle. Les modèles de surface qui utilisent un réservoir linéaire dépendent des données d'observation pour calibrer τ. Cependant, à grande échelle, les données d'observation sont rares et la calibration devient difficile. Pour surmonter le manque de données observées, une pratique courante consiste à extrapoler les valeurs d'une région calibrée, comme dans Gascoin *et al.* (2009) qui est arrivé à une constante de temps de flux de base de 700 jours. Le

problème de l'extrapolation des valeurs calibrées dans de grandes zones est qu'il ne tient pas compte des hétérogénéités hydrogéologiques, comme le fait l'équation de diffusivité. Une autre approche est d'utiliser des solutions analytiques disponibles pour certaines conditions limites, comme dans Brutsaert (2005) qui a proposé une équation linéaire pour estimer le flux de base dérivé de l'équation de Boussinesq (Boussinesq, 1877) pour des aquifères horizontaux, isotropes et homogènes. Dans ce cas, τ dépend de la porosité efficace, de la transmissivité de l'aquifère et de la densité de drainage di bassin.

L'utilisation d'une solution analytique pour estimer la constante de temps du débit de base fait le lien entre les approches simples (en considérant un réservoir linéaire) et complexe (en résolvant numériquement l'équation de diffusivité) pour l'estimation du débit de base. À partir de l'équation proposé par Brutsaert (2005) et des données à l'échelle globale de porosité efficace (données GLHYMPS de porosité total, comme une approximation), densité de drainage (calculée avec le réseau HydroSHEDS) et transmissivité (en utilisant la perméabilité de GLHYMPS pour obtenir la conductivité hydraulique et un épaisseur mouillé constante de 35 m) τ a été estimée à l'échelle globale. Cependant, τ a été surestimée de cinq ordres de grandeur en moyenne globale par rapport à des valeurs de littérature d'analyse de récession obtenues de données observées de débit de fleuves (Zecharias and Brutsaert, 1988; Eng and Milly, 2007; Brutsaert, 2008; Thomas *et al.*, 2013; Stoelzle *et al.*, 2013) et de calibration de modèles de surface avec un réservoir linéaire (Gascoin *et al.*, 2009). Cette surestimation indique que l'utilisation de données globales actuellement disponibles ne convient pas pour estimer τ dans des LSM avec des schémas simples d'eaux souterraines (réservoir linéaire), en soulevant une question : Peut-on améliorer l'estimation de la constante de temps du débit de base à l'échelle globale pour obtenir une meilleure dynamique de débit de base dans des LSM à grande échelle ?

Cette thèse aide à répondre à cette question et elle est divisée en cinq chapitres. Les chapitres deux et trois aident à trouver quels paramètres exercent un fort contrôle sur τ. Le chapitre deux montre une évaluation de la densité de drainage, car c'est l'un des paramètres qui contrôlent τ. Afin d'améliorer la représentation des flux d'eau souterrains et de surface, les modèles de surface à l'échelle globale reposent de plus en plus sur des paramètres à haute résolution. Plusieurs études se sont notamment concentrées sur le développement de modèles numériques de terrain (MNT) à haute résolution, hydrologiquement conditionnés pour améliorer la localisation des cours d'eau. Dans ce cadre, les cours d'eau sont définis comme l'ensemble des pixels rivières, eux-mêmes couramment définis comme ayant une aire contributive amont supérieure à une aire contributive seuil (ACS), nécessaire pour déclencher l'initialisation d'un cours d'eau. A l'échelle globale ou continentale, il est très fréquemment utilisé une valeur unique de l'ACS, ce qui rend compte du contrôle de premier ordre de la topographie sur le réseau de drainage, sa longueur et la densité de drainage résultante, mais néglige les effets de la lithologie et du climat locaux, alors qu'ils peuvent être importants. Ce chapitre montre une calibration de l'ACS basée sur les données de lithologie et précipitations globales. La base de données HydroSHEDS a été utilisé pour le MNT, qui propose des aires contributives aux résolutions de 500 m et 1 km, et définit en standard les cours d'eau avec un seuil de 100 pixels, donc des ACS de 25 et 100 km^2 respectivement. Pour s'affranchir de cette dépendance d'échelle, et obtenir un meilleur réalisme des réseaux hydrographiques, les ACS ont été calibrés pour correspondre localement à des réseaux de drainage observés de bonne qualité (à partir des bases de données nationales franÃ§aises et australiennes), en distinguant plusieurs classes de précipitation et de lithologie. Les seuils calibrés sont ensuite utilisés pour l'extraction du réseau de drainage mondial. Toutes les valeurs de seuil restent inférieures à 5 km^2, et les valeurs les plus élevées, qui correspondent à de faibles densités de drainage, sont dans des régions de roches carbonatées et de sédiments non consolidés, et/ou de faibles précipitations (climat arides à semi-arides). Les résultats sont obtenus à 500 m de résolution et synthétisés en termes de densités de drainage dans une grille de 7,5' (surface moyenne de 150 km^2). Celles-ci ont le même ordre de grandeur que les densités de drainage observées, avec une erreur inférieure à 1%. La validation est présentée à partir des données nationales de réseau de drainage des états-Unis, du Brésil. Le chapitre deux a été publié sous la forme d'un article dans Geophysical Research Letters (Schneider *et al.*, 2017), et il a résulté en un jeux de données globales appelé GRIN (Global

River Network) composé d'une densité de drainage (à 7,5') et d'un réseau à une échelle de 1 : 50 000 (appelé LCS), disponibles gratuitement pour le téléchargement à `http://www.metis.upmc.fr/en/node/211`.

Le chapitre trois montre une analyse de sensibilité de τ à tous les paramètres utilisés sur son estimation (porosité efficace, transmissivité, densité de drainage et pente pour une formulation qui considère un aquifère en pente), par l'utilisation de données différentes à l'échelle mondiale pour chaque paramètre. τ a été estimée en utilisant une formulation pour des aquifères libres, isotropes, homogènes et en pente à partir d'une solution de l'équation de Boussinesq linéarisée à 0,5° et 7,5'. Une analyse de sensibilité a montré que les principaux paramètres qui contrôlent τ sont la transmissivité et la densité de drainage. L'utilisation d'une formulation qui considère l'aquifère en pente a réduit les valeurs de τ. Cet effet est plus fort à plus haute résolution (vu que les pentes ont une gamme de valeurs plus large) et les valeurs sont plus proches de valeurs de référence. L'utilisation des valeurs de porosité efficace au lieu de porosité totale ont réduit les valeurs de τ. Cependant, les différentes valeurs testées de porosité, épaisseur mouillé et pente n'ont pas un fort impact sur la moyenne ou médiane de τ. L'utilisation du réseau LCS (chapitre deux) a réduit d'une ordre de grandeur les valeurs de τ. Malgré la sensibilité de δ à τ, la transmissivité peut avoir un impact plus fort sur les valeurs de τ expliqué par le fait que la gamme des valeurs de transmissivité est plus grande (varie de plusieurs ordres de grandeurs) que la gamme de δ. La combinaison de paramètres qui résulte en valeurs plus proches à des valeurs de référence est $\tau_{UG/10}$, qui utilise la porosité efficace de de Graaf *et al.* (2015) divisée par 10 (qui assume une surestimation de la porosité efficace par des mesures faites en laboratoire (Brutsaert, 2008)), la densité de drainage LCS (Schneider *et al.*, 2017) et une transmissivité calculée avec un épaisseur mouillé constante de 35 m et une conductivités hydrauliques moyenne obtenu à partir d'une combinaison de données de sol Reynolds *et al.* (2000) et GLHYMPS (Gleeson *et al.*, 2014).

Le chapitre quatre répond à la question : La constante de temps de base calculée à partir des propriétés hydrogéologiques et topographiques améliore-t-elle les simulations de décharge de la rivière des LSM avec un réservoir linéaire ? Les τ

estimées au chapitre trois sont utilisées dans le modèle ORCHIDEE. Le chapitre quatre montre une évaluation des impacts d'une échelle de temps plus longue dans le réservoir d'eau souterraine d'ORCHIDEE aux débits simulés, accompagné d'une évaluation de la sensibilité d'ORCHIDEE à τ. Une comparaison des données de débit observées GRDC et des résultats de simulation de référence (τ_{ORC}) a montré que la τ d'ORCHIDEE résulte en des résultats de débit biaisés (positivement et négativement), avec seulement 23% des 276 stations évaluées avec un coefficient de corrélation supérieur à 0,75, mais avec 64% de ces stations ont un coefficient Nash-Sutcliffe supérieur à 0,5. L'utilisation d'une constant de temps plus longue ($\tau_{UG/10}$) a amélioré le coefficient de corrélation de seulement 10% des stations et Nash-Sutcliffe de 18% des stations évaluées, mais pour 80% des stations le débit a été dégradé par l'utilisation d'une τ plus longue. Les anomalies totaux de stock d'eau de la simulation de référence résultent en un coefficient de détermination supérieur à 0,95 pour la majorité des 19 grandes bassin versants évalués par rapport aux observations de GRACE (Swenson, 2012).

Enfin, dans les conclusions, on montre que la τ proposée ne convient pas à l'utilisation dans un modèle de surface à réservoir linéaire peu profond tel qu'ORCHIDEE en raison des valeurs élevées calculées (discutées dans le chapitre 3) qui ne sont pas représentatives du flux local dominant et dégradent les résultats de débit simulés. Dans ce chapitre, on propose l'utilisation de différents jeux de données qui ont une porosité efficace plus faible, une transmissivité plus élevée et une densité de drainage plus élevée qui pourrait donner des τ qui représentent des interactions rapides entre les eaux souterraines et les eaux de surface. Comme ces données ne sont pas actuellement disponibles à l'échelle globale, l'amélioration des jeux des données est une perspective à long-terme de cette thèse. L'utilisation de différentes hypothèses (comme réservoir non linéaire et anisotrope) pourrait également améliorer τ en perspective à court terme, en introduisant des résultats et des formulations plus adéquats pour les simulations des eaux souterraines.

CONTENTS

1 Introduction **1**
1.1 Groundwater: concepts and definitions . . . 1
1.2 Base flow . . . 4
1.3 Aquifer properties . . . 7
1.4 Base flow time constant . . . 12
1.5 Groundwater schemes and base flow modeling in LSMs . . . 13
1.6 Objectives . . . 18

2 Global river network extraction **21**
2.1 Introduction . . . 21
2.2 Data description . . . 24
2.3 A_{cr} Calibration and Model Selection . . . 30
2.4 Model Results and Discussion . . . 37
2.5 Conclusions . . . 42

3 Estimation of the base flow time constant **45**
3.1 Introduction . . . 46
3.2 Data description and selection of parameters . . . 50
3.3 Base flow time constant results and discussion . . . 63
3.4 Conclusions . . . 68

4 Base flow time constant in ORCHIDEE **71**
4.1 The ORCHIDEE model . . . 72
4.2 Simulation set-up . . . 79
4.3 Validation data . . . 83
4.4 Simulation results . . . 87
4.5 Conclusions . . . 99

5 Conclusions and perspectives **103**

Bibliography **107**

6 Appendix **119**
6.1 Article published in Geophysical Research Letters . . . 121
6.2 Drainage density calibration results . . . 131
6.3 A_{cr} values for LC and LCS models . . . 135
6.4 Drainage density maps for FAO data and LC model . . . 139
6.5 Descriptive statistics of τ . . . 141
6.6 Statistical coefficients of TWS anomalies . . . 143

LIST OF FIGURES

1.1 Water cycle and the global mean storage (10^3 km^3) and annual fluxes (10^3 km^3.yr^{-1}). Adapted from Shiklomanov and Rodda (2003) and Trenberth *et al.* (2007). 2

1.2 Scheme for confined and unconfined aquifers. Adapted from Dingman (2015). 4

1.3 Groundwater and stream interactions, depending on water table depth. a) Groundwater flow towards stream; b) Stream flow towards groundwater; c) Stream is disconnected from groundwater; d) Groundwater flow towards streams in humid regions; d) Groundwater flow towards streams in arid regions. Adapted from Barlow and Leake (2012) (a, b and c) and Maxwell *et al.* (2015) (d and e). 12

1.4 Estimated base flow time constant (τ_{ini}) estimated at 0.5° resolution with HydroSHEDS and GLHYMPS data. The bottom and top of the box-plot represent first and third quartiles, the middle bar gives the median, and red diamond indicates the mean value. Whiskers show the 10th and 90th percentiles. Points represent mean values from the literature for the simple approach (linear reservoir). 14

2.1 Environmental classes evaluated in drainage density calibration. a) All lithology classes from Hartmann and Moosdorf (2012) (Pi = Intermediate plutonic rocks; Mt = Metamorphic rocks; Va = Acid volcanic rocks; Ss = Siliciclastic sedimentary rocks; Vi = Intermediate volcanic rocks; Pb = Basic plutonic rocks; Pa= Acid plutonic rocks; Vb = Basic volcanic rocks; Py = Pyroclastics; Sm = Mixed sedimentary rocks; Su = Unconsolidated sediments; Sc = Carbonate sedimentary rocks); b) Climate classes from Weedon *et al.* (2014); c) Slope classes calculated from HydroSHEDS elevation. 25

2.2 Zoom over Australian areas of Geofabric data that have high drainage densities in arid regions. A) Drainage density map viewed in Google Earth over Lake Eyre (north). B) Google Earth view without the drainage density layer. C) Drainage density map viewed in Google Earth over Lakes Eyre (south), Torrens, Frome, and Gairdner. D) Google Earth view without the drainage density layer. 27

2.3 Zoom over the French river network region with unnatural channels in an irrigated zone (Landes area). 27

2.4 Zoom over the United States river network region with unnatural channels in an agricultural zone west of San Francisco. 28

2.5 Zoom over the Brazilian river network region with variable resolution problems. 29

2.6 A_{cr} calibration results in Australia (red) and France (blue) per environmental class. a) and b) represent the lithology, c) and d) climate, and e) and f) slope classes. The bottom and top of the boxplots represent the first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles. 32

2.7 A_{cr} map converted to pixels to extract the river network from a) LC and b) LCS models. 35

2.8 Location of pixels with intermittent rivers (in blue) over Australia from: a) Geofabric, b) FAO; c) LCS with 400 mm.y^{-1} precipitation threshold ; d) LCS with 500 mm.y^{-1} precipitation threshold; e) LCS with 600 mm.y^{-1} precipitation threshold. See complementary analysis in Table 2.2. 36

2.9 Maps of drainage density (km^{-1}) for the reference river networks and the evaluated models in France (a,b,c), Australia (d,e,f), US (g,h,i), and Brazil (j,k,l): a,d,g,j) Reference data; b,e,h,k) LC; c,f,i,l) LCS. . . 37

2.10 Maps of drainage density (km^{-1}) over the US: a) Reference; b) Luo *et al.* (2016); c) LCS - Reference; d) LCS - Luo *et al.* (2016). 39

2.11 Maps of drainage density (km^{-1}) over Europe for the LCS model (left) and CCM2 (Vogt *et al.*, 2007) (right). 40

2.12 Global maps of drainage density (km^{-1}) for LCS model: a) for full network including intermittent streams; b) intermittent streams only. Percentage of intermittent streams: c) LCS; d) FAO. In each 7.5' cell, the % of intermittent streams is the ratio of the intermittent stream length to total stream length. The violet color north of 60°N indicates where HydroSHEDS data are missing. 41

3.1 Aquifer scheme proposed by a) Brutsaert (2005) and b) Zecharias and Brutsaert (1988). WT = water table; D = aquifer depth; B = aquifer length; h = water table elevation; θ = aquifer slope; z and x are the vertical and horizontal axes, respectively. 47

3.2 Total (white) and effective (gray) porosity results at 0.5° (blue) and 7.5' (red) resolution. The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles. 52

3.3 WHYMAP groundwater classification. R represents the groundwater recharge (in mm.yr^{-1}). Data from BGR and UNESCO (2008). 55

3.4 Aquifer scheme used on transmissivity estimation, assuming a linear, unconfined, homogeneous isotropic aquifer. Transmissivity is given by the product of a) saturated depth (e) and three hydraulic conductivity data (K): b) using GLHYMPS data, c) combining USDA and GLHYMPS data, and d) using USDA data. WT = water table, T = transmissivity, K = hydraulic conductivity, e = saturated thickness, p = linearization constant, D = aquifer depth, B = aquifer length, θ = slope. 56

3.5 Transmissivity results (m^2.s^{-1}) at 0.5° (blue) and 7.5' (red) resolution. GLHYMPS, USDA, and GLHYMPS-USDA represent the hydraulic conductivities, with saturated thickness 3 and 35 m for constant thickness, (7,17,35) and (17, 35, 69) m for (shallow, complex hydrogeological structures,major groundwater basins). The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles. 59

3.6 Slope results (%) at 0.5° (blue) and 7.5' (red) resolutions. Topography represents the slope calculated using HydroSHEDS hydrologically corrected elevations (Lehner *et al.*, 2008), and water table the slope derived from Fan *et al.* (2013) water table depth constrained by observations. The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles. 60

3.7 Drainage density results (δ, in km^{-1}) at 0.5° (blue) and 7.5' (red) resolution. White area shows δ from HydroSHEDS data (Lehner *et al.*, 2008), and gray region δ with LCS river network (Schneider *et al.*, 2017). The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles. 61

3.8 Base flow time constant results at global scale using parameters of Table 3.6 (in years) at 0.5°. Pink region represents a reference range from literature presented in Table 3.1 (2 to 342 days = 0.005 to 0.94 years). The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles. 64

3.9 Comparison of τ results (years) at the basin scale for τ_{ne}, τ_{UG}, and $\tau_{UG/10}$. 65

3.10 Comparison of the effect of drainage density and hydraulic conductivity on τ. Red represents the use of $K_{GLHYMPS}$ (τ_{ne}) and blue $K_{USDA-GLHYMPS}$ (τ_{UG}), both at 0.5° resolution. 66

3.11 Base flow time constant at 0.5° resolution for τ_{UG} (effective porosity, LCS river network, 35 m saturated depth, and $K_{USDA-GLHYMPS}$). Gray regions are no data values resultant from the HydroSHEDS DEM used on the river network extraction. 67

4.1 ORCHIDEE general scheme. Adapted from ORCHIDEE's training course. 74

4.2 Flow directions used on the routing scheme with a zoom over the Obidos station at the Amazon basin. Blue to red colors represent increasing upstream area of the Amazon basin. Green star show the position of the Obidos station. Green dot represents the point of river discharge into the ocean. Blue dots represent coastal flow. 76

4.3 Reservoirs of ORCHIDEE's routing scheme. P represents precipitation, ET evapotranspiration, SR surface runoff, and D drainage. Adapted from Guimberteau (2014). 77

4.4 Slow reservoir of ORCHIDEE using the standard initialization, with $\tau_{slow} = 10\tau_{ORC}$. 82

4.5 GRDC stations classified by time series length GRDC (2014). 86

4.6 Comparison of the global average results (1985-2010) for τ_{ORC} to the validation data for precipitation (blue), evaporation (red), and runoff (green), in mm.d^{-1}. The bars represent the reference data, and the lines the simulated results. 87

4.7 Evaluation of the simulated river discharge of τ_{ORC} (red) against observed river discharge from GRDC (black), between 1985-2010. Shaded areas represent values within ±1 standard deviation of the observed period. 88

4.8 Relative bias (%) calculated for the 276 GRDC stations that have an upstream area error under 5% compared to the ORCHIDEE upstream area. 89

4.9 Nash-Sutcliffe coefficient (-) on the top and correlation coefficients (-) on the bottom calculated for the 276 GRDC stations that have an upstream area error under 5% compared to the ORCHIDEE upstream area. 90

4.10 Total Water Storage anomalies (TWS) in mm (y axis) for 19 major watersheds in the world. Black lines represent GRACE data (Swenson, 2012), and red lines the reference simulation. Months (x axis) represent the averages for 2002 to 2010 period. No data values were excluded of the analysis. 91

4.11 Evaluation of the simulated river discharge of τ_{ORC} (red), τ_{1y} (green), τ_{10y} (cyan), and τ_{100y} (pink) against observed river discharge from GRDC (black), between 1985-2010. Shaded areas represent values within ±1 standard deviation of the observed discharge. 92

4.12 Evaluation of the simulated river discharge of τ_{ORC} (red), τ_{ORCx3} (green), $\tau_{UG/10}$ (cyan), τ_{UG} (pink), and τ_1 against observed river discharge from GRDC (black), between 1985-2010. Shaded areas represent values within ±1 standard deviation of the observed discharge. 93

4.13 Scatter plot of Nash-Sutcliffe coefficients, bias, and correlation coefficients calculated for river discharge measurements from GRDC and simulations with ORCHIDEE of τ_{ORC} versus τ_{ne}, τ_{UG}, $\tau_{UG/10}$. 94

4.14 Scatter plot of the mean drainage (mm.d^{-1}) versus river discharge range (m^3.s^{-1}) per basin (indicated by their first four letters). Range is given by the difference between maximum and minimum value of the river discharge (Q) shown in Figure 4.6 for the τ_{ORC} simulation, and drainage (D) is the mean value for the same time period (1985-2010). 96
4.15 Slow reservoir anomalies for τ_{ORC} (red), τ_{ORCx3} (green), $\tau_{UG/10}$ (cyan), τ_{UG} (pink), and τ_{ne} (orange) simulations. 97
4.16 TWS anomalies (mm) for τ_{ORC} (red), $\tau_{UG/10}$ (pink), τ_{UG} (cyan), and τ_{ne} (green) simulations and observations from GRACE (black). 98
4.17 Aquifer response rate (day^{-1}) versus reduction factor (%) for τ_{ne} (orange), τ_{UG} (pink), and $\tau_{UG/10}$ (cyan) for the evaluated watersheds. The black curve is from Erskine and Papaioannou (1997). Top values represent the base flow time constant. 100

6.1 Drainage density (km^{-1}) for a) LC model and b) FAO data. 139

LIST OF TABLES

1.1 Mean residence times (MRT) and water stocks (10^3km^3) per water cycle component. Data from Shiklomanov and Rodda (2003). 3

1.2 Porosities, specific yields, and specific retention values per geologic materials (Dingman, 2015). Values were measured by Morris and Johnson (1967). 8

1.3 Representative values of hydraulic conductivities for unconsolidated sedimentary materials, sedimentary rocks, and crystalline rocks. Adapted from Domenico and Schwartz (1990). 10

1.4 List of groundwater schemes recently introduced in LSMs. 16

2.1 Selected environmental parameters and their coverage area (%) used for A_{cr} calibration (all lithology, precipitation, and slope evaluated classes). Area was calculated over regions with available data (No Data values were removed from the analysis). 30

2.2 Intermittency analysis. Spatial correlations (Corr. to FAO) are defined based on the presence/absence of intermittent streams in the 7.5' pixel 36

2.3 Statistical parameters for the comparison of the calculated and reference drainage densities δ per country^1. Ref. $= \delta$ from river network reference data (AU = Geofabric and FR = CARTHAGE used for calibration, US = NHD and BR = IBGE (2015) used for validation); Others $= \delta$ from Luo *et al.* (2016) over the US and from CCM2 (Vogt *et al.*, 2007) over France and Europe; Mean = mean δ (km^{-1}); Std.Dev. = standard deviation of δ (km^{-1}); Bias = bias between model and Ref (km^{-1}); Corr. = correlation coefficient between model and Ref (dimensionless); RMSE = root mean square error between model and Ref (km^{-1}). % Int. corresponds to the ratio of total length of intermittent rivers to the total river length. 38

3.1 Summary of base flow time constant (τ) range (days) selected from the literature for recession analysis studies using graphical method. * Analytical method derived from the Boussinesq equation. ** Model calibration results. 49

3.2 Porosity values found in the literature classified by hydrolithology classes of Gleeson *et al.* (2014). Lithology classes come from Hartmann and Moosdorf (2012) (Pi = Intermediate plutonic rocks; Mt = Metamorphic rocks; Va = Acid volcanic rocks; Ss = Siliciclastic sedimentary rocks; Vi = Intermediate volcanic rocks; Pb = Basic plutonic rocks; Pa = Acid plutonic rocks; Vb = Basic volcanic rocks; Py = Pyroclastics; Sm = Mixed sedimentary rocks; Su = Unconsolidated sediments; Sc = Carbonate sedimentary rocks; Ev = Evaporites; Wb = Water Bodies; Ig = Ice and Glaciers; Nd = No Data). Unconsolidated and siliciclastic sedimentary represent the average values of c.g. (coarse grained) and f.g. (fine grained). 51
3.3 Descriptive statistics results of saturated thicknesses (e, in m) used on τ estimation, from aquifer depths (D, in m), at 7.5' and 0.5° resolution. 56
3.4 GLHYMPS permeability (k in m^2) data (Gleeson *et al.*, 2014) and calculated hydraulic conductivity ($K_{GLHYMPS}$ in m.s^{-1}) per hydrolithology class (c.g. = coarse grained and f.g. = fine grained). 58
3.5 Hydraulic conductivities (m.s^{-1}) from Carsel and Parrish (1988) per soil texture of Reynolds *et al.* (2000). Adapted from d'Orgeval (2006). 59
3.6 Summary of the data used to test the sensitivity of τ to different parameters. n = total porosity, n_e = effective porosity, K = hydraulic conductivity, e = saturated thickness, δ = drainage density, θ = slope, HS = HydroSHEDS. 62

4.1 Descriptive statistic result for ORCHIDEE's time constant per reservoir (in days), at 0.5° resolution 78
4.2 Meteorological variables corrected according to variable for WFDEI files (Weedon *et al.*, 2014). 79
4.3 Summary of the simulations performed with ORCHIDEE. The initialization Std is the ORCHIDEE standard (zero initial volume, indicated by Median V_{slow} (mm)). 81
4.4 Summary of the datasets used for validation of ORCHIDEE's results. P = total precipitation; E = total evapotranspiration; R = total runoff; Q = river discharge; TWS = total water storage anomalies; O = observed/literature values; M = model results; MM = multi-model results; * = global averages are calculated with the provided dataset for the whole time period. 86
4.5 Mean slope (-) and τ (years) per watershed for the evaluated GRDC stations. τ_{1y}, τ_{10y}, and τ_{100y} were not represented since they are constant for all basins, equals 1, 10 and 100 years, respectively. 95

6.1 Descriptive statistics for the evaluated τ (in years). Gray rows represent results at 0.5°, and white rows at 7.5'. 141
6.2 Coefficient of determination (R^2) and Nash-Sutcliffe coefficient (Nash) for TWS anomalies (mm) of the evaluated watersheds. 143

LIST OF SYMBOLS

α	Equation parameter (m^{-1})
δ	Drainage density ($m.m{-}2$)
η_e	Effective porosity ($-$)
τ	Time constant ($days$)
τ_i	Time constant of each reservoir in ORCHIDEE's routing scheme (d)
θ	Volumetric water content ($m^3.m^{-3}$)
θ_f	Relative humidity ($-$)
θ_r	Residual water content ($m^3.m^{-3}$)
θ_s	Saturated water content ($m^3.m^{-3}$)
A	Basin area ($m2$)
D	Drainage ($m.s^{-1}$)
d	Flow distance (km)
D_i	Diffusivity ($m^2.s^{-1}$)
E	Total evapotranspiration ($m.s^{-1}$)
F	Drainage coefficient ($0 \leq F \leq 1$)
f	Irrigated fraction (-)
g_i	Property constant of each reservoir in ORCHIDEE's routing scheme ($d.km^{-1}$)
H_s	Saturated hight in the beginning of the recession (m)
I_{req}	Irrigation requirement ($kg.m^{-2}.s^{-1}$)
K	Hydraulic conductivity ($m.s^{-1}$)
k	Topographic index (km)

K_s	Saturate hydraulic conductivity $(m.s^{-1})$
$K_s(z)$	Hydraulic conductivity at a z depth $(mm.d^{-1})$
L	River length (m)
n	Equation parameter $(-)$
o_i	Observed results of the time i for each location (x, y)
P	Total precipitation reaching the soil $(m.s^{-1})$
p	Empiric adjustment coefficient $(-)$
q	Flux field $(m.s^{-1})$
Q_i	Flow of each reservoir in ORCHIDEE's routing scheme $(kg.d^{-1})$
R	Runoff $(m.s^{-1})$
R_{tot}	Total runoff $(kg.m^{-2}.s^{-1})$
S	Slope (-)
s	Transpiration $(m^3.m^{-3}.s^{-1})$
s_i	Simulated results of the time i for each location (x, y)
t	Time (s)
t	Time step of ORCHIDEE's routing scheme (d)
T_e	Effective hydraulic transmissivity $(m^2.s^{-1})$
T_{pot}	Potential transpiration $(kg.m^{-2}.s^{-1})$
V_i	Volume of water inside the reservoir in ORCHIDEE's routing scheme (kg)
W	Total water content (m)
z	Soil depth below the surface (m)
z_s	Characteristic depth of exponential decay for K_s (m)
z_{lim}	Limit depth (m)
${K_s}^{CP}$	Saturate hydraulic conductivity from Carsel and Parrish (1988) $(mm.d^{-1})$

LIST OF ACRONYMS

ALMA	Assistance for Land-surface Modeling Activities
CRU	Climate Research Unit
CWRR	Center of Water Resources Research
DEM	Digital Elevation Model
DRT	Dominant River Tracing
EPA	United States Environmental Protection Agency
FAO	Food and Agriculture Organization of the United Nations
GCM	General Circulation Model
GIS	Geographical Information System
GLEAM	Global Land surface Evaporation: the Amsterdam Methodology
GLOWABO	GLObal WAter BOdies database
GLWD	Global Lakes and Wetlands Database
GPCC	Global Precipitation Climatology Center
GRACE	Gravity Recovery and Climate Experiment
GTOPO30	Global 30 arc-second elevation
HydroSHEDS	Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales
IBGE	Brazilian Institute of Geography and Statistics
IGN	L'Institut National de l'information géographique et forestière
IPSL	Institut Pierre Simon Laplace
LMD	Laboratoire de Météorologie Dynamique
LPJ	Lund-Potsdam-Jena Dynamic Global Vegetation model
LSM	Land Surface Model
MCRO	World Map of Carbonate Rock Outcrops
MPI	Max Plank Institute
NCEP/DOE	National Centers for Environmental Prediction / Department Of Energy
NHD	National Hydrography Dataset
NHDPlus	National Hydrography Dataset Plus
ORCHIDEE	ORganizing Carbon and Hydrology In Dynamics EcosystEms
PFT	Plant Functional Type
SECHIBA	Schématisation des ÉChanges Hydriques à l'Interface entre la Biosphère et l'Atmosphère
SRTM	Shuttle Radar Topography Mission

STOMATE	Saclay Toulouse Orsay Model for the Analysis of Terrestrial Ecosystems
TRIP	Total Runoff Integrating Pathways
TWS	Terrestrial Water Storage
USGS	United States Geological Survey
WATCH	WATer and global CHange
WBM	Water Balance Model
WFDEI	WATCH Forcing Data methodology applied to ERA-Interim reanalysis data
WGHM	WaterGAP Global Hydrological Model
WHYMAP	World-wide Hydrogeological Mapping and Assessment Programme
WT	Water Table
WTD	Water Table Depth

CHAPTER 1

INTRODUCTION

Contents

1.1 Groundwater: concepts and definitions 1

1.2 Base flow 4

1.3 Aquifer properties 7

1.3.1 Porosity and specific yield 7

1.3.2 Transmissivity 9

1.3.3 Aquifer length 10

1.4 Base flow time constant 12

1.5 Groundwater schemes and base flow modeling in LSMs 13

1.6 Objectives 18

1.1 Groundwater: concepts and definitions

The water availability *per capita* is reducing over the past decades, which increases a general concern for global water resources with growing population (Grafton *et al.*, 2012). Fresh water increasing consumption *versus* low availability draws attention to groundwater, as it is the most abundant source of available fresh water in the world, and it is the source of one third of global fresh water withdrawals (Döll *et al.*, 2012). Groundwater is connected to surface water, and it can interact with the atmosphere from what is called the water cycle.

The water cycle represents the water movement on Earth (Figure 1.1), in which water travels between solid, liquid, and vapor form through the main compartments:

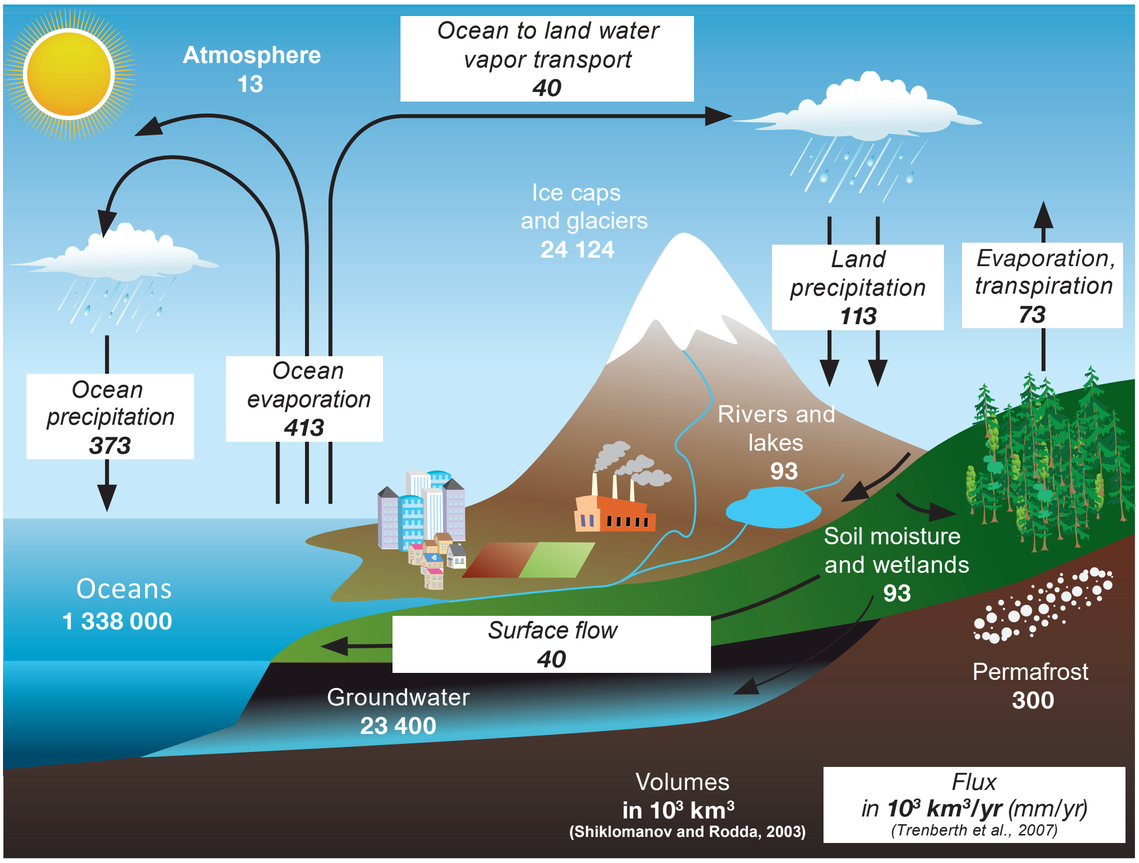

Atmosphere
13
Ocean to land water vapor transport
40
Ice caps and glaciers
24 124
Land precipitation
113
Evaporation, transpiration
73
Ocean precipitation
373
Ocean evaporation
413
Rivers and lakes
93
Soil moisture and wetlands
93
Oceans
1 338 000
Surface flow
40
Groundwater
23 400
Permafrost
300
Volumes
in 10³ km³
(Shiklomanov and Rodda, 2003)
Flux
in 10³ km³/yr (mm/yr)
(Trenberth et al., 2007)

Figure 1.1 – Water cycle and the global mean storage (10^3 km^3) and annual fluxes (10^3 km^3.yr^{-1}). Adapted from Shiklomanov and Rodda (2003) and Trenberth *et al.* (2007).

the atmosphere, lithosphere, biosphere, hydrosphere, and cryosphere. Evapotranspiration (from oceans, rivers, soils, transpiration of plants, interception, and ice and snow sublimation) and precipitation compose the vertical movement of the water cycle. Over land the precipitation exceeds evapotranspiration, while over oceans is the inverse. This disequilibrium results in two main horizontal fluxes: advection in the atmosphere (from oceans towards land), and runoff in the lithosphere (from land to oceans) (Ducharne and Laval, 2015). Although different compartments of the water cycle store large volumes of water, the water is not always available where it is needed. Most of the world's water is salty or frozen (Table 1.1), being unsuitable for agricultural, industrial, or domestic uses due to the high economic or environmental costs[1].

---

[1] Economic costs are defined by the cost in money to perform an activity. Environmental costs are defined as the actual or potential deterioration of natural assets by an activity (UNSD, 2001).

Table 1.1 – Mean residence times (MRT) and water stocks (10^3km^3) per water cycle component. Data from Shiklomanov and Rodda (2003).

Water of hydrosphere		MRT	Volume (10^3km^3)
Cryosphere	Permafrost	10,000 years	300
	Ice caps	9,700 years	24,000
	Mountain glaciers	1,600 years	124
Oceans		2,500 years	1,338,000
Groundwater	Saturated zone	1,400 years	19,800
	Soil moisture and wetlands	1 to 5 years	28
Lakes		17 years	91
Channel networks		16 days	2
Atmosphere		8 days	13
Biosphere		several hours	1

Water is renewed at different times in each compartment of the water cycle (Table 1.1). The most accessible source of fresh water (lakes and channel networks) is not the most abundant in volume, but is the most quickly renewed. Larger stocks are associated to longer residence times, as the water in the cryosphere and groundwater. Gleeson *et al.* (2015) estimated that the total groundwater volume in the first 2 km of the Earth's crust is between 16 and 30 million km^3, with 0.1 to 5 million km^3 under 50 years old, for which uncertainties are related to the porosity estimation.

Groundwater is an important part of the water cycle, and is a key element in sustaining base flow to rivers, wetlands, and lakes. Groundwater is here defined as free water below the surface in the vadose and saturated zones. A formation (rock or sediment) that can yield significant amounts of groundwater to wells and springs is called aquifer. Aquifers can be classified into confined (in which the hydraulic head is above its top), and unconfined aquifers (bounded only below by an impermeable layer). Aquitards are formations that can store groundwater but they can not transmit it in significant or economic quantities, as in low hydraulic conductivity formations.

In unconfined aquifers, the water table represents the hydraulic head (variable that relates the water table energy to its elevation), while for confined aquifers the hydraulic head is above the top elevation, and is represented by the potentiometric

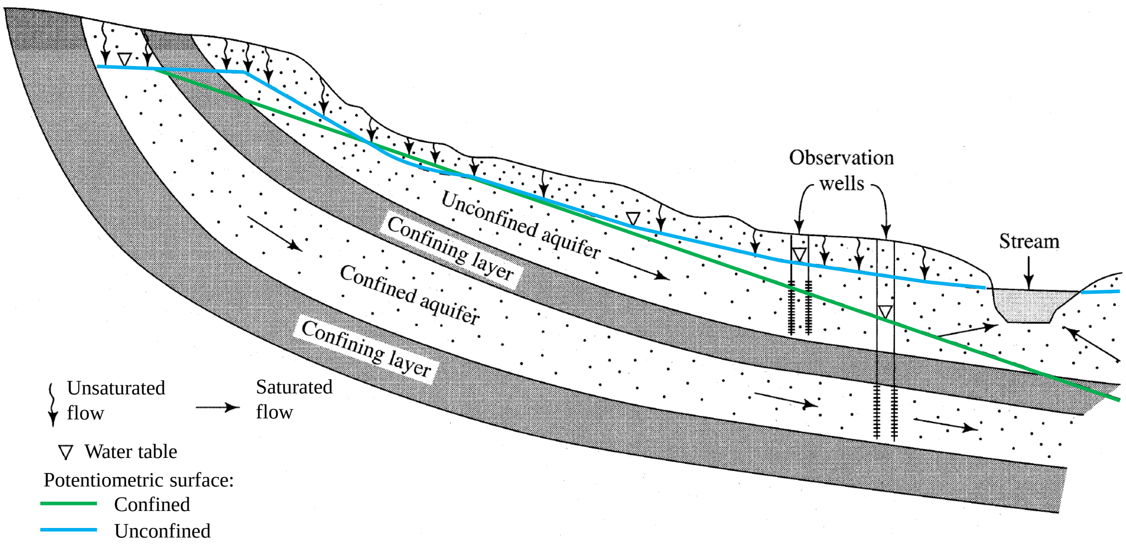

Observation wells
Stream
Unconfined aquifer
Confining layer
Confined aquifer
Confining layer
Unsaturated flow
Saturated flow
Water table
Potentiometric surface:
Confined
Unconfined

Figure 1.2 – Scheme for confined and unconfined aquifers. Adapted from Dingman (2015).

surface (Figure 1.2). The groundwater moves from a point of highest hydraulic head towards a lowest naturally discharging into streams, lakes, or oceans, at a velocity that depends on the aquifer properties.

1.2 Base flow

Base flow is the natural discharge of aquifers to rivers (Younger, 2009), or it can be defined as the part of stream flow that is sustained by groundwater discharge (Carter *et al.*, 2015). Base flow can be the only source of water to rivers, being an important component in water management, and a key variable in Land Surface Models (LSMs). Base flow quantification depends on estimation, since currently there are no techniques to directly measure it. Large scale LSMs can estimate base flow using two main approaches:

- Using a solution of the diffusion equation, here presented in the two dimensional (x and y directions) form:

$$\frac{\partial}{\partial x}\left(K_x h \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y h \frac{\partial h}{\partial y}\right) = S_y \frac{\partial h}{\partial t} + Q_w, \tag{1.1}$$

where h is the water table height (m) and t is the time step (s), K_x saturated hydraulic conductivity in the x direction (m.s^{-1}), K_y saturated hydraulic conductivity in the y direction (m.s^{-1}), S_y is the specific yield (-), and Q_w the term representing a source or a well flow per water table area (positive indicates outflow, m.s^{-1}).

- Using an one-dimensional linear reservoir, in which base flow is estimated by the ratio of the reservoir volume and a base flow time constant that represents a residence time.

The diffusion equation (Equation 1.1) is non-linear, meaning that there are no general solutions available, resulting in several methods based on different assumptions to arrive to a solution. However, the diffusion equation can be solved numerically, in which base flow depends on hydrogeological data. Numerical solutions are computational demanding, which is increased at large scales. Another approach is to use analytical solutions available for certain boundary conditions, as in Brutsaert (2005) that proposed a linear equation to estimate base flow derived from the Boussinesq equation (Boussinesq, 1877) as it follows. The Boussinesq equation is derived from the diffusion equation with Dupuit-Forchheimer assumptions (water table is a free surface, in which the capillary flow above the water table is neglected; and the distribution of water pressure in the direction normal to the flow is hydrostatic) here shown in the one-dimensional non-linear form, assuming the effective porosity equals the specific yield:

$$\frac{\partial h}{\partial t} = \frac{K}{n_e} \frac{\partial}{\partial x} \left(h \frac{\partial h}{\partial x} \right) \tag{1.2}$$

where K the saturated hydraulic conductivity (m.s^{-1}), and n_e is the effective porosity (-). Based on the Boussinesq equation (1.2), Brutsaert (2005) considered that the water table depth is never very different from an average value (h_0) remaining close to constant to have a linearized form of the Boussinesq equation, known as the linear diffusion equation (or heat equation):

$$\frac{\partial h}{\partial t} = \frac{K h_0}{n_e} \frac{\partial^2 h}{\partial x^2}, \tag{1.3}$$

considering a constant groundwater diffusivity (D_h):

$$D_h = \frac{Kh_0}{n_e} \tag{1.4}$$

To calculate base flow, Brutsaert (2005) assumed an initially saturated one-dimensional horizontal aquifer after cessation of recharge, as during recession [2]. During recession, base flow decreases exponentially (Chapman, 1999):

$$Q_t = Q_0 e^{-\frac{t}{\tau}} \tag{1.5}$$

where Q_t is the base flow at the time step t ($\mathrm{m^3.s^{-1}}$), Q_0 is the initial [3] base flow ($\mathrm{m^3.s^{-1}}$), and τ is the base flow time constant (s). Brutsaert (2005) classified the recession into two regimes: short-term, observed in the beginning of the recession, and a long-term, after the beginning of the recession. Brutsaert (2005) found that for the long-term solution, the base flow time constant is given by:

$$\tau = \frac{4n_e B^2}{\pi^2 T,} \tag{1.6}$$

where T is the aquifer transmissivity ($\mathrm{m^2.s^{-1}}$), and B is the aquifer length. Several studies use river discharge measurements to obtain base flow time constant during recession based on Equation 1.5, resulting in values that range from 5 to 250 days for watersheds between 7 and 200 $\mathrm{km^2}$ (Zecharias and Brutsaert, 1988; Brandes *et al.*, 2005; Eng and Milly, 2007; Brutsaert, 2008; Stoelzle *et al.*, 2013; Thomas *et al.*, 2013). However, at large scales observation data is scarce, and calibration becomes difficult. To overcome the lack of observed data, a common practice is to extrapolate the values from a calibrated region, as in Gascoin *et al.* (2009) that arrived to a base flow time constant of 700 days. The problem with extrapolation of calibrated values to large areas is that it does not take into account hydrogeological heterogeneities, as the diffusion equation does. The base flow time constant given by Equation 1.6 represents an aquifer response i.e., the time the aquifer will take to achieve equilibrium after a change (e.g., sudden recharge). For LSMs that use a linear reservoir, the base flow time constant can also represent the mean residence time (mean amount of time base flow takes to travel from recharge to discharge areas). Assuming no recharge for a homogeneous linear reservoir, the continuity equation assumes the form of:

---

[2] Recession is the period of time river discharge decreases, as in periods of no precipitation or after a flood.

[3] Q_0 is the initial base flow for the regime: for the short-term is the initial base flow, and for the long-term is the initial base flow of the long-term regime (after the short-term).

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -Q_t, \tag{1.7}$$

where V is the volume of the linear reservoir (m^3). The integration of equation 1.7, considering the base flow as the volume of the groundwater reservoir divided by the mean residence time ($Q_t = -V/\tau$), results in:

$$V_t = V_0 e^{-\frac{t}{\tau}} \tag{1.8}$$

In the case of equation 1.8, the base flow time constant equal the mean residence time, which allows to account for hydrogeological properties instead of only relying on calibration and extrapolation of the calibrated values by using τ obtained from equation 1.6 in equation 1.8. Both numerical and analytical approaches rely on a good characterization of aquifer properties, and available data to estimate base flow time constant.

1.3 Aquifer properties

1.3.1 Porosity and specific yield

Porosity is given by the ratio of void volume in a sediment or rock to the total volume of material ($\mathrm{m}^3.\mathrm{m}^{-3}$), representing the total amount of water that can be present in the void spaces, also called total porosity (n). Effective porosity (n_e) is a measure of the available spaces that groundwater can move. To better understand effective porosity, two properties need to be defined: specific yield (S_y) is the ratio of water volume that can be freely drained due to gravity from rocks or sediments to the total volume of rock or soil, representing the porosity available to contribute to the water flow, and specific retention is the amount of water that is retained in

Table 1.2 – Porosities, specific yields, and specific retention values per geologic materials (Dingman, 2015). Values were measured by Morris and Johnson (1967).

Material	**Porosity**			**Specific yield**			**Specific retention**		
	Min.	**Avg.**	**Max.**	**Min.**	**Avg.**	**Max.**	**Min.**	**Avg.**	**Max.**
Unconsolidated Alluvial Deposits									
Clay	0.34	0.42	0.57	0.01	0.06	0.18	0.25	0.38	0.47
Silt	0.34	0.46	0.61	0.01	0.20	0.39	0.03	0.28	0.45
Fine sand	0.26	0.43	0.53	0.01	0.33	0.46	0.03	0.08	0.43
Medium sand	0.29	0.39	0.49	0.16	0.32	0.46	0.01	0.04	0.18
Coarse sand	0.31	0.39	0.46	0.18	0.30	0.43	0.05	0.07	0.18
Fine gravel	0.25	0.34	0.39	0.13	0.28	0.40	0.00	0.07	0.17
Medium gravel	0.24	0.32	0.44	0.17	0.24	0.44	0.01	0.07	0.15
Coarse gravel	0.24	0.28	0.37	0.13	0.21	0.25	0.03	0.09	0.14
Unconsolidated Glacial Deposits									
Silty till	0.30	0.34	0.41	0.01	0.06	0.13	0.23	0.28	0.33
Sandy till	0.22	0.31	0.37	0.02	0.16	0.31	0.03	0.14	0.29
Gravelly till	0.22	0.26	0.30	0.05	0.16	0.34	0.01	0.12	0.25
Unconsolidated Aeolian Deposits									
Loess	0.44	0.49	0.57	0.14	0.18	0.22	0.22	0.27	0.30
Aeolian sand	0.40	0.45	0.51	0.32	0.38	0.47	0.01	0.03	0.06
Unconsolidated Biogenic Deposits									
Peat		0.92			0.44			0.44	
Weathered Rock (Saprolites)									
Granite	0.34	0.45	0.57						
Gabbro	0.42	0.43	0.45						
Clastic Sedimentary Rocks									
Fine sandstone	0.14	0.33	0.49	0.02	0.21	0.40	0.01	0.13	0.31
Med. sandstone	0.30	0.37	0.44	0.12	0.27	0.41	0.05	0.10	0.19
Siltstone	0.29	0.35	0.48	0.01	0.12	0.33	0.05	0.29	0.45
Claystone	0.41	0.43	0.45						
Shale	0.01	0.06	0.10						
Carbonate Rocks									
Limestone	0.07	0.30	0.56	0.02	0.14	0.36	0.05	0.13	0.29
Dolomite	0.19	0.26	0.33						
Igneous and Metamorphic Rocks									
Basalt	0.03	0.17	0.35						
Volcanic tuff	0.07	0.41	0.55	0.02	0.21	0.47	0.06	0.21	0.38
Schist	0.04	0.38	0.49	0.22	0.26	0.33	0.22	0.26	0.33

the formation. The sum of both terms represent total porosity. For unconfined aquifers, the specific yield can be approximated to effective porosity.

The specific yield values vary up to one order of magnitude not only according to different formations, from 0.06 to 0.44 in average (Table 1.2), but also within the same material, as in fine sand (from 0.01 to 0.46) or volcanic tuff (from 0.02 to 0.47). Total porosity and specific yield are not directly related, as clays have high total porosity, however they yield little water, since they are a fine-grained formation (resulting in high specific retention). The opposite behavior is observed for medium sands (Table 1.2), which have little specific retention, resulting in specific yield close to the total porosity values.

1.3.2 Transmissivity

Transmissivity is given by the product of saturated hydraulic conductivity and saturated thickness (part of the aquifer thickness that is saturated), and it describes the ability of the aquifer to transmit groundwater through the saturated thickness. For confined aquifers, the saturated thickness equals the aquifer thickness. The saturated thickness represents the saturated depth of an aquifer from the top of the water table to the bottom impermeable layer. Saturated thickness can vary from a few meters up to hundreds of meters, and for unconfined aquifers it varies with the water table height (Borman and Meredith, 1983). Brutsaert (2005) estimated saturated thickness as the average value of the water table height.

Saturated hydraulic conductivity is a measure of how fast the aquifer can transmit water. Hydraulic conductivities can vary among several orders of magnitude depending on the aquifer formation, as between gravel and unfractured igneous metamorphic rocks this difference can reach twelve orders (Table 1.3). However, hydraulic conductivity variations are observed not only with different materials, but also within a same formation due to natural heterogeneities in function of their degree of fracture and alteration, as in permeable basalt that the difference can reach 5 orders of magnitude.

Table 1.3 – Representative values of hydraulic conductivities for unconsolidated sedimentary materials, sedimentary rocks, and crystalline rocks. Adapted from Domenico and Schwartz (1990).

Material	**Hydraulic conductivity range (m.s^{-1})**
Unconsolidated sedimentary materials	
Gravel	3.0E-4 - 3.0 E-2
Coarse sand	9.0E-7 - 6.0E-3
Medium sand	9.0E-7 - 5.0E-4
Fine sand	2.0E-7 - 2.0E-4
Silt, loess	1.0E-9 - 2.0E-5
Till	1.0E-12 - 2.0E-6
Clay	1.0E-11 - 4.7E-9
Unweathered marine clay	8.0E-13 - 2.0E-9
Sedimentary rocks	
Karst and reef limestone	1.0E-6 - 2.0E-2
Limestone, dolomite	1.0E-9 - 6.0E-6
Sandstone	3.0E-10 - 6.0E-6
Siltstone	1.0E-11 - 1.4E-8
Salt	1.0E-12 - 1.0E-10
Anhydrite	4.0E-13 - 2.0E-8
Shale	1.0E-13 - 2.0E-9
Crystalline rocks	
Permeable basalt	4.0E-7 - 2.0E-2
Fractured igneous and metamorphic rocks	8.0E-9 - 3.0E-4
Weathered granite	3.3E-6 - 5.2E-5
Weathered gabbro	5.5E-7 - 3.8E-6
Basalt	2.0E-11 - 4.2E-7
Unfractured igneous and metamorphic rocks	3.0E-14 - 2.0E-10

1.3.3 Aquifer length

Aquifer length represents the distance from the river (or discharge zone) to the groundwater divide. Aquifer length is not an easy parameter to quantify due to water table variations and seasonality (represented by ephemeral streams). At large scales the difficulty is increased by natural heterogeneities that are not always depicted. However, Horton (1945) presented a relationship to estimate the average aquifer length inside a basin from drainage density, given by:

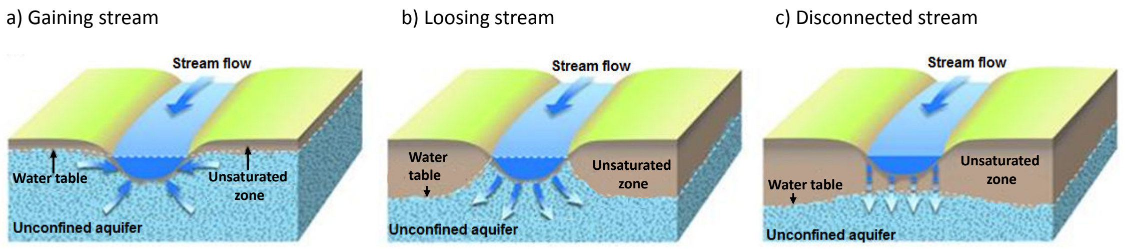
$$B = \frac{1}{2\delta}, \tag{1.9}$$

where B is the aquifer length (in m), and δ is the drainage density (in m^{-1}). This simple equation allowed to estimate the mean aquifer length from a parameter that is easier to acquire at small and larger scales (δ):

$$\delta = \frac{\sum L}{A}, \tag{1.10}$$

where L is the total river length inside a basin (in m), and A is the basin's area (m^2). Drainage density is a property that is used to quantitatively evaluate a watershed, giving a macro-scale measure of a basin stream length (Strahler, 1957; Vogt *et al.*, 2007; Dingman, 2015).

Groundwater can be connected to rivers, depending on the water table depth. In regions where the water table depth is low (Figure 1.3.a), groundwater can sustain the stream. For deeper water table depths, as in long periods without recharge or in case of flood, groundwater can be recharged by the stream (Figure 1.3.b), or even be disconnected from it for long periods without recharge (Figure 1.3.c), in the case of deep water table. When groundwater sustain streams, the aquifer length is usually shorter in humid regions, where drainage densities are higher (Figure 1.3.d), representing streams closer from one another. In arid regions, the aquifer length can be significantly longer (Figure 1.3.e), as streams are far from one another (representing low drainage densities) and groundwater travels longer to discharge to the stream. Topography and hydraulic conductivity are related to groundwater connection to rivers, as drainage density also depends on slope and lithology (Vogt *et al.*, 2003; Colombo *et al.*, 2007), which was represented by Haitjema and Mitchell-Bruker (2005) and Gleeson *et al.* (2011b) by a property called water table ratio (WTR). WTR depends on topography, hydraulic conductivity, and recharge, resulting in topography-controlled regions for log(WTR)>0 and recharge-controlled regions for log(WTR)<0.


a) Gaining stream
Stream flow
Water table
Unsaturated zone
Unconfined aquifer
b) Loosing stream
Stream flow
Water table
Unsaturated zone
Unconfined aquifer
c) Disconnected stream
Stream flow
Water table
Unsaturated zone
Unconfined aquifer

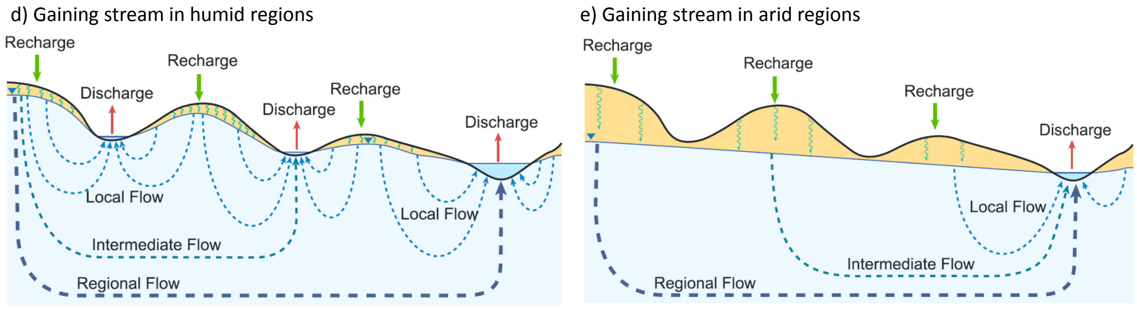

d) Gaining stream in humid regions
Recharge
Recharge
Recharge
Discharge
Discharge
Discharge
Local Flow
Local Flow
Intermediate Flow
Regional Flow
e) Gaining stream in arid regions
Recharge
Recharge
Recharge
Discharge
Local Flow
Intermediate Flow
Regional Flow

Figure 1.3 – Groundwater and stream interactions, depending on water table depth. a) Groundwater flow towards stream; b) Stream flow towards groundwater; c) Stream is disconnected from groundwater; d) Groundwater flow towards streams in humid regions; d) Groundwater flow towards streams in arid regions. Adapted from Barlow and Leake (2012) (a, b and c) and Maxwell *et al.* (2015) (d and e).

1.4 Base flow time constant

Base flow time constant is a key parameter in estimating base flow in LSMs that have a simple representation of groundwater (as a linear reservoir). The use of an analytical solution to estimate the base flow time constant makes the link between simple (considering a linear reservoir) and complex (as solving numerically the diffusion equation) approaches for base flow estimation. It is possible to estimate the base flow time constant at global scale with available hydrogeological data using the analytical solution proposed by Brutsaert (2005) represented by Equation 1.6 combined with Equation 1.9:

$$\tau = \frac{S_y}{\pi^2 T \delta^2}, \tag{1.11}$$

where S_y is the specific yield (-), T is the aquifer transmissivity ($\mathrm{m}^2.\mathrm{s}^{-1}$), and δ is the drainage density (m^{-1}). GLHYMPS and HydroSHEDS data were used to estimate τ at 0.5° resolution (common resolution used in LSMs). GLHYMPS data (Gleeson *et al.*, 2014) provide global values for total porosity (associating hydrolithology classes with Morris and Johnson (1967) values), and permeability (from hydrogeological modeling) that represent the mean value over 100 m depth. Total porosity values were used instead of specific yield, since no S_y global data are currently available. Transmissivity was estimated using hydraulic conductivity derived from GLHYMPS permeability data[4], and a constant saturated depth of 34.65 m. HydroSHEDS (Lehner *et al.*, 2008) is a global hydrographic information product providing a hydrologically corrected Digital Elevation Model (DEM) and derived products (river networks, flow accumulations, flow directions) at 15" and 30". Drainage density was calculated using 15" HydroSHEDS river network for the 0.5° grid.

Base flow time constant is overestimated by almost five orders of magnitude in average when compared to literature values from recession analysis results derived from observed river discharge data (Zecharias and Brutsaert, 1988; Eng and Milly, 2007; Brutsaert, 2008; Thomas *et al.*, 2013; Stoelzle *et al.*, 2013) and calibration of groundwater model with linear reservoir (Gascoin *et al.*, 2009) shown in Figure 1.4 and Table 3.1 (in chapter 3). This overestimation indicates that the use of currently available global data are not suitable to estimate τ to be used in LSMs with simple groundwater schemes, raising a question: *Can the estimation of the base flow time constant be improved?*

1.5 Groundwater schemes and base flow modeling in LSMs

LSMs can be used as a tool to better understand the water cycle, quantify water budgets, evaluate anthropogenic effects, and predict future changes. When coupled to Global Climate Models (GCMs), they can also provide an input of the energy and water budgets to predict climate change, and provide direct impacts on water resources.

---

[4] Hydraulic conductivity relates to permeability by $K = \frac{perm \cdot \rho \cdot g}{\mu}$, where *perm* is GLHYMPS permeability, ρ is water density (1,000 kg.m^{-3}), g is gravitational acceleration (10 m.s^{-2}), and μ is water viscosity (0.001 kg.m^{-1}.s^{-1}).

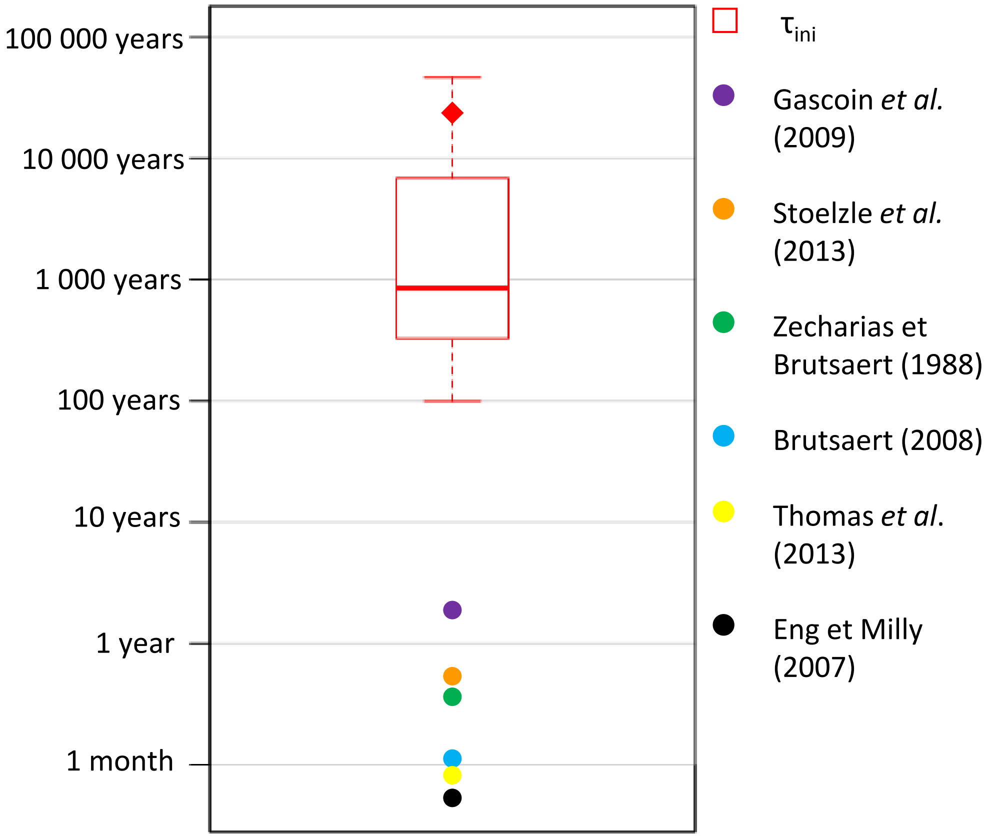


Figure 1.4 – Estimated base flow time constant (τ_{ini}) estimated at 0.5° resolution with HydroSHEDS and GLHYMPS data. The bottom and top of the box-plot represent first and third quartiles, the middle bar gives the median, and red diamond indicates the mean value. Whiskers show the 10th and 90th percentiles. Points represent mean values from the literature for the simple approach (linear reservoir).

LSMs were initially developed to be coupled with Global Climate Models (GCMs) and they depend on simplifications in the calculations of energy and water fluxes, as neglecting local geology or anthropogenic effects (as irrigation). These simplifications evolved over the years, and models can be classified based on the simplicity as first, second, and third generation models (Pitman, 2003). The first generation of LSMs was represented by a bucket model, where the precipitation feeds a reservoir representing the soil moisture with no drainage (as a bucket); when the reservoir is filled (saturated soil) the precipitation will create the surface runoff, and evaporation depends on the water available in the bucket.

The second generation of LSMs improved the bucket model by considering vegetation impacts on the energy and water budgets, and better representing soil moisture. In these models the soil is discretized into several layers, using the Richards equation to calculate the water transfer. They also use separated reservoirs to stock water, and the vegetation can interact with the soil and atmosphere (Pitman, 2003). The third generation of LSMs includes the carbon cycle in addition to the second generation models.

Groundwater started to be represented in LSMs of second and third generations, motivated by water table interactions with soil moisture that can influence the surface energy and water exchanges with the lower atmosphere (Koster *et al.*, 2000; Ducharne *et al.*, 2000; York *et al.*, 2002; Yeh and Eltahir, 2005; Maxwell and Miller, 2005; Fan *et al.*, 2007; Gascoin *et al.*, 2009; Lo *et al.*, 2010; Vergnes and Decharme, 2012), in which a non-exhaustive list of these LSMs is presented in Table 1.4. CLSM (Catchment Land Surface Model) model had a shallow groundwater scheme introduced by Ducharne *et al.* (2000) and Koster *et al.* (2000) simulating base flow based on TOPMODEL[5] equations. Gascoin *et al.* (2009) added a deep groundwater reservoir to CLSM to account for longer timescales, resulting in a model with a groundwater outflow that is composed of two reservoirs (shallow, from Ducharne *et al.* (2000), and deep). This new scheme impacted both soil moisture and land surface fluxes, significantly improving river discharges, maintaining low flow and reducing overestimated peak flows due to the buffer effect inherent to a long timescale reservoir. Ngo-Duc *et al.* (2007) observed that the introduction of a transfer scheme in ORCHIDEE model (by linear reservoirs, in which groundwater is represented by a linear reservoir) improved simulated land water storage, when compared to GRACE observations. However, ORCHIDEE does not represent capillary rise, in contrast with most models.

Yeh and Eltahir (2005) coupled the LSM called LSX to an unconfined aquifer model (LSXGW) that simulates base flow using as the observed water table depth for model calibration. The introduction of a groundwater scheme improved the representation of hydrological processes that are attributed to good calibration of

---

[5] TOPography based hydrological MODEL (Beven and Kirby, 1979).

Table 1.4 – List of groundwater schemes recently introduced in LSMs.

Model	Type	Groundwater scheme	Reference
CLSM	1D	Shallow groundwater reservoir based on TOPMODEL	Ducharne *et al.* (2000) and Koster *et al.* (2000)
CLSM	1D	Inclusion of a deep groundwater reservoir with longer timescales	Gascoin *et al.* (2009)
ORCHIDEE	1D	Linear reservoir that delays routed water discharging into the stream	Polcher (2003) and Ngo-Duc *et al.* (2007)
LSXGW	1D	Water table depth is given by a multicolumn nonlinear reservoir	Yeh and Eltahir (2005)
SIMGM	1D	Reservoir at the bottom of the soil column that simulates base flow using TOPMODEL equations	Niu *et al.* (2007)
LEAF2-Hydro	2D	Water table depth is obtained through recharge and lateral flow	Fan *et al.* (2007) and Miguez-Macho *et al.* (2007)
TRIP	2D	Groundwater reservoir is based on MODCOU, solving water table depth and base flow	Vergnes and Decharme (2012)
ParFlow	3D	Physically-based groundwater scheme to solve water table depth, base flow, and lateral fluxes	Maxwell and Miller (2005)

the water table depth. However, Yeh and Eltahir (2005) acknowledge the difficulties in using the coupled model at large scales, due to lack of data to calibrate the LSXGW model.

Using another approach, Maxwell and Miller (2005) coupled the groundwater model Parflow with the LSM CLM (Common Land Model) at fine spatial resolution, by introducing a physically based groundwater scheme that solves two sets of equations under steady (for fully saturated flow) and transient (for flows partially saturated) states. The water table depth is defined by the region with 100% saturation and water pressure greater than zero. They observed that the new groundwater scheme provides simulation results that have a memory of water stored in the deep subsurface, with no significant changes in the evaporation results, but with improved runoff and soil moisture. This model was also applied at continental scales to estimate groundwater residence times (Maxwell *et al.*, 2015) at 1 km resolution, observing that residence times are mainly controlled by hydraulic conductivity.

Fan *et al.* (2007) and Miguez-Macho *et al.* (2007) developed a simple groundwater flow model to estimate equilibrium water table, by considering only the lateral flow at steady state. Fan *et al.* (2007) focused on the equilibrium water table to be used in hydrological models, due to the lack of water table data at large scales that can be used for calibration or model initialization, also observed by Yeh and Eltahir (2005). Miguez-Macho *et al.* (2007) observed that the new groundwater scheme improved stream discharge simulations, and it allows to depict seasonal variations in both river discharge and evapotranspiration.

Later in the same year, Niu *et al.* (2007) proposed a simple groundwater scheme that predicts groundwater storage and water table depth in a reservoir at the bottom of the soil column of the CLM model, called SIMGM. Base flow is simulated using formulations that are based on TOPMODEL. As in Fan *et al.* (2007), SIMGM does not have lateral flow between grid cells. Niu *et al.* (2007) observed an improvement in simulated runoff and water storage due to the inclusion of the groundwater scheme, which increased evapotranspiration due to higher values of soil moisture.

Vergnes and Decharme (2012) introduced a new groundwater scheme in the Total Runoff Integrated Pathways (TRIP) routing model based on the MODCOU (Ledoux *et al.*, 2007) hydrogeological model that accounts for water table dynamics and groundwater-river exchanges. The groundwater reservoir is fed by drainage produced from the ISBA (Interactions between Soil, Biosphere, and Atmosphere) LSM, with an empirical parameter that controls the transfer time between river and groundwater. They observed an improvement of simulated river discharges due to the introduction of a new groundwater scheme over France.

These studies showed that introducing a better characterization of the groundwater in LSMs allows to have more realistic results of the river discharge (as base flow discharge into rivers), total water storage, and evapotranspiration (in which the water table affects soil moisture, impacting both soil evaporation and transpiration from plants). However, large-scale LSMs with simple groundwater schemes rely on calibration and extrapolation of τ, which can neglect natural heterogeneities due to the lack of observed data to perform calibration globally. The introduction of a base flow time constant that depends on hydrogeological and topography data in LSMs with groundwater represented by a linear reservoir may allow to account for heterogeneities naturally found in aquifers.

1.6 Objectives

This thesis was motivated by the lack of observed data to calibrate base flow time constant in LSMs with a linear reservoir. Such models currently rely on extrapolation of the results for a calibrated region to the global scale, neglecting aquifer heterogeneities. The use of a base flow time constant that is function of hydrogeological and topographical properties can introduce natural heterogeneities found in aquifers. However, when using available global data to estimate the base flow time constant, the results are overestimated by more than two orders of magnitude, raising the main question: *Can the estimation of the base flow time constant at global scale be improved to obtain a better base flow dynamics in large-scale LSMs?* More specifically:

1. *Which parameters exert a strong control on* τ*?*
2. *Does the base flow time constant estimated from hydrogeological and topographical properties improve river discharge simulations of LSMs with linear reservoir?*

This thesis helps to answer these research questions, and it is divided in five chapters. Chapters two and three help to answer the first research question. The second chapter shows an evaluation of the drainage density, since it is one of the parameters that control τ. Chapter two was published in the form of an article in Geophysical Research Letters (Schneider *et al.*, 2017), and it resulted in a global data set called GRIN (Global River Network) composed by drainage density (at

7.5') and river network (at 1:50,000 scale) data freely available for download at `http://www.metis.upmc.fr/en/node/211`.

The third chapter shows a sensitivity analysis of τ to all parameters used on its estimation (effective porosity, transmissivity, drainage density, and slope for a formulation that considers a sloping aquifer), by the use of different data at global scale for each parameter. τ results are compared to reference data from recession analysis at the basin scale to validate these estimations. The second research question is answered in chapter fourth, in which τ results of chapter three are used in the ORCHIDEE model. Chapter fourth shows an assessment of the impacts of a longer timescale in ORCHIDEE's groundwater reservoir to river discharge, accompanied by an evaluation of ORCHIDEE's sensitivity to τ. Finally, in the conclusions, the main research question is answered showing the conclusions that emerged from this thesis followed by the perspectives for this work.

This thesis was funded by the European Institute of Innovation and Technology (Climate KIC - Knowledge and Innovation Community), by the French *Agence Nationale de la Recherche* (ANR Grant No. ANR-14-CE01-00181-01), and the French national program LEFE/INSU.

CHAPTER 2

GLOBAL RIVER NETWORK EXTRACTION

Contents

2.1 Introduction 21
2.2 Data description 24
2.3 A_{cr} Calibration and Model Selection 30
2.4 Model Results and Discussion 37
2.5 Conclusions 42

Scope

The available global river networks result in constant and underestimated drainage density when compared to reference data at local scales, raising the research question: *Can drainage density estimation be improved at global scale?* This chapter describes the extraction of a river network at global scale that resulted in drainage density values much closer to reference data both in terms of spatial match and bias. This chapter was published in the form of an article in Geophysical Research Letters (Schneider *et al.*, 2017), available in Appendix 6.1.

2.1 Introduction

Precise characterization of river geometry is crucial for many applications related to river hydraulics, and has gained a lot from the advances in digital elevation models (DEMs) processing since (Hutchinson, 1989) first proposed the “stream burning” technique to correct the location of extracted streams. At large-scales, due to the advent of very high resolution global DEMs such as HYDRO1k at 1 km (Verdin and

Greenlee, 1998) and HydroSHEDS at 30" and 15" (Lehner *et al.*, 2008), priority has been given to producing correct flow direction maps, including coarser resolutions used in global runoff routing models (Graham *et al.*, 1999; Fekete *et al.*, 2001; Döll and Lehner, 2002; Wu *et al.*, 2012).

Proper characterization of stream length from DEMs has received less attention despite its broad influence on water sciences (e.g., on erosion and sediment transport (Moore and Burch, 1986); riverine water quality, which is strongly controlled by the upstream residence time (Billen *et al.*, 2009); and the buffering effect of groundwater on extreme flows, which partly depends on the distance from the recharge zones to the rivers (Brutsaert and Lopez, 1998)). A major issue is identifying the smallest streams, called the "Aqua incognita" by Bishop *et al.* (2008). This is true in the field because of stream intermittency, artificial ditches, or hydraulic obstructions in flat areas. Additional problems arise when implementing traditional methods to map the "blue lines", usually from airborne or satellite imagery, because of insufficient resolution, vegetation masking, or scarce field data (Tarboton *et al.*, 1991; Lehner and Döll, 2004; Benstead and Leigh, 2012; Persendt and Gomez, 2016). It is believed that river networks extracted from very high resolution DEMs are more comprehensive, assuming they are properly calibrated (Benstead and Leigh, 2012; Stein *et al.*, 2014).

Many methods exist to extract river networks from DEMs, mostly from a geomorphological reference (Tarboton *et al.*, 1992; Montgomery and Foufoula-Georgiou, 1993; Heine *et al.*, 2004; Pelletier, 2013). However, the most widely used method at small and regional scales, by far, simply relies on flow direction and a critical drainage area (A_{cr}) that represents the minimum upstream drainage area required to initiate a river (O'Callaghan and Mark, 1984). In this classical framework, extracted rivers are shorter if the A_{cr} is larger, which results in a smaller drainage density (δ), defined as the total river length inside a watershed divided by its area (Horton, 1932).

Drainage density, gives a macro-scale measure of stream length and is often used to quantitatively evaluate watershed properties (Strahler, 1957; Vogt *et al.*, 2007; Dingman, 2015). In natural river networks, δ is spatially variable (Tucker *et al.*, 2001; Vogt *et al.*, 2007), since river initiation depends on climate, slope, lithology, soil properties, and vegetation cover (Tarboton *et al.*, 1992; Montgomery and Foufoula-Georgiou, 1993; Vogt *et al.*, 2003; Colombo *et al.*, 2007; Sangireddy *et al.*, 2016). However, given the lack of sufficient information to constrain A_{cr}, it is common to use a single value for river network extraction in a given domain, resulting in a rather uniform δ given the long-established link between A_{cr} and δ (Tarboton *et al.*, 1992).

In global hydrographic datasets such as HYDRO1k, HydroSHEDS and AQUAMAPS (a river network product derived from 15" HydroSHEDS; FAO (2014)), the A_{cr} used to extract river networks are uniform across continents, and their values are arbitrarily chosen to limit the number of small streams for large-scale applications: 1000 km^2 in HYDRO1k and the global version of AQUAMAPS, 100 km^2 for the continental version of AQUAMAPS, and approximately 25 km^2 (at the equator) in HydroSHEDS (100 15" pixels). These discrepancies highlight the fact that each of the corresponding networks is a "hypothetical" river network, constrained in its extent by the chosen A_{cr}. The above values all result in δ under 0.2 km^{-1}, which is much lower than the values obtained from detailed observations in small watersheds (40 up to 1000 km^2), which show drainage densities between 0.5 and 1.6 km^{-1} (Horton, 1945; Brutsaert and Nieber, 1977; Zecharias and Brutsaert, 1988; Brutsaert and Lopez, 1998). This led Raymond *et al.* (2013) to define their own global river network based on the 15" HydroSHEDS and a smaller A_{cr} of 10 km^2 for estimating carbon sinks and sources for inland waters (resulting in global mean δ of 0.28 km^{-1}).

Similarly, several studies have analyzed the spatial dependence of δ on environmental parameters at the continental scale (Colombo *et al.*, 2007; Vogt *et al.*, 2007; Luo *et al.*, 2016). Vogt *et al.* (2007) evaluated drainage density over Europe (CCM2 dataset, second version of Colombo *et al.* (2007) work), with a mean value of 0.31 km^{-1} based on a 100 m resolution DEM and 10 different A_{cr} values, ranging from 0.72 to 12.80 km^2, defined from the association of landscape classes to subjective scores of valley dissection potential. Luo *et al.* (2016) used a geomorphological detection method and a 30 m resolution DEM to extract the land dissection density (considered equivalent to drainage density) over the United States, with values ranging from 0 to 5 km^{-1}, and a clear spatial dependence on climate, lithology and several terrain-based attributes.

Based on this analysis, our goal is to propose spatially variable A_{cr} values for a simple extraction of river networks from large-scale DEMs, as well as a first-order assessment of intermittent streams, using the 15" HydroSHEDS data as an example. The heterogeneities of A_{cr} and the resulting drainage density, δ, are linked to environmental parameters (i.e., slope, lithology, climate [1]) to match drainage densities from reference river networks at the 1:50,000 scale in France and Australia. An independent evaluation conducted against reference river networks from the United States and Brazil is discussed before generalizing the method to extract 15" river networks consistent with 1:50,000 blue lines across the continents.

2.2 Data description

ArcGIS (version 10.3.1) tools was used to process several hydrologically corrected layers from the 15" (ca. 500 m at the equator) resolution HydroSHEDS database (Lehner *et al.*, 2008) including flow directions and flow accumulation for river network extraction, and elevation to calculate the local slope, using the neighborhood method (Burrough and McDonell, 1998) corrected for latitudinal distortions (show in Figure 2.1). Global lithology data are from Hartmann and Moosdorf (2012), with an average scale of 1:3,750,000 and 12 classes mapped in Figure 2.1a. Climate is represented by the multi-annual mean of total precipitation (1980-2009) raster at 0.5° resolution from the WFDEI (WATCH Forcing Data methodology applied to ERA-Interim) dataset, including correction by the GPCC (Global Precipitation Climatology Centre) data (Weedon *et al.*, 2014) (Figure 2.1b).

---

[1] In this chapter, climate is represented by total precipitation.

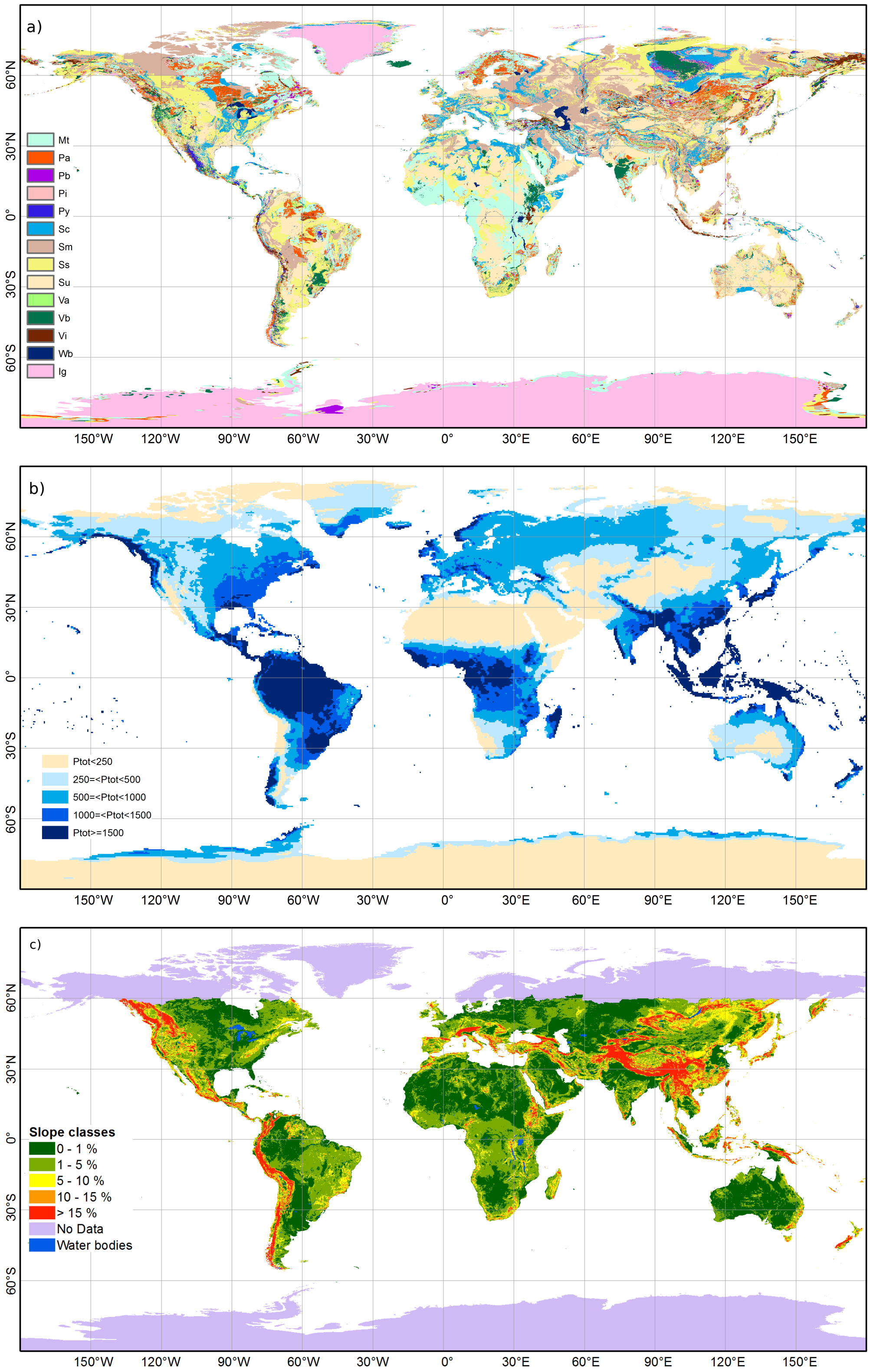

a)
60°N
30°N
0°
30°S
60°S
Mt
Pa
Pb
Pi
Py
Sc
Sm
Ss
Su
Va
Vb
Vi
Wb
Ig
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
b)
Ptot<250
250=<Ptot<500
500=<Ptot<1000
1000=<Ptot<1500
Ptot>=1500
c)
Slope classes
0 - 1 %
1 - 5 %
5 - 10 %
10 - 15 %
> 15 %
No Data
Water bodies

Figure 2.1 – Environmental classes evaluated in drainage density calibration. a) All lithology classes from Hartmann and Moosdorf (2012) (Pi = Intermediate plutonic rocks; Mt = Metamorphic rocks; Va = Acid volcanic rocks; Ss = Siliciclastic sedimentary rocks; Vi = Intermediate volcanic rocks; Pb = Basic plutonic rocks; Pa= Acid plutonic rocks; Vb = Basic volcanic rocks; Py = Pyroclastics; Sm = Mixed sedimentary rocks; Su = Unconsolidated sediments; Sc = Carbonate sedimentary rocks); b) Climate classes from Weedon *et al.* (2014); c) Slope classes calculated from HydroSHEDS elevation.

The reference river networks were acquired in vector format from four national databases, from Australia (OSDM, 2015) (called Geofabric) and France (IGN, 2014) (known as CARTHAGE) at a 1:50,000 scale, from the United States (USGS, 2015) (called NHD) at a 1:24,000 scale, from Brazil (IBGE, 2015) at variable scales between 1:25,000 and 1:250,000. They were chosen for their quality and reported scale of approximately 1:50,000, despite some inconsistencies. For additional comparison with a "hypothetical" global river network, the global AQUAMAPS (FAO, 2014) network was used, further called FAO, based on the 15" HydroSHEDS data and a constant A_{cr} of 100 km^2, which also provides an intermittency assessment.

The Australian river network was acquired from the Australian Hydrological Geospatial Fabric (Geofabric) (OSDM, 2015), a derived product of AusHydro cartographic streams. The river network is derived from a 1" resolution DEM, resulting in high resolution mapped features for rivers, lakes, water tanks, and dams (OSDM, 2015). This product uses the same methodology as its previous version at the 9" resolution DEM (1:250,000 stream map scale), described by Stein *et al.* (2014), where rivers are initiated by the gridded channel heads from AusHydro cartographic streams (from AHGF mapped streams). Stein *et al.* (2014) stated that the locations of the Geofabric streams accurately reflect the locations of the AusHydro cartographic streams, since only 5% of 1000 randomly sampled points of the Geofabric streams are located more than 125 m from an AusHydro watercourse (with all of them under 500 m). No significant problems were observed in this data, since it was possible to restrict our reference river network to the natural channels only, excluding the artificial segment layers. Although it is documented that variable resolutions could be found in some areas, no unusual density changes were observed. Figure 2.2 shows high values of drainage density as red pockets in some arid areas of Australia; a detailed analysis reveals that they correspond to sources of Lake Eyre, Torrens, Frome, and Gairdner, which are classified as non-perennial streams in the Geofabric dataset, like most of Australia (98% in length, Table 2.2).

The French river network is version 3.1 of the 2014 national cartography database known as CARTHAGE (IGN, 2014), at a 1:50,000 resolution. Although these are

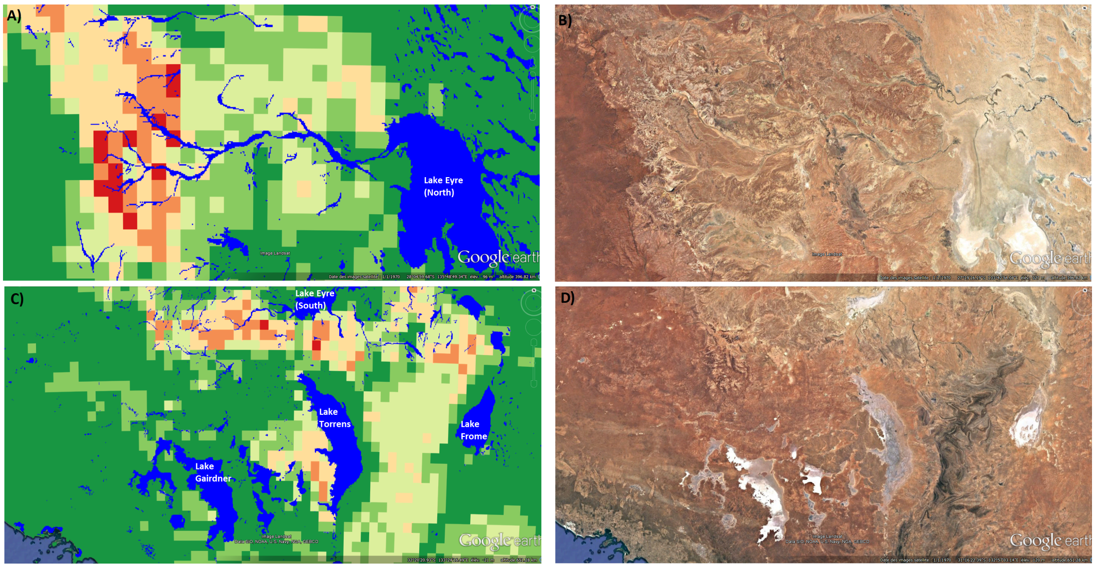

A)
Lake Eyre (North)
B)
C)
Lake Eyre (South)
Lake Torrens
Lake Frome
Lake Gairdner
D)
Google earth

Figure 2.2 – Zoom over Australian areas of Geofabric data that have high drainage densities in arid regions. A) Drainage density map viewed in Google Earth over Lake Eyre (north). B) Google Earth view without the drainage density layer. C) Drainage density map viewed in Google Earth over Lakes Eyre (south), Torrens, Frome, and Gairdner. D) Google Earth view without the drainage density layer.

good quality river network data, it was not possible to separate natural from unnatural channels. The CARTHAGE network, however, includes artificial networks, for example, irrigation ditches connected to hillside storage ponds, as shown in Figure 2.3. This results in overestimated drainage densities in some areas compared to what would be expected from the natural stream network only, but this overestimation cannot be quantified.

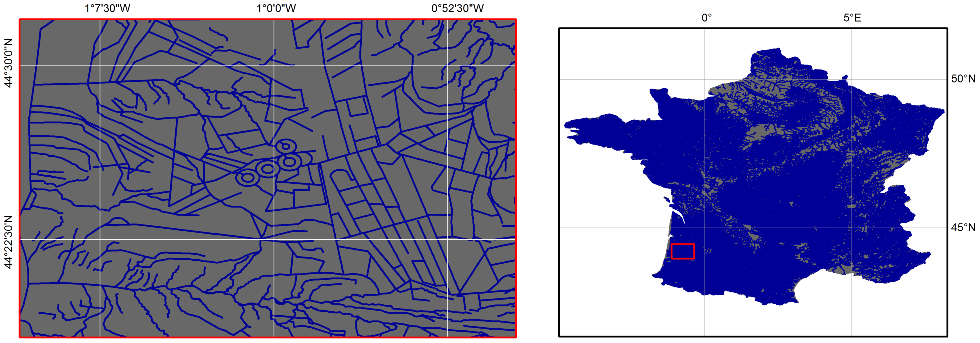

1°7'30"W
1°0'0"W
0°52'30"W
44°30'0"N
44°22'30"N
0°
5°E
50°N
45°N

Figure 2.3 – Zoom over the French river network region with unnatural channels in an irrigated zone (Landes area).

The United States river network was acquired from the National Hydrography Dataset (NHD) at the best resolution available (1:24,000) (USGS, 2015), but it is documented that mixed resolutions (coarser than 1:24,000) can be found in certain regions (USGS, 2015), resulting in a "lack of consistency between the different quadrangles needed to form a single, country-wide map" of drainage density according to Luo *et al.* (2016). Several features were found that are similar to unnatural channels, as shown in Figure 2.4. Although only one example is presented here, the same kind of features were found in several regions of the country.

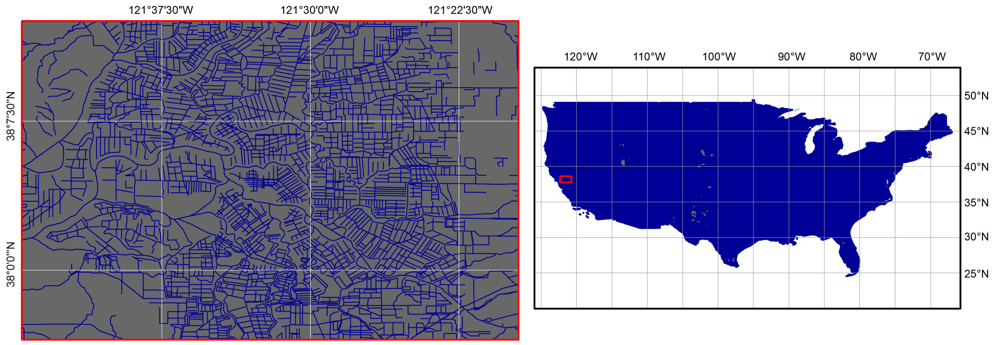


Figure 2.4 – Zoom over the United States river network region with unnatural channels in an agricultural zone west of San Francisco.

The Brazilian river network was acquired from the national database (IBGE, 2015), using the December 2015 dataset version. These data combine multiple scales, varying from under 1:25,000 up to 1:250,000, as shown by the tiles corresponding to the different scales in Figure 2.5. Although this dataset gives a good representation of the main river network, it is likely inaccurate for headwater streams.

The HydroSHEDS data used in this study are described according to the documentation (version 1.1) of HydroSHEDS (Lehner *et al.*, 2008):

- Flow direction: the 15" flow direction raster data were used, described as "drainage directions". They represent the 8 single possible flow directions and are provided in the ESRI format (values from 1 to 128).

- Flow accumulation: the 15" flow accumulation raster data were used, described as "flow accumulation". They represent the upstream drained cells, and they

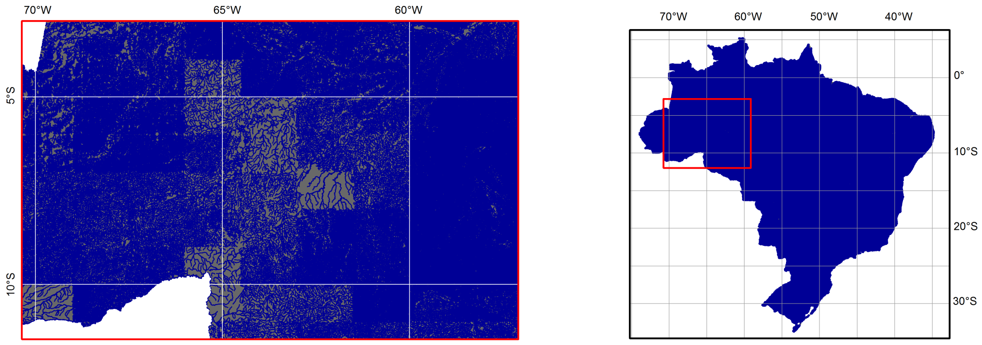


Figure 2.5 – Zoom over the Brazilian river network region with variable resolution problems.

can be translated as the upstream catchment area after multiplication by the pixels' area.

The AQUAMAPS river network (FAO, 2014) was also used for evaluation in this study. It is also derived from the 15" HydroSHEDS, using a single A_{cr} of 100 km^2 for each of the continental datasets. A global layer is available, but it was not used here as it is based on a very high A_{cr} of 1000 km^2. The AQUAMAPS river networks offer a regime classification (perennial and intermittent streams), which is based on the Strahler order and the aridity index (annual precipitation divided by annual evapotranspiration).

In order to process a DEM in a decimal degree coordinate systems, such as HydroSHEDS, ArcGIS requires A_{cr} to be expressed in a number of upstream pixels (commensurate to flow accumulation). Since the area of 15" pixels decreases with latitude, keeping A_{cr} in pixels means the corresponding area varies with latitude. To correct this fact, we accounted for pixel areas for both A_{cr} calibration and global applications. To this end, we calculated the area of each pixel, a spherical quadrangle with sides of 15" (1°/240), using a simple formula considering the latitude λ of the pixel's center:

$$A_{pixel} \approxeq \left(\frac{2\pi R}{360} \cdot \frac{1}{240}\right)^2 \cdot cos\lambda \tag{2.1}$$

where A is in km^2 and R = 6378.137 km (the radius of Earth). As expected, the area is maximal at the equator (latitude = 0°, $A_{pixel} = 0.215$ km^2) and decreases with latitude. All drainage density analyses and calculations were made using a 7.5' grid (grid-cells ca. 150 km^2 at equator) over the continents. Drainage densities (km^{-1}) were calculated as the total river length (km) inside a 7.5' grid-cell, divided by the corresponding land area (km^2). Each 7.5' grid-cell was characterized by a single value for δ and each of the selected environmental parameters (dominant class for lithology, areal average of total precipitation (mm.y^{-1}), and slope (%), each reclassified into 5 classes) (Table 2.1 and Figure 2.1).

Table 2.1 – Selected environmental parameters and their coverage area (%) used for A_{cr} calibration (all lithology, precipitation, and slope evaluated classes). Area was calculated over regions with available data (No Data values were removed from the analysis).

Lithology		Precipitation		Slope	
Classes (-)	Area (%)	Classes mm.y^{-1}	Area (%)	Classes (%)	Area (%)
Metamorphic rocks	14.7	$[0, 250[$	21.0	$[0, 1[$	37.5
Acid plutonic rocks	6.4	$[250, 500[$	19.5	$[1, 5[$	38.0
Basic plutonic rocks	0.5	$[500, 1000[$	25.1	$[5, 10[$	11.5
Intermediate plutonic rocks	0.3	$[1000, 1500[$	16.3	$[10, 15[$	5.9
Unconsolidated sediments	30.3	$[1500, \infty[$	18.2	$[15, \infty[$	7.2
Siliciclastic sedimentary rocks	17.7				
Mixed sedimentary rocks	14.5				
Carbonate sedimentary rocks	8.4				
Pyroclastics	0.8				
Acid volcanic rocks	1.1				
Basic volcanic rocks	3.4				
Intermediate volcanic rocks	1.8				

2.3 A_{cr} Calibration and Model Selection

The spatial intersection of the selected environmental parameters (12 lithology classes, 5 precipitation classes, and 5 slope classes) defines 300 environmental classes. In each of them, A_{cr} was calibrated by minimizing the bias between the corresponding drainage density (δ) and δ_{Ref} (drainage density with a reference river network).

This minimization was performed independently in each country, by testing 300 extracted river networks, defined by a wide range of A_{cr} (from 0.3 to 200 km^2). To avoid calibration errors due to significant heterogeneities inside the 7.5' grid-cells, calibration was only performed in grid-cells where the dominant lithology class covered more than 70% of the grid-cell (representing 81% of the continents).

To permit independent validation, calibration was restricted to France and Australia, in which the reference river networks share the same 1:50,000 scale and show few inconsistencies. Together, these two countries also encompass all the precipitation, lithology, and slope classes of the currently available HydroSHEDS domain (56°S to 60°N). The resulting calibrated values in the two countries are given in dataset Appendix 6.2.

Much smaller A_{cr} were found than the values used to define stream networks in the global-scale databases (i.e., HYDRO1k, AQUAMAPS, HydroSHEDS, with A_{cr} values between 25 and 1000 km^2), and a similar dependency on lithology, climate, and slope in both countries (Figure 2.6). In agreement with previous studies that were reviewed in the Introduction, the calibrated A_{cr} increases and δ_{Ref} decreases when precipitation decreases (arid and semi-arid climates, Figure 2.6.d) for permeable rocks (unconsolidated sediments and carbonated rocks, Figure 2.6.b) and when the slope decreases (consistent with stronger erosive power in steeper watersheds). The calibrated A_{cr} also tends to be larger in Australia (especially for classes with the highest values which permit the largest variability), which can be attributed largely to arid and semi-arid climates (65% of Australia), where rivers are rare and often intermittent (the latter amounting to 69% of total stream length in FAO, and to 98% in Geofabric, Table 2.2).

In the second step, statistical models were developed to produce a set of A_{cr} values for the different environmental classes, to be used to extract river networks worldwide. The rationale was to use the rules emerging from Figure 2.6 regarding the effects of lithology, slope and climate to create consistency between the two countries and define one single A_{cr} value for a given environmental class. In doing so, the effects of latitude on pixel area were considered, and the environmental

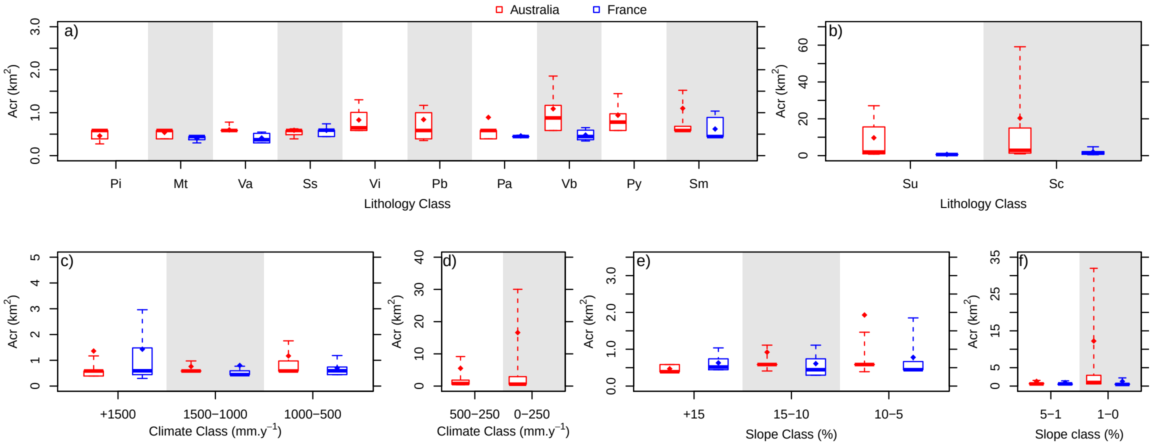


Figure 2.6 – A_{cr} calibration results in Australia (red) and France (blue) per environmental class. a) and b) represent the lithology, c) and d) climate, and e) and f) slope classes. The bottom and top of the boxplots represent the first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

classes were grouped to limit the number of different A_{cr} values.

Two models of different complexity were eventually constructed to evaluate the effects of slope on river network extraction by considering only lithology and climate parameters (LC), and adding the slope (LCS). The corresponding A_{cr} values are given in the Appendix 6.3. The LC model comprises 11 different A_{cr} values varying from 0.3 km^2 to 97 km^2 for 35 environmental classes, derived from 7 lithology groups (Mt+Pi+Pa+Py, Pb, Sc, Su, Sm, Ss+Va, Vb+Vi) and 5 climate classes (Table 2.1). The LCS model has 29 different A_{cr} values varying from 0.3 to 193 km^2 corresponding to 120 environmental classes that combine 6 lithology classes (Mt+Pi+Pa+Py, Pb, Sc, Su, Ss, Sm+Va+Vb+Vi), 5 climate classes, and 4 slope classes (under 1%, between 1 and 5%, between 5 and 10%, and over 10%).

The A_{cr} calibration was performed for France and Australia based on the dominant lithology, total precipitation, and slope (Figure 2.1) of each 7.5' vector grid-cell according to the following steps:

1. Many river networks were extracted using the 15" HydroSHEDS flow directions and flow accumulations and as many A_{cr} values. From 1 to 100 pixels (ca. 0.25

to 25 km^2 at the equator) in increments of 1 pixel were used, and from 100 to 1000 pixels (ca. 25 to 250 km^2) in increments of 5 pixels were used.

2. For each river network extracted in step 1, the drainage densities were calculated as the sum of total river length inside the 7.5' grid-cells divided by its land area.

3. We defined 300 uniform zones corresponding to the 300 environmental classes resulting from the intersection of the 12 lithology classes (keeping only the grid-cells where the dominant lithology covers at least 70% of the grid-cell area), 5 precipitation classes, and 5 slope classes (the classes are described in Table 2.1).

4. In each of these 300 zones with data in the processed country, we selected one A_{cr} value (converted to km^2 using the mean pixel area over France and Australia, i.e., 0.148 km^2 and 0.193 km^2 respectively), minimizing the difference in mean δ between the river networks extracted from HydroSHEDS in step 1 and the reference river network. Thus, the selected A_{cr} corresponds to the smallest drainage density bias in each of the uniform zones.

The A_{cr} calibration results are shown in Appendix 6.2, with 300 values (in km^2) both in France and Australia. Zero A_{cr} corresponds to cases where rivers are absent from all the 7.5' grid-cells of an environmental class, which therefore has a zero drainage density. Lack of values corresponds to environmental classes that are absent in the country.

To produce the resultant global river network and drainage density map, the following steps were performed:

1. For each of the selected models (LC or LCS), a global A_{cr} map was produced by converting the A_{cr} given by the models in km^2 to a number of pixels $n_{cr} = A_{cr}/A_{pixel}$, where A_{pixel} comes from Equation 2.1, and n_{cr} is rounded to the closest integer).

2. The resulting maps from step 1 (Figure 2.7) were used to extract the river

pixels, where local flow accumulation is bigger than the local n_{cr} from the map, the model's A_{cr} in km^2.

3. The river layer from step 2 was further verified to assure that all initiated rivers continued to exist downstream, independent from A_{cr} values. A FORTRAN script was developed to check and correct eventual cuts in the river network due to abrupt changes in A_{cr}.

4. The river layer from step 3 was then classified with Strahler orders using ArcGIS 10.3 hydrology tools.

5. The rivers where converted to vectors with ArcGIS hydrology tools and HydroSHEDS flow directions at 15”.

6. These vectors were then intersected with the 7.5’ grid to split the vectors at the 7.5’ intersections, and their lengths were calculated (in meters) with an ArcGIS python script.

7. For the drainage density, all river lengths inside a 7.5’ grid were summed (spatial join tool), and the result was divided by the grid-cell area recognized as land based on the 15” HydroSHEDS (in m^2, calculated with an ArcGIS Python script).

Drainage density from AQUAMAPS, CCM2 (Colombo *et al.*, 2007), and reference river networks were calculated according to steps 6 and 7 described above, since the original data does not provide this information. Only Luo *et al.* (2016) provided drainage density data for comparison.

The Australian case was used to propose a first-order method to identify intermittent rivers explained by the aridity of climates. The principle is that rivers initiated in regions where precipitation is smaller than a calibrated threshold are defined as intermittent until they reach a pixel where precipitation exceeds this value. Once a stream becomes perennial, all downstream river pixels remain as such (see the Nile in Figure 2.12), so this strategy cannot reproduce cases of disrupted con-

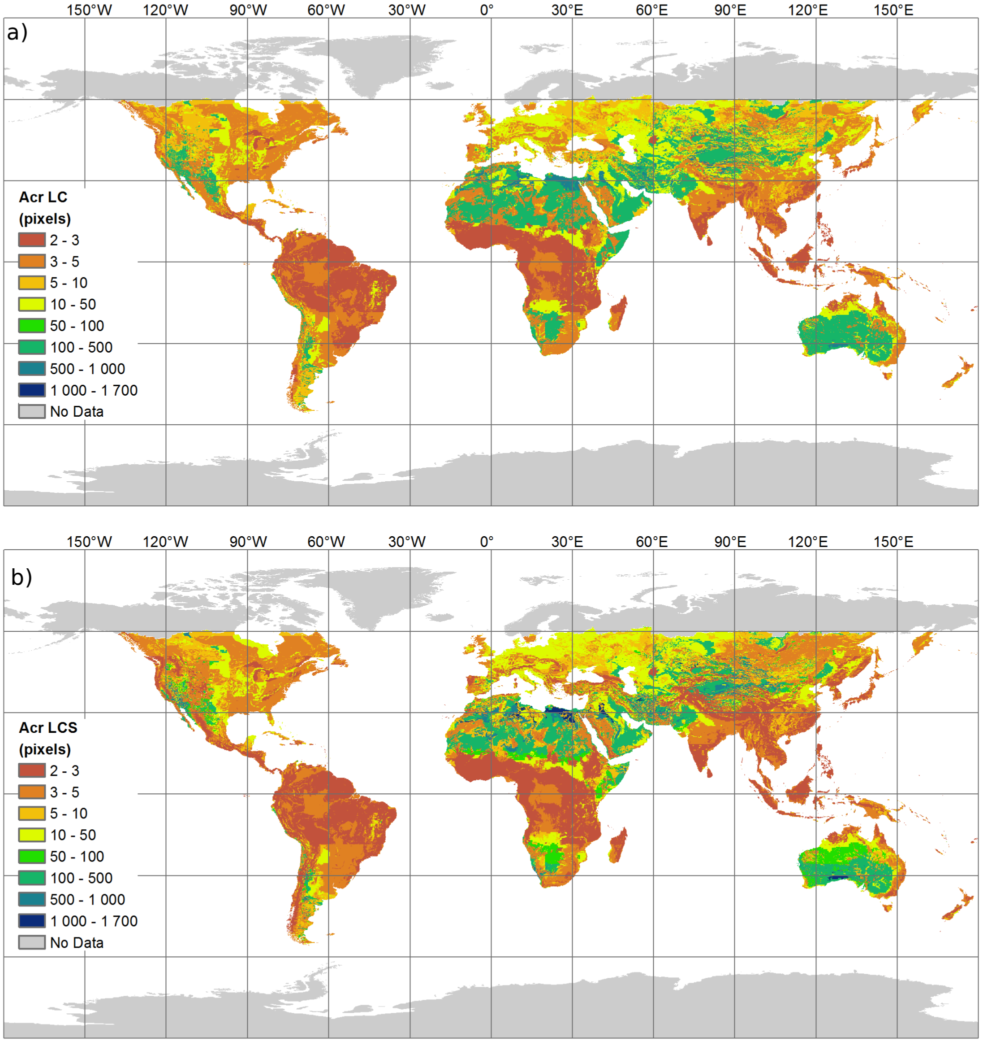

a)
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
60°N
30°N
0°
30°S
60°S
Acr LC
(pixels)
2 - 3
3 - 5
5 - 10
10 - 50
50 - 100
100 - 500
500 - 1 000
1 000 - 1 700
No Data
b)
Acr LCS
(pixels)
2 - 3
3 - 5
5 - 10
10 - 50
50 - 100
100 - 500
500 - 1 000
1 000 - 1 700
No Data

Figure 2.7 – A_{cr} map converted to pixels to extract the river network from a) LC and b) LCS models.

nectivity. The precipitation threshold was calibrated to get the best overlap with areas of high intermittency in FAO, which defines 69% of intermittent rivers in Australia (% of total length). This defined a threshold of 500 mm.y^{-1}, i.e., the classical upper bound of semi-arid climates, according to which 42% of the rivers from the LC or LCS models in Australia are intermittent. This percentage is smaller than in FAO because our models predict higher densities, so the total length of intermittent streams is much higher with our models, ca. 420,000 km in FAO; 1,483,000 km in LCS; 3,242,000 km in Geofabric (Table 2.2). The latter dataset was disregarded as it classifies 98% of streams as intermittent, which exceeds the

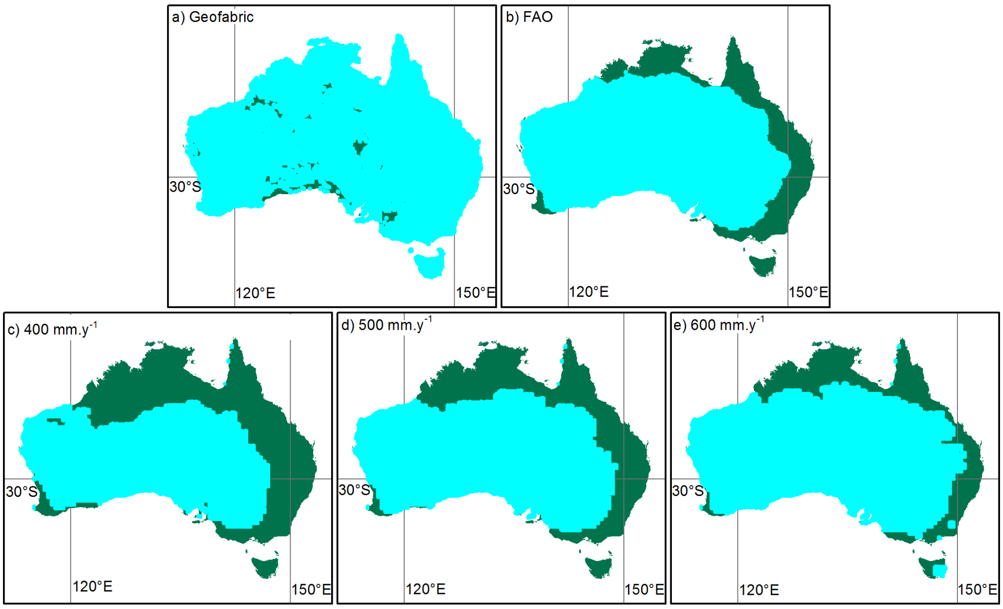


Figure 2.8 – Location of pixels with intermittent rivers (in blue) over Australia from: a) Geofabric, b) FAO; c) LCS with 400 mm.y^{-1} precipitation threshold ; d) LCS with 500 mm.y^{-1} precipitation threshold; e) LCS with 600 mm.y^{-1} precipitation threshold. See complementary analysis in Table 2.2.

maximum of 90% used in Raymond *et al.* (2013). Matching this 98% with our uniform precipitation threshold would also constrain the predicted perennial streams to very humid climates (annual mean precipitation > 1500 mm.y^{-1}).

Table 2.2 – Intermittency analysis. Spatial correlations (Corr. to FAO) are defined based on the presence/absence of intermittent streams in the 7.5' pixel

	Australia				
	Geofabric	**FAO**	**LCS 400 mm.y^{-1}**	**LCS 500 mm.y^{-1}**	**LCS 600 mm.y^{-1}**
% in length	98	69	34	42	53
% in area	100	74	54	66	74
Corr. to FAO	-	-	0.68	0.77	0.76
δ_{int} (km^{-1})	0.42	0.06	0.15	0.19	0.24
L_{int} (10^6km)	3.31	0.42	1.16	1.48	1.83
	Global				
% in length	-	34	-	29	36
% in area	-	36	-	39	48
Corr. to FAO	-	-	-	0.65	0.54
δ_{int} (km^{-1})	-	0.03	-	0.22	0.29
L_{int} (10^6km)	-	3.09	-	25.1	31.4

2.4 Model Results and Discussion

The resulting drainage densities were first evaluated against the four national reference hydrography datasets (Figure 2.9, Table 2.3). The LC model captures the main features of the reference δ, in particular the low values characterizing areas with carbonate rocks in France, and arid to semi-arid climates in Australia.

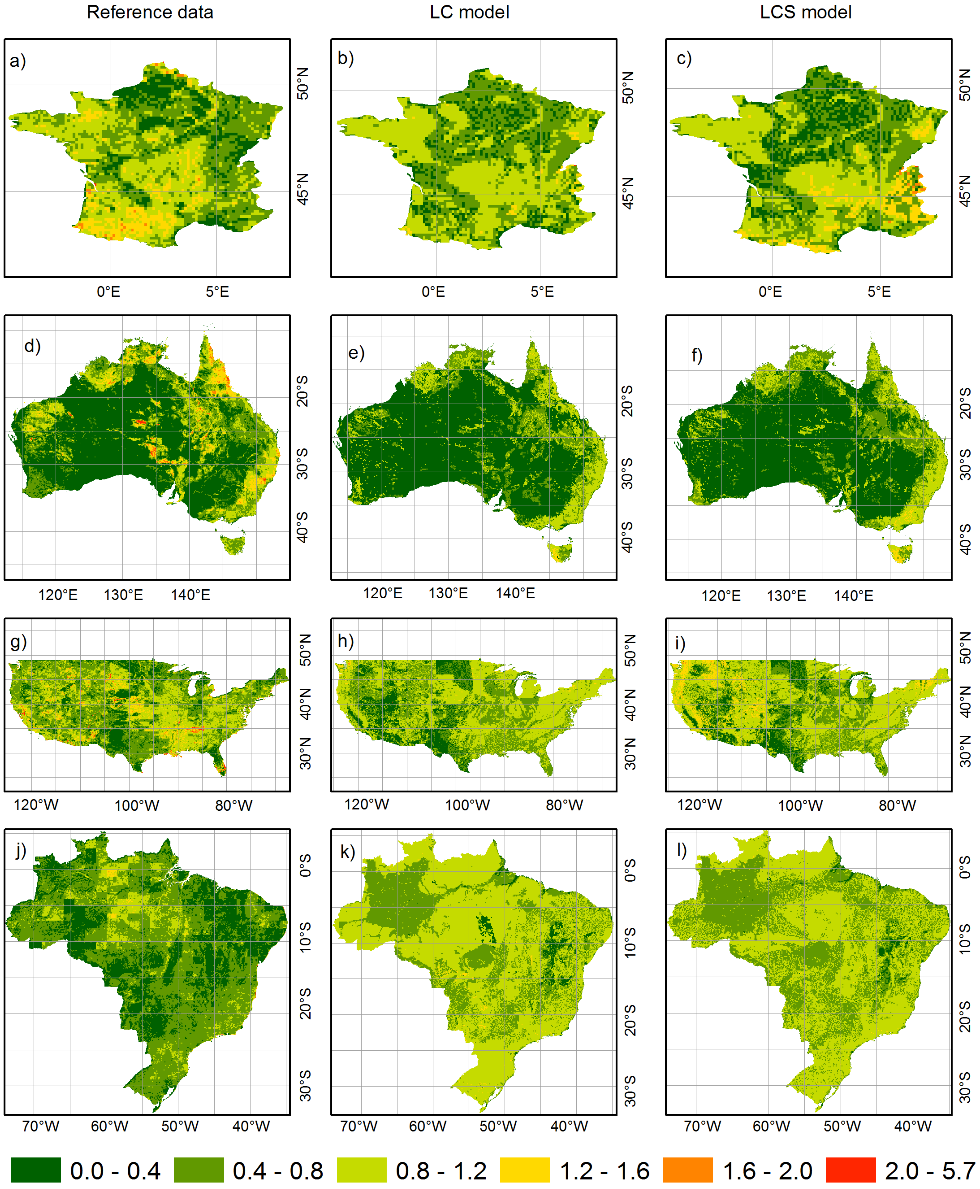

Reference data
LC model
LCS model
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
50°N
45°N
0°E
5°E
20°S
30°S
40°S
120°E
130°E
140°E
30°N
40°N
120°W
100°W
80°W
0°S
10°S
70°W
60°W
50°W
40°W
0.0 - 0.4
0.4 - 0.8
0.8 - 1.2
1.2 - 1.6
1.6 - 2.0
2.0 - 5.7

Figure 2.9 – Maps of drainage density (km^{-1}) for the reference river networks and the evaluated models in France (a,b,c), Australia (d,e,f), US (g,h,i), and Brazil (j,k,l): a,d,g,j) Reference data; b,e,h,k) LC; c,f,i,l) LCS.

However, it underestimates the high δ values, mostly found in mountainous areas in both countries. These are better represented by LCS, which accounts for the increase of δ with slope, but this model still underestimates some very high values, such as in Australia, where siliciclastic rocks are present (Ss in Figure 2.1a), and

Table 2.3 – Statistical parameters for the comparison of the calculated and reference drainage densities δ per country[1]. Ref. = δ from river network reference data (AU = Geofabric and FR = CARTHAGE used for calibration, US = NHD and BR = IBGE (2015) used for validation); Others = δ from Luo *et al.* (2016) over the US and from CCM2 (Vogt *et al.*, 2007) over France and Europe; Mean = mean δ (km^{-1}); Std.Dev. = standard deviation of δ (km^{-1}); Bias = bias between model and Ref (km^{-1}); Corr. = correlation coefficient between model and Ref (dimensionless); RMSE = root mean square error between model and Ref (km^{-1}). % Int. corresponds to the ratio of total length of intermittent rivers to the total river length.

Region	Parameter	FAO	LC	LCS	Ref.	Others
AU	Mean	0.08	0.42	0.45	0.43	-
	Std.Dev.	0.07	0.32	0.32	0.43	-
	Bias	-0.36	-0.01	0.02	0.00	-
	Corr.	-0.01	0.57	0.59	1.00	-
	RMSE	0.56	0.36	0.34	0.00	
	Mean	0.08	0.68	0.71	0.75	0.34
Std.Dev.	0.08	0.24	0.30	0.36	0.18	
FR	Bias	-0.66	-0.07	-0.04	0.00	-0.41
	Corr.	0.01	0.30	0.32	1.00	0.19
RMSE	0.81	0.40	0.39	0.00	0.56	
US	Mean	0.07	0.72	0.78	0.75	1.30
	Std.Dev.	0.07	0.28	0.30	0.37	0.84
	Bias	-0.68	-0.03	0.03	0.00	0.55
	Corr.	0.09	0.09	0.11	1.00	0.24
	RMSE	0.77	0.47	0.47	0.00	1.00
BR	Mean	0.08	0.93	0.92	0.52	-
	Std.Dev.	0.07	0.16	0.16	0.29	-
	Bias	-0.44	0.41	0.40	0.00	-
	Corr.	0.12	0.14	0.14	1.00	-
	RMSE	0.49	0.49	0.48	0.00	-
Europe	Mean	0.08	0.59	0.63	-	0.31
	Std.Dev.	0.08	0.29	0.32	-	0.15
Global	Mean	0.08	0.69	0.74	-	-
	Std.Dev.	0.08	0.32	0.34	-	-
	% Int.	34	27	29	-	-

in the south-western part of France (Landes), even though it is a rather flat area (Figure 2.1c) with permeable rocks (Su, Sc and Sm in Figure 2.1a). The latter is likely due to artificial drainage densities in the reference data because of irrigation ditches in this area of intensive agriculture.

Similar behaviors are found in the US and Brazil, where the inclusion of a slope constraint in the LCS model also leads to slightly better results than LC, but both models exhibit poor correlations with the reference datasets, as quantified in Table 2.3. Nevertheless, the excessive underestimation of δ by FAO is markedly improved by the variable A_{cr} models and general spatial patterns are well depicted. In the US, France, and Australia, the biases approach zero (less than 5% absolute value). They remain negative over France, which is mainly attributed to the overestimation of the highest δ_{Ref} compared to a "pristine" case because of human-made networks, as discussed previously. In Brazil, in contrast, the proposed method leads to high positive biases; this likely results from the reference river network being based on multiple scaled data, up to 1:250,000, which alters the natural variability of δ (Figure 2.9) and leads to underestimation of δ_{Ref} compared to what would prevail at the 1:50,000 scale used for the A_{cr} calibration in France and Australia. The incon-

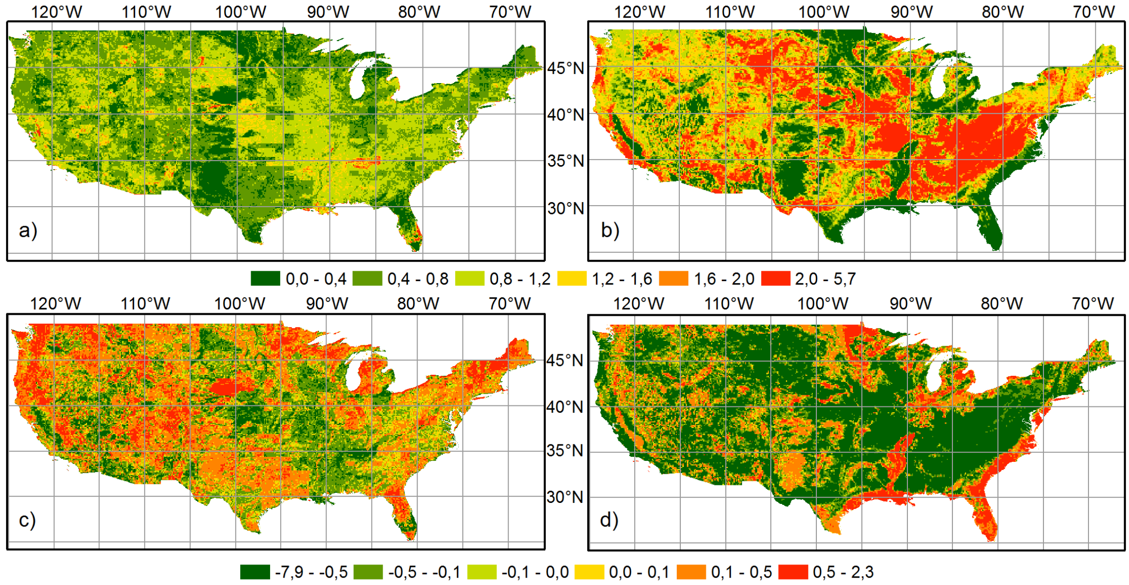


Figure 2.10 – Maps of drainage density (km^{-1}) over the US: a) Reference; b) Luo *et al.* (2016); c) LCS - Reference; d) LCS - Luo *et al.* (2016).

sistencies of the reference river networks largely explain the poor spatial fit to δ_{Ref} as revealed by the correlation coefficients and RMSE.

These inconsistencies prevented Luo *et al.* (2016) from analyzing the relationships between drainage densities from NHD and possible explanatory factors and led them to construct their own drainage density map. Table 1 shows that this latter map overestimates the mean δ compared to NHD and our two estimates, which suggests that the 30-m DEM of Luo *et al.* (2016) corresponds to a finer scale than both 1:24,000 and 1:50,000. As expected, the correlation coefficients increased (doubled, as $Corr_{LCS,Luo} = 0.24$) when comparing δ of LC and LCS model to δ_{Luo} rather than δ_{Ref} (Figure 2.10, see also the map of δ_{LC} in Appendix 6.4). However, this correlation remains low, and the weak performance of the proposed models in the US could be due to the quality of the lithological map (Figure 2.1a), which explains the main discontinuities along the Great Plains from North Dakota to Texas, and around Lake Michigan (anticlockwise from Wisconsin to Michigan). It is worth noting that the geological map of Schruben *et al.* (1994) does not exhibit these discontinuities.

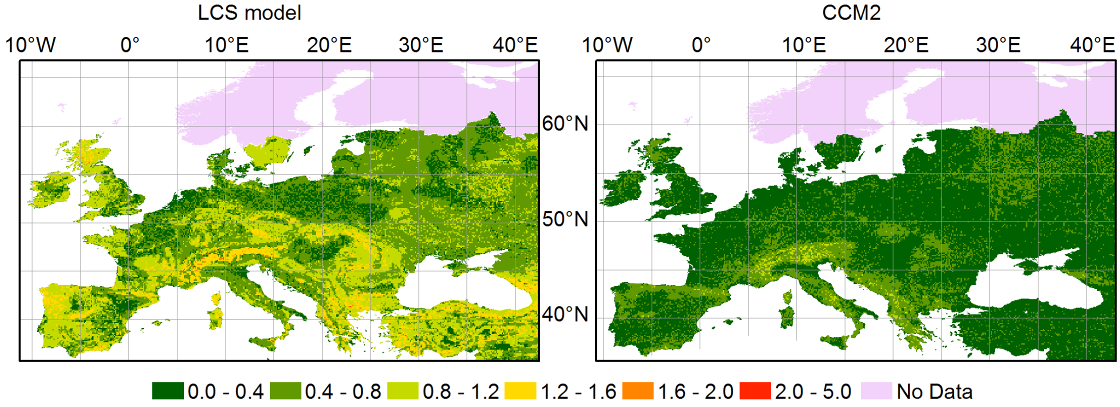


Figure 2.11 – Maps of drainage density (km^{-1}) over Europe for the LCS model (left) and CCM2 (Vogt *et al.*, 2007) (right).

The proposed δ were also compared to values extracted from Vogt *et al.* (2007), which exhibited weak spatial variations throughout Europe, strongly controlled by topography (Figure 2.11) with values below 0.4 km^{-1} in most lowland areas, which is below the range of 0.5 to 1.6 km^{-1} from the small scale studies reported in the Introduction. As a result, δ_{Ref} is much closer to δ from LC and LCS than to δ_{CCM2}

over France (Table 2.3), both in terms of spatial match and bias.

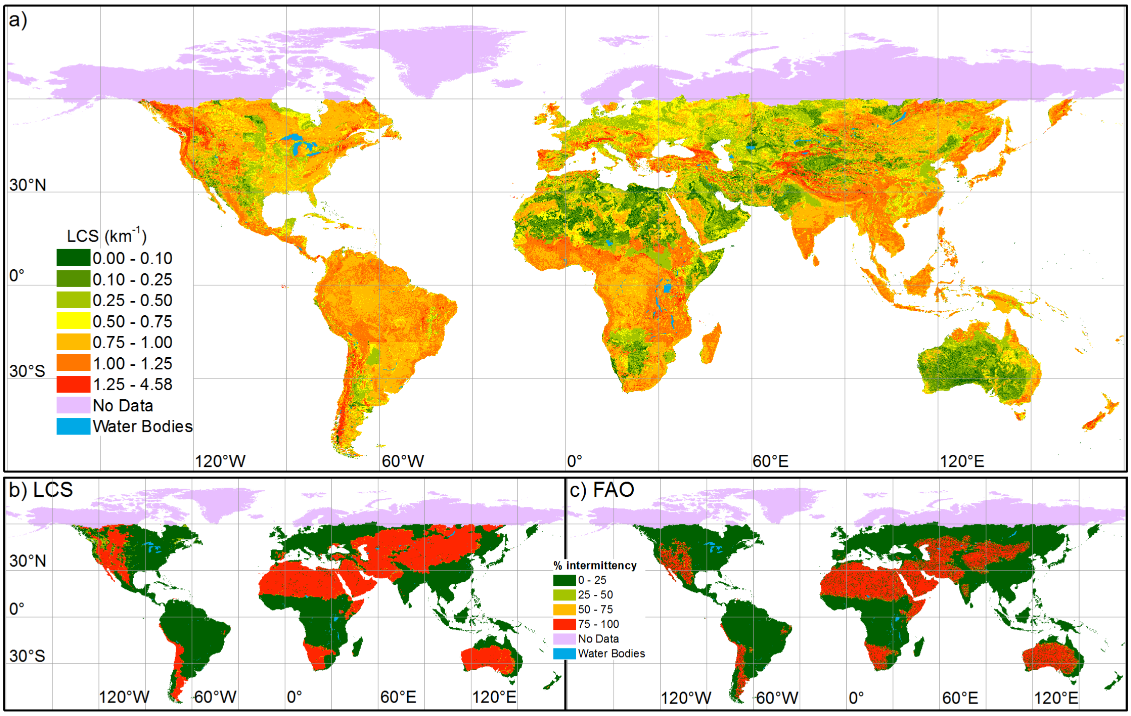

a)
30°N
0°
30°S
LCS (km^{-1})
0.00 - 0.10
0.10 - 0.25
0.25 - 0.50
0.50 - 0.75
0.75 - 1.00
1.00 - 1.25
1.25 - 4.58
No Data
Water Bodies
120°W
60°W
0°
60°E
120°E
b) LCS
c) FAO
% intermittency
0 - 25
25 - 50
50 - 75
75 - 100
No Data
Water Bodies

Figure 2.12 – Global maps of drainage density (km^{-1}) for LCS model: a) for full network including intermittent streams; b) intermittent streams only. Percentage of intermittent streams: c) LCS; d) FAO. In each 7.5' cell, the % of intermittent streams is the ratio of the intermittent stream length to total stream length. The violet color north of 60°N indicates where HydroSHEDS data are missing.

Figure 2.12a shows the global drainage density map from the LCS model, which exhibits complex patterns arising from the combined controls by lithology, climate, and slope (see also LC and FAO in Appendix 6.4). The perennial streams show the same density patterns (Figure 2.12b), but are absent from arid and semi-arid areas (39% of global land based on our precipitation dataset), except for rivers initiated as perennial which conserve this feature when crossing arid regions, as the Nile for instance. By design, the complementary intermittent streams are concentrated in arid and semi-arid areas both in FAO and LCS (Figures 2.12c,d), with a rather similar proportion of the full network (around one third, Table 2.3), and a correlation coefficient of 0.65 over land. The main difference is found north of 45°N (North America and eastern Asia), with a larger fraction of intermittent streams according to LCS than FAO. This is consistent with the use of an aridity index combining both precipitation and evapotranspiration to discriminate intermittent streams in FAO/AQUAMAPS. Note the total length of intermittent streams is much higher with our models than with FAO (Table 2.2), primarily due to differences in drainage density.

2.5 Conclusions

This study presents a method to obtain multiple critical drainage areas (A_{cr}) that are spatially dependent on lithology, climate, and slope. This new method addresses an important component of the large scale river delineation process, in combination with proper DEM hydrologic conditioning, obtained here from HydroSHEDS. The A_{cr} values were calibrated against national hydrography data at a 1:50,000 scale over France and Australia, resulting in two models of increasing complexity: LC (using 7 lithology, and 5 climate classes), and LCS (using 6 lithology, 5 climate, and 4 slope classes). This work is based on the 15" hydrologically conditioned version of HydroSHEDS, but the proposed A_{cr} values are a priori suitable to constrain river network extractions from any DEM with similar or higher resolutions.

Both models show fair performance compared to the reference river networks, with better agreement in the countries used for calibration. The inclusion of slope in the model improves the performance criteria in the evaluated countries, but the effect is modest. Combined with an intermittency assessment solely based on mean precipitation, the proposed variable A_{cr} models give similar locations and percentages of intermittent streams as FAO/AQUAMAPS, but with higher and more spatially variable drainage densities. The limitations of this first-order classification underline the need for better description of the multiple controls of intermittency.

The two proposed models predict global mean drainage density to reach ca. 0.70 km^{-1}, with a precision of approximately 5% compared to the reference data. Drainage density and scale are tightly linked (Tarboton *et al.*, 1992), so the proposed value is consistent with the 1:50,000 scale that prevailed for A_{cr} calibration. It is also higher than the mean densities derived from classical single A_{cr} river networks, which are thus shorter and should be used with caution for fine scale applications, as previously reported by Raymond *et al.* (2013).

As previously discussed, the quality of A_{cr} calibration and validation in our methodology strongly depends on the reference data. The main uncertainties seem to come from unnatural channels and scale inconsistencies. The quality of the input DEM and flow accumulations is also important, although this is difficult to evaluate. As shown, environmental input parameters can induce uncertainties, particularly with respect to the lithological map. As a result, any improved lithology, either at the global scale or over a specific region, could assist with the estimation of drainage density, provided calibration is updated accordingly. Eventually, another way to improve drainage density estimates would be to include more control factors at the calibration step, e.g. more complex geomorphologic information or vegetation parameters (Colombo *et al.*, 2007; Luo *et al.*, 2016), which were not addressed here for the sake of simplicity. Another approach is to use hyper-resolution hydrological modeling to define the locations where streams initiate, as recommended by Lehner and Grill (2013), and recently achieved by Maxwell *et al.* (2015) over the United States based on the HydroSHEDS DEM. Land use information could also be used to generate artificialized river networks if the reference data for calibration includes information on anthropogenic pressures. However, when dealing with a surveyed river network with no such information, the drainage density difference with the LCS network may be used as a first-order indicator of anthropogenic impacts.

Taking the above caveats into consideration, river networks and corresponding drainage density maps that can be constructed from the proposed A_{cr} models have the advantage of describing the main heterogeneities of natural river networks with a uniform scale of 1:50,000 across all continents. This feature is important to support water management in regions with limited observations and to provide consistent information to large-scale models seeking higher resolution, which is an important evolution of both land surface and hydrological models (Wood *et al.*, 2011).

CHAPTER 3

ESTIMATION OF THE BASE FLOW TIME CONSTANT

Contents

3.1 Introduction 46
3.2 Data description and selection of parameters 50
  3.2.1 Porosity 50
  3.2.2 Transmissivity 53
  3.2.3 Slope 60
  3.2.4 Drainage density 61
  3.2.5 Selection of the parameters 62
3.3 Base flow time constant results and discussion 63
3.4 Conclusions 68

Scope

Base flow time constant is used to estimate base flow from analytical solutions of the Boussinesq equation in the linearized form, and it depends on effective porosity, aquifer transmissivity, and drainage density. Base flow time constant (τ) is overestimated using the available global hydrogeological data compared to references obtained from observed river discharge at the basin scale for recession analysis, raising the research question: *Which parameters exert a strong control on τ?* In this chapter a sensitivity analysis to the parameters that control τ is performed showing that the main uncertainty of τ is linked to drainage density and hydraulic conductivity. τ is reduced by two orders of magnitude in average by using the drainage density calculated in chapter two, with the effective porosity values from de Graaf *et al.* (2015), and a combination of hydraulic conductivities from Reynolds *et al.* (2000) and Gleeson *et al.* (2014).

3.1 Introduction

Base flow estimation depends on several assumptions and boundary conditions due to the non-linearity of groundwater flow equations, as described in the introduction. One common approach to estimate base flow relies on linearization of the Boussinesq equation, resulting in a solution that represents a direct link between the physical mechanisms of the Boussinesq equation and the mathematical aspects of linear systems (Brutsaert, 2005). The linearization considers small water table variations over time, in which the water table remains close to a constant average value (Brutsaert, 2005). Base flow estimation from the linearized form of the Boussinesq equation can have different formulations based on the assumptions made, as for unconfined, horizontal or sloping aquifers, and if the system is in steady (with no variations of the storage over time) or in transient state (as after a sudden hydraulic change). Brutsaert and Nieber (1977) showed that base flow from a horizontal, unconfined aquifer fully discharging into a stream follow a general power function:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -aQ^b, \tag{3.1}$$

where Q ($\mathrm{m}^3.\mathrm{s}^{-1}$) represents base flow, t is time (s), and a (s^{-1}) and b (dimensionless) are constants. When integrating equation 3.1 for a linear reservoir ($b = 1$), the base flow is given by:

$$Q = Q_0 e^{-at}, \tag{3.2}$$

and the base flow time constant is given by the inverse of the constant a ($\tau = a^{-1}$). In transient states, this time constant is related to the time to reach equilibrium, characterizing the base flow rate of recession (Erskine and Papaioannou, 1997; Eng and Milly, 2007).

As described in Chapter 1.2, Brutsaert (2005) classified the base flow recession into short and long-term. The short-term is observed in the beginning of the recession, defined by $t < 0.5\tau$ (Brutsaert, 2005), and is followed by the long-term regime that Rousseau-Gueutin *et al.* (2013) called near-steady state. The long-term

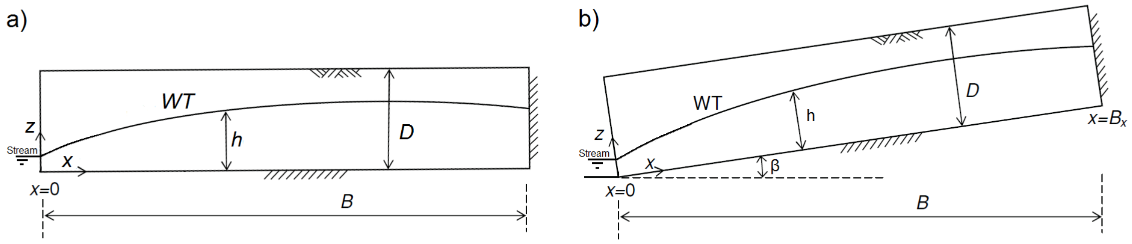


Figure 3.1 – Aquifer scheme proposed by a) Brutsaert (2005) and b) Zecharias and Brutsaert (1988). WT = water table; D = aquifer depth; B = aquifer length; h = water table elevation; θ = aquifer slope; z and x are the vertical and horizontal axes, respectively.

solution considers that the water table has a curvilinear shape observed only after drainage has occurred for some time, without any major water table variations over time (constant water table depth). For the long-term solution of an unconfined, homogeneous, isotropic aquifer represented by Figure 3.1, and using drainage density (δ) to obtain the mean aquifer length ($\delta = 1/2B$) proposed by Horton (1945), τ is given by equation 3.3 (Brutsaert, 2005):

$$\tau = \frac{n_e}{\pi^2 \delta^2 T}, \tag{3.3}$$

where T is the aquifer transmissivity ($\mathrm{m}^2.\mathrm{s}^{-1}$), and δ is the drainage density (m^{-1}). Zecharias and Brutsaert (1988) developed a solution for the Boussinesq equation in the linearized form for homogeneous, isotropic, unconfined, sloping aquifers at quasi-steady state (Figure 3.1.b), resulting in an equation to describe base flow that depends on the same parameters of Equation 3.3, but also on the aquifer slope. Dupuit-Forchheimer assumptions used by Zecharias and Brutsaert (1988) do not introduce significant errors when aquifer slopes are lower than 20% Dingman (2015). For a long-term solution, Zecharias and Brutsaert (1988) obtained the base flow time constant (τ_{Zech}) as:

$$\tau_{Zech} = \frac{n_e}{8\delta^2 T \cos^2\theta + 4\delta K \sin\theta}, \tag{3.4}$$

where θ is the aquifer slope ($\mathrm{m}.\mathrm{m}^{-1}$). To obtain the formulation for the horizontal case (Equation 3.3) as the limit of the slope formulation when θ tends to zero, i.e.,

τ_{Zech} for $\theta = 0$ equals τ with Equation 3.3, it was assumed the factor 8 in Equation 3.4 is approximately π^2, resulting in:

$$\tau = \frac{n_e}{\pi^2 \delta^2 T \cos^2\theta + \frac{\pi^2}{2} \delta K \sin\theta} \tag{3.5}$$

The use of Equation 3.5 results in smaller values of τ compared to Equation 3.4 (the proposed formulation results in $\tau = 0.81\tau_{Zech}$). The chosen formulations to estimate τ for the long-term regime were Equation 3.5 for the sloping aquifer, and Equation 3.3 for the horizontal case.

Zecharias and Brutsaert (1988) estimated τ for nineteen small basins in the United States using Equation 3.4 and two methodologies: the graphical method based on Equation 3.1 (that depends on observed river discharges), and the Equation 3.5. They found τ values ranging from 4 to 10 days using the graphical method, against 42 to 248 days for the analytical one. This difference resulted from the subjectiveness inherent to the graphical method and the assumptions of the physically-based equation. The uncertainties associated to the graphical method are related to the choice of periods with low flow, and to the discontinuity of discharge measurements (Vogel and Kroll, 1995). The physically-based equation method considers a homogeneous and isotropic aquifer, which neglects natural heterogeneities and preferential flow paths (as in fractured aquifers). Zecharias and Brutsaert (1988) used hydraulic conductivities obtained from laboratory tests, which are not representative for the regional scale use on τ estimation. Based on the graphical method, Eng and Milly (2007) estimated τ for 93 basins in the United States (with area ranging between 4 and 829 km^2), arriving to the same order of magnitude (5 to 44 days) observed by Zecharias and Brutsaert (1988) using the same approach (Table 3.1).

Brandes *et al.* (2005) estimated τ using the graphical method from river discharge measurements during recession, in which τ remained under two months. Brandes *et al.* (2005) showed that drainage density is the most important parameter on τ estimation, as river emergence is related to both aquifer permeability and geometry within a watershed. τ remained under two years for all references presented in Table

3.9. Gascoin *et al.* (2009) calibrated τ by comparing simulated river discharge and total runoff to observations at the Somme river basin in France, obtaining τ of 700 days.

Table 3.1 – Summary of base flow time constant (τ) range (days) selected from the literature for recession analysis studies using graphical method. * Analytical method derived from the Boussinesq equation. ** Model calibration results.

Reference	τ range (days)	Area range (km^2)	Location
Zecharias and Brutsaert (1988)	4 - 10	17 - 205	United States
Brandes *et al.* (2005)	4 - 37	20 - 120	United States
Eng and Milly (2007)	5 - 44	-	United States
Brutsaert (2008)	37 - 46	25 - 69	United States
Ries and Eng (2010)	25 - 59	5 - 80	United states
Jimenez-Martinez *et al.* (2013)	20 - 76	-	United States
Botter *et al.* (2013)	2 - 25	3 - 1,766	United States
Thomas *et al.* (2013)	4 - 32	14 - 906	United States
Stoelzle *et al.* (2013)	62 - 342	26 - 954	Germany
Zecharias and Brutsaert (1988)	42 - 238*	17 - 205	United States
Gascoin *et al.* (2009)	700**	5,566	France

Graphical methods to estimate τ depend on longterm and uninterrupted river discharge observations that are not available at global scale. For analytical solutions as shown in Equations 3.5 and 3.3, τ estimation relies on a good characterization of the watershed. As shown in the introduction of this thesis, τ is overestimated by five orders of magnitude in average by using available global data with Equation 3.3. Brutsaert (2008) showed that one factor that can lead to overestimated τ is porosity, as he obtained τ longer than 500 days (in the same expected range shown in Table 3.9) by using effective porosity values from Johnson (1967), raising the research question: *Which parameters exert a strong control on* τ*?*

In this chapter, base flow time constant is estimated with the long-term solution of the linearized form of the Boussinesq equation for homogeneous, isotropic, unconfined aquifers at global scale. Two cases are considered for slope, in which for horizontal (slope is zero) aquifers τ is estimated from Equation 3.3, while for sloping aquifers Equation 3.5 is used. The effect of the scale is evaluated by using two resolutions (7.5’ and 0.5°), from available global datasets described in detail in section 3.2. A sensitivity analysis is performed in section 3.3 for all parameters (porosity, transmissivity, saturated thickness, and drainage density), and τ results are compared to reference data presented in Table 3.9.

3.2 Data description and selection of parameters

The parameters used to estimate τ were converted to 7.5’ and 0.5° resolutions, where each grid-cell contains: the weighted arithmetic mean for porosity and slope, the weighted geometric mean for hydraulic conductivity (used to calculate transmissivity), and the calculated drainage density (sum of total river length inside the grid-cell by its area), as described in details in the following subsections.

3.2.1 Porosity

By definition, porosity is a property that describes the amount of void spaces in a material, and it can be divided into total porosity (n, representing the total amount of voids) and effective porosity (n_e, porosity available for fluid flow), i.e., effective porosity will be always less or equal to total porosity. However, no effective porosity data are currently available at global scale, despite the recent efforts of producing high resolution data such as Gleeson *et al.* (2011a) for global hydraulic conductivity, and the GLHYMPS dataset for global permeability and total porosity (Gleeson *et al.*, 2014). Gleeson *et al.* (2011a) classified similar geologic units (that have close permeability values) from several available geologic maps and hydrogeological models into seven hydrolithologic categories. Each hydrolithology contains a single permeability value. GLHYMPS total porosity data was produced by associating the hydrolithologies classes of Gleeson *et al.* (2011a) with literature values (small scale laboratory measurements from Morris and Johnson (1967)) of total porosity. Gleeson *et al.* (2014) updated the hydrolithologies categories by using the global lithology map of Hartmann and Moosdorf (2012) (detailed in Chapter 2), with a finer division based on the grain size for unconsolidated and consolidated siliciclastic sediments, and identification of the permafrost areas from Gruber (2012) data. GLHYMPS resulted in total porosity data representing the average values of each hydrolithology class up to 100 m depth.

de Graaf *et al.* (2015) presented a table with a single specific yield (S_y) value for each hydrolithology identified by Gleeson *et al.* (2011a). For unconfined aquifers, S_y can be approximated to n_e. A global effective porosity data was created by associating each S_y from de Graaf *et al.* (2015) to the correspondent hydrolithology from Gleeson *et al.* (2014). The resultant n and n_e per hydrolithology is shown in Table 3.2, showing that the n_e from de Graaf *et al.* (2015) is not always coherent with the expected values for each class. For unconsolidated and carbonate rocks n_e has higher values than n, possibly explained by the large range of values that can be found for this class. From Table 1.2, n of carbonate rocks are usually in average 0.30 (varying from 0.07 to 0.56), and 0.14 for S_y (range from 0.02 to 0.36), suggesting

Table 3.2 – Porosity values found in the literature classified by hydrolithology classes of Gleeson *et al.* (2014). Lithology classes come from Hartmann and Moosdorf (2012) (Pi = Intermediate plutonic rocks; Mt = Metamorphic rocks; Va = Acid volcanic rocks; Ss = Siliciclastic sedimentary rocks; Vi = Intermediate volcanic rocks; Pb = Basic plutonic rocks; Pa = Acid plutonic rocks; Vb = Basic volcanic rocks; Py = Pyroclastics; Sm = Mixed sedimentary rocks; Su = Unconsolidated sediments; Sc = Carbonate sedimentary rocks; Ev = Evaporites; Wb = Water Bodies; Ig = Ice and Glaciers; Nd = No Data). Unconsolidated and siliciclastic sedimentary represent the average values of c.g. (coarse grained) and f.g. (fine grained).

Hydrolithology	Gleeson *et al.* (2014)		de Graaf *et al.* (2015)	
	Lithology	n_t	Lithology	n_e
Unconsolidated	Su, Wb	0.220	Su	0.235
c.g. unconsolidated	Su	0.280	Su	0.360
f.g. unconsolidated	Su	0.150	Su	0.110
Siliciclastic sedimentary	Ss, Sm	0.190	Ss	0.055
c.g. siliciclastic sedimentary	Ss, Sm	0.270	Ss	0.100
f.g. siliciclastic sedimentary	Ss, Sm, Ev	0.120	Ss	0.010
Carbonate	Sc	0.060	Sm, Sc, Ev	0.140
Crystalline	Mt, Pa, Pb, Pi	0.010	Va, Vb, Vi	0.010
Volcanic	Va, Vb, Vi, Py	0.090	Mt, Pa, Pb, Pi	0.050
Not assigned	Nd, Ig	-	Nd, Ig	-

that Gleeson *et al.* (2014) n are maybe underestimated for carbonate sedimentary rocks. For unconsolidated rocks the major inconsistency is due to the coarse grained values. When comparing with coarse grained unconsolidated rocks of Table 1.2, they both fit into the range 0.24 - 0.46 for n, and 0.13 - 0.43 for S_y. Even if the values of de Graaf *et al.* (2015) for S_y per hydrolithology are coherent with literature values, the lithology classes that compose each hydrolithology worth attention. The inclusion of mixed sedimentary rocks and evaporites rocks in carbonate category (instead of siliciclastic sedimentary); or acid, basic and intermediate volcanic rocks in crystalline category (instead of volcanic); and metamorphic and plutonic rocks in volcanic category (instead of crystalline) suggest a mistype error by the author for the lithology classes, supported by the fact that they use the same hydrolithology classes as Gleeson *et al.* (2011a).

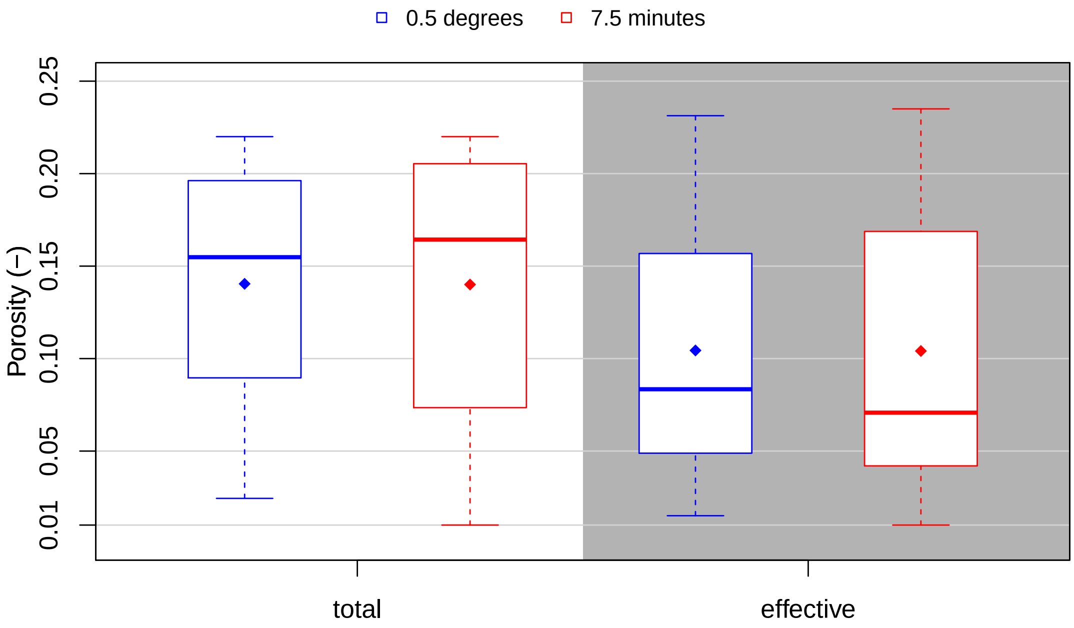


Figure 3.2 – Total (white) and effective (gray) porosity results at 0.5° (blue) and 7.5' (red) resolution. The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

Total and effective porosities shown in Table 3.2 were converted to the 7.5' and 0.5° grid, with a single value of porosity per grid-cell (assuming homogeneous and isotropic aquifers) that corresponds to the weighted arithmetic mean (by area).

Although for carbonate and unconsolidated sedimentary rocks n_e have higher values than n_t, n_e is smaller in global average than n_t for both resolutions (Figure 3.2). As expected, the 7.5' resolution resulted in wider range for both n_e and n_t, since coarser resolutions tend to average the results by the methodology used.

Effective porosity can vary not only for different hydrolithologies (as shown in Table 3.2) but also within a same hydrolithology (as in the case of a fractured aquifer), being one of the reasons that both Gleeson *et al.* (2014) and de Graaf *et al.* (2015) used values from small scale laboratory results to estimate total porosity and specific yield. However, the use of laboratory results are not always representative for regional scale measurements, as Brutsaert (2008) showed that they can overestimate effective porosity by one order of magnitude.

3.2.2 Transmissivity

As stated in the introduction, transmissivity represents the rate that groundwater can flow horizontally through an aquifer, and it is equal to the product of hydraulic conductivity and saturated thickness. The saturated thickness was estimated from the aquifer depth by an equation that arose from the linearization process that assumes rather small variations of the water table level over time, which allows linearizing the diffusivity (Brutsaert, 2005):

$$e = pD, \tag{3.6}$$

where e is the saturated thickness (m), D is the active aquifer depth (m), and p is a parameter used to compensate for the linearization ($0 < p \leq 1$, dimensionless). As stated by Rupp and Selker (2006), linear equations are appropriate when changes in the water table height are small (Figure 3.4.a), which is observed for long-term solutions as used in Equation 3.3 and 3.5. For the assumptions used on τ estimation in this thesis, and because there is not yet a theory for determining p, the saturated thickness (e) was estimated using $p = 0.3465$ (Brutsaert and Nieber, 1977; Brutsaert, 2005; Rupp and Selker, 2006; Brutsaert, 2008).

The aquifer depth D (Figure 3.1) is not an easy parameter to determine at large scales (Yeh and Eltahir, 2005), since it depends on the knowledge of initial water table depth, and the depth to bedrock. Recent studies tried to estimate the depth to bedrock, as in Pelletier *et al.* (2016) and Shangguan *et al.* (2017); however they have significant uncertainties associated to the methodology used. In Pelletier *et al.* (2016) the depth to bedrock data does not include regolith thickness due to significant uncertainties associated with this data. Another uncertainty comes to the fact that Pelletier *et al.* (2016) fixed the limit of bed to bedrock to 50 m, even when observations exceeded this value. Shangguan *et al.* (2017) also acknowledge the difficulties of estimating depth to bedrock at large scales due to lack of observations. Shangguan *et al.* (2017) provided a depth to bedrock data that is based on statistical modeling results, which could not be evaluated over Latin America, Asia, and Africa due to the lack of observations in those regions. In this thesis, the definition of bedrock is the bottom impermeable layer of the unconfined aquifer, which is not the case for Pelletier *et al.* (2016) (that fixed a limit of 50 m) and Shangguan *et al.* (2017) (that defined as the consolidated solid rock underlying unconsolidated surface materials). These uncertainties and the different definitions led us to use two aquifer depths to estimate saturated thickness by using a single value over the globe, and per groundwater class defined in the following paragraph, in a way that the impact of this parameter on τ estimation could be evaluated.

Three groundwater classes were used from the World-wide Hydrogeological Mapping assessment Program (WHYMAP) to estimate saturated thickness. WHYMAP was developed to help manage and protect major groundwater resources of the world BGR and UNESCO (2008), and it provides a global aquifer classification, at 1:25,000,000 resolution, into shallow aquifers, complex hydrogeological structures, and major groundwater basins, classified by recharge rates (Figure 3.3). To evaluate the sensitivity of τ to saturated thickness, four values were tested (summarized in Figure 3.4.a):

1. a single aquifer depth of 10 m, chosen arbitrarily to represent the region where the groundwater could interact with the surface, resulting in 3.46 m of satu-

rated thickness;

2. a single aquifer depth of 100 m, as the depth that Gleeson *et al.* (2014) considered to estimate total porosity and hydraulic conductivity. This depth resulted in 34.65 m of saturated thickness, representing deeper groundwater;

3. the association of 20, 50, and 100 m to WHYMAP shallow aquifers, complex hydrogeological structures, and major groundwater basins, respectively. This resulted in three values of e: 6.93 m for shallow aquifers, 17.32 m for complex hydrogeological structures, and 34.65 m for major groundwater basins (represented by orange, green, and blue in Figure 3.4.a);

4. and the combination of 50, 100, and 200 m, resulting in a saturated thickness of 17.32 m for shallow aquifers, 34.65 m for complex hydrogeological structures, and 69.30 m for major groundwater basins (represented by orange, green, and blue in Figure 3.4.a).

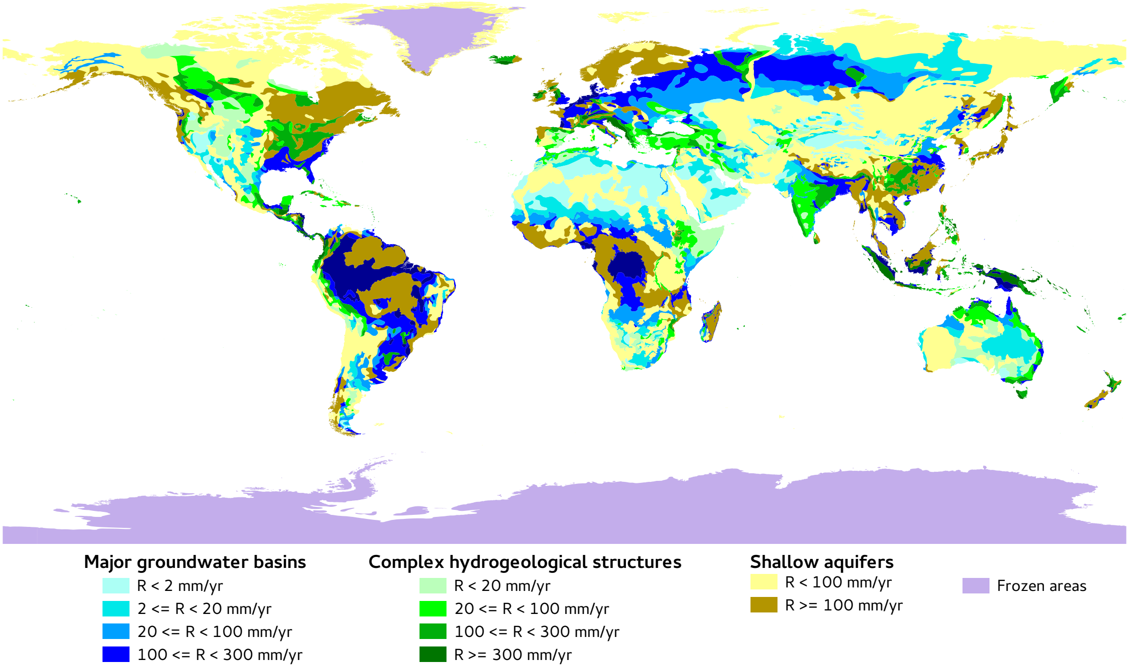

Major groundwater basins
R < 2 mm/yr
2 <= R < 20 mm/yr
20 <= R < 100 mm/yr
100 <= R < 300 mm/yr
Complex hydrogeological structures
R < 20 mm/yr
20 <= R < 100 mm/yr
100 <= R < 300 mm/yr
R >= 300 mm/yr
Shallow aquifers
R < 100 mm/yr
R >= 100 mm/yr
Frozen areas

Figure 3.3 – WHYMAP groundwater classification. R represents the groundwater recharge (in mm.yr^{-1}). Data from BGR and UNESCO (2008).

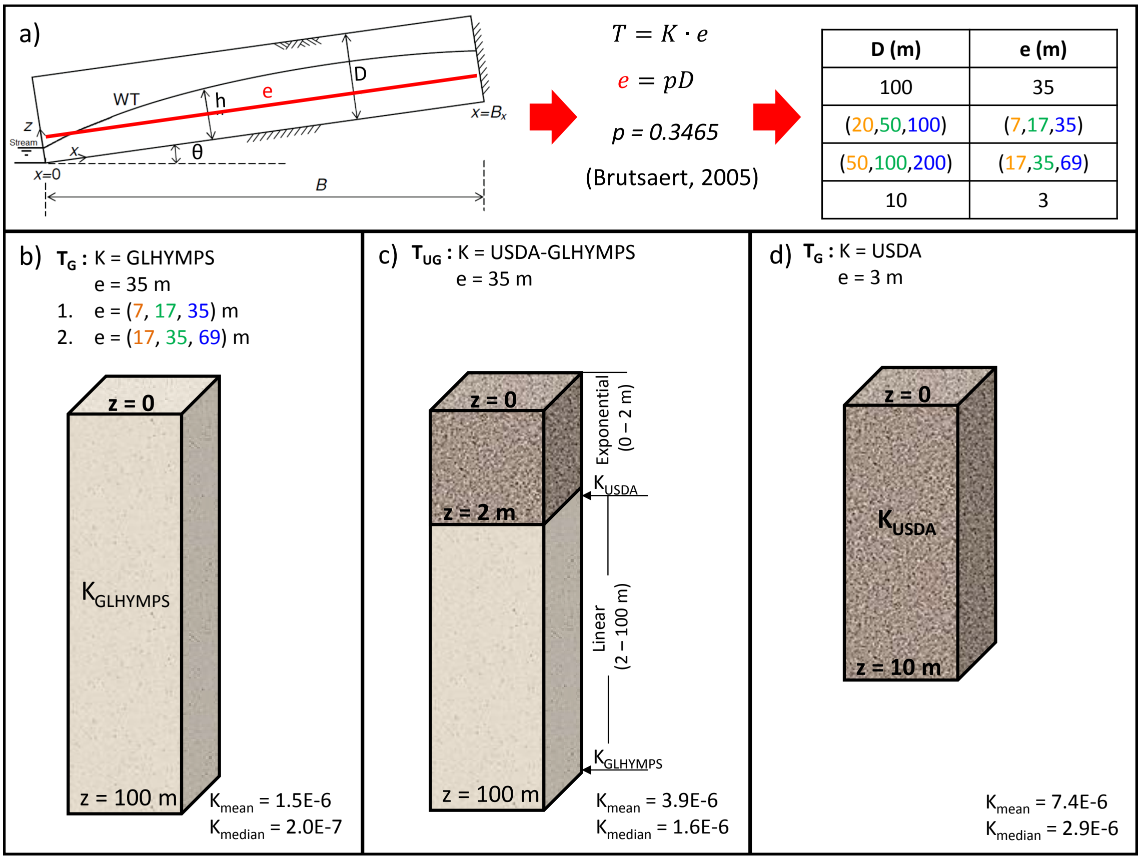

a)
WT
h
e
D
z
Stream
x
θ
x=0
x=B_x
B
T = K · e
e = pD
p = 0.3465
(Brutsaert, 2005)
D (m) | e (m)
100 | 35
(20,50,100) | (7,17,35)
(50,100,200) | (17,35,69)
10 | 3
b) T_G : K = GLHYMPS
e = 35 m
1. e = (7, 17, 35) m
2. e = (17, 35, 69) m
z = 0
K_GLHYMPS
z = 100 m
K_mean = 1.5E-6
K_median = 2.0E-7
c) T_UG : K = USDA-GLHYMPS
e = 35 m
z = 0
Exponential (0 – 2 m)
K_USDA
z = 2 m
Linear (2 – 100 m)
K_GLHYMPS
z = 100 m
K_mean = 3.9E-6
K_median = 1.6E-6
d) T_G : K = USDA
e = 3 m
z = 0
K_USDA
z = 10 m
K_mean = 7.4E-6
K_median = 2.9E-6

Figure 3.4 – Aquifer scheme used on transmissivity estimation, assuming a linear, unconfined, homogeneous isotropic aquifer. Transmissivity is given by the product of a) saturated depth (e) and three hydraulic conductivity data (K): b) using GLHYMPS data, c) combining USDA and GLHYMPS data, and d) using USDA data. WT = water table, T = transmissivity, K = hydraulic conductivity, e = saturated thickness, p = linearization constant, D = aquifer depth, B = aquifer length, θ = slope.

Table 3.3 – Descriptive statistics results of saturated thicknesses (e, in m) used on τ estimation, from aquifer depths (D, in m), at 7.5’ and 0.5° resolution.

Resolution	D	1st Q	Median	Mean	3rd Q
7.5’	10 m	3.46	3.46	3.46	3.46
0.5°	10 m	3.46	3.46	3.46	3.46
7.5’	100 m	34.65	34.65	34.65	34.65
0.5°	100 m	34.65	34.65	34.65	34.65
7.5’	(20,50,100) m	6.93	6.93	17.32	18.95
0.5°	(20,50,100) m	6.93	6.93	15.42	17.32
7.5’	(50,100,200) m	17.32	17.32	34.65	39.45
0.5°	(50,100,200) m	17.32	17.32	32.89	34.65

To produce the saturated thickness layer at 7.5’ and 0.5°, the groundwater classification of WHYMAP was represented by the majority class in each grid-cell. The classes were then converted to saturated thickness by associating each class to the values described in items three and four above. This process resulted in four saturated thickness layers per resolution, summarized in Table 3.3. As expected, lower resolution reduced the thickness range (Table 3.3), observed by variations in the quartiles, but with no major changes to the mean values.

At global scale, two datasets are available for hydraulic conductivity: GLHYMPS (as the average permeability over 100 m depth) and USDA (as soil hydraulic conductivity). In this thesis, three datasets (Figure 3.4) were tested to evaluate τ sensitivity to saturated hydraulic conductivity (K):

1. **GLHYMPS**: with K calculated using GLHYMPS permeability, in which K is given by:

$$K = \frac{perm \cdot \rho \cdot g}{\mu}, \tag{3.7}$$

   where $perm$ is GLHYMPS permeability, ρ is the density of the fluid, g is the acceleration due to gravity, and μ is the viscosity of the fluid (Gleeson *et al.*, 2014). The permeability provided by Gleeson *et al.* (2014) represents the average values over 100 m depth per hydrolithology class (same classes as the porosity product). Permanent permafrost and no data regions have permeability of 10^{-20} m^2. $K_{GLHYMPS}$ was calculated with Equation 3.7 using GLHYMPS permeability, gravitational acceleration of 9.81 m.s^{-2}, water density of 1,000 kg.m^{-3} and viscosity of 0.001 kg.m^{-1}.s^{-1}. The resultant $K_{GLHYMPS}$ has a wide range, varying up to six orders of magnitude among hydrolithologies (Table 3.4), with unconsolidated sedimentary rocks as the most permeable material, and fine grained siliciclastic sedimentary rocks as the least permeable.

2. **USDA**: K resulted from the association of hydraulic conductivities given by Carsel and Parrish (1988) to the USDA soil texture classes of Reynolds *et al.* (2000) at 5' resolution (d'Orgeval, 2006), also used by Fan *et al.* (2007), and in ORCHIDEE. The choice of using soil hydraulic conductivities (K_{USDA}) instead of rock ($K_{GLHYMPS}$) has only scientific meaning when applied in the soil layer (i.e., for the proposed saturated thickness of 3.46 m). K_{USDA} has lower range than $K_{GLHYMPS}$ (three orders of magnitude, Tables 3.4 and 3.5).

Table 3.4 – GLHYMPS permeability (k in m^2) data (Gleeson *et al.*, 2014) and calculated hydraulic conductivity ($K_{GLHYMPS}$ in m.s^{-1}) per hydrolithology class (c.g. = coarse grained and f.g. = fine grained).

Hydrolithology	k (m^2)	K (m.s^{-1})
Unconsolidated	1.0E-13	9.8E-7
c.g. unconsolidated	1.3E-11	1.2E-4
f.g. unconsolidated	1.0E-14	9.8E-8
Siliciclastic sedimentary	6.3E-16	6.2E-9
c.g. siliciclastic sedimentary	3.2E-13	3.1E-6
f.g. siliciclastic sedimentary	3.2E-17	3.1E-10
Carbonate	1.6E-12	1.6E-5
Crystalline	7.9E-15	7.8E-8
Volcanic	3.2E-13	3.1E-6
Not assigned	1.0E-20	9.8E-14

3. **USDA-GLHYMPS**: $K_{USDA-GLHYMPS}$ was obtained by combining $K_{GLHYMPS}$ and K_{USDA} for both 7.5' and 0.5° resolution. In this scenario, it was assumed a vertical profile where each grid-cell has 100 m depth. K decreases from the surface following an exponential decay (Equation 3.8) until 2 m depth (with $K_{z=2m} = K_{USDA}$), and then decreases linearly until 100 m (where $K_{z=100m} = K_{GLHYMPS}$). $K_{USDA-GLHYMPS}$ is given by the vertical average of each grid-cell, where K at the surface was calculated from Equation 3.8, using a decay factor (f) of 1 m^{-1}, chosen arbitrarily (but in the same order of magnitude of d'Orgeval (2006)). In regions where $K_{GLHYMPS}$ is higher than K_{USDA}, it was assumed a constant K between 2 and 100 m (i.e., $K_{z=100m} = K_{USDA}$).

$$K_{USDA} = K_{z=0} \cdot e^{-fz} \tag{3.8}$$

Transmissivity (m^2.s^{-1}) was then obtained by the product of hydraulic conductivity and saturated thickness, resulting in five transmissivity combinations (Figure 3.4, with three resultant transmissivities for the scenario b, one for c and one for

Table 3.5 – Hydraulic conductivities (m.s^{-1}) from Carsel and Parrish (1988) per soil texture of Reynolds *et al.* (2000). Adapted from d'Orgeval (2006).

Soil texture	K (m.s^{-1})
Sand	8.25E-5
Loamy sand	4.05E-5
Sandy loam	1.23E-5
Silt loam	1.25E-6
Silt	6.94E-7
Medium loam	2.89E-6
Sandy clay loam	3.64E-6
Silty clay loam	1.94E-7
Clay loam	7.22E-7
Sandy clay	3.33E-7
Silty clay	5.56E-8
Clay	5.56E-7

d). Figure 3.5 shows that higher resolution (in blue) increases the range of transmissivities, as coarser resolutions tend to average the results. However, changes in saturated thicknesses resulted in no significant transmissivity variations when com-

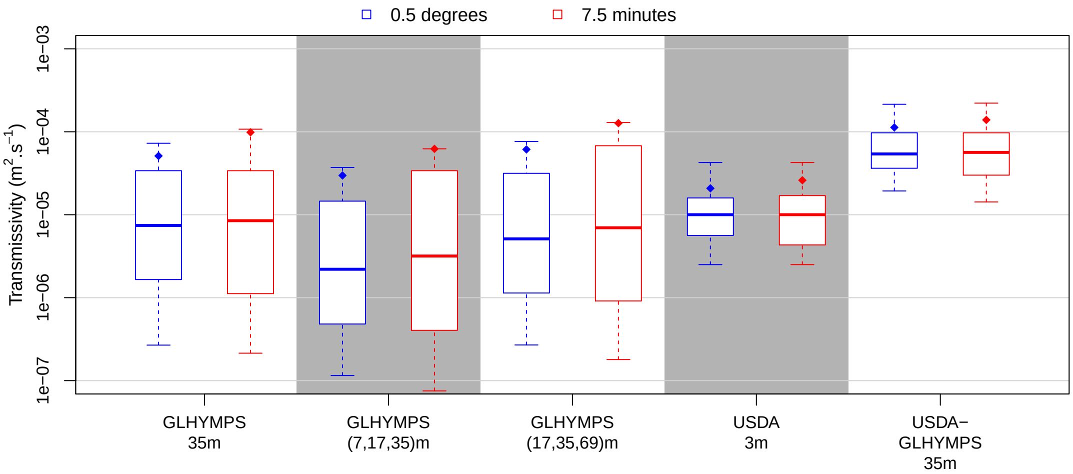


Figure 3.5 – Transmissivity results (m^2.s^{-1}) at 0.5° (blue) and 7.5' (red) resolution. GLHYMPS, USDA, and GLHYMPS-USDA represent the hydraulic conductivities, with saturated thickness 3 and 35 m for constant thickness, (7,17,35) and (17, 35, 69) m for (shallow, complex hydrogeological structures,major groundwater basins). The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

pared to the use of different hydraulic conductivities. Using a saturated thickness ten times lower with K_{USDA} (Figure 3.5 - USDA 3 m) still have higher median values than with $K_{GLHYMPS}$, but with lower average (Figure 3.5 - GLHYMPS 35 m). The combined product $K_{USDA-GLHYMPS}$ resulted in transmissivities not lower than 10^{-5} $m^2.s^{-1}$, where K_{USDA} significantly impacts $K_{USDA-GLHYMPS}$ results.

3.2.3 Slope

Aquifer slope was calculated for two datasets: the 15" hydrologically corrected elevation from HydroSHEDS (Lehner *et al.*, 2008), assuming that the aquifer will have a slope that follows the topography, like in TOPMODEL (Beven and Kirby, 1979), and the water table depth from Fan *et al.* (2013) at 30' (used for the 7.5' grid) and 0.25° (used for the 0.5° grid), assuming that the aquifer will have a slope that follows the water table. The two resultant 7.5' and 0.5° resolutions contain the mean slope value per grid-cell, of the topography (from Lehner *et al.* (2008) elevations) and water table (for Fan *et al.* (2013) water table depths).

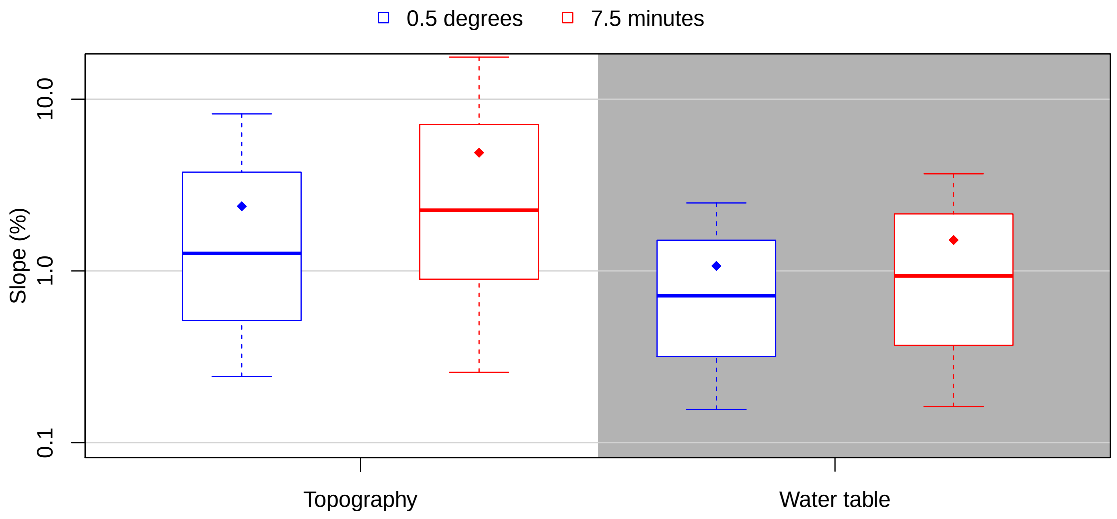


Figure 3.6 – Slope results (%) at 0.5° (blue) and 7.5' (red) resolutions. Topography represents the slope calculated using HydroSHEDS hydrologically corrected elevations (Lehner *et al.*, 2008), and water table the slope derived from Fan *et al.* (2013) water table depth constrained by observations. The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

Topography and water table slopes have the same order of magnitude on average for both resolutions, with a wider range for 7.5' (Figure 3.6). As expected, water table slopes have lower values than topography. Therefore, 98% of the topography, but 100% of the water table slopes are under the 20% limit that would not lead to significant errors with Dupuit assumptions (Dingman, 2015), making water table slopes more suitable for τ estimation. Yet, topography slopes were still used to evaluate τ sensitivity to slope.

3.2.4 Drainage density

Drainage density (δ) calculations at 7.5' resolution were discussed in chapter two. Drainage densities for 0.5° and 7.5' resolution were calculated from the sum of the total river length inside a grid-cell, divided by its area. To evaluate τ sensitivity to δ, two river networks were chosen: HydroSHEDS river network (Lehner *et al.*, 2008), and the LCS river network (chapter two, Schneider *et al.* (2017)), both at 15" resolution.

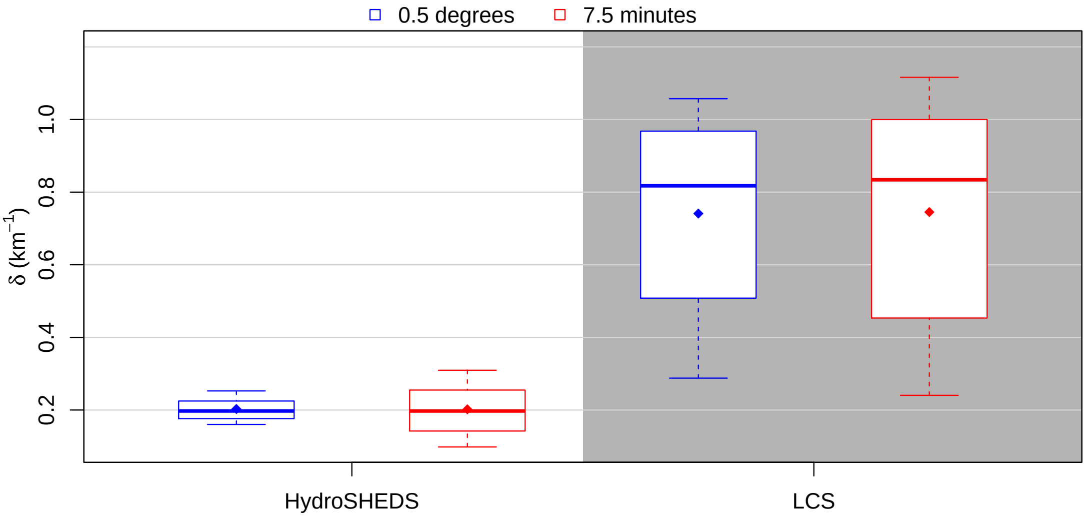


Figure 3.7 – Drainage density results (δ, in km^{-1}) at 0.5° (blue) and 7.5' (red) resolution. White area shows δ from HydroSHEDS data (Lehner *et al.*, 2008), and gray region δ with LCS river network (Schneider *et al.*, 2017). The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

As discussed in chapter two, drainage densities from HydroSHEDS river network have lower values than from LCS, with a significantly smaller range that was magnified with the use of a coarser resolution (Figure 3.7). However, no major variations were observed for the mean or median values (Figure 3.7).

3.2.5 Selection of the parameters

To evaluate the sensitivity of τ to porosity, transmissivity, drainage density, and slope on τ estimation, several combinations of the data described in this section were used, as summarized in Table 3.6. τ was estimated using Equation 3.5 for sloping aquifers, and Equation 3.3 for horizontal aquifers. Since the use of different resolutions (7.5' and 0.5°) do not result in major differences to mean and median values for all evaluated parameters used to estimate τ, the sensitivity analysis was perform at 0.5° resolution (a common resolution used for large scale LSMs). The main criteria was to evaluate the use of different data sets from τ_{ini} presented in the introduction of this thesis in order to reduce τ values, since τ_{ini} is overestimated when compared to literature values presented in Table 3.1. For each evaluated τ a single parameter was changed between porosity, transmissivity, drainage density, and slope, described in Table 3.6.

Table 3.6 – Summary of the data used to test the sensitivity of τ to different parameters. n = total porosity, n_e = effective porosity, K = hydraulic conductivity, e = saturated thickness, δ = drainage density, θ = slope, HS = HydroSHEDS.

	n	K	e	δ	θ	Formulation
τ_{ini}	n	$K_{GLHYMPS}$	35	HS	-	horizontal
τ_{G1}	n	$K_{GLHYMPS}$	(7,17,35)	HS	-	horizontal
τ_{G2}	n	$K_{GLHYMPS}$	(17,35,69)	HS	-	horizontal
τ_{WT}	n	$K_{GLHYMPS}$	35	HS	water table	sloping
τ_{TP}	n	$K_{GLHYMPS}$	35	HS	topography	sloping
τ_{LCS}	n	$K_{GLHYMPS}$	35	LCS	water table	sloping
τ_{ne}	n_e	$K_{GLHYMPS}$	35	LCS	water table	sloping
τ_{U}	n_e	K_{USDA}	3	LCS	water table	sloping
τ_{UG}	n_e	$K_{USDA-GLHYMPS}$	35	LCS	water table	sloping

3.3 Base flow time constant results and discussion

τ results were first evaluated at the global scale based on the variation of each parameter presented in Table 3.6 (Figure 3.8). The use of different saturated thicknesses resulted in τ results in the same order of magnitude in global average (Figure 3.8, τ_{ini} versus τ_{G1} and τ_{G2}). The use of a formulation that includes aquifer slope reduced τ results, but they remained in the same order of magnitude both for topography or water table slope data. τ is mostly sensitive to both drainage density and hydraulic conductivity. The use of LCS river network (δ_{mean} = 0.74 km^{-1}) resulted in τ one order of magnitude lower in average than HydroSHEDS river network (δ_{mean} = 0.20 km^{-1}). A river network that represent the heterogeneities found in natural river networks can have a major effect, since τ is proportional to δ^2. For all cases presented in Figure 3.6, mean values of τ are significantly higher than the median, with distributions with high positive skewness (approximately 100 for horizontal formulations and 10 for sloping, Appendix 6.1), indicating that the majority of values are lower than the mean. Positive skewness is also observed for hydraulic conductivities (13 for GLHYMPS data and 4 for USDA and GLHYMPS-USDA) and slope (around 2), while drainage density and porosity have negative skewness close to zero (approximately -0.5).

Transmissivity have a stronger effect on τ than drainage density. τ estimated using $K_{GLHYMPS}$ had the highest values (Appendix 6.1), with one order of magnitude higher on average than with $K_{USDA-GLYMPS}$. Even for a saturated thickness ten times lower (that would result in τ ten times higher), τ is in average four times higher than with K_{USDA} (Figure 3.8 for τ_{ne} and τ_U, and Appendix 6.1) due to the higher values of transmissivity. However, independently of the parameters used, τ is higher than the results observed by the references listed in Table 3.1 on global average for all evaluated combinations of parameters from Table 3.6. This overestimation can be attributed to the high variability of the results, which can vary up to 8 orders of magnitude at the global scale (as in τ_{G1} and τ_{G2}, Appendix 6.1), suggesting that τ should be compared locally (as references are for local scales, and not global averages).

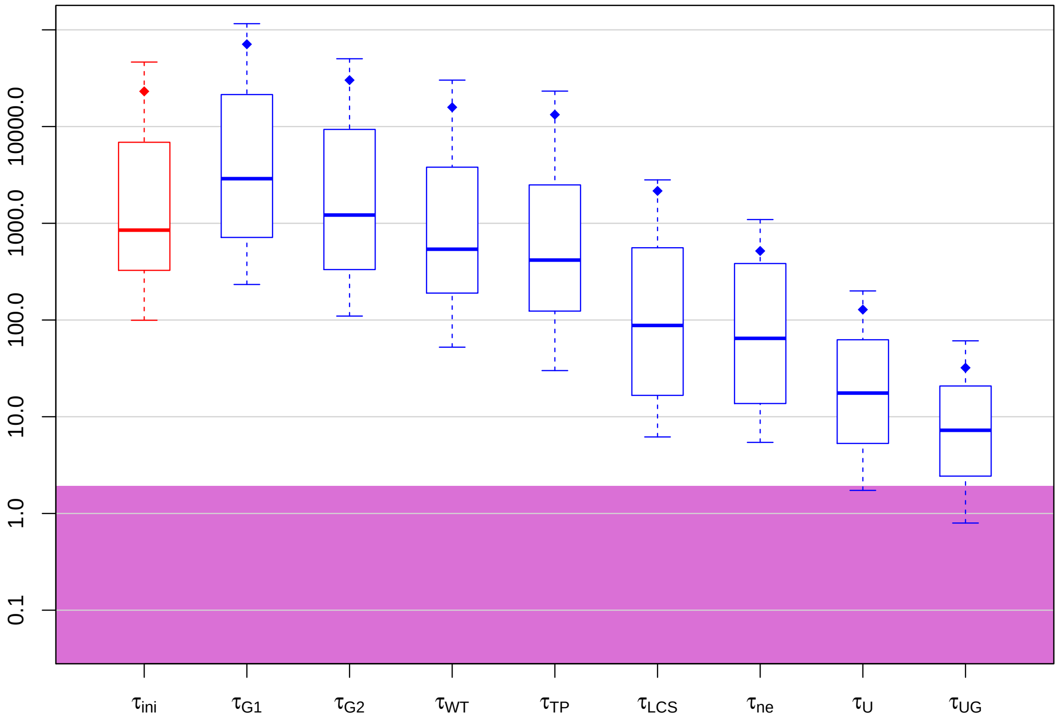


Figure 3.8 – Base flow time constant results at global scale using parameters of Table 3.6 (in years) at 0.5°. Pink region represents a reference range from literature presented in Table 3.1 (2 to 342 days = 0.005 to 0.94 years). The bottom and top of the box-plots represent first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

In Figure 3.9, τ was calculated at the basin scale (Table 3.1, between 14 and 906 km^2), using the average parameters (arithmetic for porosity and geometric for transmissivity) inside each basin (instead of 0.5° and 7.5' grids). At the basin scale, reference values showed in Figure 3.9 have a higher range than the estimated values, suggesting that the parameters used on τ estimation do not represent accurately τ variability found at small scales. The combination of parameters that resulted in τ values closer to the references from Table 3.1 are for τ_{UG}, which uses effective porosity, drainage density with LCS river network, $K_{USDA-GLHYMPS}$ with the sloping formulation. τ_{ne} resulted in values over 100 years, two orders of magnitude higher than τ_{UG} (1 year). However, τ_{UG} still remained overestimated when compared to references, which could be attributed to effective porosity and transmissivity. Brutsaert (2008) observed that literature values for effective porosity (as the ones from

Johnson (1967)) are overestimated by at least one order of magnitude when compared to values measured in field. When dividing τ_{UG} by a factor ten ($\tau_{UG/10}$), τ becomes much closer to literature values at the basin scale, suggesting that effective porosity data used are overestimated. The assumption of an homogeneous and isotropic aquifer per basin neglects preferential flows, which results in incorrect parameter estimation that is no appropriate for an effective τ, leading to overestimated effective porosity, underestimated transmissivities (as it neglects preferential flows), and underestimated drainage density (as the shortest flow paths are more important than the long ones to local flow).

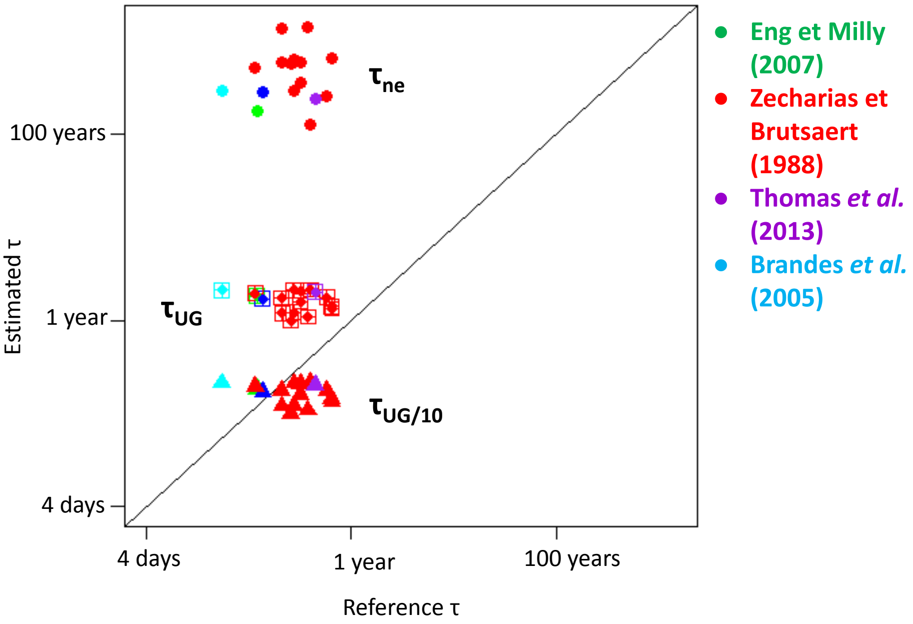


Figure 3.9 – Comparison of τ results (years) at the basin scale for τ_{ne}, τ_{UG}, and $\tau_{UG/10}$.

Lower drainage densities resulted in higher τ (Figure 3.10), as expected. However, the effect of the hydraulic conductivity is stronger than the drainage density, as observed in Figure 3.10. In regions with average drainage density for τ_{ne}, high τ values were observed (Figure 3.10, in red), which was not the case for τ_{UG} (Figure 3.10, in blue). High τ values for regions of average drainage density are explained by the significant low K found in $K_{GLHYMPS}$. Such low hydraulic conductivity could be justified by the depth chosen by Gleeson *et al.* (2014). Deeper formations tend to have lower hydraulic conductivities that can result in a low average τ for the 100 m depth chosen by Gleeson *et al.* (2014), resulting in higher τ results.

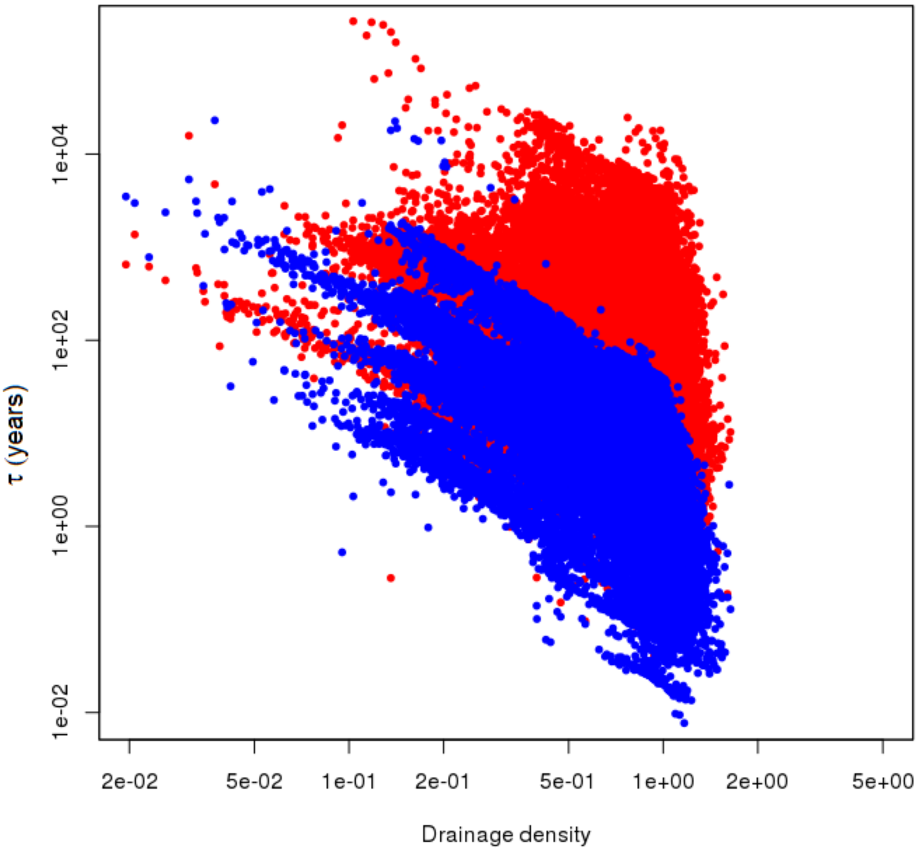


Figure 3.10 – Comparison of the effect of drainage density and hydraulic conductivity on τ. Red represents the use of $K_{GLHYMPS}$ (τ_{ne}) and blue $K_{USDA-GLHYMPS}$ (τ_{UG}), both at 0.5° resolution.

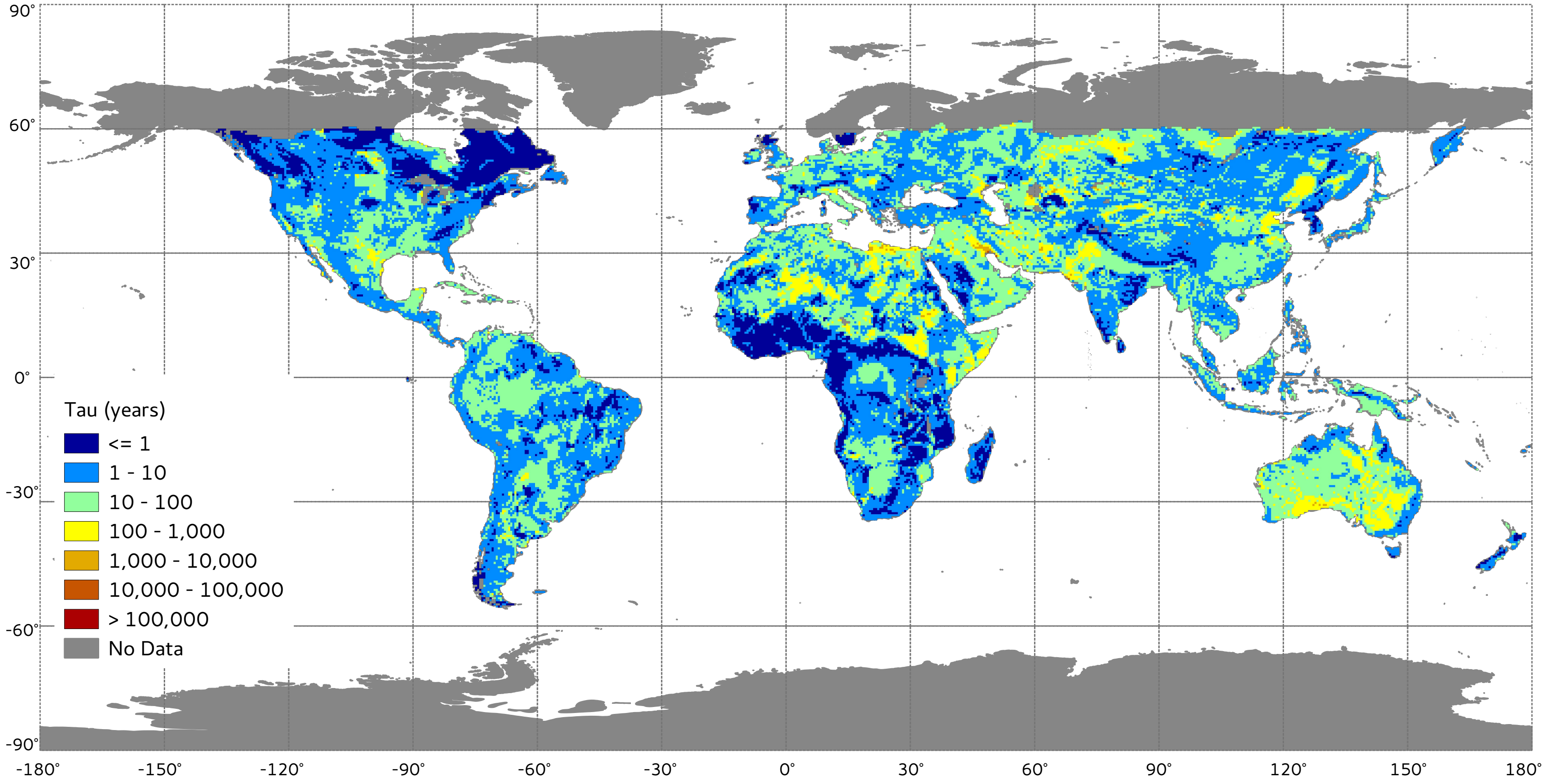

90°
60°
30°
0°
-30°
-60°
-90°
-180°
-150°
-120°
-90°
-60°
-30°
0°
30°
60°
90°
120°
150°
180°
Tau (years)
<= 1
1 - 10
10 - 100
100 - 1,000
1,000 - 10,000
10,000 - 100,000
> 100,000
No Data

Figure 3.11 – Base flow time constant at 0.5° resolution for τ_{UG} (effective porosity, LCS river network, 35 m saturated depth, and $K_{USDA-GLHYMPS}$). Gray regions are no data values resultant from the HydroSHEDS DEM used on the river network extraction.

Figure 3.11 shows the base flow time constant for τ_{UG} (effective porosity, LCS river network, 35 m saturated depth, and $K_{USDA-GLHYMPS}$) at 0.5° as the closest estimated value to reference data. Most of τ values remained under 100 years. Although a constant depth was used for the estimation of τ_{UG}, similarities are observed for τ patterns when compared to WHYMAP (Figure 3.3) shallow aquifers and major groundwater basins (as in center Africa or north of South America). Errors inherent to the input datasets used can be seen in the North Dakota, where the state line is visible. These errors come from the hydraulic conductivity data, since the hydrolithologies used to produce K are based on lithology classes of Hartmann and Moosdorf (2012) that the same pattern is observed, also shown in Chapter 2 for the drainage density.

3.4 Conclusions

In this Chapter, the base flow time constant was estimated at 0.5° resolution based on global available datasets. τ was estimated using a linearized solution of the Boussinesq equation, considering unconfined, isotropic, homogeneous sloping aquifers. τ depends on effective porosity, hydraulic conductivity, saturated thickness, and drainage density. A series of datasets were tested in order to evaluate τ sensitivity to effective porosity, hydraulic conductivity, saturated thickness, drainage density, and a higher resolution (7.5'). This methodology allows to estimate τ at global scale even in regions without discharge measurements, for use as an input to improve the representation of base flow in models with simple groundwater schemes based in linear reservoirs (as ORCHIDEE, CLM, among others).

The sensitivity analysis showed that the main properties that control τ are transmissivity and drainage density. The use of a formulation for sloping aquifers reduced τ results. This reduction is stronger at higher resolutions (as slopes are less averaged and have a higher range), being closer to reference data. As expected, the use of effective porosity instead of total porosity had a positive effect on τ (as effective porosity is always lower than total porosity by definition) resulting in a better agreement to reference data. However, τ reduction was not as strong as expected due to the use of effective porosity, due to inconsistencies observed in the data used. Changes in slope, saturated thickness, and porosity alone had no major impacts on average or median values, remaining in the same order of magnitude, possibly due to the resolution used, as coarser resolution result in lower ranges closer to the average. The use of a river network that represents the natural heterogeneities, with river initiation that are in agreement with observations (LCS river network) resulted in lower τ (due to the elevated δ values). Although τ is sensitive to δ, hydraulic conductivity can have a stronger impact on τ results, as they have a wider range not only for different hydrolithologies, but also within the same hydrolithology class (explained by fractures and porosity variations).

For all combinations tested (Table 3.6), τ resulted in higher values than reference data (Figure 3.8). The overestimation could be explained by the assumptions used on τ estimation. To consider an homogeneous, isotropic aquifer neglects preferential flows and natural horizontal and vertical variations for the hydrogeological properties. These assumptions lead to underestimated transmissivities, by neglecting preferential flow, and drainage density, as shortest flow paths have stronger impact to local flow than long ones. The use of an effective porosity data that comes from laboratory analysis result in overestimated values when compared to local scale field results (Brutsaert, 2008) by at least one order of magnitude, resulting in overestimated τ. The τ that resulted in values closest to reference data ($\tau_{UG}/10$) was estimated using the effective porosity from de Graaf *et al.* (2015) divided by a factor 10, drainage density from Schneider *et al.* (2017) (LCS river network), and a combination of hydraulic conductivities from USDA (Reynolds *et al.*, 2000) and GLHYMPS (Gleeson *et al.*, 2014) for a 35 m saturated depth.

CHAPTER 4

BASE FLOW TIME CONSTANT IN ORCHIDEE

Contents

4.1 The ORCHIDEE model 72
4.1.1 Overview 72
4.1.2 SECHIBA 73
4.1.3 Routing scheme 75
4.2 Simulation set-up 79
4.2.1 Forcing data 79
4.2.2 Sensitivity experiments 80
4.2.3 Initialization 82
4.3 Validation data 83
4.4 Simulation results 87
4.4.1 Evaluation of the reference simulation τ_{ORC} 87
4.4.2 Sensitivity analysis 92
4.5 Conclusions 99

Scope

Land Surface Models (LSMs) at global scale with simple groundwater schemes (linear reservoir) usually rely on calibration over a region and extrapolation over the entire globe to simulate base flow, due to lack of observation data at large scales. The use of a base flow time constant (τ) estimated from an analytical solution of the Boussinesq equation in the ORCHIDEE model raises the research question: *Does the base flow time constant estimated from hydrogeological and topographical properties improve river discharge simulations of LSMs with linear reservoir?* Simulation results with τ resulted from chapter three showed that the use of a physically based τ deteriorated ORCHIDEE river discharge results. A reference simulation showed the use of a higher τ (as the ones tested) tends to attenuate hydrographs amplitude and retard their peaks.

4.1 The ORCHIDEE model

4.1.1 Overview

ORganizing Carbon and Hydrology In Dynamics EcosystEms (ORCHIDEE) (Krinner *et al.*, 2005) is a Land Surface Model (LSM) developed in the *Institut Pierre Simon Laplace* (IPSL)), with the objective to simulate the energy and water balance of terrestrial ecosystems. ORCHIDEE is in constant development, and is divided in three modules (summarized in Figure 4.1):

1. **SECHIBA:** *Schématisation des EChanges Hydriques à l'interface entre la Biosphère et l'Atmosphère* (SECHIBA) was initially developed by Ducoudré *et al.* (1993). SECHIBA simulates energy and water balances considering soil-vegetation-atmosphere interactions, with a 30 minute time step to describe the diurnal cycle of the energy fluxes. Two hydrological parametrizations are available to solve the vertical water balance: the so-called Choisnel scheme, with two layers working as a bucket model, with no drainage from the soil (Ducoudré *et al.*, 1993; Ducharne *et al.*, 1998; De Rosnay and Polcher, 1998), and the so-called CWRR scheme, using a physically-based description of soil-water fluxes with Richards equation, based on the Center for Water Resources and Research model, with 11 layers and free drainage at the bottom (De Rosnay *et al.*, 2002; d'Orgeval *et al.*, 2008; Campoy *et al.*, 2013). SECHIBA also includes a routing scheme to solve horizontal water flow, and vegetation mechanisms (as photosynthesis and transpiration).

2. **STOMATE:** *Saclay Toulouse Orsay Model for the Analysis of Terrestrial Ecosystems* (STOMATE) was developed in 1996 (Viovy, 1996), and it simulates the soil carbon cycle and vegetation biomass, providing carbon flow within the soil-plant-atmosphere continuum, with a daily time step. Phenology and car-

bon dynamics are simulated considering the processes of photosynthesis, soil carbon dynamics, respiration, and vegetation growth.

3. **LPJ:** *Lund Potsdam Jena* (LPJ) was developed in 2003 (Sitch *et al.*, 2003) to describe the vegetation dynamics in the simulation, through ORCHIDEE's 13 Plant Functional Types (PFTs), with 1 year time step. This module allows to simulate biochemical and biophysical interactions between ecosystems and the atmosphere, considering soil biogeochemistry, physiology and vegetations dynamics, climate conditions, and competition among the PFTs, changing the PFT maps over time.

ORCHIDEE can be used coupled to an atmospheric model to obtain atmospheric inputs, or in off-line mode (configuration used in this chapter), in which the atmospheric inputs come from a forcing dataset. The atmospheric forcing data dictate ORCHIDEE's spatial output resolution. Figure 4.1 shows a graphic representation of ORCHIDEE and its modules, and the input parameters vary in function of the activated modules and configuration used. In this chapter, we used the revision r3970 of the standard version of ORCHIDEE (called trunk). The STOMATE module was activated, as well as the CWRR soil hydrology scheme in the SECHIBA module. The routing sub-module (4.1.3) and the 3-layer snow scheme of Wang *et al.* (2013) were activated. However, floodplains, swamps, ponds, and soil freezing were not activated. SECHIBA was set for the CWRR soil hydrology scheme. The routing sub-module (Section 4.1.3) and the snow scheme were activated. However, floodplains, swamps, ponds, and soil freezing were not activated.

4.1.2 SECHIBA

SECHIBA computes water and energy budgets within rectangular grid-cells, assuming energy and mass conservation. The energy budget is calculated per grid-cell with no vertical discretization, depending on climate factors (net radiation, air temperature, wind speed), surface characteristics (given by albedo and roughness), and water availability. The surface water budget is computed separately in three

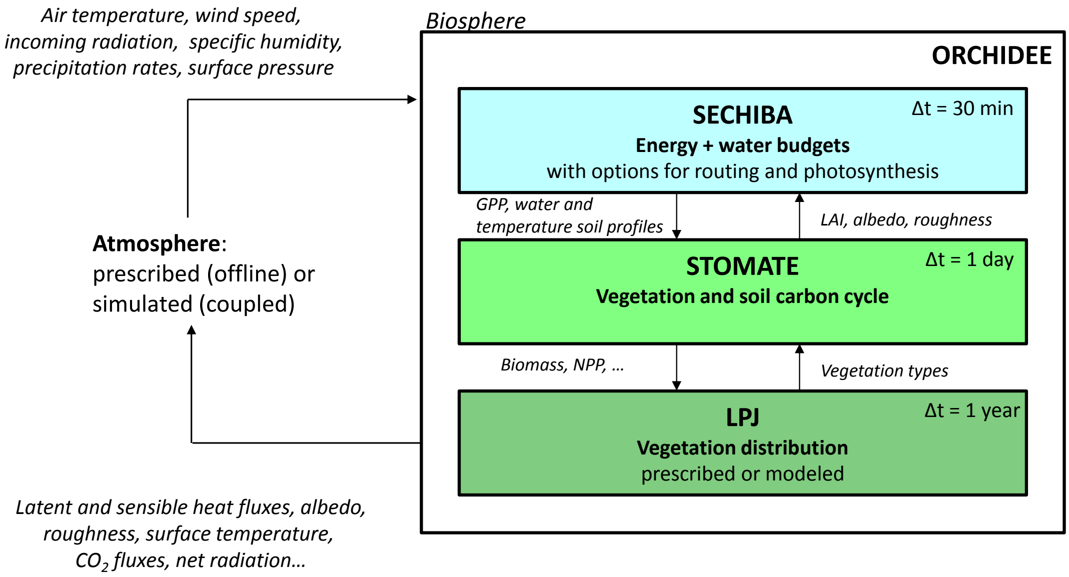

Air temperature, wind speed,
incoming radiation, specific humidity,
precipitation rates, surface pressure
Biosphere
ORCHIDEE
SECHIBA
Δt = 30 min
Energy + water budgets
with options for routing and photosynthesis
GPP, water and temperature soil profiles
LAI, albedo, roughness
Atmosphere: prescribed (offline) or simulated (coupled)
STOMATE
Δt = 1 day
Vegetation and soil carbon cycle
Biomass, NPP, …
Vegetation types
LPJ
Δt = 1 year
Vegetation distribution
prescribed or modeled
Latent and sensible heat fluxes, albedo,
roughness, surface temperature,
CO2 fluxes, net radiation…

Figure 4.1 – ORCHIDEE general scheme. Adapted from ORCHIDEE's training course.

main reservoirs: canopy interception (composed by leaf area index), snow pack, and soil. The water budget is then computed by:

$$\frac{\partial W}{\partial t} = P_{tot} - E_{tot} - R_{tot}, \tag{4.1}$$

where W is the soil moisture, and total precipitation (P_{tot}), total evaporation (E_{tot}) and total runoff (R_{tot}) are given in mm.s^{-1}. Total precipitation is given by the sum of rainfall and snowfall. Total evaporation is composed by evapotranspiration (bare soil evaporation, transpiration of vegetation, interception loss of canopy), snow sublimation, and potential floodplains evaporation (in this thesis this latter term is not accounted, since floodplains module was not activated). Total runoff is the sum of surface runoff and drainage that resulted from the soil column infiltration.

The soil hydrology used for all ORCHIDEE simulations was the 11-layer scheme called CWRR described in details on the user guide developed by Ducharne (2017). In SECHIBA, the vegetation is represented by thirteen Plant Functional Types (PFTs), composed of one bare soil, eight forest, two grassland, and two cropland. The water budget is computed using a maximum of three separated soil columns per grid-cell: all forest PFT, all grass and crops PFT, and a no vegetation PFT (including bare soil, ice, free water, etc. surface types) (Ducharne, 2017). Evapotranspiration makes the link between energy and water budgets, and is calculated per soil column.

Water diffusion and redistribution in the soil use 2 m soil depth for each soil column, which have a vertical discretization in 11 nodes (with a geometric progression of the internode distance), using a free drainage boundary condition (Ducharne, 2017). This scheme uses a physically-based description introduced by De Rosnay *et al.* (2002), based on a one dimensional Fokker-Plank equation. All variables are assumed horizontally homogeneous, and lateral fluxes are neglected.

4.1.3 Routing scheme

ORCHIDEE's routing scheme is part of the SECHIBA module, introduced in ORCHIDEE by Polcher (2003) with a routine called `routing.f90`, based on existing routing models from the work of Ledoux (1980), Miller *et al.* (1994), Hagemann and Dümenil (1998), Fekete *et al.* (2000), and Ducharne *et al.* (2003). Total runoff is transported horizontally from one grid-cell to another across major river basins towards oceans (blue points in Figure 4.2) or endorrheic lakes (inland sinks). In each grid-cell, the resultant flow is transported to a single downstream grid-cell, according to eight possible flow directions (as shown in Figure 4.2, with a zoom over the Obidos GRDC station at the Amazon basin), using Vörösmarty *et al.* (2000) data at 0.5° (for both flow directions and watersheds). The main watersheds are identified by a basin map, with a single code per basin. In this thesis, all simulations were performed at 0.5°, in which each grid-cell contains a single basin code.

The water is routed in through a stream reservoir, and locally delayed by two reservoirs called fast and slow (Figure 4.3) that do not interact with the atmosphere (Guimberteau *et al.*, 2012a). In each grid-cell, the slow reservoir represents local groundwater, and is fed by the drainage produced by the soil hydrology scheme, while the fast reservoir describes the delay between surface runoff production and the arrival of overland flow to the river system. Both reservoirs discharge into the stream of the next grid-cell, according to:

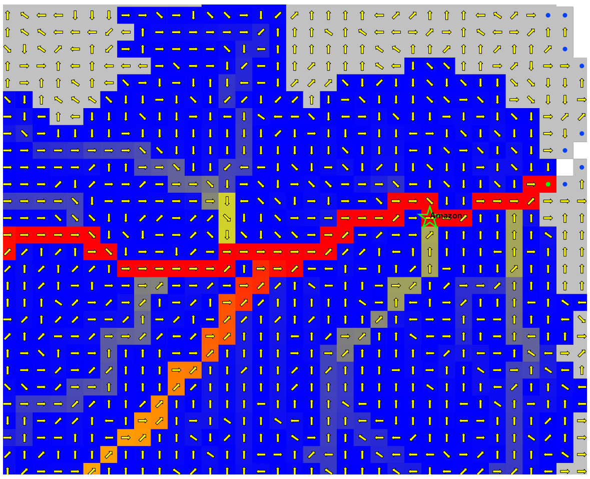


Figure 4.2 – Flow directions used on the routing scheme with a zoom over the Obidos station at the Amazon basin. Blue to red colors represent increasing upstream area of the Amazon basin. Green star show the position of the Obidos station. Green dot represents the point of river discharge into the ocean. Blue dots represent coastal flow.

$$\frac{\mathrm{d}V_i}{\mathrm{d}t} = Q_i^{in} - Q_i^{out}, \tag{4.2}$$

where V represents the storage (kg) of the reservoir i, Q is the input (in) or output (out) flow (kg.day^{-1}). For the stream reservoir, Q_{stream}^{in} is given by the sum of the upstream reservoirs (stream, fast, and slow), while Q_{fast}^{in} by the surface runoff, and Q_{slow}^{in} by the drainage. The outflow (Q_i^{out}) is given by:

$$Q_i^{out} = \frac{V_i}{\tau_i}, \tag{4.3}$$

where τ_i represents the residence time of the reservoir i. Equations 4.2 and 4.3 are integrated using and explicit finite difference scheme, with a time step of six hours. In ORCHIDEE, the residence time depends on the product of a topographic index (parameter independent of the reservoir, here called k, in km), and a reservoir constant called g (days.km^{-1}). The parameter g was obtained by calibration of river discharge over the Senegal basin (Ngo-Duc *et al.*, 2007), resulting in $0.24{\cdot}10^{-3}$, $3.0{\cdot}10^{-3}$, $25{\cdot}10^{-3}$ (days.km^{-1}) for stream, fast, and slow, respectively. The topographic index is spatially variable, and it was derived from the work of Ducharne *et al.* (2003), based on a simplification of Manning's equation, that depends on the

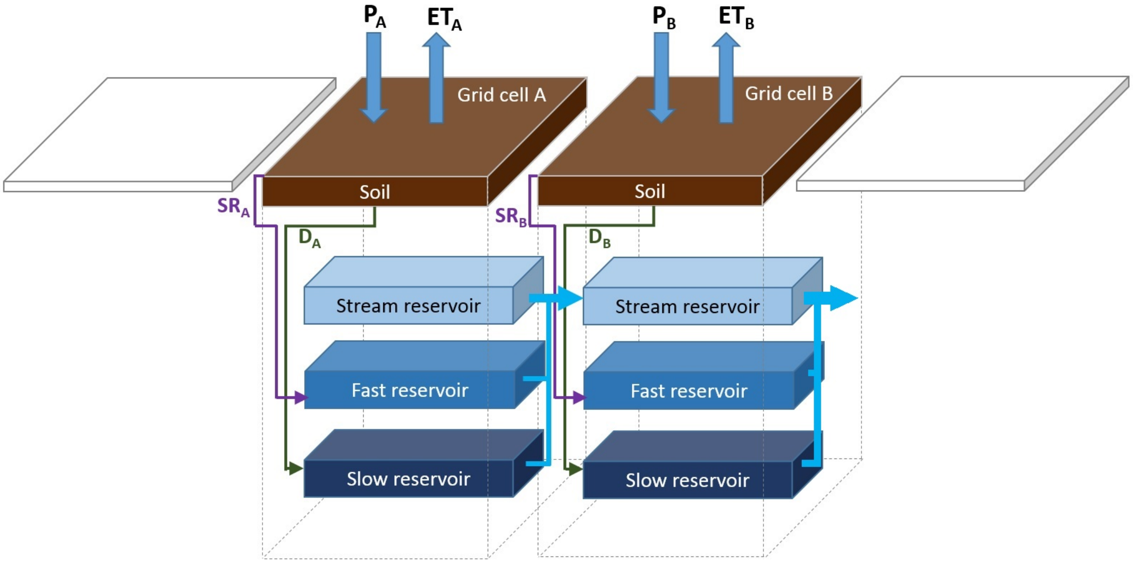

P_A
ET_A
Grid cell A
P_B
ET_B
Grid cell B
Soil
Soil
SR_A
D_A
SR_B
D_B
Stream reservoir
Fast reservoir
Slow reservoir
Stream reservoir
Fast reservoir
Slow reservoir

Figure 4.3 – Reservoirs of ORCHIDEE's routing scheme. *P* represents precipitation, *ET* evapotranspiration, *SR* surface runoff, and *D* drainage. Adapted from Guimberteau (2014).

distance (d) and the slope (θ) between two grid cells:

$$k = \frac{d}{\sqrt{\theta}} \tag{4.4}$$

$$\tau_i = k \cdot g_i \tag{4.5}$$

Since in ORCHIDEE the topographic index depends only on slope, high slope values result in small time constants. Therefore, plane areas will have the highest τ, while mountain regions the lowest. Although τ in ORCHIDEE depends only on the topography, τ_{slow} values shown in Table 4.1 (for grid-cell areas between 14 and 3,098 km^2) are in the same order of magnitude of literature values presented in Table 3.1 for recession analysis results (between 2 and 342 days for basins between 3 and 1,766 km^2). However, Gascoin *et al.* (2009) obtained calibration results of 700 days in the Somme basin (5,566 km^2) that are over ten times higher than the ORCHIDEE median results for τ_{slow}.

Table 4.1 – Descriptive statistic result for ORCHIDEE's time constant per reservoir (in days), at 0.5° resolution

Reservoir	Min	1st Q.	Median	Mean	3rd Q.	Max
Stream	0.004	0.24	0.43	0.64	1.07	2.37
Fast	0.05	2.94	5.34	7.99	13.33	29.58
Slow	0.43	24.48	44.55	66.56	111.10	246.50

The scheme shown in Figure 4.3 was described for the standard configuration of the routing scheme (as used in this chapter), that can be altered when other sub-modules are activated (as irrigation, floodplains, swamps, and ponds), shown in with a brief description below:

- Irrigation: in irrigated regions (indicated by an irrigation map (Guimberteau *et al.*, 2012b)), the water is taken successively from the stream, fast, and slow reservoir according to the available water. In case of water scarcity (all reservoirs are empty), water will be taken from the stream reservoir of a neighbor basin (for resolutions higher than 1°, and the irrigation is ceased for lower resolutions).

- Floodplains: floodplains are identified by regions where GLWD map (Global Lakes and Wetlands Database from Lehner and Döll (2004)) are reservoirs, wetlands, and floodplains, converted to 0.5°. In floodplain regions, the water from the stream reservoir goes to the floodplain reservoir, where it can be re-infiltrated to the soil or potentially evaporated, or returned to the stream reservoir.

- Swamps: regions with swamps are represented by swamp or flooded forests in the GLWD map (Lehner and Döll, 2004), also converted to 0.5°, which 20% of the stream inflow times the swamp fraction of the grid-cell are re-infiltrated to the same grid-cell.

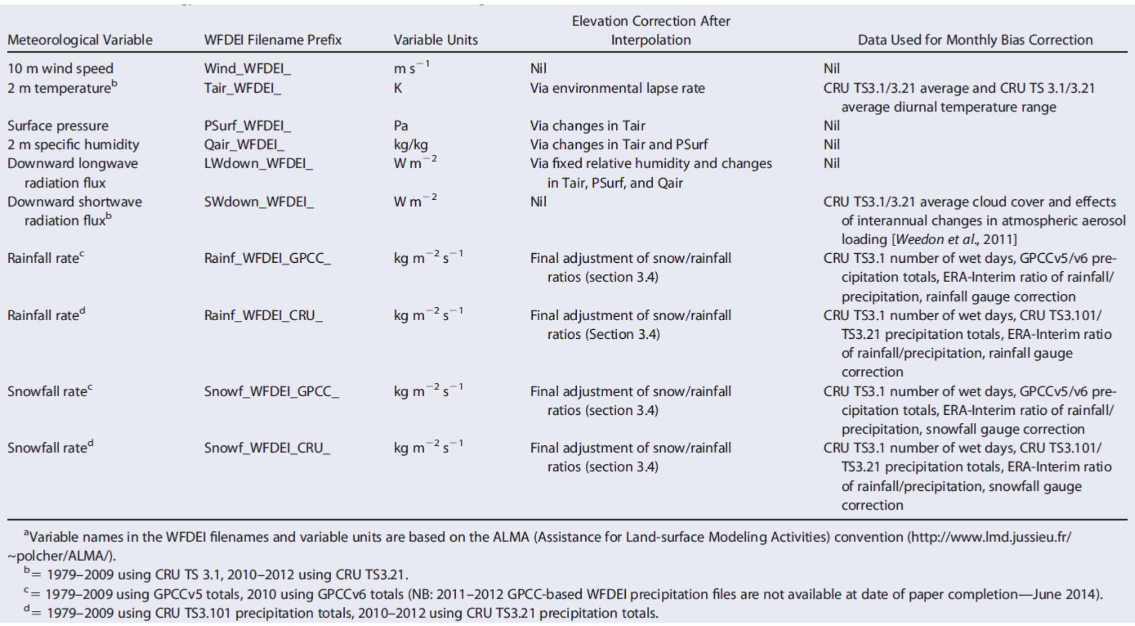
- Ponds: intermittent wetlands or lakes in the GLWD map (at 0.5°), representing small and shallow lakes. They are represented by a pond reservoir fed by the fast reservoir, where the water can be re-infiltrated into the soil at a rate equal to the hydraulic conductivity of the first 2 cm of soil times the pond fraction in the grid-cell (at the same rate as in the floodplains).

4.2 Simulation set-up

4.2.1 Forcing data

The atmospheric forcing data used for all simulations come from the WFDEI (WATCH[1] Forcing Data methodology applied to ERA-Interim reanalysis data) meteorological dataset at 0.5°. Weedon *et al.* (2014) corrected WFDEI meteorological variables for elevation after interpolation and bias correction (using CRU (Climate Research Unit) and GPCC (Global Precipitation Climatology Center) gridded observations), resulting in averages for the three hours before the given time step.

Table 4.2 – Meteorological variables corrected according to variable for WFDEI files (Weedon *et al.*, 2014).



Meteorological Variable	WFDEI Filename Prefix	Variable Units	Elevation Correction After Interpolation	Data Used for Monthly Bias Correction
10 m wind speed	Wind_WFDEI_	$m\,s^{-1}$	Nil	Nil
2 m temperature[b]	Tair_WFDEI_	K	Via environmental lapse rate	CRU TS3.1/3.21 average and CRU TS 3.1/3.21 average diurnal temperature range
Surface pressure	PSurf_WFDEI_	Pa	Via changes in Tair	Nil
2 m specific humidity	Qair_WFDEI_	kg/kg	Via changes in Tair and PSurf	Nil
Downward longwave radiation flux	LWdown_WFDEI_	$W\,m^{-2}$	Via fixed relative humidity and changes in Tair, PSurf, and Qair	Nil
Downward shortwave radiation flux[b]	SWdown_WFDEI_	$W\,m^{-2}$	Nil	CRU TS3.1/3.21 average cloud cover and effects of interannual changes in atmospheric aerosol loading [*Weedon et al.*, 2011]
Rainfall rate[c]	Rainf_WFDEI_GPCC_	$kg\,m^{-2}\,s^{-1}$	Final adjustment of snow/rainfall ratios (section 3.4)	CRU TS3.1 number of wet days, GPCCv5/v6 precipitation totals, ERA-Interim ratio of rainfall/precipitation, rainfall gauge correction
Rainfall rate[d]	Rainf_WFDEI_CRU_	$kg\,m^{-2}\,s^{-1}$	Final adjustment of snow/rainfall ratios (Section 3.4)	CRU TS3.1 number of wet days, CRU TS3.101/TS3.21 precipitation totals, ERA-Interim ratio of rainfall/precipitation, rainfall gauge correction
Snowfall rate[c]	Snowf_WFDEI_GPCC_	$kg\,m^{-2}\,s^{-1}$	Final adjustment of snow/rainfall ratios (section 3.4)	CRU TS3.1 number of wet days, GPCCv5/v6 precipitation totals, ERA-Interim ratio of rainfall/precipitation, snowfall gauge correction
Snowfall rate[d]	Snowf_WFDEI_CRU_	$kg\,m^{-2}\,s^{-1}$	Final adjustment of snow/rainfall ratios (section 3.4)	CRU TS3.1 number of wet days, CRU TS3.101/TS3.21 precipitation totals, ERA-Interim ratio of rainfall/precipitation, snowfall gauge correction

[a] Variable names in the WFDEI filenames and variable units are based on the ALMA (Assistance for Land-surface Modeling Activities) convention (http://www.lmd.jussieu.fr/~polcher/ALMA/).

[b] = 1979–2009 using CRU TS 3.1, 2010–2012 using CRU TS3.21.

[c] = 1979–2009 using GPCCv5 totals, 2010 using GPCCv6 totals (NB: 2011–2012 GPCC-based WFDEI precipitation files are not available at date of paper completion—June 2014).

[d] = 1979–2009 using CRU TS3.101 precipitation totals, 2010–2012 using CRU TS3.21 precipitation totals.

---

[1] WATCH (Water and Global Change) project was developed to help the scientific community analyze and quantify the components of the current and future water cycles, as well the vulnerability of the global water resources. This project resulted on a better understanding of climate change and land use impacts of the global water cycle. Their final report is available at `http://www.eu-watch.org`.

Although GPCC includes around 3-4 times more stations to measure precipitation than CRU (Schneider *et al.*, 2013), they both have good results for precipitation monthly totals when evaluated at regional scale (Weedon *et al.*, 2014). However, when locally assessed, CRU presented lower than expected snowfall rates in mountainous areas, while GPCC showed a good agreement, being a more reliable and accurate forcing dataset (Schneider *et al.*, 2013). For wind speed, air temperature and humidity, and downwards shortwave radiation fluxes showed similarities in means and distributions when compared to near-global daily observations. Based on these results, the forcing datasets used for all simulations were WFDEI corrected by GPCC (further called WFDEI GPCC). The choice of WFDEI GPCC was also confirmed by the analysis of river discharge results from WFDEI CRU, in which WFDEI GPCC produced better river discharge results.

The soil forcing used was the soil texture map called USDA (soil texture from Reynolds *et al.* (2000)), from which saturated hydraulic conductivity values from Carsel and Parrish (1988) are associated with the 12 classes. The vegetation forcing was the standard PFT and LAI (Leaf Area Index) maps associated to the trunk version of ORCHIDEE.

4.2.2 Sensitivity experiments

As described in the preview subsections, the slow reservoir in ORCHIDEE represents groundwater, and base flow is computed using a residence time (same as τ, as it is a linear reservoir) that depends only on topography. To evaluate ORCHIDEE's sensitivity to a change of τ in the slow reservoir, and the effect of τ on simulated river discharge, sensitivity experiments were performed by using the base flow time constants estimated in chapter 3, including additional values described below:

- τ_{ORC}: currently implemented time constant in ORCHIDEE's groundwater reservoir;

- τ_{ORCx3}: using three times τ_{ORC}, chosen arbitrarily as a higher value close to minimum values of τ results obtained in chapter 3;

- τ_{ne}: estimated base flow time constant from chapter 3, using effective porosity, GLHYMPS hydraulic conductivity, 35 m saturated depth, and LCS drainage density, chosen as the closest to references using $K_{GLHYMPS}$;
- τ_{UG}: estimated base flow time constant from chapter 3, using effective porosity, USDA-GLHYMPS hydraulic conductivity, 35 m saturated depth, and LCS drainage density, chosen as the closest τ to references;
- $\tau_{UG/10}$: estimated base flow time constant from chapter 3, using effective porosity, USDA-GLHYMPS hydraulic conductivity, 35 m saturated depth, and LCS drainage density divided by ten;
- τ_{1y}: using a spatially constant value of $\tau_{slow} = 1$ year;
- τ_{10y}: using a spatially constant value of $\tau_{slow} = 10$ years;
- τ_{100y}: using a spatially constant value of $\tau_{slow} = 100$ years.

Table 4.3 – Summary of the simulations performed with ORCHIDEE. The initialization Std is the ORCHIDEE standard (zero initial volume, indicated by Median V_{slow} (mm)).

	τ_{ORC}	τ_{ORCx3}	τ_{ne}	τ_{UG}	$\tau_{UG/10}$	τ_{1y}	τ_{10y}	τ_{100y}
Median τ_{slow} (years)	0.12	0.37	64.63	7.11	0.71	1	10	100
Median V_{slow} (mm)	0	7.81	2,297	207	20.7	34.0	340	3,402
Initialization	Std	Eq.4.6	Eq.4.6	Eq.4.6	Eq.4.6	Eq.4.6	Eq.4.6	Eq.4.6
Spin up	6 years							
Trunk version	3970							
Hydrology scheme	CWRR (11 layers, 2 m depth)							
Routing	Activated							
Floodplains	Deactivated							
Routing time step	6 hours (21,600 s)							
Time step	30 minutes							
Study period	1979 - 2010 (evaluations made for 1985 - 2010)							
Soil forcing	USDA (soil texture from Reynolds *et al.* (2000))							
Atmospheric forcing	WFDEI GPCC (at 0.5°)							
STOMATE	Activated							
Explicit snow	Activated							
Soil freeze	Deactivated							

The standard parameters of ORCHIDEE's trunk version were used, except for the routing time step (six hours, instead of the one day standard, for better numerical accuracy), and τ (since is the evaluated parameter). Table 4.3 summarizes the relevant configuration and parameters used on all simulations, in which they were identified by the τ used.

4.2.3 Initialization

The standard initialization of ORCHIDEE uses the volume of all reservoirs equals zero, that start to fill based on the water balance and soil hydrology described on this chapter, depending on τ (Equation 4.3). The initialization of the slow reservoir is an important step, as the tested τ showed in Table 4.3 are higher than the ORCHIDEE standard, resulting in longer period to achieve stationary state. To illustrate that, an arbitrary simulation was performed with the standard initialization of ORCHIDEE, with τ equals ten times τ_{ORC} (median = 1.22 years). Figure 4.4 shows that it takes circa 13 years for the stationary state to be achieved, which can be even longer for higher τ (as some of the evaluated τ can reach several thousands of years), which consumes important computing time.

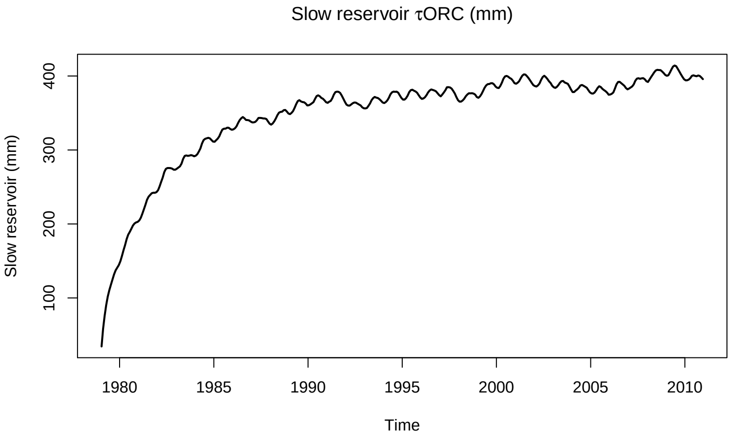


Figure 4.4 – Slow reservoir of ORCHIDEE using the standard initialization, with $\tau_{slow} = 10\tau_{ORC}$.

In order to avoid a long period of spin up, all simulations were initialized based on a six-year spin up (from 1979 to 1984). Simulations that used a time constant different from τ_{ORC} had the volume of the slow reservoir manually corrected using the average drainage of the τ_{ORC} at the stationary state (that is independent of τ):

$$V_{ss-slow} = D_r \cdot \tau_{slow} \cdot 86400 \cdot A, \tag{4.6}$$

where $V_{ss-slow}$ is the volume of the slow reservoir (in kg) at steady state, D_r is the average drainage (kg.m^{-2}.s^{-1}) of a 26 years simulation using τ_{ORC} (Table 4.3), τ_{slow} is the time constant used in the simulation (days), A is the grid-cell area (since at 0.5° the land area corresponds to the grid-cell area), and 86400 is a unit conversion factor (days into seconds). This procedure allowed to reduce the spin up period, and an additional six years (1979 - 1984) were used for all simulations to assure that the seasonal disequilibrium was captured (Figure 4.4).

4.3 Validation data

The global datasets used to validate ORCHIDEE simulation results are described in detail below, and summarized in Table 4.4, separated by type of data (evapotranspiration (E), precipitation (P), runoff (R), river discharge (Q), and total water storage anomalies (TWS)), and type of product (model (M) or multi-model results (MM), and observations (O)). In the case where modeling results were provided (indicated by an *), the resultant values represent the global average (including Antarctica, with the same coverage of all simulations) of the given time period.

- **Precipitation, evaporation, and runoff:**

  - de Marsily (1995): provides an estimation of the global water cycle: total evaporation, total runoff, and total precipitation. Values are based on global fluxes from *World Resources 1990-1991*, excluding Antarctica.

  - Dirmeyer *et al.* (2005): Second Global Soil Wetness Project (GSWP-2) is a multi-model (15 models[2]) analysis of the land surface state variables and fluxes. The results are provided as a 10 years time series, at 1° resolution. The atmospheric forcing used is from NCEP/DOE (National Centers for Environmental Prediction / Department Of Energy) at 1° resolution, with 3 h time step. The data are available for download, as monthly averages from 1986 to 1995.

- Oki and Kanae (2006): give an estimation of the world water resources and fluxes by combining the information of several references, excluding Antarctica.

- Trenberth *et al.* (2007): estimate the storage and fluxes of the hydrological cycle by comparing several references, and reanalysis of simulation results.

- Harding *et al.* (2011): WATer and global CHange (WATCH) is a project created to better understand the water cycle. It provides multi-model[3] average results (1901-2001) of evaporation, runoff, and precipitation, at 0.5° resolution.

- Getirana *et al.* (2014): they used simulation with ORCHIDEE: without routing, STOMATE, irrigation, floodplains, and swamps (all modules deactivated); using the CWRR scheme (11 layers, 2 m depth), Zobler soil forcing, and Princeton with GPCC correction atmospheric forcing, at 1° resolution. The data used for the comparison were the monthly averaged values of the 30 years simulation (from 1979 to 2008). Although this reference used a zoom over the Amazon region, the global simulation results were used on the comparisons.

- Rodell *et al.* (2015): they quantified mean annual and monthly fluxes of the water cycle (at global scale), based on satellite measurements and data-integrating model results. They provide optimized estimates using observation data on water budget equations (to close water budget), that are consistent with observations and associated uncertainties. The values used for comparison are the optimized data.

---

[2] The 15 models used on GSWP-2 project are: CLM2-TOP, HY-SSiB, ISBA, Mosaic, MOSES2, NOAH, NSIPP-Catchment, SiBUC, SSiBCOLA, SWAP, VISA, LaD, ORCHIDEE, Sland, and BucketIIS.

[3] The models used on WATCH project are: JULES, ORCHIDEE, HTessel, H08, LPJmL, MacPDM, Matsiro, GWAVA, MPI-HM, VIC, WaterGAP.

- **Precipitation:**

  - Schneider *et al.* (2011): Global Precipitation Climatology Center (GPCC) provides a dataset that contains precipitation data from 67,200 stations worldwide, at 0.5° resolution. The dataset downloaded is a monthly averaged precipitation time series (v.56) from 1981 to 2010 (30 years), at 0.5° resolution.

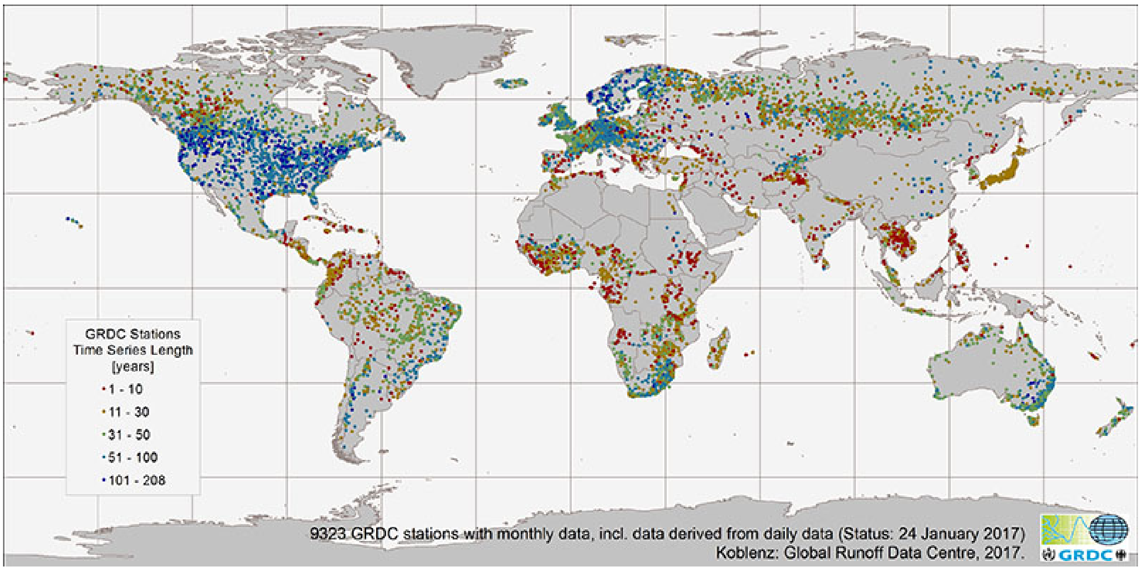
  - Jones and Harris (2013): The Climate Research Unit (CRU) provides a dataset that contains precipitation information from monitoring stations around the world since 1982 to 2010.

- **Total water storage anomalies:**

  - Swenson (2012): Gravity Recovery and Climate Experiment (GRACE) observes temporal variations of Earth's gravitational potential (Landerer and Swenson, 2012) from which three products of total water storage are derived (GFZ - Geoforschungs Zentrum Potsdam, CSR - Center for Space Research at University of Texas, and JPL -Jet Propulsion Laboratory). GRACE total water storage anomalies were obtained by the average value of the three products (JPL, GFZ, and CSR) (Sakumura *et al.*, 2014). The evaluated period goes from 2002 to 2010, in which GRACE anomalies were provided (average between 2004 and 2009 removed).

- **River discharge:**

  - GRDC (2014): this data contains monthly averaged river discharge information of 5723 stations, with global coverage. Since the data is monitored with variable times of data acquisition start and end (Figure 4.5), and monitoring interruptions, the time record contains many missing values. The time series goes from 1807 to 2014, with monthly time step. From all available GRDC stations, only 276 stations have upstream area error compared to ORCHIDEE under five percent that were used for statistical analysis of river discharge results.



9323 GRDC stations with monthly data, incl. data derived from daily data (Status: 24 January 2017)
Koblenz: Global Runoff Data Centre, 2017.

Figure 4.5 – GRDC stations classified by time series length GRDC (2014).

Table 4.4 – Summary of the datasets used for validation of ORCHIDEE's results. P = total precipitation; E = total evapotranspiration; R = total runoff; Q = river discharge; TWS = total water storage anomalies; O = observed/literature values; M = model results; MM = multi-model results; * = global averages are calculated with the provided dataset for the whole time period.

Reference	Data	Product	Resolution	Period
de Marsily (1995)	P, E, R	O	Global	-
Dirmeyer *et al.* (2005)	P, E, R	MM*	1°	1986-1995
Oki and Kanae (2006)	P, E, R	O	Global	-
Trenberth *et al.* (2007)	P, E, R	MM	Global	-
Harding *et al.* (2011)	P, E, R	MM	0.5°	1901-2001
Getirana *et al.* (2014)	P, E, R	M*	1°	1979-2008
Rodell *et al.* (2015)	P, E, R	MM	Global	-
Schneider *et al.* (2011)	P	M*	0.5°	1981-2010
Jones and Harris (2013)	P	M*	0.5°	1982-2010
GRDC (2014)	Q	O	Stations	1807-2014
Swenson (2012)	TWS	O	1°	2002 - 2010

4.4 Simulation results

4.4.1 Evaluation of the reference simulation τ_{ORC}

The simulation τ_{ORC} (Table 4.3) was used as a reference to evaluate how the changes in the groundwater reservoir (by using different values of τ) can affect simulated river discharge in ORCHIDEE. On global average, total precipitation results of τ_{ORC} are in agreement with the validation data (Table 4.6) as expected since the precipitation data comes from an atmospheric forcing that is corrected to be in agreement with observations. Larger differences for the three components of the water budget observed in de Marsily (1995) and Oki and Kanae (2006) in Figure 4.6 are explained by the exclusion of Antarctica used in these observations, that are reduced to under 5% when comparing with the same extent. Total precipitation, evaporation and runoff are in agreement with the validation data, with differences that are in average 5%. The major differences are observed for Trenberth *et al.* (2007) (in which they can reach 30% for total runoff), explained by the use of different time periods for the collected observed datasets for precipitation and evaporation. They also alert for the use of total runoff, in which they state that their values are not quantitatively correct, and should not be used for assessments. τ_{ORC} has a closed water budget, with a 0.007 mm.d^{-1} residual.

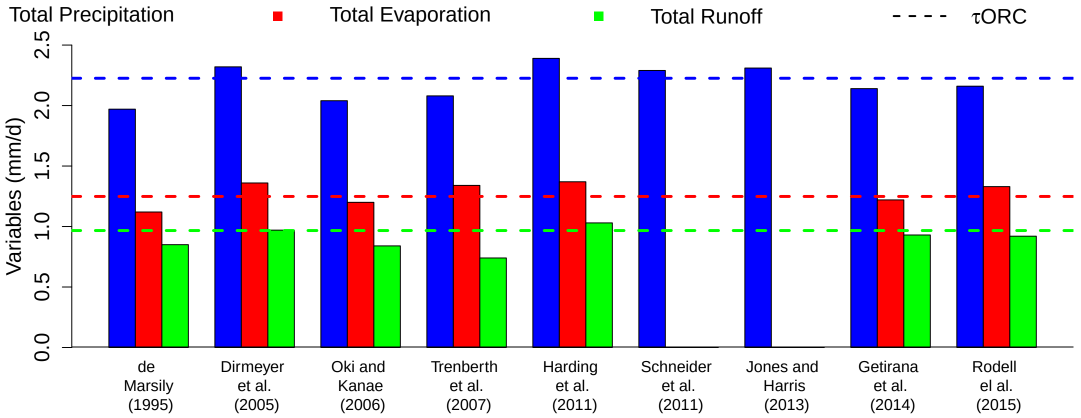


Figure 4.6 – Comparison of the global average results (1985-2010) for τ_{ORC} to the validation data for precipitation (blue), evaporation (red), and runoff (green), in mm.d^{-1}. The bars represent the reference data, and the lines the simulated results.

Nineteen major watersheds were chosen to evaluate river discharge, comparing to GRDC observations (Figure 4.7). The evaluated period was between 1985-2010 for both simulations and observation. Periods without observations were excluded from the evaluation. τ_{ORC} resulted in several hydrographs with water balance problems (Figure 4.7 for stations 1, 2, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 19). In northern rivers (1, 2, 9, 10, and 11 in Figure 4.7), the simulated river discharge was underestimated, possibly due to the soil freeze that was not activated. For the majority of the evaluated watersheds, τ_{ORC} river discharge is not in agreement with observed data.

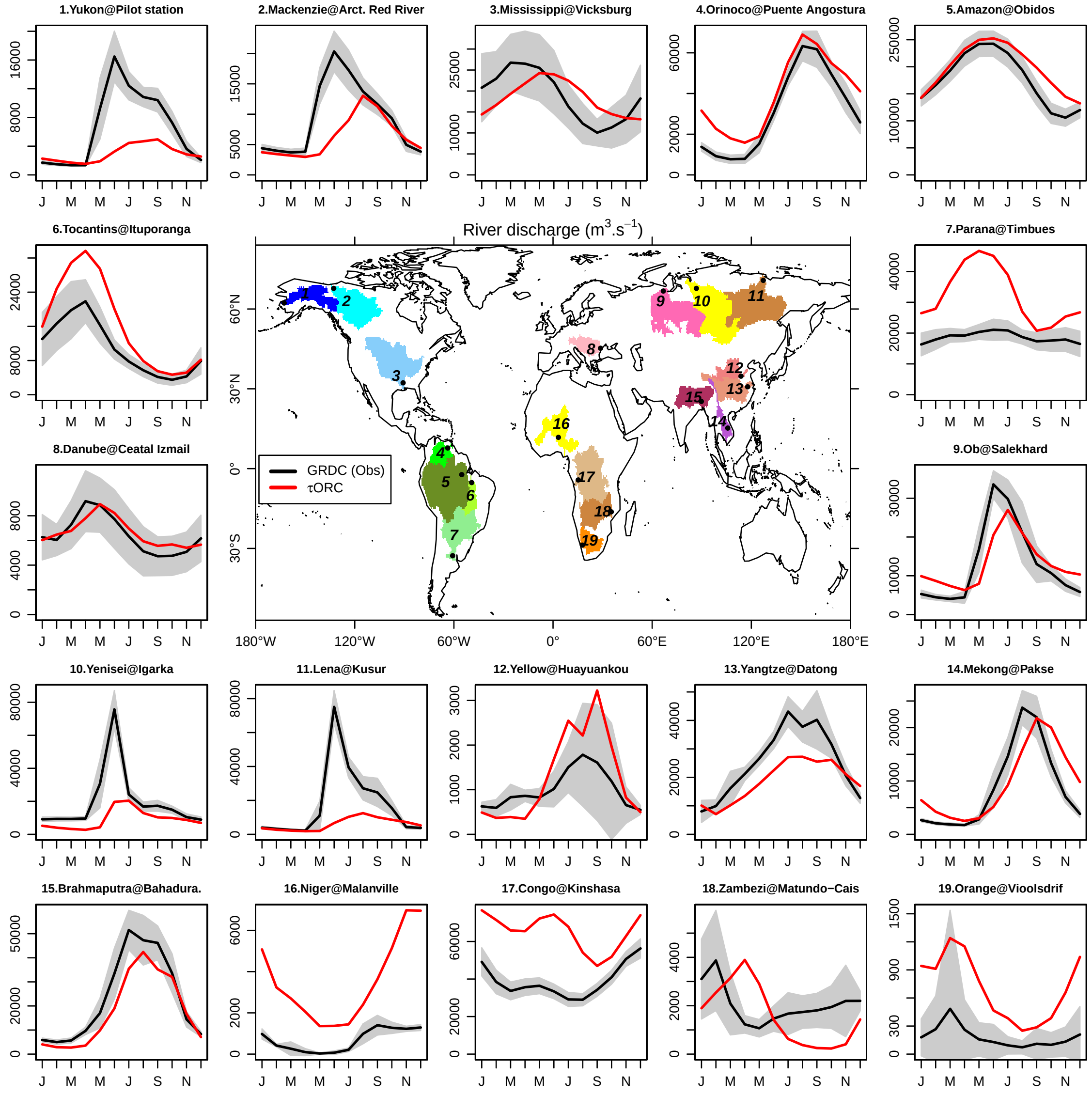


Figure 4.7 – Evaluation of the simulated river discharge of τ_{ORC} (red) against observed river discharge from GRDC (black), between 1985-2010. Shaded areas represent values within ±1 standard deviation of the observed period.

When the 276 selected GRDC stations were evaluated, it was observed that river discharge results were biased (positively and negatively, depending on the station), with the majority of the stations reaching more than 100%, with only seven stations in the range of -1% and 1% relative bias (Figure 4.8). The use of a higher τ (as the ones tested) can not improve the relative bias observed in the τ_{ORC} simulation results, since higher τ results in a reduction of the hydrograph amplitude and a delay in the peak of the hydrograph (Erskine and Papaioannou, 1997).

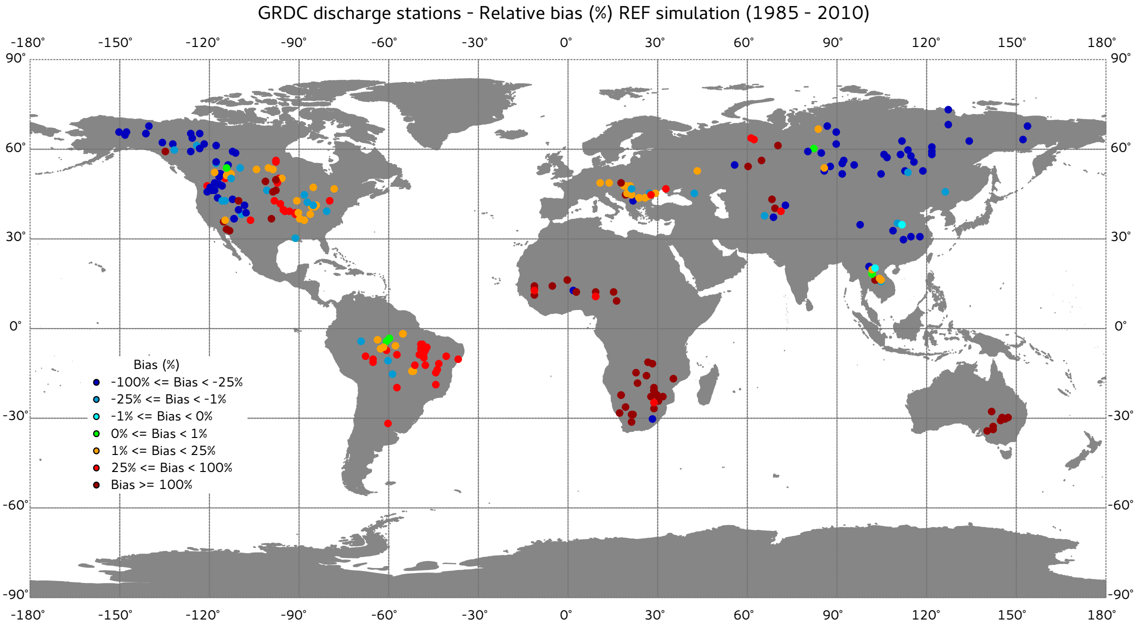

GRDC discharge stations - Relative bias (%) REF simulation (1985 - 2010)
-180°
-150°
-120°
-90°
-60°
-30°
0°
30°
60°
90°
120°
150°
180°
90°
60°
30°
0°
-30°
-60°
-90°
Bias (%)
● -100% <= Bias < -25%
● -25% <= Bias < -1%
● -1% <= Bias < 0%
● 0% <= Bias < 1%
● 1% <= Bias < 25%
● 25% <= Bias < 100%
● Bias >= 100%

Figure 4.8 – Relative bias (%) calculated for the 276 GRDC stations that have an upstream area error under 5% compared to the ORCHIDEE upstream area.

Although most of the evaluated stations are biased for river discharge, 64% of the evaluated GRDC stations have correlation coefficients over 0.75 (Figure 4.9), indicating that most of the evaluated stations have hydrographs positively correlated to observations. However, only 23% of the stations have a Nash-Sutcliffe coefficient over 0.5 (Figure 4.9), which indicates that the model does not simulate river discharge accurately for the majority of the stations.

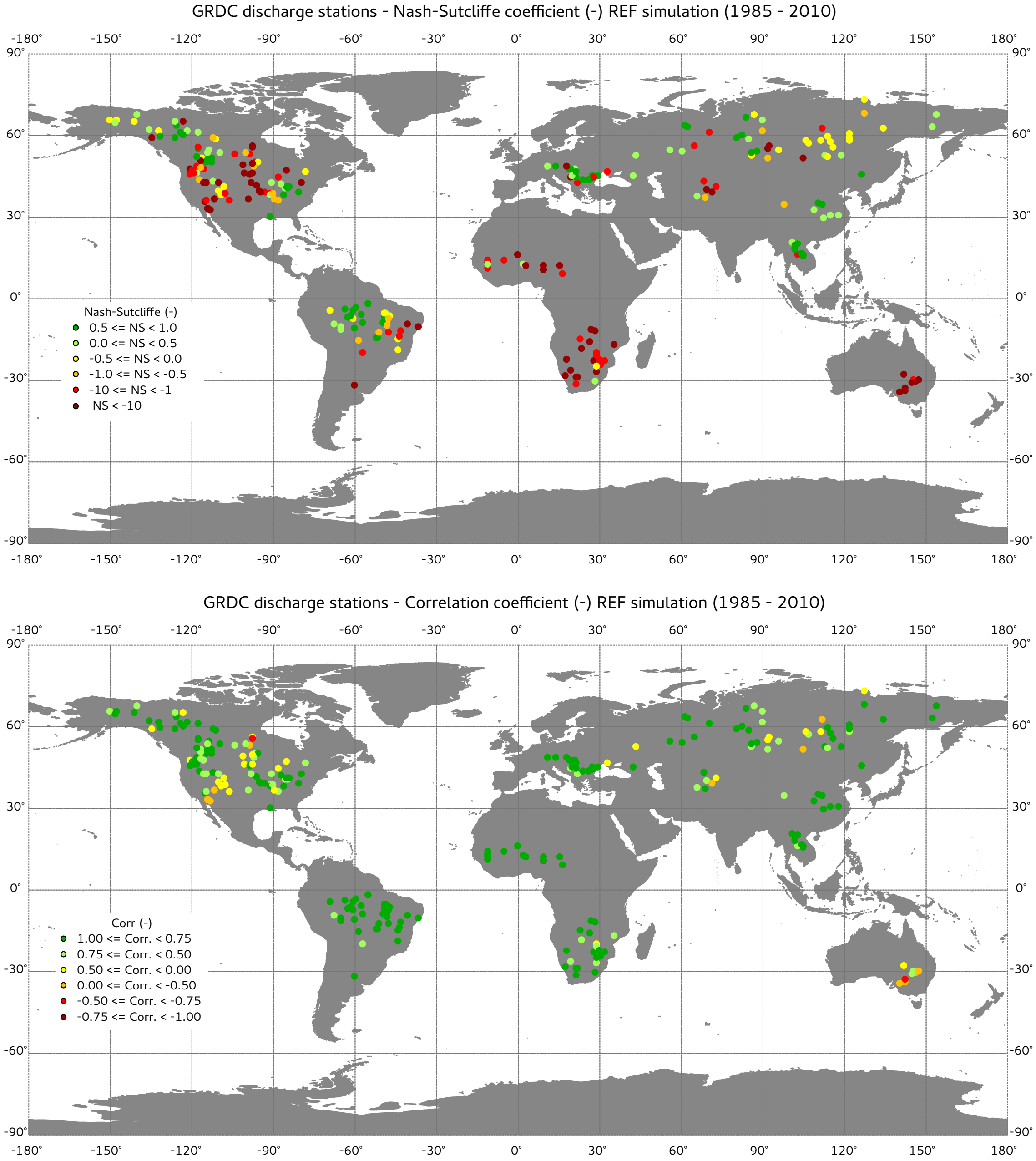

GRDC discharge stations - Nash-Sutcliffe coefficient (-) REF simulation (1985 - 2010)
Nash-Sutcliffe (-)
● 0.5 <= NS < 1.0
● 0.0 <= NS < 0.5
● -0.5 <= NS < 0.0
● -1.0 <= NS < -0.5
● -10 <= NS < -1
● NS < -10
GRDC discharge stations - Correlation coefficient (-) REF simulation (1985 - 2010)
Corr (-)
● 1.00 <= Corr. < 0.75
● 0.75 <= Corr. < 0.50
● 0.50 <= Corr. < 0.00
● 0.00 <= Corr. < -0.50
● -0.50 <= Corr. < -0.75
● -0.75 <= Corr. < -1.00
-180° -150° -120° -90° -60° -30° 0° 30° 60° 90° 120° 150° 180°
90° 60° 30° 0° -30° -60° -90°

Figure 4.9 – Nash-Sutcliffe coefficient (-) on the top and correlation coefficients (-) on the bottom calculated for the 276 GRDC stations that have an upstream area error under 5% compared to the ORCHIDEE upstream area.

Total water storage anomalies (TWS) results of the 19 major watersheds chosen to evaluate river discharge were compared to GRACE (Swenson, 2012) data for the same period described in Table 4.4. Figure 4.10 showed that for most of the evaluated watersheds ORCHIDEE results are in agreement with observed data from GRACE, with $R^2 > 0.90$ (Amazon, Congo, Mississippi, Parana, Yenisei, Ob, Niger, Zambezi, Yangtze, Mackenzie, Brahmaputra, Orinoco, Yukon, Mekong, and Tocantins). Only Lena and Orange had $R^2 < 0.70$.

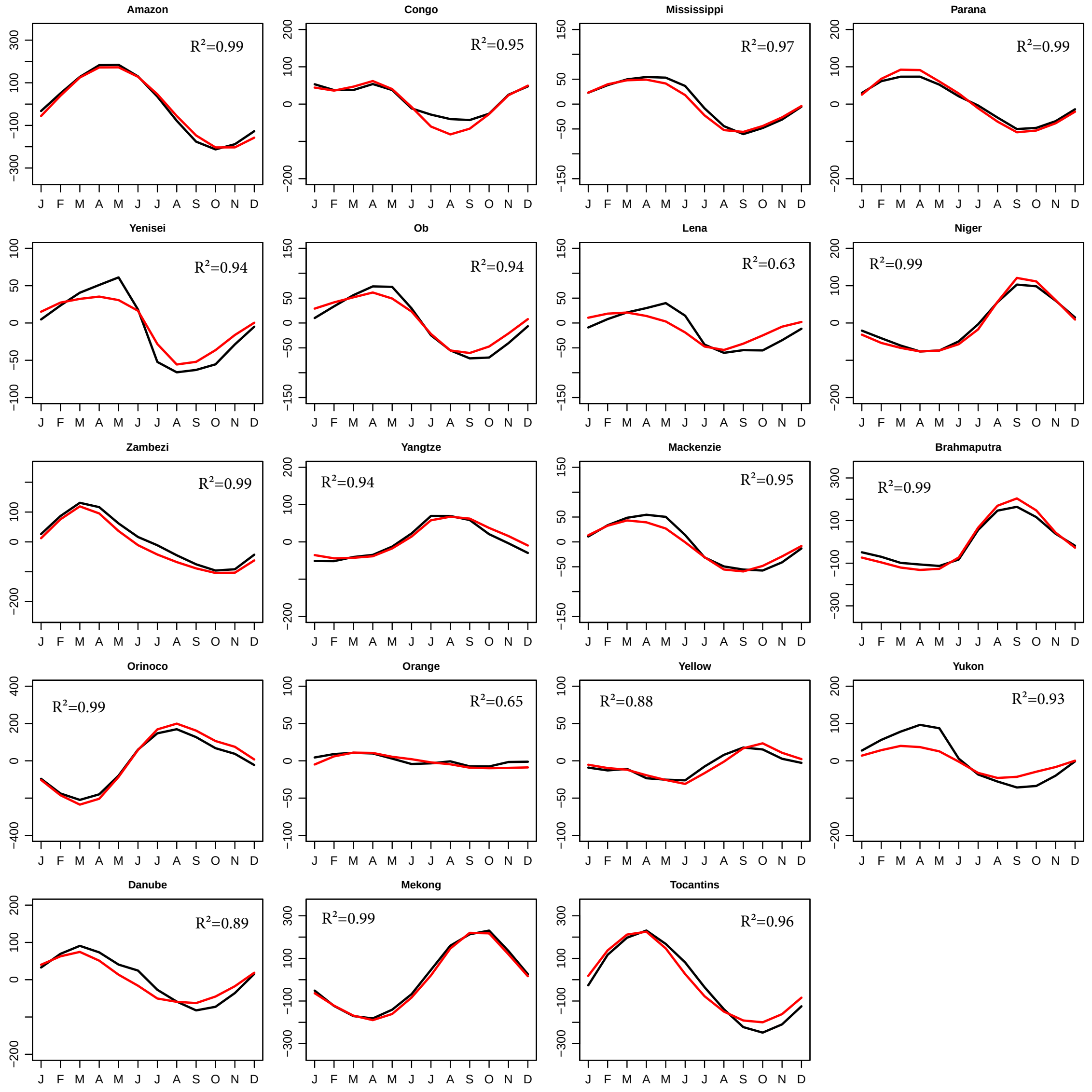


Figure 4.10 – Total Water Storage anomalies (TWS) in mm (y axis) for 19 major watersheds in the world. Black lines represent GRACE data (Swenson, 2012), and red lines the reference simulation. Months (x axis) represent the averages for 2002 to 2010 period. No data values were excluded of the analysis.

4.4.2 Sensitivity analysis

Sensitivity analyses of τ were performed by evaluating the river discharge and the groundwater reservoir with different base flow time constants described on Table 4.3. In a first step, river discharge was evaluated by using different τ increasing in time (τ_{ORCx3}, τ_{1y}, τ_{10y}, τ_{100y}) compared to the reference simulation. The buffer effect of the river discharge increased with τ, as higher τ values result in smoother (lower

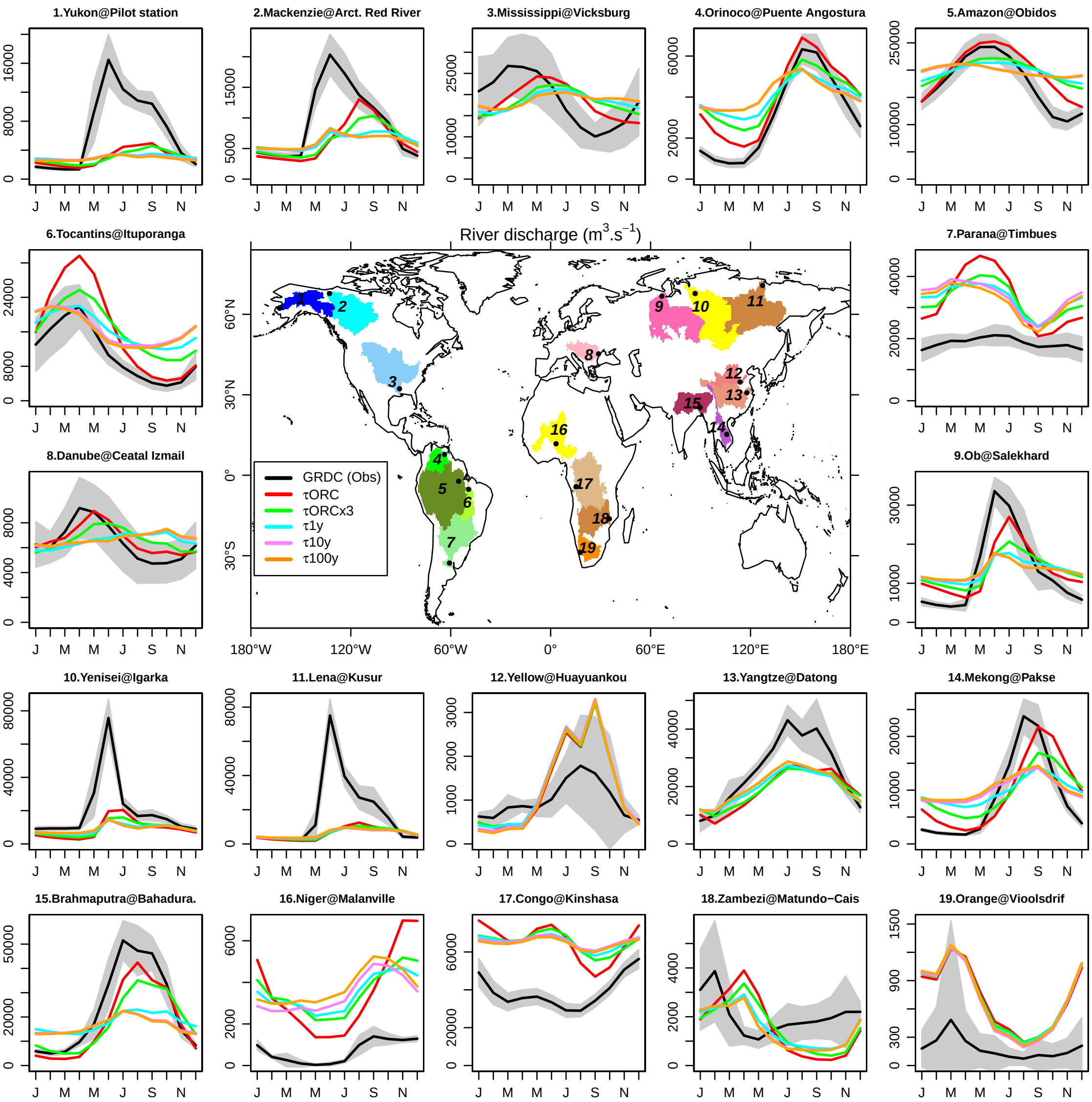


Figure 4.11 – Evaluation of the simulated river discharge of τ_{ORC} (red), τ_{1y} (green), τ_{10y} (cyan), and τ_{100y} (pink) against observed river discharge from GRDC (black), between 1985-2010. Shaded areas represent values within ±1 standard deviation of the observed discharge.

in range) curves (Figure 4.11), since the drainage that arrives to the groundwater will take longer time to arrive to the stream. This is coherent with many other studies, as in Erskine and Papaioannou (1997), Ducharne *et al.* (2003), Gascoin *et al.* (2009), among others. The river discharge showed highly sensitive to changes in τ for the majority of the evaluated watersheds, and two main behaviors emerged from Figure 4.12: there are a couple of basins where the buffer effect is not observed (river discharge curves for all simulations are almost perfectly overlapped, represented by

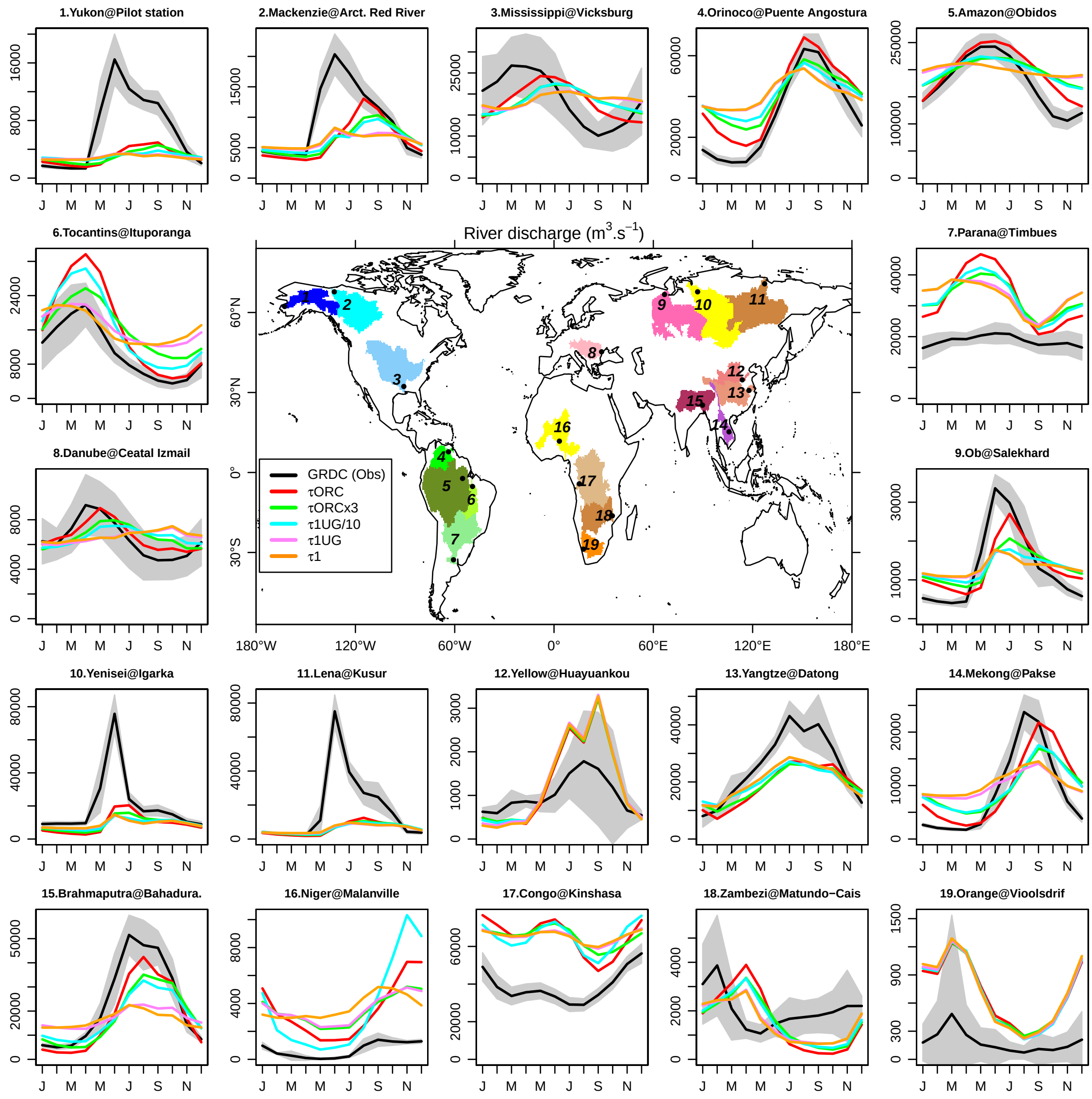


Figure 4.12 – Evaluation of the simulated river discharge of τ_{ORC} (red), τ_{ORCx3} (green), $\tau_{UG/10}$ (cyan), τ_{UG} (pink), and τ_1 against observed river discharge from GRDC (black), between 1985-2010. Shaded areas represent values within ±1 standard deviation of the observed discharge.

Figure 4.11.12 and 4.11.19); but the majority of basins exhibit a significant buffer effect. When evaluating river discharge with higher τ (Figure 4.12), no major differences in the river discharge were observed for watersheds with elevated τ (Table 4.5, τ_{UG}, τ_{ne}), in which the same behavior was observed for τ_{10y} and τ_{100y} in Figure 4.11. Although τ_{UG} is significantly smaller than τ_{ne} in Yukon and Mackenzie (Figure 4.12.1 and 4.12.2), the differences in the river discharge are not observed due to the high pick of the observed data (as in Figure 4.11 for τ_{1y} and τ_{10y}). These results indicate that the proposed τ is probably not suitable for shallow groundwater reservoirs, that have faster interactions with the surface. A possible solution would be the introduction of a second compartment in the groundwater scheme (as in Chapman (1999)), representing both fast and slow interactions with the surface, with a more realistic representation of the groundwater. However, other variables in the model can affect the river discharge. It is recommended to test other model configurations (as in activating the floodplains module, that can ameliorate the water balance, or using the soil freeze, that should improve river discharge simulations in the northern rivers).

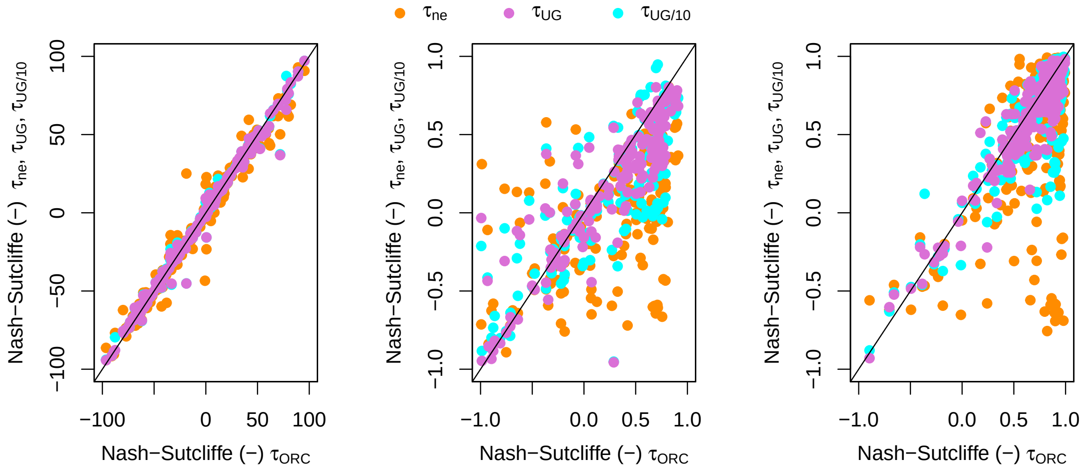


Figure 4.13 – Scatter plot of Nash-Sutcliffe coefficients, bias, and correlation coefficients calculated for river discharge measurements from GRDC and simulations with ORCHIDEE of τ_{ORC} versus τ_{ne}, τ_{UG}, $\tau_{UG/10}$.

When comparing statistical analysis results to the reference simulation τ_{ORC}

(Figure 4.13), the use of a longer time constant improved only 10% of simulated river discharges for τ_{ne} and 18% for $\tau_{UG/10}$ when evaluating Nash-Sutcliffe coefficients, which was under 10% for bias and correlation coefficients.

Table 4.5 – Mean slope (-) and τ (years) per watershed for the evaluated GRDC stations. τ_{1y}, τ_{10y}, and τ_{100y} were not represented since they are constant for all basins, equals 1, 10 and 100 years, respectively.

Basin	τ_{ORC}	τ_{ORCx3}	τ_{ne}	τ_{UG}	$\tau_{UG/10}$	Slope
1.Yukon	0.11	0.33	4,968.5	0.71	0.07	0.03
2.Mackenzie	0.12	0.36	1,992.5	5.36	0.54	0.03
3.Mississippi	0.17	0.51	651.1	13.27	1.33	0.01
4.Orinoco	0.24	0.72	71.69	18.60	1.86	0.01
5.Amazon	0.28	0.84	156.9	12.48	1.25	0.01
6.Tocantins	0.20	0.60	66.35	10.90	1.09	0.01
7.Parana	0.22	0.66	579.0	15.44	1.54	0.05
8.Danube	0.14	0.42	321.3	19.08	1.91	0.09
9.Ob	0.22	0.66	2,572.3	88.27	8.83	0.05
10.Yenisei	0.15	0.45	250.7	34.77	3.48	0.02
11.Lena	0.14	0.42	584.1	36.99	3.70	0.02
12.Yellow River	0.14	0.42	238.5	45.45	4.54	0.08
13.Yangtze	0.16	0.48	180.2	17.48	1.75	0.06
14.Mekong	0.21	0.63	188.1	9.35	0.93	0.08
15.Brahmaputra	0.18	0.54	237.1	15.73	1.57	0.02
16.Niger	0.23	0.69	381.1	36.33	3.63	0.04
17.Congo	0.22	0.66	46.5	5.60	0.56	0.06
18.Zambezi	0.20	0.60	2,255.1	15.52	1.55	0.02
19.Orange	0.14	0.42	687.4	16.05	1.60	0.03

As described in section 4.1, the slow reservoir in ORCHIDEE has drainage as a single input, discharging in the stream with a delay represented by τ. This configuration indicates that changes on river discharge could also depend on drainage. The amplitude difference of the river discharge between τ_{ORC} and τ_{ne} (representing the smallest and highest τ, Table 4.5) was compared to the average drainage for the reference simulation of each watershed (Figure 4.14). Basins that have low drainages are mainly governed by surface processes, and base flow is too small to effectively contribute to the stream flow when compared to the surface runoff. For basins with

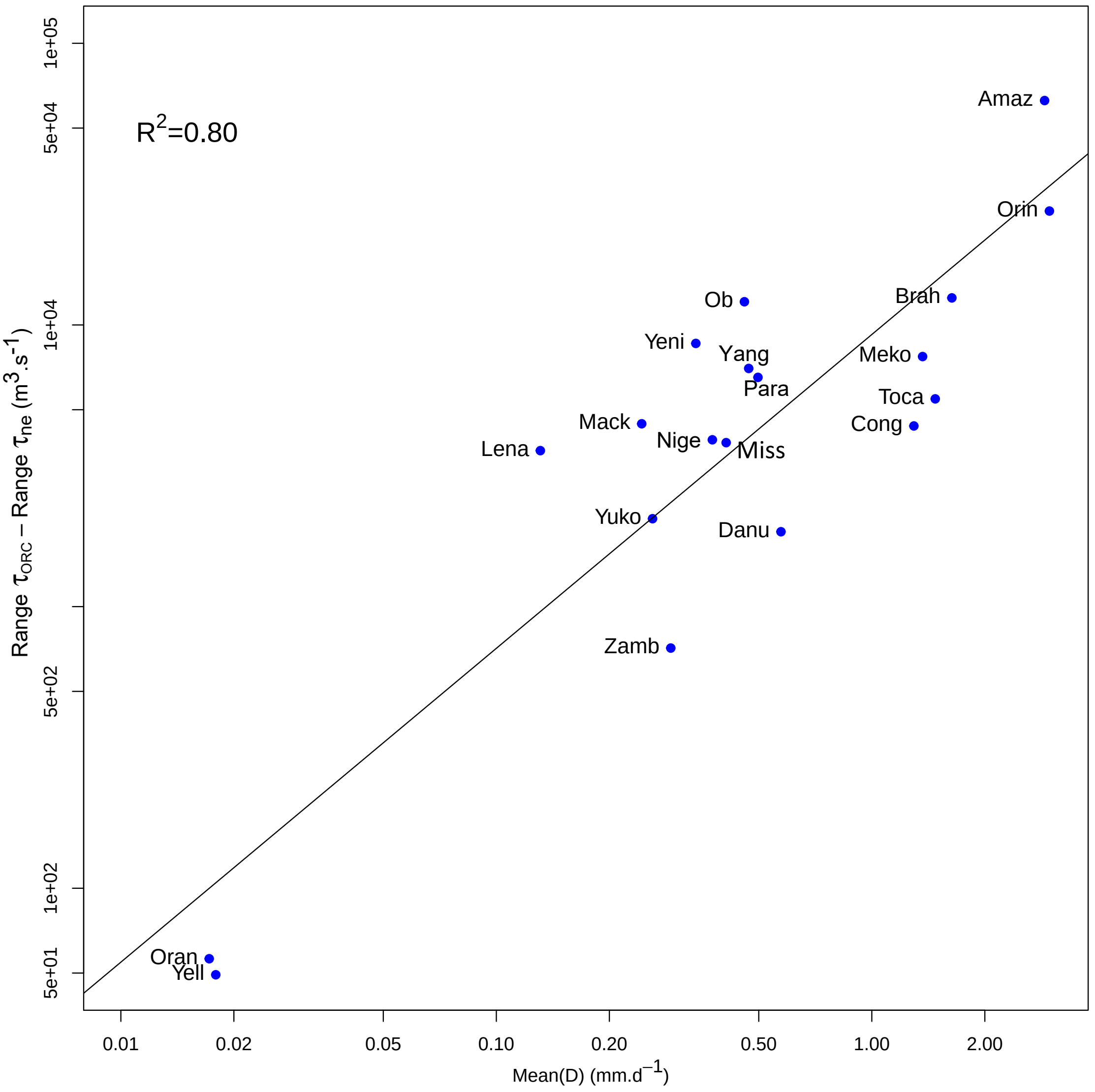


Figure 4.14 – Scatter plot of the mean drainage (mm.d^{-1}) versus river discharge range (m^3.s^{-1}) per basin (indicated by their first four letters). Range is given by the difference between maximum and minimum value of the river discharge (Q) shown in Figure 4.6 for the τ_{ORC} simulation, and drainage (D) is the mean value for the same time period (1985-2010).

mean drainage under 0.05 mm.d^{-1}, the buffer effect was not observed for the river discharge (Figures 4.11.12, 4.11.19, 4.12.12, and 4.11.19). A log-linear relationship was observed between the discharge and the difference of river discharge amplitude (Figure 4.14), with a R^2 of 0.80. The difference of river discharge amplitude increases with drainage. Attention should be paid to the scale used for each hydrograph, as in Niger the large differences among τ_{1y}, τ_{10y}, and τ_{100y} hydrographs are related to small amplitudes (up to 4,000 $m^3.s^{-1}$), and in Lena the small differences are due to the high observed data values (up to 80,000 $m^3.s^{-1}$).

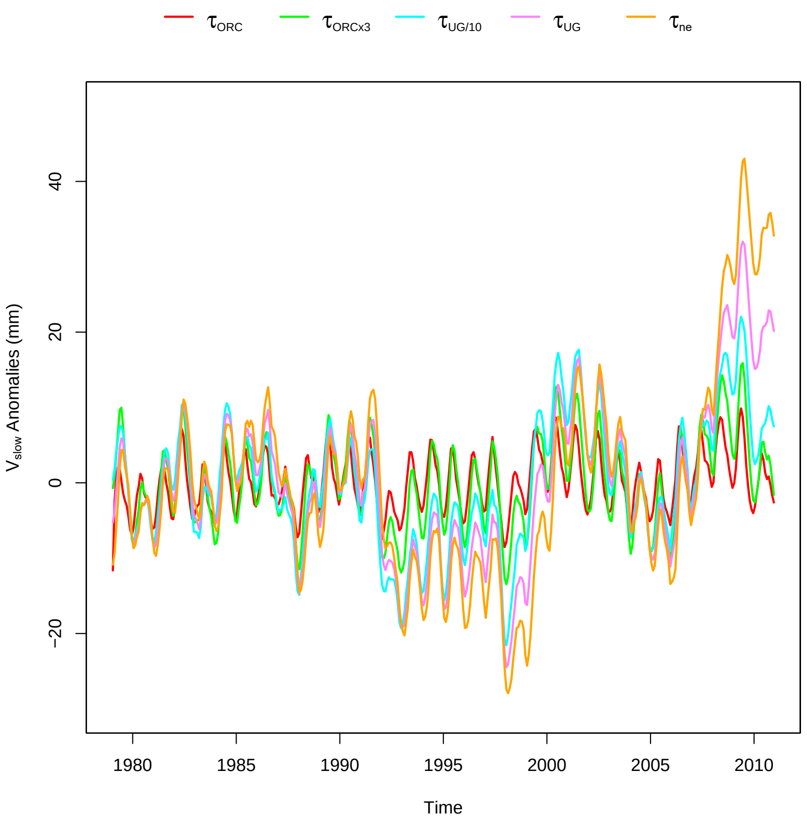


Figure 4.15 – Slow reservoir anomalies for τ_{ORC} (red), τ_{ORCx3} (green), $\tau_{UG/10}$ (cyan), τ_{UG} (pink), and τ_{ne} (orange) simulations.

In a second step ORCHIDEE was evaluated by comparing the global average slow reservoir anomalies (global spatial average of the slow reservoir volume to the temporal average for the evaluated time series) among each other for τ_{ORC}, τ_{ORCx3}, $\tau_{UG/10}$, τ_{UG}, and τ_{ne} (Figure 4.15). The amplitude of the anomalies increased with τ (Figure 4.15, Table 4.3), showing differences among the evaluated simulations that could not be observed when evaluating only the river discharge (Figure 4.12).

ORCHIDEE results of TWS anomalies were compared to GRACE observations (Figure 4.16), resulting in little seasonal variations of TWS anomalies with τ among the evaluated simulations. The use of a longer τ deteriorates simulations results (represented by τ_{ne} in green) for most of the evaluated watersheds (Table 6.2), compared to the reference simulation τ_{ORC}. For the Mississippi basin a little improvement was

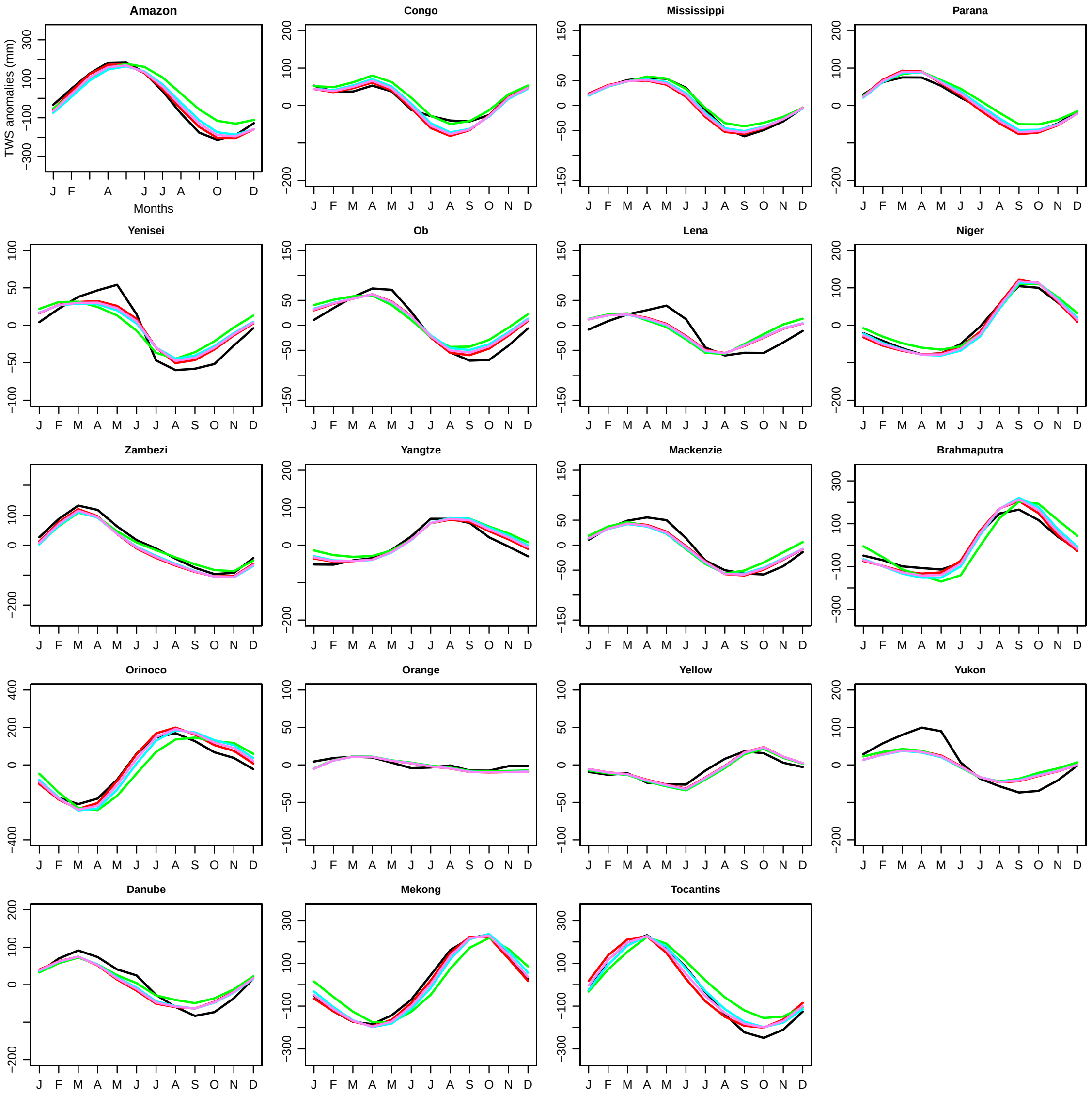


Figure 4.16 – TWS anomalies (mm) for τ_{ORC} (red), $\tau_{UG/10}$ (pink), τ_{UG} (cyan), and τ_{ne} (green) simulations and observations from GRACE (black).

observed with increasing τ for the coefficient of determination ($R^2_{ORC} = 0.97$ against $R^2_{ne} = 0.99$), but the highest Nash-Sutcliffe coefficient was observed for τ_{UG} (0.98, Table 6.2). For the Danube and Tocantins the improvement of both statistical coefficients was observed for τ_{UG} and $\tau_{UG/10}$ simulations compared to τ_{ORC}, while for τ_{ne} the Nash-Sutcliffe coefficient was reduced (Table 6.2). Although lower base flow time constants (as the currently implemented in ORCHIDEE, τ_{ORC}) have better simulation results not only for river discharge, but also for the TWS anomalies in the most of the evaluated watersheds.

Erskine and Papaioannou (1997) evaluated the aquifer response rate (given as $1/\tau$), finding a relationship that describes the buffer effect (that they represented by a reduction factor) as function of the aquifer response rate, using an analytical approach. The reduction factor is expressed by the ratio of minimum to the average base flow, and the relationship to the aquifer response rate is shown in Figure 4.17. They observed that the river discharge buffer effect to τ increased for higher τ values, and the river discharge amplitude was reduced by the reduction factor. When plotting the calculated reduction factor for τ_{ne}, τ_{UG}, and $\tau_{UG/10}$ simulations for all evaluated basins versus the average τ (Table 4.5), and comparing to the curve given by Erskine and Papaioannou (1997) (Figure 4.17), it was observed that the calculated values for τ_{ne} and τ_{UG} are shifted towards longer τ, with smaller aquifer response rates than the curve proposed by Erskine and Papaioannou (1997). For $\tau_{UG/10}$, the relationship between aquifer response rate and reduction factor do not follow the slope of the curve proposed by Erskine and Papaioannou (1997). However for τ_{UG} and τ_{ne}, this could be explained by the transient behavior imposed by the atmospheric forcing used in the simulations, combined with the assumptions inherent to τ formulations. Both τ_{UG} and τ_{ne} were estimated considering sloping aquifers, resulting in higher values of τ (consequently lower aquifer response rate) for lower slopes.

4.5 Conclusions

This Chapter evaluated the use of a physically based base flow time constant (τ) estimated in Chapter 3 on ORCHIDEE Land Surface Model, which uses a simple groundwater scheme. ORCHIDEE routes the water horizontally, through a stream reservoir that is locally delayed by a fast (representing the delay between surface runoff production and the arrival of overland flow to the stream) and a slow reservoir (that represents the groundwater). A reference simulation τ_{ORC} was performed with the standard ORCHIDEE configurations and parameters that resulted in a closed water budget in agreement with available references. A total of eight time constants were used to evaluate the sensitivity of ORCHIDEE river discharge and

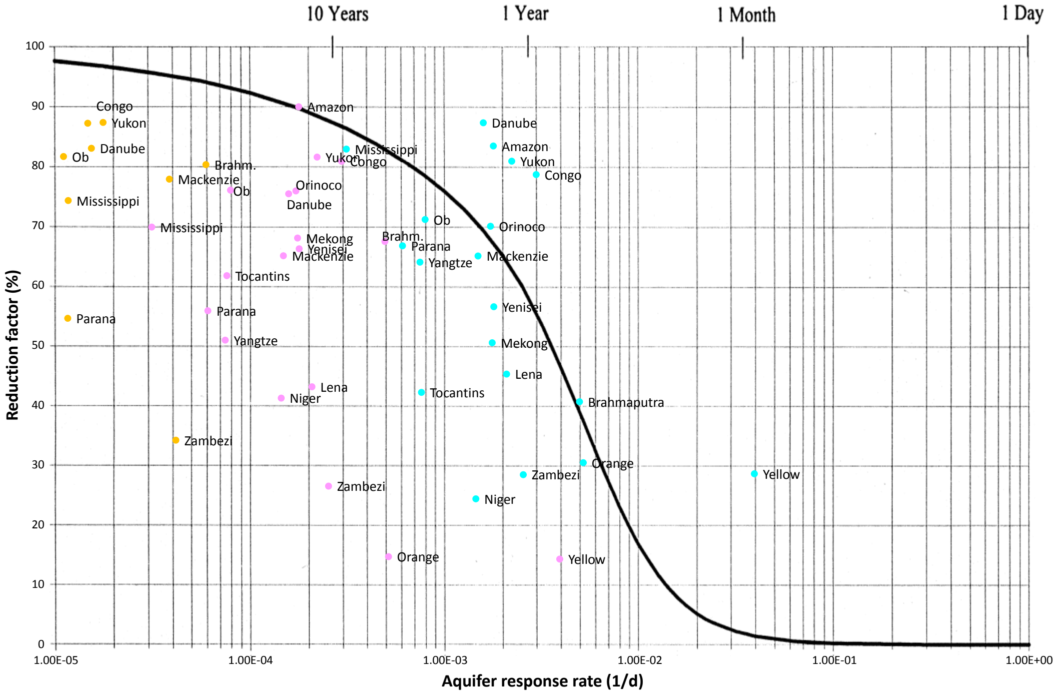


Figure 4.17 – Aquifer response rate (day^{-1}) versus reduction factor (%) for τ_{ne} (orange), τ_{UG} (pink), and $\tau_{UG/10}$ (cyan) for the evaluated watersheds. The black curve is from Erskine and Papaioannou (1997). Top values represent the base flow time constant.

slow reservoir anomalies to τ: τ_{ORC} (ORCHIDEE standard), τ_{ORCx3} (three times ORCHIDEE standard), τ_{1y}, τ_{10y}, and τ_{100y} (as constant high values), τ_{ne}, τ_{UG} (using two sets of hydraulic conductivity, since is the major uncertainty factor of τ), and $\tau_{UG/10}$ (as the physically-based τ with the "best" parameter setting obtained in Chapter 3 that is the closest to reference data). For river discharge and total water storage anomalies, nineteen major watersheds in the world were evaluated, while the slow reservoir volume was globally evaluated.

In a first step river discharge was evaluated, showing that this variable is highly sensible to τ for the majority of evaluated watersheds, and the buffer effect of the river discharge increased with τ. Low drainages result in little water entering the slow reservoir, that combined by the use of a high residence time (that results in bigger reservoir volumes), result in small impacts on the river discharge. It was observed a log-linear relationship between the amplitude difference of river discharges (with different τ) and drainage, indicating that for watersheds with average drainage under 0.05 mm.d^{-1}, river discharge is not sensitive to τ.

In a second step, the slow reservoir volume of ORCHIDEE was evaluated by the global spatially averaged anomalies for the five spatially variable τ (τ_{ORC}, τ_{ORCx3}, τ_{ne}, τ_{UG}, and $\tau_{UG/10}$). The amplitude anomalies increased with τ, showing significant variations in the reservoir (that increased with τ) that could not be observed by evaluating river discharge. The comparison of TWS anomalies between GRACE observations and simulation results for τ_{ORC}, τ_{ne}, τ_{UG}, and $\tau_{UG/10}$ showed a better agreement between simulated and observed data for τ_{ORC}, with R^2 over 0.95 for the majority of 19 evaluated stations. The use of a longer τ (τ_{ne}) slightly improved the coefficient of determination (from R^2 = 0.88 to R^2 =0.96) for the Danube station, but with no improvements of the Nash-Sutcliffe coefficient, remaining under 0.7.

With another approach, the simulations results using a physically-based τ were used to compare with the methodology proposed by Erskine and Papaioannou (1997), who found a relationship between the reduction factor (as a measure of the river discharge buffer effect), and the aquifer response rate (given by $1/\tau$). It was observed that the evaluated τ had a higher reduction factor for higher τ values, in agreement with Erskine and Papaioannou (1997). However, the plotted data was shifted in the direction of higher τ in comparison with the curve proposed by them, probably resultant of the transient behavior imposed by the atmospheric forcing used in the simulations.

The introduction of a longer τ in ORCHIDEE slow reservoir deteriorated river discharge simulations for most of the evaluated watersheds, compared to 276 GRDC stations all over the world. ORCHIDEE reference simulation showed strongly biased (positively and negatively), with only 23% of correlation coefficients over 0.75, but with 64% of Nash-Sutcliffe coefficient over 0.5. When comparing τ_{ne} and $\tau_{UG/10}$ with the reference simulation, it was observed an improvement of the Nash-Sutcliffe coefficient for only 10% of the stations, and this value reached 18% for $\tau_{UG/10}$, while τ_{ne} remained under those values. However, Nash-Sutcliffe and correlation coefficients improvement remained under 10% for both τ_{ne} and $\tau_{UG/10}$ compared to the reference simulation.

The proposed τ is not suitable for a shallow groundwater reservoir, due to the high values calculated (discussed in Chapter 3) that are not representative for the dominant local flow. However, more investigations should be made. It is recommended to test other model configurations that can ameliorate the water balance (as activating floodplains, and the soil freeze), that could result in better river discharges correcting the underestimated hydrographs for northern watersheds (as Ob, Lena, Yenisei, Mackenzie and Yukon), and overestimated for humid regions (as Parana and Tocantins), since the reference simulation showed short-comings when compared to river discharge observations. Another recommendation would be to test different atmospheric forcing files, that can directly affect river discharge results, as well for the buffer effect introduced by τ (that was not observed in regions with low drainage). The introduction of a second (deeper) compartment representing deep groundwater in ORCHIDEE should be also tested to evaluate if τ is suitable (or not) for models with groundwater represented by a linear reservoir. Such approach should have a deep (high values of τ, representing intermediate and regional flows) and shallow (with short τ, representing local flows) that implies in reviewing τ estimation from parameters that are more adequate to represent local flows. The use of a base flow time constant from parameters that are more adequate to represent rapid interactions of surface water and groundwater depends on having values closer to reference data for recession analysis, as discussed in Chapter 3.

CHAPTER 5

CONCLUSIONS AND PERSPECTIVES

In this thesis the base flow time constant (τ) was estimated at the global scale, by using a solution of the linearized Boussinesq equation for homogeneous, isotropic, unconfined sloping aquifers. This time constant depends on effective porosity, drainage density (δ), transmissivity, and aquifer slope. The lack of data at global scales to calibrate the parameters used in LSMs to simulate base flow and the overestimation of base flow time constant by the currently available global datasets motivated this thesis, to answer the main question *Can the estimation of the base flow time constant at global scale be improved to obtain a better base flow dynamics in large-scale LSMs?* To answer this question, drainage density, transmissivity, and porosity were evaluated to define which parameter exert a strong control of τ. τ was estimated using a formulation proposed by Zecharias and Brutsaert (1988), adapted in a way that the limit of this equation resulted in the long-term formulation for horizontal aquifers proposed by Brutsaert (2005). τ was improved by two orders of magnitude in average by changing the parameters used on its estimation (drainage density, effective porosity, and transmissivity) when compared to reference data for recession analysis and model calibration results from the literature. τ is significantly sensitive to drainage density (as it depends on δ^2) and transmissivity (that can vary from several orders of magnitude, depending on the aquifer formation), while other parameters can have local effects, but no major impacts on the global mean and median values were observed.

Differences between estimated τ and references are explained by the methodology and scale used. Comparing to literature values for recession analysis, all proposed τ were overestimated. When assessing a parameter that significantly impact τ results, τ is more sensitive to transmissivity than to δ in regions with low transmissivities (as the ones found in fine grained siliciclastic sedimentary, and crystalline rocks). This was explained by the much lower transmissivity in those regions compared to the δ range, even if τ depends on δ^2. Although global estimation of the saturated aquifer depth is uncertain, the main uncertainty of transmissivity comes from hydraulic conductivity data, as it varies over several orders of magnitude. It was observed that τ uncertainties are associated not only to the input data used on its estimation, but also to the assumptions used for τ formulation. To consider an homogeneous, isotropic aquifer neglects preferential flows and natural horizontal and vertical variations for the hydrogeological properties that can lead to underestimated transmissivities (by neglecting preferential flow) and drainage density (shortest flow paths showed to have a stronger impact on local flow than long ones). The overestimation could also be related to effective porosity, as the use of laboratory analyses results for effective porosity are approximately ten times overestimated (Brutsaert, 2008).

The evaluation of δ showed that traditional methods for river network extraction at large scales resulted in underestimated and spatially constant δ, compared to reference river networks, producing overestimated τ. This motivated us to extract a new river network, constrained by lithology, climate, slope, and observed drainage density. This river network depicts natural spatial patterns and results in higher δ, closer to reference values (Schneider *et al.* (2017), published in Geophysical Research Letters, with data freely available for download). The main uncertainties associated with this new river network are related to reference data. Unnatural channels and scale inconsistencies found in the data used for calibration and validation can induce uncertainties, and a new calibration should be performed when updated datasets become available. Environmental parameters used for calibration are also a source of uncertainties, specially related to the lithological map, in which the calibration should be updated with any improved lithology maps at global or regional scales.

Human pressures, such as artificial ditches for irrigation or navigation channels, are commonly observed in reference datasets, but they can not be estimated by only using hydrogeological properties. These man-made networks increase drainage density results, that affect drainage density calibration, and should not be accounted when calibrating river network with the proposed methodology.

This study also evaluated the interest of the estimated τ in the ORCHIDEE model. It is recommended to use a good initialization when using high τ values, as the use of an average half year τ needed approximately 13 years to achieve stationary state, reducing the spin-up time. A sensitivity analysis to the river discharge showed a high sensitivity of this variable to τ for the majority of evaluated watersheds, that had increased buffer effect of river discharge with high τ values. Compared to the 276 evaluated watersheds from GRDC observed river discharge data, the reference simulation showed biased for the majority of the stations (positively and negatively), with correlation coefficients over 0.75 for only 23% of the stations, and Nash-Sutcliffe coeffcient over 0.5 for 64% of the stations. The use of a longer τ deteriorated river discharge simulation results, with improvement of the Nash-Sutcliffe coefficient for only 10% and 18% of correlation coefficients for $\tau_{UG/10}$ simulation. The use of a different τ can not correct water balance and bias issues as observed with the reference simulations for most of evaluated watersheds. The watersheds that are not sensitive to τ variations have average drainage under 0.05 mm.d^{-1}. A log-linear relationship with drainage was observed, affecting the sensitivity of the buffer effect to the river discharge. For the groundwater reservoir volume, the amplitude anomalies increased with τ. In comparison with GRACE total water storage anomalies, the use of higher values of τ reduced the coefficient of determination, while the reference simulation presented a R^2 of 0.95 for the majority of evaluated watersheds. For only the Danube watershed R^2 increased from 0.88 (for τ_{ORC}) to 0.96 (for τ_{ne}), but with no improvements for the Nash-Sutcliffe coefficient (remaining under 0.7).

When comparing the simulations results to the relationship between reduction factor and aquifer response rates proposed by Erskine and Papaioannou (1997), it was observed that higher τ resulted in higher reduction factors. Although the results showed the same shape of the curve, the position was shifted in the direction of longer base flow time constants, possibly explained by the transient behavior imposed by the atmospheric forcing.

A perspective that emerged from this thesis is the use of different parameters to estimate τ to obtain a base flow time constant that is representative of the local flow, as the ones found in the literature for recession analysis and calibration results. Simulation results showed that shorter τ provides better simulated river discharge than longer ones, indicating that local flow is dominant for LSMs that have groundwater represented by a linear reservoir, as in ORCHIDEE. The use of different datasets that have lower effective porosity, higher transmissivity, and higher drainage density could result in τ that represent rapid interactions between groundwater and surface water. As such data are not currently available at global scale, the improvement of global datasets are a long-term perspective. The use of different assumptions (as non-linear and anisotropic reservoir) could also improve τ as a short-term perspective, by introducing more adequate results and formulations for the groundwater simulations.

BIBLIOGRAPHY

Barlow, P.M., Leake, S.A. Streamflow depletion by wells-Understanding and managing the effects of groundwater pumping on streamflow. *U.S. Geological Survey Circular 1376, 84p.*, 2012. `https://pubs.usgs.gov/circ/1376/`.

Benstead, Jonathan P, Leigh, David S. An expanded role for river networks. *Nature Geoscience*, 2012, 5, 10, p. 678–679.

Beven, K. J., Kirby, M. J. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 1979, 24, 1, p. 43–69, doi: 10.1080/02626667909491834.

BGR, UNESCO. Groundwater Resources of the World 1 : 25 000 000. *Hannover, Paris*, 2008.

Billen, Gilles, Thieu, Vincent, Garnier, Josette, *et al.* Modelling the N cascade in regional watersheds: The case study of the Seine, Somme and Scheldt rivers. *Agriculture, ecosystems & environment*, 2009, 133, 3, p. 234–246.

Bishop, K., Buffam, I., Erlandsson, M., *et al.* Aqua Incognita: the unknown headwaters. *Hydrological Processes*, 2008, 22, 8, p. 1239–1242, ISSN 1099-1085, doi:10.1002/hyp.7049. `http://dx.doi.org/10.1002/hyp.7049`.

Borman, R.G., Meredith, Thomas S. Geology, altitude, and depth of the bedrock surface beneath the Ogallala formation in the northern high plains of Colorado. *United States Geological Survey*, 1983.

Botter, Gianluca, Basso, Stefano, Rodriguez-Iturbe, Ignacio, *et al.* Resilience of river flow regimes. *PNAS*, 2013, doi:www.pnas.org/cgi/doi/10.1073/pnas.1311920110.

Boussinesq, Joseph. *Essai sur la théorie des eaux courantes.* Imprimerie nationale, 1877. t. 2.

Brandes, David, Hoffmann, Justin G., Mangarillo, James T. Base flow recession rate, lowflows, and hydrologic features of small watersheds in Pensylvania, USA. *Journal of the American Water Resources Association*, 2005, , 04058, p. 1177–1186.

BRUTSAERT, W. *Hydrology: An Introduction.* Cambridge University Press, 2005, 605 p.

BRUTSAERT, W. Long-term groundwater storage trends estimated from streamflow records: Climatic perspective. *Water Resources Research*, 2008, 44, W02409.

BRUTSAERT, W., NIEBER, J.L. Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, 1977, 13, p. 637–643.

BRUTSAERT, Wilfried, LOPEZ, James P. Basin-scale geohydrologic drought flow features of riparian aquifers in the southern Great Plains. *Water Resources Research*, 1998, 34, 2, p. 233–240.

BURROUGH, P. A., MCDONELL, R.A. *Principles of Geographical Information Systems.* Oxford University Press, 1998, 190 p.

CAMPOY, A., DUCHARNE, A., CHERUY, F., *et al.* Response of land surface fluxes and precipitation to different soil bottom hydrological conditions in a general circulation model. *Journal of Geophysical Research: Atmospheres*, 2013, 118, 19, p. 10 725 – 10 739, ISSN 2169-8996.

CARSEL, R. F., PARRISH, R. S. Developing joint probability distributions of soil water retention characteristics. *Water Resources Research*, 1988, 24, 5, p. 755–769, ISSN 1944-7973.

CARTER, Janet M., DRISCOLL, Daniel G., WILLIAMSON, Joyce E., *et al. Atlas of water resources in the Black Hills area, South Dakota.* United States Geological Survey, 2015. ISBN 0607975105. https://pubs.usgs.gov/ha/ha747/pdf/definition.pdf.

CHAPMAN, Tom. A comparison of algorithms for stream flow recession and base flow separation. *Hydrological Processes*, 1999.

COLOMBO, Roberto, VOGT, Jürgen V, SOILLE, Pierre, *et al.* Deriving river networks and catchments at the European scale from medium resolution digital elevation data. *Catena*, 2007, 70, 3, p. 296–305.

DE GRAAF, I. E. M., SUTANUDJAJA, E. H., VAN BEEK, L. P. H., *et al.* A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences*, 2015, 19, 2, p. 823–837, doi:10.5194/hess-19-823-2015.

DE MARSILY, G. *L'eau.* Flammarion, 1995, 128 p. ISBN 2-08-035162-1.

DE ROSNAY, P., POLCHER, J., BRUEN, M., *et al.* Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes. *Journal of Geophysical Research: Atmospheres*, 2002, 107, D11, p. ACL3:1–ACL3:19.

DE ROSNAY, Patricia, POLCHER, Jan. Modelling root water uptake in a complex land surface scheme coupled to a GCM. *Hydrology and Earth System Sciences*, 1998, 2, p. 239–255.

DINGMAN, S Lawrence. *Physical hydrology.* Waveland Press, 2015, 643 p. t. 575.

DIRMEYER, P., GAO, X., ZHAO, M., *et al. The Second Global Soil Wetness Project (GSWP-2): Multi-Model Analysis and Implications for our Perception of the Land Surface.* Rap. tech., may 2005.

DÖLL, P., HOFFMANN-DOBREV, H., PORTMANN, F.T., *et al.* Impact of water withdrawals from groundwater and surface water on continental water storage variations. *Journal of Geodynamics*, 2012, 59-60, p. 143 – 156, ISSN 0264-3707, doi:http://dx.doi.org/10.1016/j.jog.2011.05.001.

DÖLL, P., LEHNER, B. Validation of a new global 30-min drainage direction map. *Journal of Hydrology*, 2002, 258, p. 214–231.

DOMENICO, P.A., SCHWARTZ, F.W. *Physical and chemical HYdrogeology.* John Wiley & Sons, 1990.

D'ORGEVAL, T., POLCHER, J., DE ROSNAY, P. Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes. *Hydrology and Earth System Sciences*, 2008, 12, p. 1387–1401.

D'ORGEVAL, Tristan. *Impact du changement climatique sur le cycle de l'eau en Afrique de l'Ouest : Modélisation et incertitudes.* Thèse de doctorat, UniversitṔierre et Marie Curie, 2006.

DUCHARNE, A. *The hydrol module of ORCHIDEE: scientific documentation [rev 3977] and on, work in progress !!* Rap. tech., Université Pierre et Marie Curie, January 2017. `forge.ipsl.jussieu.fr/orchidee/raw-attachment/wiki/Documentation/UserGuide/eqs_hydrol.pdf`.

DUCHARNE, A., LAVAL, K., POLCHER, J. Sensitivity of the hydrological cycle to the parametrization of soil hydrology in a GCM. *Climate dynamics*, 1998, 14, p. 307–327.

DUCHARNE, Agnes, GOLAZ, Catherine, LEBLOIS, Etienne, *et al.* Development of a high resolution runoff routing model, calibration and application to assess runoff from the LMD GCM. *Journal of hydrology*, 2003, 280, 1, p. 207–228.

DUCHARNE, Agnès, LAVAL, Katia. *L'eau à découvert (Le cycle de l'eau).* CNRS Editions, 2015, 26-27 p. ISBN 978-2-271-08829-1.

DUCHARNE, AgnÃ¨s, KOSTER, Randal D., SUAREZ, Max J., *et al.* A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration. *Journal of Geophysical Research: Atmospheres*, 2000, 105, D20, p. 24 823–24 838, ISSN 2156-2202, doi:10.1029/2000JD900328.

DUCOUDRÉ, Nathale I, LAVAL, Katia, PERRIER, Alain. SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model. *Journal of Climate*, 1993, 6, 2, p. 248–273.

ENG, K., MILLY, P. Relating low-flow characteristics to the base flow recession time constant at partial record stream gauges. *Water Resources Research*, 2007, 43, p. W01 201, doi:10.1029/2006WR005293.

ERSKINE, AD, PAPAIOANNOU, A. The use of aquifer response rate in the assessment of groundwater resources. *Journal of hydrology*, 1997, 202, 1, p. 373–391.

FAN, Y., LI, H., MIGUEZ-MACHO, G. Global Patterns of Groundwater Table Depth. *Science*, 2013, 339, 6122, p. 940–943.

FAN, Ying, MIGUEZ-MACHO, Gonzalo, WEAVER, Christopher P., *et al.* Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations. *Journal of Geophysical Research: Atmospheres*, 2007, 112, D10, ISSN 2156-2202, doi:10.1029/2006JD008111. `http://dx.doi.org/10.1029/2006JD008111`.

FAO. AQUAMAPS Global spatial database on water and agriculture. *Food and Organization of the United Nations*, 2014. `http://www.fao.org/nr/water/aquamaps/`.

FEKETE, B. M., CHARLES, V., GRABS, W. *Global, composite runoff fields based on observed river discharge and simulated water balances.* Rap. tech., Global Runoff Data Centre (GRDC), 2000.

FEKETE, Balàzs M., VÖRÖSMARTY, Charles J., LAMMERS, Richard B. Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error. *Water Resources Research*, 2001, 37, 7, p. 1955–1967, ISSN 1944-7973, doi:10.1029/2001WR900024. `http://dx.doi.org/10.1029/2001WR900024`.

GASCOIN, Simon, DUCHARNE, Agnès, RIBSTEIN, Pierre, *et al.* Adaptation of a catchment-based land surface model to the hydrogeological setting of the Somme River basin (France). *Journal of Hydrology*, 2009, 368, 1–4, p. 105–116, ISSN 0022-1694, doi:10.1016/j.jhydrol.2009.01.039. `http://www.sciencedirect.com/science/article/pii/S0022169409000626`.

GETIRANA, A.C.V., DUTRA, E., GUIMBERTEAU, M., *et al.* Water Balance in the Amazon Basin from a Land Surface Model Ensemble. *Journal of Hydrometeorology*, 2014, 15, 6, p. 2586–2614.

GLEESON, T., SMITH, L., JANSEN, N., *et al.* Mapping permeability over the surface of the earth. *Geophysical Research Letters*, 2011a, 38, L02401.

GLEESON, Tom, BEFUS, Kevin M, JASECHKO, Scott, *et al.* The global volume and distribution of modern groundwater. *Nature Geoscience*, 2015, 9, 2, p. 161–167.

GLEESON, Tom, MARKLUND, Lars, SMITH, Leslie, *et al.* Classifying the water table at regional to continental scales. *Geophys. Res. Lett.*, 2011b, 38, L05401, p. 1–6.

Gleeson, Tom, Moosdorf, Nils, Hartmann, Jens, *et al.* A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, 2014, 41, 11, p. 3891–3898, ISSN 1944-8007.

Grafton, R. Quentin, Pittock, Jamie, Davis, Richard, *et al.* Global insights into water resources, climate change and governance. *Nature Clim. Change*, 2012, 3, 4, p. 315–321, doi:http://dx.doi.org/10.1038/nclimate1746.

Graham, S. T., Famiglietti, J. S., Maidment, D. R. Five-minute, 1/2°, and 1° data sets of continental watersheds and river networks for use in regional and global hydrologic and climate system modeling studies. *Water Resources Research*, 1999, 35, 2, p. 583–587, ISSN 1944-7973, doi:10.1029/1998WR900068. http://dx.doi.org/10.1029/1998WR900068.

GRDC. Global Runoff Data Centre: Monthly river discharge data from 1807 to 2014. *Federal Institute of Hydrology (BfG)*, 2014.

Gruber, S. Derivation and analysis of a high-resolution estimate of global permafrost zonation. *The Cryosphere*, 2012, 6, 1, p. 221–233, doi:10.5194/tc-6-221-2012. http://www.the-cryosphere.net/6/221/2012/.

Guimberteau, M. ORCHIDEE training course: The routing scheme in ORCHIDEE. *Orsay*, July 2014.

Guimberteau, M, Drapeau, G, Ronchail, J, *et al.* Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets. *Hydrology and Earth System Sciences*, 2012a, 16, p. 911–935.

Guimberteau, M., Laval, K., Perrier, A., *et al.* Global effect of irrigation and its impact on the onset of the Indian summer monsoon. *Climate dynamics*, 2012b, 39, p. 1329–1348.

Hagemann, S., Dümenil, L. A parametrization of the lateral waterflow on the global scale. *Climate dynamics*, 1998, 14, p. 17–31.

Haitjema, Henk M., Mitchell-Bruker, Sherry. Are Water Tables a Subdued Replica of the Topography? *Ground Water*, 2005, 43, 6, p. 781–786, ISSN 1745-6584, doi:10.1111/j.1745-6584.2005.00090.x. http://dx.doi.org/10.1111/j.1745-6584.2005.00090.x.

Harding, R., Warnaars, T., Weedon, G., *et al.* *Technical Report N. 56: Executive Summary of the Completed WATCH Project.* Rap. tech., NERC National Centre for Earth Observation, september 2011.

Hartmann, Jens, Moosdorf, Nils. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, 2012, 13, 12, ISSN 1525-2027, doi:10.1029/2012GC004370. http://dx.doi.org/10.1029/2012GC004370.

HEINE, Reuben A, LANT, Christopher L, SENGUPTA, Raja R. Development and comparison of approaches for automated mapping of stream channel networks. *Annals of the Association of American Geographers*, 2004, 94, 3, p. 477–490.

HORTON, R.E. Drainage basin characteristics. *Transactions, American Geophysical Union*, 1932, p. 12.

HORTON, R.E. Erosional development of streams and their drainage basins : hydrophysical approach to quantitative morphology. *Geological Society of America*, 1945, 56, p. 275–370.

HUTCHINSON, MF. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *journal of Hydrology*, 1989, 106, 3-4, p. 211–232.

IBGE. *Base continua - Documentaçao Tecnica Geral.* ftp://geoftp.ibge.gov.br/: Insituto Brasileiro de Geografia e Estatistica, 2015, 32 p.

IGN. Base de Données sur la CARtographie THématique des AGences de l'eau et du ministère chargé de l'environnement. *Institut National de L'Information Géographique et Forestière*, 2014. http://professionnels.ign.fr/bdcarthage.

JIMENEZ-MARTINEZ, J., LONGUEVERGNE, L., LE BORGNE, T., *et al.* Temporal and spatial scaling of hydraulic response to recharge in fractured aquifers: insights from a frequency domain analysis. *Water Resources Research*, 2013, 49, p. 3007–3023.

JOHNSON, A.I. Specific yield: compilation of specific yields for various materials. *USGS Water Supply Paper*, 1967, , 1662, p. 74.

JONES, P.D., HARRIS, I. CRU TS3.21: Climatic Research Unit (CRU) Time-Series (TS) Version 3.21 of High Resolution Gridded Data of Month-by-month Variation in Climate (Jan. 1901- Dec. 2012). *NCAS British Atmospheric Data Centre*, september 2013. http://dx.doi.org/10.5285/D0E1585D-3417-485F-87AE-4FCECF10A992.

KOSTER, Randal D., SUAREZ, Max J., DUCHARNE, AgnÃ¨s, *et al.* A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres*, 2000, 105, D20, p. 24 809–24 822, ISSN 2156-2202, doi:10.1029/2000JD900327.

KRINNER, G., VIOVY, Nicolas, DE NOBLET-DUCOUDRÉ, Nathalie, *et al.* A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 2005, 19, 1, ISSN 1944-9224.

LANDERER, F. W., SWENSON, S. C. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 2012, 48, 4, ISSN 1944-7973, doi: 10.1029/2011WR011453. http://dx.doi.org/10.1029/2011WR011453.

LEDOUX, E. *Modélisation intégré des écoulements de surface et des écoulements souterrains sur un bassin hydrologique.* Thèse de doctorat, Ecole Nationale Supérieure des Mines de Paris, 1980.

LEDOUX, E., GOMEZ, E., MONGET, J.M., *et al.* Agriculture and groundwater nitrate contamination in the Seine basin. The STICS-MODCOU modelling chain. *Science of The Total Environment*, 2007, 375, 1, p. 33 – 47, ISSN 0048-9697, doi:http://dx.doi.org/10.1016/j.scitotenv.2006.12.002. Human activity and material fluxes in a regional river basin: the Seine River watershed, `http://www.sciencedirect.com/science/article/pii/S0048969706009284`.

LEHNER, B., DÖLL, P. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of hydrology*, 2004, 296, 1-4, p. 1–22.

LEHNER, B., GRILL, G. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 2013, 27, 15, p. 2171–2186.

LEHNER, B., VERDIN, K., JARVIS, A. New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 2008, 89, 10, p. 93–94.

LO, Min-Hui, FAMIGLIETTI, James S., YEH, P. J.-F., *et al.* Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. *Water Resources Research*, 2010, 46, 5, ISSN 1944-7973, doi:10.1029/2009WR007855. `http://dx.doi.org/10.1029/2009WR007855`.

LUO, Wei, JASIEWICZ, Jaroslaw, STEPINSKI, Tomasz, *et al.* Spatial association between dissection density and environmental factors over the entire conterminous United States. *Geophysical Research Letters*, 2016, 43, 2, p. 692–700, ISSN 1944-8007, doi:10.1002/2015GL066941. 2015GL066941, `http://dx.doi.org/10.1002/2015GL066941`.

MAXWELL, Reed M., MILLER, Norman E. Development of a coupled land surface and groundwater model. *Journal of Hydrometeorology*, 2005.

MAXWELL, RM, CONDON, LE, KOLLET, SJ. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geoscientific Model Development*, 2015, 8, 3.

MIGUEZ-MACHO, Gonzalo, FAN, Ying, WEAVER, Christopher P., *et al.* Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation. *Journal of Geophysical Research: Atmospheres*, 2007, 112, D13, p. 2156–2202.

MILLER, J. R., RUSSEL, G. L., CALIRI, G. Continental-scale river flow in climate models . *Journal of Climate*, 1994, 7, p. 914–928.

MONTGOMERY, D.R., FOUFOULA-GEORGIOU, E. Channel network source representation using digital elevation models. *Water Resources Research*, 1993, 29, 12, p. 3925–3934.

MOORE, I. D., BURCH, G. J. Physical basis of the length-slope factor in the Universal Soil Loss Equation. *Soil Science Society of America Journal*, 1986, 50, 5, p. 1294–1298.

MORRIS, D., JOHNSON, A. Summary of hydrological and physical properties of rrock and soil materials, as analyzed by the U.S. Geological Survey 1948-1960. *Water -Supply paper, U.S. Geological Survey*, 1967.

NGO-DUC, T, LAVAL, K, RAMILLIEN, G, *et al.* Validation of the land water storage simulated by Organising Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) with Gravity Recovery and Climate Experiment (GRACE) data. *Water resources research*, 2007, 43, 4.

NIU, Guo-Yue, YANG, Zong-Liang, DICKINSON, Robert E., *et al.* Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research: Atmospheres*, 2007, 112, D7, p. 2156–2202, doi:10.1029/2006JD007522.

O'CALLAGHAN, J.F., MARK, D.M. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, 1984, , 28, p. 323–344.

OKI, T., KANAE, S. Global Hydrological Cycles and World Water Resources. *Science*, 2006, 313, 1068, doi:10.1126/science.1128845.

OSDM. *Geofabric Surface Cartography - V.3.0.* http://www.bom.gov.au/water/geofabric/download.shtml: Office of Spatial Data Management, Commonwealth of Australia (Bureau of Meteorology), 2015.

PELLETIER, J.D. A robust, two-parameter method for the extraction of drainage networks from high-resolution digital elevation models (DEMs): Evaluation using synthetic and real-world DEMs. *Water Resources Research*, 2013, 49, 1, p. 75–89.

PELLETIER, Jon D., BROXTON, Patrick D., HAZENBERG, Pieter, *et al.* A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems*, 2016, 8, 1, p. 41–65, ISSN 1942-2466, doi:10.1002/2015MS000526.

PERSENDT, FC, GOMEZ, C. Assessment of drainage network extractions in a low-relief area of the Cuvelai Basin (Namibia) from multiple sources: LiDAR, topographic maps, and digital aerial orthophotographs. *Geomorphology*, 2016, 260, p. 32–50.

PITMAN, A. The evolution of, and revolution in, land surface schemes dedesign for climate models. *International Journal of Climatology*, 2003.

POLCHER, J. Les processus de surface à l'échelle globale a e et leurs interactions avec l'atmosphère. *Habilitation à diriger des recherches*, 2003.

RAYMOND, Peter A, HARTMANN, Jens, LAUERWALD, Ronny, *et al.* Global carbon dioxide emissions from inland waters. *Nature*, 2013, 503, 7476, p. 355–359.

REYNOLDS, C., JACKSON, T., RAWLS, W. Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resources Research*, 2000, 36, p. 3653–3662.

RIES, Kernell G., ENG, Ken. *Estimation of selected streamflow statistics for a network of low-flow partial-record stations in aarea affected by base realignment and closure (BRAC) in Maryland.* Rap. tech., USGS, 2010.

RODELL, M., BEAUDOING, H. K., L'ECUYER, T. S., *et al.* The Observed State of the Water Cycle in the Early Twenty-First Century. *Journal of Climate*, 2015, 28, 21, p. 8289–8318, doi:10.1175/JCLI-D-14-00555.1. `http://dx.doi.org/10.1175/JCLI-D-14-00555.1`, `http://dx.doi.org/10.1175/JCLI-D-14-00555.1`.

ROUSSEAU-GUEUTIN, P., LOVE, A. J., VASSEUR, G., *et al.* Time to reach near-steady state in large aquifers. *Water Resources Research*, 2013, 49, 10, p. 6893–6908, ISSN 1944-7973, doi:10.1002/wrcr.20534. `http://dx.doi.org/10.1002/wrcr.20534`.

RUPP, David E., SELKER, John S. analysis of the Boussinesq equation for interpreting recession hydrographs from sloping aquifers. *Water Resources Research*, 2006.

SAKUMURA, C., BETTADPUR, S., BRUINSMA, S. Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.*, 2014.

SANGIREDDY, Harish, CAROTHERS, Richard A, STARK, Colin P, *et al.* Controls of climate, topography, vegetation, and lithology on drainage density extracted from high resolution topography data. *Journal of Hydrology*, 2016, 537, p. 271–282.

SCHNEIDER, A., JOST, A., COULON, C., *et al.* Global scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density. *Geophysical Research Letters*, 2017, 44, ISSN 1944-8007, doi:10.1002/2016GL071844. 2016GL071844, `http://dx.doi.org/10.1002/2016GL071844`.

SCHNEIDER, U., BECKER, A., FINGER, P., *et al.* GPCC Full Data Reanalysis Version 6.0 at 0.5Â°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data, 2011. `ftp://ftp.dwd.de/pub/data/gpcc/html/fulldata_v6_doi_download.html`.

SCHNEIDER, Udo, BECKER, Andreas, FINGER, Peter, *et al.* GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, 2013, 115, 1-2, p. 15–40, ISSN 0177-798X, doi:10.1007/s00704-013-0860-x.

SCHRUBEN, Paul G Arndt, BAWIEC, Raymond E, KING, Walter J, *et al. Geology of the Conterminous United States at 1: 2,500,000 Scale–A Digital Representation of the 1974 PB King and HM Beikman Map.* Peter N. Schweitzer, 1994.

SHANGGUAN, Wei, HENGL, Tomislav, MENDES DE JESUS, Jorge, *et al.* Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, 2017, ISSN 1942-2466, doi:10.1002/2016MS000686.

SHIKLOMANOV, I. A., RODDA, J. C. *World water resources at the begining of the twenty-first century.* Cambridge University Press, 2003. ISBN 0 521 82085 5.

SITCH, S, SMITH, B, PRENTICE, I Colin, *et al.* Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 2003, 9, 2, p. 161–185.

STEIN, JL, HUTCHINSON, MF, STEIN, JA. A new stream and nested catchment framework for Australia. *Hydrology and Earth System Sciences*, 2014, 18, 5, p. 1917–1933.

STOELZLE, M, STAHL, K, WEILER, M. Are streamflow recession characteristics really characteristic? *Hydrology and Earth System Sciences*, 2013, 17, 2, p. 817–828.

STRAHLER, A. N. Quantitative analysis of watershed geomorphology. *Transactions, American Geophysical Union*, 1957, 38, 6, p. 913–920.

SWENSON, S. GRACE monthly land water mass grids NETCDF RELEASE 5.0. Ver. 5.0. *PO.DAAC, CA, USA.*, 2012, doi:http://dx.doi.org/10.5067/TELND-NC005.

TARBOTON, David G., BRAS, Rafael L., RODRIGUEZ-ITURBE, Ignacio. On the extraction of channel networks from digital elevation data. *Hydrological Processes*, 1991, 5, 1, p. 81–100, ISSN 1099-1085, doi:10.1002/hyp.3360050107. `http://dx.doi.org/10.1002/hyp.3360050107`.

TARBOTON, D.G., BRAS, R.L., RODRIGUEZ-ITURBE, I. A physical basis for drainage density. *Geomorphology*, 1992, 5, p. 59–76.

THOMAS, Brian F., VOGEL, Richard M., KROLL, Charles N., *et al.* Estimation of the base flow recession constant under human interference. *Water Resources Research*, 2013, 49, p. 7366–7379.

TRENBERTH, K. E., SMITH, L., QIAN, T., *et al.* Estimates of the global water budget and its annual cycle using observational and model data. *Journal of Hydrometeorology*, 2007, 8, 4, p. 758–769.

TUCKER, G.E., CATANI, F., RINALDO, A., *et al.* Statistical analysis of drainage density from digital terrain data. *Geomorphology*, 2001, 36, p. 187–202.

UNSD. *Glossary of environment statistics.* Department for Economic and Social Information and Policy Analysis, 2001.

USGS. *National Hydrography Dataset at 1:24000, for all available states.* http://nhd.usgs.gov/data.html, accessed in December 2015: United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), the United States Geological Survey (USGS), and the Environmental Protection Agency (EPA), 2015.

VERDIN, KL, GREENLEE, SK. HYDRO1k documentation. *Sioux Falls, ND, US Geological Survey, EROS Data Center, http://edcdaac. usgs. gov/gtopo30/hydro/readme. html*, 1998.

VERGNES, J.-P., DECHARME, B. A simple groundwater scheme in the TRIP river routing model: global off-line evaluation against GRACE terrestrial water storage estimates and observed river discharges. *Hydrology and Earth System Sciences*, 2012, 16, 10, p. 3889–3908, doi:10.5194/hess-16-3889-2012. `http://www.hydrol-earth-syst-sci.net/16/3889/2012/`.

VIOVY, N. Interannuality and CO2 sensitivity of the SECHIBA-BGC coupled SVAT-BGC model. *Physics and Chemistry of the Earth*, 1996, 21, 5–6, p. 489–497, ISSN 0079-1946, doi:10.1016/S0079-1946(97)81147-0. Ocean and Atmosphere, `http://www.sciencedirect.com/science/article/pii/S0079194697811470`.

VOGEL, R. M., KROLL, C. Estimation of base flow recession constants. *Water Resources Management*, 1995.

VOGT, J.V., COLOMBO, R., BERTOLO, F. Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation data and environmental characteristics. *Geomorphology*, 2003, 53, 3, p. 281–298.

VOGT, JV, SOILLE, P, JAGER, AD, *et al.* A pan-European river and catchment database. European Commission, Joint Research Center. *Institute for Environment and Sustainability, JRC, Luxembourg*, 2007, , EUR 22920 EN, p. 120.

VÖRÖSMARTY, C., FEKETE, B., MEYBECK, M., *et al.* Global system of rivers : Its role in organizing continental land mass and defining land-to- ocean linkages. *Global Biogeochemical Cycles*, 2000, 14, 2, p. 599–621.

WANG, Tao, OTTLÃ©, Catherine, BOONE, Aaron, *et al.* Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model. *Journal of Geophysical Research: Atmospheres*, 2013, 118, 12, p. 6064–6079, ISSN 2169-8996, doi:10.1002/jgrd.50395. `http://dx.doi.org/10.1002/jgrd.50395`.

WEEDON, Graham P., BALSAMO, Gianpaolo, BELLOUIN, Nicolas, *et al.* The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 2014, 50, 9, p. 7505–7514, ISSN 1944-7973, doi:10.1002/2014WR015638.

WOOD, Eric F, ROUNDY, Joshua K, TROY, Tara J, *et al.* Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 2011, 47, 5.

WU, H., KIMBALL, J. S., LI, H., *et al.* A New Global River Network Database for Macroscale Hydrologic modeling. *Water Resources Research*, 2012, 48, W0970, ISSN 1944-7973, doi:10.1029/2009WR008871. http://dx.doi.org/10.1029/2009WR008871.

YEH, P. J.-F., ELTAHIR, E. A. B. Representation of Water Table Dynamics in a Land Surface Scheme. Part I: Model development. *Journal of Climate*, 2005, 18, 12, p. 1881–1901, doi:10.1175/JCLI3331.1.

YORK, J. P., PERSON, M., GUTOWSKI, W.J., *et al.* Putting aquifers into atmospheric simulation models: an example from the Mill Creek watershed, northeasten Kansas. *Advances in Water Resources*, 2002, 25, 2, p. 221–238.

YOUNGER, Paul L. *Groundwater in the environment: an introduction.* John Wiley & Sons, 2009.

ZECHARIAS, Yemane B, BRUTSAERT, Wilfried. Recession characteristics of groundwater outflow and base flow from mountainous watersheds. *Water Resources Research*, 1988, 24, 10, p. 1651–1658.

CHAPTER 6

APPENDIX

6.1 Article published in Geophysical Research Letters

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2016GL071844

Key Points:

- We propose simple models for global river network extraction coherent with a 1:50,000 scale and first-order intermittency assessment
- The resulting drainage density is spatially variable and offers an acceptable fit with observation-based patterns
- On average, globally, the predicted drainage density is approximately 0.70 km^{-1}, with approximately 30% intermittent streams

Supporting Information:

- Supporting Information S1

Correspondence to:
A. Schneider,
ana.schneider@upmc.fr

Citation:
Schneider, A., A. Jost, C. Coulon, M. Silvestre, S. Théry, and A. Ducharne (2017), Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density, *Geophys. Res. Lett.*, *44*, doi:10.1002/2016GL071844.

Received 7 NOV 2016
Accepted 19 FEB 2017
Accepted article online 21 FEB 2017

©2017. American Geophysical Union.
All Rights Reserved.

Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density

A. Schneider[1], **A. Jost**[1], **C. Coulon**[1], **M. Silvestre**[2], **S. Théry**[2], **and A. Ducharne**[1]

[1]Sorbonne Universités, UPMC Univ Paris 06, CNRS, EPHE, UMR 7619 METIS, Paris, France, [2]Sorbonne Universités, UPMC Univ Paris 06, CNRS, FR 3020 FIRE (Fédération Île-de-France de Recherche sur l'Environnement), Paris, France

Abstract To improve the representation of surface and groundwater flows, global land surface models rely heavily on high-resolution digital elevation models (DEMs). River pixels are routinely defined as pixels with drainage areas that are greater than a critical drainage area (A_{cr}). This parameter is usually uniform across the globe, and the dependence of drainage density on many environmental factors is often overlooked. Using the 15″ HydroSHEDS DEM as an example, we propose the calibration of a spatially variable A_{cr} as a function of slope, lithology, and climate, to match drainage densities from reference river networks at a 1:50,000 scale in France and Australia. Two variable A_{cr} models with varying complexities were derived from the calibration, with satisfactory performances compared to the reference river networks. Intermittency assessment is also proposed. With these simple tools, river networks with natural heterogeneities at the 1:50,000 scale can be extracted from any DEM.

1. Introduction

Precise characterization of river geometry is crucial for many applications related to river hydraulics and has gained a lot from the advances in digital elevation models (DEMs) processing since *Hutchinson* [1989] first proposed the "stream burning" technique to correct the location of extracted streams. At large scales, due to the advent of very high resolution global DEMs such as HYDRO1k at 1 km [*Verdin and Greenlee*, 1998] and HydroSHEDS at 30″ and 15″ [*Lehner et al.*, 2008], priority has been given to producing correct flow direction maps, including coarser resolutions used in global runoff routing models [*Graham et al.*, 1999; *Fekete et al.*, 2001; *Döll and Lehner*, 2002; *Wu et al.*, 2012].

Proper characterization of stream length from DEMs has received less attention despite its broad influence on water sciences (e.g., on erosion and sediment transport [*Moore and Burch*, 1986]; riverine water quality, which is strongly controlled by the upstream residence time [*Billen et al.*, 2009]; and the buffering effect of groundwater on extreme flows, which partly depends on the distance from the recharge zones to the rivers [*Brutsaert and Lopez*, 1998]). A major issue is identifying the smallest streams, called the "aqua incognita" by *Bishop et al.* [2008]. This is true in the field because of stream intermittency, artificial ditches, or hydraulic obstructions in flat areas. Additional problems arise when implementing traditional methods to map the "blue lines," usually from airborne or satellite imagery, because of insufficient resolution, vegetation masking, or scarce field data [*Tarboton et al.*, 1991; *Lehner and Döll*, 2004; *Benstead and Leigh*, 2012; *Persendt and Gomez*, 2016]. It is believed that river networks extracted from very high resolution DEMs are more comprehensive, assuming that they are properly calibrated [*Benstead and Leigh*, 2012; *Stein et al.*, 2014].

Many methods exist to extract river networks from DEMs, mostly from a geomorphological reference [*Tarboton et al.*, 1992; *Montgomery and Foufoula-Georgiou*, 1993; *Heine et al.*, 2004; *Pelletier*, 2013]. However, the most widely used method at small and regional scales, by far, simply relies on flow direction and a critical drainage area (A_{cr}) that represents the minimum upstream drainage area required to initiate a river [*O'Callaghan and Mark*, 1984]. In this classical framework, extracted rivers are shorter if the A_{cr} is larger, which results in a smaller drainage density (δ), defined as the total river length inside a watershed divided by its area [*Horton*, 1932].

Drainage density, δ, gives a macroscale measure of stream length and is often used to quantitatively evaluate watershed properties [*Strahler*, 1957; *Vogt et al.*, 2007; *Dingman*, 2015]. In natural river networks, δ is spatially variable [*Tucker et al.*, 2001; *Vogt et al.*, 2007], since river initiation depends on climate, slope, lithology, soil properties, and vegetation cover [*Tarboton et al.,* 1992; *Montgomery and Foufoula-Georgiou*, 1993; *Vogt et al.,* 2003; *Colombo et al.,* 2007; *Sangireddy et al.,* 2016]. However, given the lack of sufficient information to constrain A_{cr}, it is common to use a single value for river network extraction in a given domain, resulting in a rather uniform δ given the long-established link between A_{cr} and δ [*Tarboton et al.,* 1992].

In global hydrographic data sets such as HYDRO1k, HydroSHEDS, and AQUAMAPS (a river network product derived from 15″ HydroSHEDS) [*Food and Agriculture Organization* (*FAO*), 2014], the A_{cr} used to extract river networks are uniform across continents, and their values are arbitrarily chosen to limit the number of small streams for large-scale applications: 1000 km^2 in HYDRO1k and the global version of AQUAMAPS, 100 km^2 for the continental version of AQUAMAPS, and approximately 25 km^2 (at the equator) in HydroSHEDS (100 15″ pixels). These discrepancies highlight the fact that each of the corresponding networks is a "hypothetical" river network, constrained in its extent by the chosen A_{cr}. The above values all result in δ under 0.2 km^{-1}, which is much lower than the values obtained from detailed observations in small watersheds (40 up to 1000 km^2), which show drainage densities between 0.5 and 1.6 km^{-1} [*Horton*, 1945; *Brutsaert and Nieber*, 1977; *Zecharias and Brutsaert*, 1988; *Brutsaert and Lopez*, 1998]. This led *Raymond et al.* [2013] to define their own global river network based on the 15″ HydroSHEDS and a smaller A_{cr} of 10 km^2 for estimating carbon sinks and sources for inland waters (resulting in global mean δ of 0.28 km^{-1}).

Similarly, several studies have analyzed the spatial dependence of δ on environmental parameters at the continental scale [*Colombo et al.,* 2007; *Vogt et al.,* 2007; *Luo et al.,* 2016]. *Vogt et al.* [2007] evaluated drainage density over Europe (CCM2 data set, second version of *Colombo et al.* [2007] work), with a mean value of 0.31 km^{-1} based on a 100 m resolution DEM and 10 different A_{cr} values, ranging from 0.72 to 12.80 km^2, defined from the association of landscape classes to subjective scores of valley dissection potential. *Luo et al.* [2016] used a geomorphological detection method and a 30 m resolution DEM to extract the land dissection density (considered equivalent to drainage density) over the United States, with values ranging from 0 to 5 km^{-1}, and a clear spatial dependence on climate, lithology and several terrain-based attributes.

Based on this analysis, our goal is to propose spatially variable A_{cr} values for a simple extraction of river networks from large-scale DEMs, as well as a first-order assessment of intermittent streams, using the 15″ HydroSHEDS data as an example. The heterogeneities of A_{cr} and the resulting drainage density, δ, are linked to environmental parameters (i.e., slope, lithology, and climate) to match drainage densities from reference river networks at the 1:50,000 scale in France and Australia. An independent evaluation conducted against reference river networks from the United States and Brazil is discussed before generalizing the method to extract 15″ river networks consistent with 1:50,000 blue lines across the continents.

2. Data Description

We used ArcGIS (version 10.3.1) tools to process several hydrologically corrected layers from the 15″ (approximately 500 m at the equator) resolution HydroSHEDS database [*Lehner et al.,* 2008] including flow directions and flow accumulation for river network extraction, and elevation to calculate the local slope, using the neighborhood method [*Burrough and McDonell*, 1998] corrected for latitudinal distortions. Global lithology data are from *Hartmann and Moosdorf* [2012], with an average scale of 1:3,750,000 and 12 classes mapped in Figure S4 in the supporting information (Pi = Intermediate plutonic rocks; Mt = Metamorphic rocks; Va = Acid volcanic rocks; Ss = Siliciclastic sedimentary rocks; Vi = Intermediate volcanic rocks; Pb = Basic plutonic rocks; Pa = Acid plutonic rocks; Vb = Basic volcanic rocks; Py = Pyroclastics; Sm = Mixed sedimentary rocks; Su = Unconsolidated sediments; and Sc = Carbonate sedimentary rocks). Climate is represented by the multiannual mean of total precipitation (1980–2009) raster at 0.5° resolution from the WFDEI Water and Global Change Forcing Data methodology applied to ERA-Interim data set, including correction by the GPCC (Global Precipitation Climatology Centre) data [*Weedon et al.,* 2014].

The reference river networks were acquired in vector format from four national databases, from Australia [*OSDM*, 2015] (Geofabric) and France [*IGN*, 2014] (CARTHAGE) at a 1:50,000 scale, from the United States [*USGS*, 2015] (NHD) at a 1:24,000 scale, from Brazil [Instituto Brasileiro de Geografia e Estatística (*IBGE)*, 2015] at variable scales between 1:25,000 and 1:250,000. They were chosen for their quality and reported scale of approximately 1:50,000, despite some inconsistencies, as analyzed in the supporting information Text S1. For additional comparison with a "hypothetical" global river network, we used the global

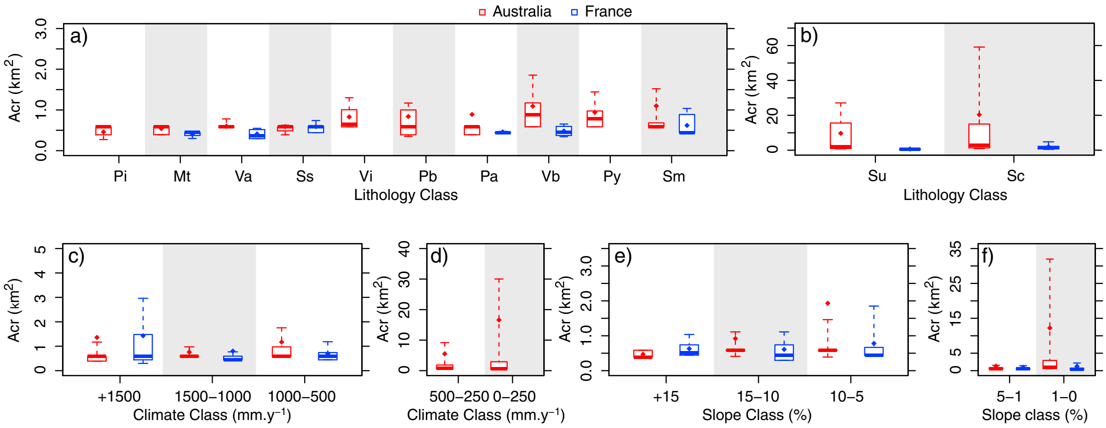


Figure 1. A_{cr} calibration results in Australia (red) and France (blue) per environmental class. (a and b) The lithology, (c and d) climate, and (e and f) slope classes. The bottom and top of the box plots represent the first and third quartiles, the middle bar gives the median, and dots indicate the mean values. Whiskers show the 10th and 90th percentiles.

AQUAMAPS [*FAO*, 2014] network, further called Food and Agriculture Organization (FAO), based on the 15″ HydroSHEDS data and a constant A_{cr} of 100 km^2, which also provides an intermittency assessment (details in Text S1).

As detailed in Text S2, all drainage density analyses and calculations were made using a 7.5′ grid (grid cells approximately 150 km^2 at equator) over the continents. Drainage densities (km^{-1}) were calculated as the total river length (km) inside a 7.5′ grid cell, divided by the corresponding land area (km^2). Each 7.5′ grid cell was characterized by a single value for δ and each of the selected environmental parameters (dominant class for lithology, areal average of total precipitation (mm yr^{-1}), and slope (%), each reclassified into five classes) (Table S1 and Figures 1c, 1d, and S6 for precipitation and Figures 1e, 1f, and S7 for the slope).

3. A_{cr} Calibration and Model Selection in France and Australia

The spatial intersection of the selected environmental parameters (12 lithology classes, 5 precipitation classes, and 5 slope classes) defines 300 environmental classes. In each of them, A_{cr} was calibrated by minimizing the bias between the corresponding drainage density (δ) and δ_{Ref} (drainage density with a reference river network). This minimization was performed independently in each country, by testing 300 extracted river networks, defined by a wide range of A_{cr} (from 0.3 to 200 km^2, see Text S2). To avoid calibration errors due to significant heterogeneities inside the 7.5′ grid cells, calibration was only performed in grid cells where the dominant lithology class covered more than 70% of the grid cell (representing 81% of the continents).

To permit independent validation, calibration was restricted to France and Australia, in which the reference river networks share the same 1:50,000 scale and show few inconsistencies. Together, these two countries also encompass all the precipitation, lithology, and slope classes of the currently available HydroSHEDS domain (56°S to 60°N). The resulting calibrated values in the two countries are given in Data Set S1.

We found much smaller A_{cr} values than the values used to define stream networks in the global-scale databases (i.e., HYDRO1k, AQUAMAPS, and HydroSHEDS, with A_{cr} values between 25 and 1000 km^2), and a similar dependency on lithology, climate, and slope in both countries (Figure 1). In agreement with previous studies that were reviewed in the Introduction, the calibrated A_{cr} increases and δ_{Ref} decreases when precipitation decreases (arid and semiarid climates, Figure 1d) for permeable rocks (unconsolidated sediments and carbonated rocks, Figure 1b) and when the slope decreases (consistent with stronger erosive power in steeper watersheds). The calibrated A_{cr} also tends to be larger in Australia (especially for classes with the highest values which permit the largest variability), which can be attributed largely to arid and semiarid climates (65% of Australia), where rivers are rare and often intermittent (the latter amounting to 69% of total stream length in FAO and to 98% in Geofabric, Table S2).

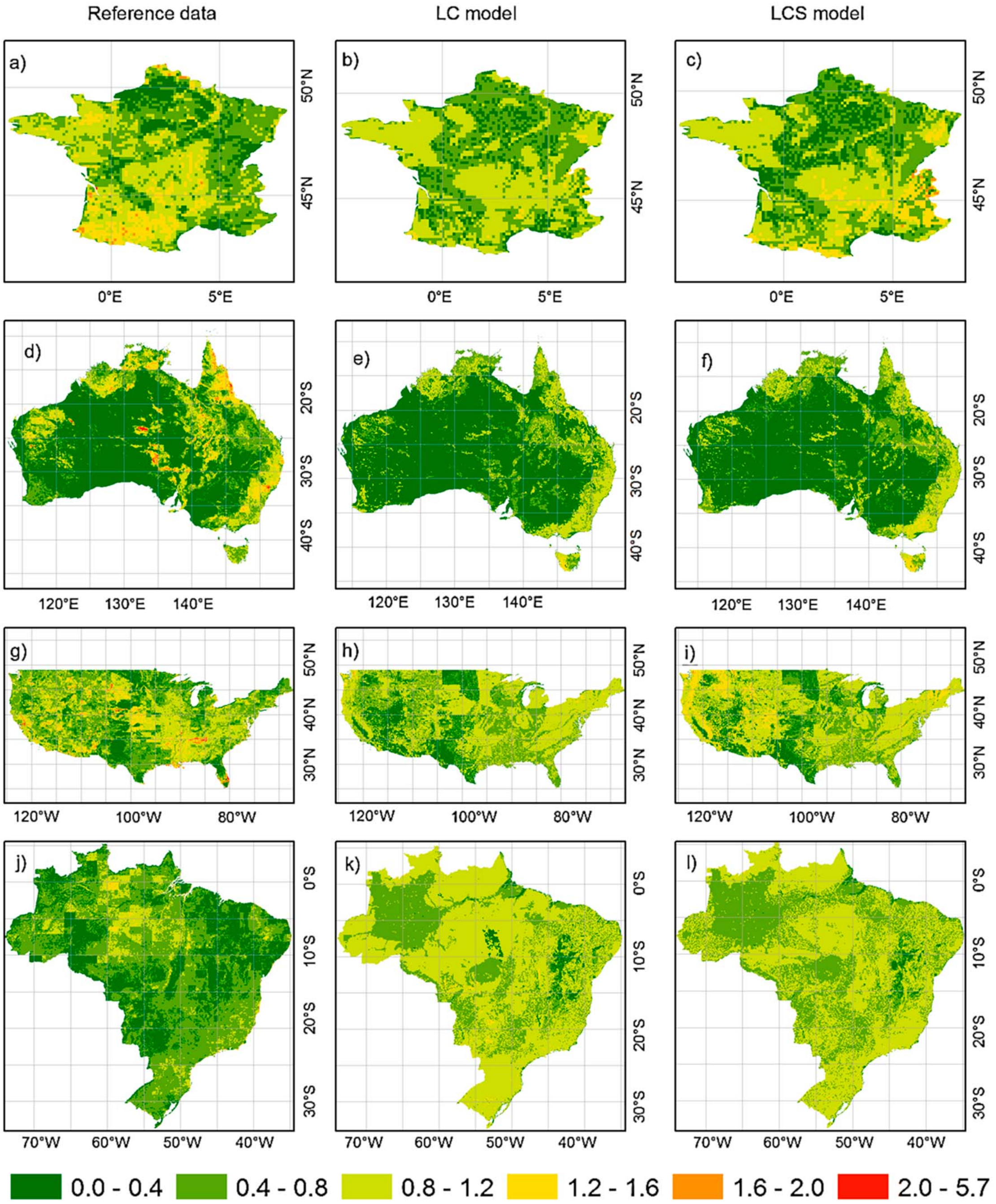

Reference data
LC model
LCS model
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
50°N
45°N
0°E
5°E
20°S
30°S
40°S
120°E
130°E
140°E
50°N
40°N
30°N
120°W
100°W
80°W
0°S
10°S
20°S
30°S
70°W
60°W
50°W
40°W
0.0 - 0.4
0.4 - 0.8
0.8 - 1.2
1.2 - 1.6
1.6 - 2.0
2.0 - 5.7

Figure 2. Maps of drainage density (km^{-1}) for the reference river networks and the evaluated models in (a–c) France, (d–f) Australia, (g–i) U.S., and (j–l) Brazil: Reference data (Figures 2a, 2d, 2g, and 2j); LC (Figures 2b, 2e, 2h, and 2k); and LCS (Figures 2c, 2f, 2i, and 2l).

In the second step, we developed statistical models to produce a set of A_{cr} values for the different environmental classes, to be used to extract river networks worldwide. The rationale was to use the rules emerging from Figure 1 regarding the effects of lithology, slope, and climate to create consistency between the two countries and define one single A_{cr} value for a given environmental class. In doing so, we ensured that we accounted for the effects of latitude on pixel area (Text S2), and we tried to group some environmental classes to limit the number of different A_{cr} values.

Table 1. Statistical Parameters for the Comparison of the Calculated and Reference Drainage Densities δ per Country[a]

Region	Parameter	FAO	LC	LCS	Ref.	Others
AU	Mean	0.08	0.42	0.45	0.43	-
	SD	0.07	0.32	0.32	0.43	-
	Bias	−0.36	−0.01	0.02	0.00	-
	Corr.	−0.01	0.57	0.59	1.00	-
	RMSE	0.56	0.36	0.34	0.00	-
FR	Mean	0.08	0.68	0.71	0.75	0.34
	SD	0.08	0.24	0.30	0.36	0.18
	Bias	−0.66	−0.07	−0.04	0.00	−0.41
	Corr.	0.01	0.30	0.32	1.00	0.19
	RMSE	0.81	0.40	0.39	0.00	0.56
US	Mean	0.07	0.72	0.78	0.75	1.30
	SD	0.07	0.28	0.30	0.37	0.84
	Bias	−0.68	−0.03	0.03	0.00	0.55
	Corr.	0.09	0.09	0.11	1.00	0.24
	RMSE	0.77	0.47	0.47	0.00	1.00
BR	Mean	0.08	0.93	0.92	0.52	-
	SD	0.07	0.16	0.16	0.29	-
	Bias	−0.44	0.41	0.40	0.00	-
	Corr.	0.12	0.14	0.14	1.00	-
	RMSE	0.49	0.49	0.48	0.00	-
Europe	Mean	0.08	0.59	0.63	-	0.31
	SD	0.08	0.29	0.32	-	0.15
Global	Mean	0.08	0.69	0.74	-	-
	SD	0.08	0.32	0.34	-	-
	% Intermittent	34	27	29	-	-

[a] Ref. = δ from river network reference data (AU = Geofabric and FR = CARTHAGE used for calibration and US = NHD and BR = *IBGE* [2015] used for validation); Others = δ from *Luo et al.* [2016] over the U.S. and from CCM2 [*Vogt et al.*, 2007] over France and Europe; Mean = mean δ (km^{-1}); SD = standard deviation of δ (km^{-1}); Bias = bias between model and Ref. (km^{-1}); Corr. = correlation coefficient between model and Ref. (dimensionless); RMSE = root-mean-square error between model and Ref. (km^{-1}). % Intermittent corresponds to the ratio of total length of intermittent rivers to the total river length.

Two models of different complexity were eventually constructed to evaluate the effects of slope on river network extraction by considering only lithology and climate parameters (LC) and adding the slope (LCS). The corresponding A_{cr} values are given in the supporting information Tables S3 and S4. The LC model comprises 11 different A_{cr} values varying from 0.3 km^2 to 97 km^2 for 35 environmental classes, derived from 7 lithology groups (Mt + Pi + Pa + Py, Pb, Sc, Su, Sm, Ss + Va, and Vb + Vi) and 5 climate classes (Table S1). The LCS model has 29 different A_{cr} values varying from 0.3 to 193 km^2 corresponding to 120 environmental classes that combine 6 lithology classes (Mt + Pi + Pa + Py, Pb, Sc, Su, Ss, and Sm + Va + Vb + Vi), 5 climate classes, and 4 slope classes (under 1%, between 1 and 5%, between 5 and 10%, and over 10%).

We also used the Australian case to propose a first-order method to identify intermittent rivers explained by the aridity of climates. The principle is that rivers initiated in regions where precipitation is smaller than a calibrated threshold are defined as intermittent until they reach a pixel where precipitation exceeds this value. Once a stream becomes perennial, all downstream river pixels remain as such (see the Nile in Figure 3b), so this strategy cannot reproduce cases of disrupted connectivity. As detailed in Text S2, the precipitation threshold was calibrated to get the best overlap with areas of high intermittency in FAO, which defines 69% of intermittent rivers in Australia (% of total length). This defined a threshold of 500 mm yr^{-1}, i.e., the classical upper bound of semiarid climates, according to which 42% of the rivers from the LC or LCS models in Australia are intermittent. This percentage is smaller than in FAO because our models predict higher densities, so the total length of intermittent streams is much higher with our models, approximately 420,000 km in FAO; 1,483,000 km in LCS; and 3,242,000 km in Geofabric. We disregarded the latter data set as it classifies 98% of streams as intermittent, which exceeds the maximum of 90% used in *Raymond et al.* [2013]. Matching this 98% with our uniform precipitation threshold would also constrain the predicted perennial streams to very humid climates (annual mean precipitation > 1500 mm yr^{-1}).

4. Model Results and Discussion

The resulting drainage densities were first evaluated against the four national reference hydrographic data sets (Figure 2 and Table 1). The LC model captures the main features of the reference δ, in particular, the low values characterizing areas with carbonate rocks in France, and arid to semiarid climates in Australia. However, it underestimates the high δ values, mostly found in mountainous areas in both countries. These are better represented by LCS, which accounts for the increase of δ with slope, but this model still underestimates some very high values, such as in Australia, where siliciclastic rocks are present (Ss in Figure S5), and

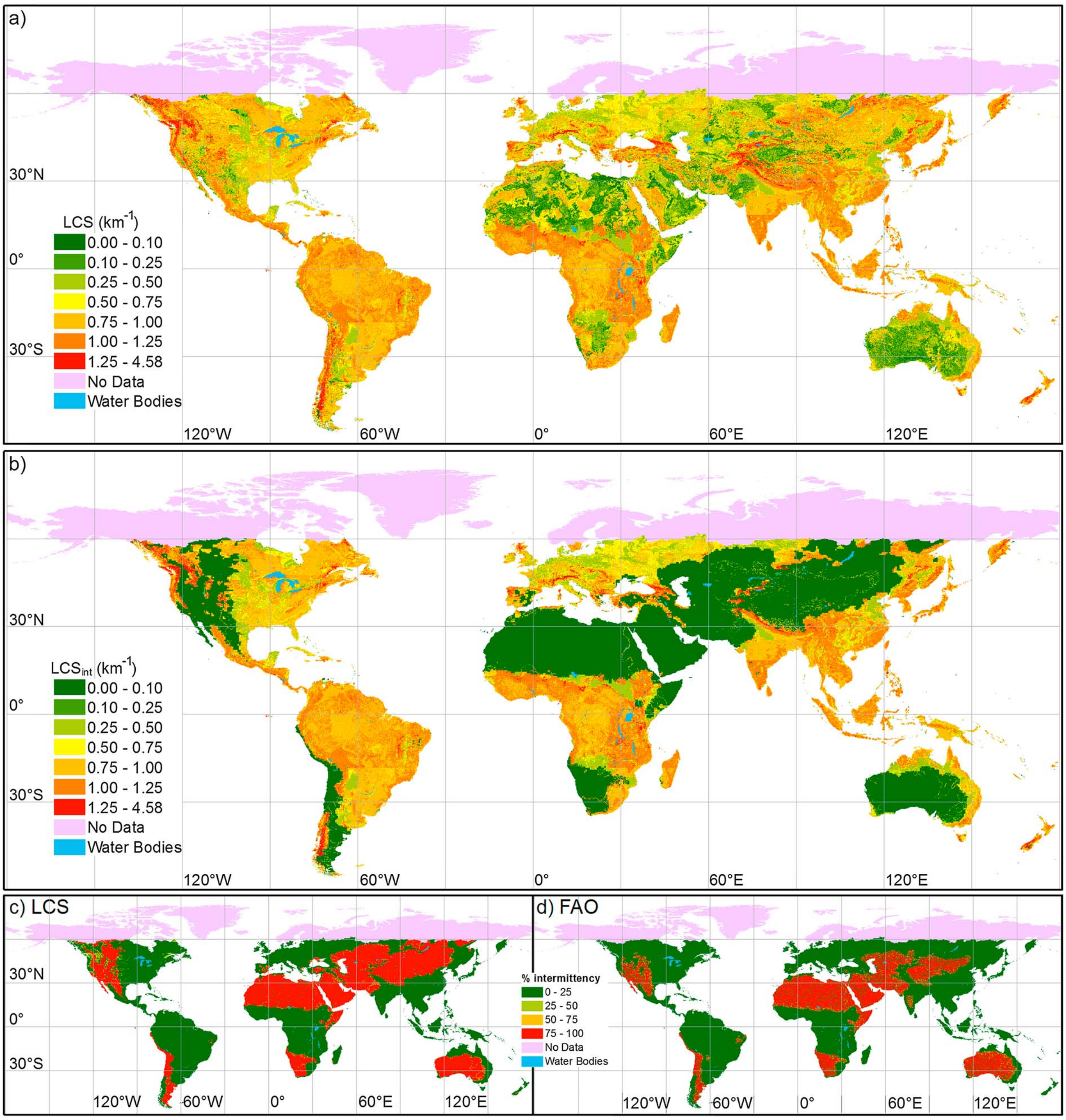

a)
30°N
0°
30°S
LCS (km^{-1})
🟩 0.00 - 0.10
🟩 0.10 - 0.25
🟩 0.25 - 0.50
🟨 0.50 - 0.75
🟧 0.75 - 1.00
🟧 1.00 - 1.25
🟥 1.25 - 4.58
No Data
Water Bodies
120°W
60°W
0°
60°E
120°E
b)
30°N
0°
30°S
LCS$_{int}$ (km^{-1})
🟩 0.00 - 0.10
🟩 0.10 - 0.25
🟩 0.25 - 0.50
🟨 0.50 - 0.75
🟧 0.75 - 1.00
🟧 1.00 - 1.25
🟥 1.25 - 4.58
No Data
Water Bodies
120°W
60°W
0°
60°E
120°E
c) LCS
30°N
0°
30°S
120°W
60°W
0°
60°E
120°E
d) FAO
% intermittency
🟩 0 - 25
🟩 25 - 50
🟧 50 - 75
🟥 75 - 100
No Data
Water Bodies
120°W
60°W
0°
60°E
120°E

Figure 3. Global maps of drainage density (km^{-1}) for LCS model: (a) for full network including intermittent streams and (b) intermittent streams only. Percentage of intermittent streams: (c) LCS and (d) FAO. In each 7.5′ cell, the % of intermittent streams is the ratio of the intermittent stream length to total stream length. The violet color north of 60°N indicates where HydroSHEDS data are missing.

in the southwestern part of France (Landes), even though it is a rather flat area (Figure S7) with permeable rocks (Su, Sc, and Sm in Figure S5). The latter is likely due to artificial drainage densities in the reference data because of irrigation ditches in this area of intensive agriculture (Text S1).

Similar behaviors are found in the U.S. and Brazil, where the inclusion of a slope constraint in the LCS model also leads to slightly better results than LC, but both models exhibit poor correlations with the reference data sets, as quantified in Table 1. Nevertheless, the excessive underestimation of δ by FAO is markedly improved by the variable A_{cr} models, and the general spatial patterns are well depicted. In the U.S., France, and Australia, the biases approach zero (less than 5% absolute value). They remain negative over France, which is mainly attributed to the overestimation of the highest δ_{Ref} compared to a "pristine" case because of human-made networks, as discussed previously. In Brazil, in contrast, the proposed method leads to high positive biases; this likely results from the reference river network being based on multiple scaled data, up to 1:250,000, which alters the natural variability of δ (Figure 2) and leads to underestimation of δ_{Ref} compared to what would prevail at the 1:50,000 scale used for the A_{cr} calibration in France and Australia. The inconsistencies of the reference river networks, discussed in Text S1 for both Brazil and the U.S., largely explain the poor spatial fit to δ_{Ref} as revealed by the correlation coefficients and RMSE.

These inconsistencies prevented *Luo et al.* [2016] from analyzing the relationships between drainage density from NHD and possible explanatory factors and led them to construct their own drainage density map. Table 1 shows that this latter map overestimates the mean δ compared to NHD and our two estimates, which suggests that the 30 m DEM of *Luo et al.* [2016] corresponds to a finer scale than both 1:24,000 and 1:50,000. As expected, the correlation coefficients increased (doubled, as $\mathrm{Corr}_{(\mathrm{LCS,Luo})} = 0.24$) when comparing δ of LC and LCS model to δ_{Luo} rather than δ_{Ref} (see also the maps of δ_{Ref}, δ_{Luo} and δ_{LC} in Figures S11 and S13). However, this correlation remains low, and the weak performance of the proposed models in the U.S. could be due to the quality of the lithological map (Figure S6), which explains the main discontinuities along the Great Plains from North Dakota to Texas and around Lake Michigan (anticlockwise from Wisconsin to Michigan). It is worth noting that the geological map of *Schruben et al.* [1994] does not exhibit these discontinuities.

We also compared the proposed δ to values extracted from *Vogt et al.* [2007], which exhibited weak spatial variations throughout Europe, strongly controlled by topography (Figure S12) with values below 0.4 km^{-1} in most lowland areas, which is below the range of 0.5 to 1.6 km^{-1} from the small-scale studies reported in section 1. As a result, δ_{Ref} is much closer to δ from LC and LCS than to δ_{CCM2} over France (Table 1), both in terms of spatial match and bias.

Figure 3a shows the global drainage density map from the LCS model, which exhibits complex patterns arising from the combined controls by lithology, climate, and slope (see also LC and FAO in Figures S13 and S14). The perennial streams show the same density patterns (Figure 3b) but are absent from arid and semiarid areas (39% of global land based on our precipitation data set), except for rivers initiated as perennial which conserve this feature when crossing arid regions, as the Nile, for instance. By design, the complementary intermittent streams are concentrated in arid and semiarid areas both in FAO and LCS (Figures 3c and 3d), with a rather similar proportion of the full network (around one third, Table 1), and a correlation coefficient of 0.65 over land. The main difference is found north of 45°N (North America and eastern Asia), with a larger fraction of intermittent streams according to LCS than FAO. This is consistent with the use of an aridity index combining both precipitation and evapotranspiration to discriminate intermittent streams in FAO/AQUAMAPS. Note that the total length of intermittent streams is much higher with our models than with FAO (Table S2), primarily due to differences in drainage density.

5. Conclusions

This study presents a method to obtain multiple critical drainage areas (A_{cr}) that are spatially dependent on lithology, climate, and slope. This new method addresses an important component of the large-scale river delineation process, in combination with proper DEM hydrologic conditioning, obtained here from HydroSHEDS. The A_{cr} values were calibrated against the national hydrographic data at a 1:50,000 scale over France and Australia, resulting in two models of increasing complexity: LC (using seven lithology and five climate classes), and LCS (using six lithology, five climate, and four slope classes). This work is based on the 15″ hydrologically conditioned version of HydroSHEDS, but the proposed A_{cr} values are a priori suitable to constrain river network extractions from any DEM with similar or higher resolutions.

Both models show fair performance compared to the reference river networks, with better agreement in the countries used for calibration. The inclusion of slope in the model improves the performance criteria in the evaluated countries, but the effect is modest. Combined with an intermittency assessment solely based on mean precipitation, the proposed variable A_{cr} models give similar locations and percentages of intermittent streams as FAO/AQUAMAPS, but with higher and more spatially variable drainage densities. The limitations of this first-order classification underline the need for better description of the multiple controls of intermittency.

The two proposed models predict global mean drainage density to reach approximately 0.70 km^{-1}, with a precision of approximately 5% compared to the reference data. Drainage density and scale are tightly linked [*Tarboton et al.*, 1992], so the proposed value is consistent with the 1:50,000 scale that prevailed for A_{cr} calibration. It is also higher than the mean densities derived from classical single A_{cr} river networks, which are thus shorter and should be used with caution for fine-scale applications, as previously reported by *Raymond et al.* [2013].

As previously discussed, the quality of A_{cr} calibration and validation in our methodology strongly depends on the reference data. The main uncertainties seem to come from unnatural channels and scale inconsistencies. The quality of the input DEM and flow accumulations is also important, although this is difficult to evaluate. As shown, environmental input parameters can induce uncertainties, particularly with respect to the lithological map. As a result, any improved lithology, either at the global scale or over a specific region, could assist with the estimation of drainage density, provided that calibration is updated accordingly. Eventually, another way to improve drainage density estimates would be to include more control factors at the calibration step, e.g., more complex geomorphologic information or vegetation parameters [*Colombo et al.,* 2007; *Luo et al.,* 2016], which were not addressed here for the sake of simplicity. Another approach is to use hyperresolution hydrological modeling to define the locations where streams initiate, as recommended by *Lehner and Grill* [2013] and recently achieved by *Maxwell et al.* [2015] over the United States based on the HydroSHEDS DEM. Land use information could also be used to generate artificialized river networks if the reference data for calibration includes information on anthropogenic pressures. However, when dealing with a surveyed river network with no such information, the drainage density difference with the LCS network may be used as a first-order indicator of anthropogenic impacts.

Taking the above caveats into consideration, river networks and corresponding drainage density maps that can be constructed from the proposed A_{cr} models have the advantage of describing the main heterogeneities of natural river networks with a uniform scale of 1:50,000 across all continents. This feature is important to support water management in regions with limited observations and to provide consistent information to large-scale models seeking higher resolution, which is an important evolution of both land surface and hydrological models [*Wood et al.,* 2011].

Acknowledgments
Raster data extracted from LC and LCS models are freely available for download at http://www.metis.upmc.fr/en/node/211. They include the 15″ resolution river networks (with intermittency and Strahler order attribute for each river pixel) and drainage densities at 7.5′ resolution. This work is part of the PhD thesis of Ana Schneider, funded by the European Institute of Innovation and Technology (Climate KIC—Knowledge and Innovation Community), the French Agence Nationale de la Recherche (ANR grant ANR-14-CE01-00181-01), and the French national programme LEFE/INSU. The authors are grateful to Wei Luo for providing drainage density data and to Bernard Lehner and another anonymous reviewer for their very constructive comments, which helped to significantly improve the present manuscript.

References

Benstead, J. P., and D. S. Leigh (2012), An expanded role for river networks, *Nat. Geosci.*, *5*, 678–679.

Billen, G., V. Thieu, J. Garnier, and M. Silvestre (2009), Modelling the N cascade in regional watersheds: The case study of the Seine, Somme and Scheldt rivers, *Agric. Ecosyst. Environ.*, *133*(3), 234–246.

Bishop, K., I. Buffam, M. Erlandsson, J. Fölster, H. Laudon, J. Seibert, and J. Temnerud (2008), Aqua incognita: The unknown headwaters, *Hydrol. Process.*, *22*, 1239–1242.

Brutsaert, W., and J. P. Lopez (1998), Basin-scale geohydrologic drought flow features of riparian aquifers in the southern Great Plains, *Water Resour. Res.*, *34*, 233–240, doi:10.1029/97WR03068.

Brutsaert, W., and J. L. Nieber (1977), Regionalized drought flow hydrographs from a mature glaciated plateau, *Water Resour. Res.*, *13*, 637–643, doi:10.1029/WR013i003p00637.

Burrough, P. A., and R. A. McDonell (1998), *Principles of Geographical Information Systems*, Oxford Univ. Press, New York.

Colombo, R., J. V. Vogt, P. Soille, M. L. Paracchini, and A. de Jager (2007), Deriving river networks and catchments at the European scale from medium resolution digital elevation data, *Catena*, *70*(3), 296–305.

Dingman, S. L. (2015), *Physical Hydrology*, Waveland Press, Long Grove, IL.

Döll, P., and B. Lehner (2002), Validation of a new global 30-min drainage direction map, *J. Hydrol.*, *258*, 214–231.

Food and Agriculture Organization (FAO) (2014), AQUAMAPS Global spatial database on water and agriculture, Food and Agriculture Organization of the United Nations. [Available at http://www.fao.org/nr/water/aquamaps/.]

Fekete, B. M., C. J. Vörösmarty, and R. Lammers (2001), Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error, *Water Resour. Res.*, *37*, 1955–1967, doi:10.1029/2001WR900024.

Graham, S. T., J. S. Famiglietti, and D. R. Maidment (1999), Five-minute, 1/2°, and 1° data sets of continental watersheds and river networks for use in regional and global hydrologic and climate system modeling studies, *Water Resour. Res.*, *35*, 583–587, doi:10.1029/1998WR900068.

Hartmann, J., and N. Moosdorf (2012), The new global lithological map database GLiM: A representation of rock properties at the Earth surface, *Geochem. Geophys. Geosyst.*, *13*, Q12004, doi:10.1029/2012GC004370.

Heine, R. A., C. L. Lant, and R. R. Sengupta (2004), Development and comparison of approaches for automated mapping of stream channel networks, *Ann. Assoc. Am. Geogr.*, *94*(3), 477–490.

Horton, R. E. (1932), Drainage basin characteristics, *Eos Trans. AGU*, *13*, 350–361.

Horton, R. E. (1945), Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology, *Geol. Soc. Am.*, *56*, 275–370.

Hutchinson, M. (1989), A new procedure for gridding elevation and stream line data with automatic removal of spurious pits, *J. Hydrol.*, *106*(3–4), 211–232.

Instituto Brasileiro de Geografia e Estatística (IBGE) (2015), Insituto Brasileiro de Geografia e Estatistica. [Available at ftp://geoftp.ibge.gov.br/.]

IGN (2014), Base de Données sur la CARtographie THématique des AGences de l'eau et du ministère chargé de Institut national de l'information géographique et forestière.

Lehner, B., and P. Döll (2004), Development and validation of a global database of lakes, reservoirs and wetlands, *J. Hydrol.*, *296*(1–4), 1–22.

Lehner, B., and G. Grill (2013), Global river hydrography and network routing: Baseline data and new approaches to study the world\r qs large river systems, *Hydrol. Process.*, *27*(15), 2171–2186.

Lehner, B., K. Verdin, and A. Jarvis (2008), New global hydrography derived from spaceborne elevation data, *Eos Trans. AGU*, *89*(10), 93–94, doi:10.1029/2008EO100001.

Luo, W., J. Jasiewicz, T. Stepinski, J. Wang, C. Xu, and X. Cang (2016), Spatial association between dissection density and environmental factors over the entire conterminous United States, *Geophys. Res. Lett.*, *43*, 692–700, doi:10.1002/2015GL066941.

Maxwell, R., L. Condon, and S. Kollet (2015), A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3, *Geosci. Model Dev.*, *8*(3), 923–937.

Montgomery, D. R., and E. Foufoula-Georgiou (1993), Channel network source representation using digital elevation models, *Water Resour. Res.*, *29*, 3925–3934, doi:10.1029/93WR02463.

Moore, I. D., and G. J. Burch (1986), Physical basis of the length-slope factor in the universal soil loss equation, *Soil Sci. Soc. Am. J.*, *50*(5), 1294–1298.

O'Callaghan, J. F., and D. M. Mark (1984), The extraction of drainage networks from digital elevation data, *Comput. Vis. Graph. Image Process.*, *28*, 323–344.

OSDM (2015), Geofabric Surface Cartography—V.3.0, Office of Spatial Data Management, Commonwealth of Australia (Bureau of Meteorology). [Available at http://www.bom.gov.au/water/geofabric/download.shtml.]

Pelletier, J. D. (2013), A robust, two-parameter method for the extraction of drainage networks from high-resolution digital elevation models (DEMs): Evaluation using synthetic and real-world DEMs, *Water Resour. Res.*, *49*, 75–89, doi:10.1029/2012WR012452.

Persendt, F. C., and C. Gomez (2016), Assessment of drainage network extractions in a low-relief area of the Cuvelai Basin (Namibia) from multiple sources: LiDAR, topographic maps, and digital aerial orthophotographs, *Geomorphology*, *260*, 32–50, doi:10.1016/j.geomorph.2015.06.047.

Raymond, P. A., et al. (2013), Global carbon dioxide emissions from inland waters, *Nature*, *503*, 355–359, doi:10.1038/nature12760.

Sangireddy, H., R. A. Carothers, C. P. Stark, and P. Passalacqua (2016), Controls of climate, topography, vegetation, and lithology on drainage density extracted from high resolution topography data, *J. Hydrol.*, *537*, 271–282, doi:10.1016/j.jhydrol.2016.02.051.

Schruben, P. G. A., R. E. Bawiec, W. J. King, P. B. Beikman, and M. Helen (1994), Geology of the conterminous United States at 1: 2,500,000 scale—A digital representation of the 1974 PB King and HM Beikman map, Peter N. Schweitzer.

Stein, J. L., M. F. Hutchinson, and J. A. Stein (2014), A new stream and nested catchment framework for Australia, *Hydrol. Earth Syst. Sci.*, *18*, 1917–1933, doi:10.5194/hess-18-1917-2014.

Strahler, A. N. (1957), Quantitative analysis of watershed geomorphology, *Eos Trans. AGU*, *38*(6), 913–920.

Tarboton, D. G., R. L. Bras, and I. Rodroguez-Iturbe (1991), On the extraction of channel networks from digital elevation data, *Hydrol. Process.*, *5*, 81–100.

Tarboton, D. G., R. L. Bras, and I. Rodriguez-Iturbe (1992), A physical basis for drainage density, *Geomorphology*, *5*, 59–76.

Tucker, G. E., F. Catani, A. Rinaldo, and R. L. Bras (2001), Statistical analysis of drainage density from digital terrain data, *Geomorphology*, *36*, 187–202.

USGS (2015), National Hydrography Dataset at 1:24000, for all available states, United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), the United States Geological Survey (USGS), and the Environmental Protection Agency (EPA). [Available at http://nhd.usgs.gov/data.html], accessed in December 2015.

Verdin, K., and S. Greenlee (1998), HYDRO1k documentation, Sioux Falls ND US Geol. Surv. EROS Data Cent. Httpedcdaac Usgs Govgtopo30hydroreadme Html.

Vogt, J., et al. (2007), A pan-European river and catchment database. European Commission, Joint Research Center, Inst. Environ. Sustain. JRC Luxemb., 120 pp.

Vogt, J. V., R. Colombo, and F. Bertolo (2003), Deriving drainage networks and catchment boundaries: A new methodology combining digital elevation data and environmental characteristics, *Geomorphology*, *53*(3), 281–298.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo (2014), The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, *50*, 7505–7514, doi:10.1002/2014WR015638.

Wood, E. F., et al. (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, *47*, W05301, doi:10.1029/2010WR010090.

Wu, H., J. S. Kimball, H. Li, M. Huang, L. R. Leung, and R. F. Adler (2012), A new global river network database for macroscale hydrologic modeling, *Water Resour. Res.*, *48*, W0970, doi:10.1029/2009WR008871.

Zecharias, Y. B., and W. Brutsaert (1988), Recession characteristics of groundwater outflow and base flow from mountainous watersheds, *Water Resour. Res.*, *24*, 1651–1658, doi:10.1029/WR024i010p01651.

6.2 Drainage density calibration results

Total precipitation ($mm.y^{-1}$)	Lithology class	A_{cr} over Australia (km^2)						A_{cr} over France (km^2)					
		No Slope	S < 1	1 ≤ S < 5	5 ≤ S < 10	10 ≤ S < 15	S ≥ 15	No Slope	S < 1	1 ≤ S < 5	5 ≤ S < 10	10 ≤ S < 15	S ≥ 15
$P_{tot} < 250$	mt	0.6	0.6	0.6									
	pa	0.6	0.6	0.6									
	pb												
	pi												
	py	0.8		0.8									
	sc	97.5	195.0	1.4	29.2								
	sm	1.8	2.9	0.6	0.6								
	ss	0.6	0.6	0.6	0.6								
	su	29.2	31.2	15.6									
	va	0.8		0.8									
	vb												
	vi												
$250 \leq P_{tot} < 500$	mt	0.6	0.6	0.6	0.4								
	pa	0.6	0.6	0.6									
	pb	1.0		0.8									
	pi	0.4		0.4									
	py	0.6		0.6	0.6								
	sc	39.0	54.6	1.9									
	sm	2.1	6.8	1.0	0.6								
	ss	0.8	1.0	0.6	0.4								
	su	29.2	39.0	9.7	1.8								
	va	0.8		0.8									
	vb	0.8	2.7	0.8									
	vi												
$500 \leq P_{tot} < 1000$	mt	0.6	0.8	0.6	0.6	0.6		0.4	0.4	0.4	0.4	0.4	0.4
	pa	0.6	0.6	0.6	0.6	0.6	0.4	0.4	0.4	0.4	0.4	0.4	
	pb	1.2		0.8	1.2	1.2							
	pi	0.6		0.6	0.6	0.6							
	py	1.4			1.0	1.8							
	sc	2.9	4.9	0.8				1.8	2.2	1.8	1.5	1.2	0.7
	sm	0.8	1.2	0.8	0.6	0.6		0.4	0.4	0.4	0.4	1.0	1.0

	ss	0.6	1.4	0.6	0.6	0.4	0.4	0.6	0.6	0.6	0.6	0.4	1.0
	su	9.7	12.7	1.9	0.6			0.6	0.6	0.6	0.6		
	va	0.6		0.6	0.6	0.6		0.4		0.6	0.4	0.3	
	vb	1.0	1.8	1.0	0.6	0.6		0.7			0.7		
	vi	0.8	1.4		0.6	0.6							
$1000 \leq P_{tot} < 1500$	mt	0.6	0.6	0.6	0.4	0.6		0.4		0.4	0.4	0.3	0.4
	pa	0.6	0.6	0.6	0.4	0.6		0.4	0.4	0.4	0.4	0.4	
	pb	1.4		3.9	0.6	0.6							
	pi	0.6		0.6									
	py												
	sc	1.0		1.0				2.1	7.4	0.7	2.2	0.6	0.4
	sm	0.6	0.6	0.6	0.6	0.6	0.6	0.4	0.4	0.4	0.4	0.9	0.6
	ss	0.6	0.6	0.6	0.6	0.6	0.6	0.4	0.4	0.6	0.4	0.6	0.7
	su	1.0	1.0	0.8	1.9			0.6	0.6	0.6	0.6		
	va	0.6		0.6	0.6			0.3				0.3	
	vb	1.0		0.8	1.0			0.4		0.4	0.4	0.3	
	vi	0.6			0.4								
$P_{tot} \geq 1500$	mt	0.4	0.4	0.4	0.4	0.6		0.3		0.3		0.3	
	pa	0.6		0.4	0.4	6.8	0.4	0.6				0.6	
	pb	0.6		0.6	0.6								
	pi	0.4		0.4		0.6							
	py												
	sc							3.6		5.9	2.2	1.5	0.4
	sm	0.6	0.6	0.6	0.6	0.6							
	ss	0.6	0.6	0.6	0.4	0.4		0.4					0.4
	su	0.8	1.2	0.8	17.5			1.2		1.2			
	va	0.4		0.6	0.6	0.4							
	vb	0.6		0.6	1.2								
	vi	0.6				0.6							

6.3 A_{cr} values for LC and LCS models

A_{cr} values defining the LC model (values in km^2).

Lithology groups	**Climate classes (mm.y^{-1})**				
	[0,250[	**[250,500[**	**[500,1000[**	**[1000,1500[**	**[1500,∞[**
Mt, Pa, Pi, Py	0.6	0.6	0.5	0.5	0.4
Pb	1.2	1.2	1.2	1.2	0.6
Sc	96.3	38.5	2.6	1.7	1.5
Sm	1.9	1.9	0.7	0.5	0.5
Su	28.9	28.9	3.4	0.7	0.7
Ss, Va	0.8	0.8	0.5	0.5	0.3
Vb, Vi	1.0	1.0	0.8	0.6	0.4

A_{cr} values defining the LCS model (values in km^2).

Lithology groups	**Slope classes (%)**				**Climate classes**
	S < 1	**1 ≤ S < 5**	**5 ≤ S < 10**	**S ≥ 10**	**(mm.y^{-1})**
Mt, Pa, Pi, Py	0.6	0.6	0.6	0.4	**Ptot < 250**
	0.6	0.6	0.6	0.4	**250 ≤ Ptot < 500**
	0.5	0.5	0.5	0.3	**500 ≤ Ptot < 1000**
	0.5	0.5	0.3	0.3	**1000 ≤ Ptot < 1500**
	0.3	0.3	0.3	0.3	**Ptot ≥ 1500**
Ss	1.0	0.6	0.6	0.4	**Ptot < 250**
	1.0	0.6	0.6	0.4	**250 ≤ Ptot < 500**
	0.9	0.5	0.5	0.3	**500 ≤ Ptot < 1000**
	0.5	0.5	0.5	0.3	**1000 ≤ Ptot < 1500**
	0.5	0.5	0.3	0.3	**Ptot ≥ 1500**
Sm, Va, Vb, Vi	2.9	0.8	0.6	0.4	**Ptot < 250**
	2.9	0.8	0.6	0.4	**250 ≤ Ptot < 500**
	1.4	0.7	0.5	0.3	**500 ≤ Ptot < 1000**
	0.5	0.5	0.5	0.3	**1000 ≤ Ptot < 1500**
	0.5	0.5	0.5	0.3	**Ptot ≥ 1500**
Su	30.8	15.4	3.1	1.5	**Ptot < 250**
	19.3	9.6	1.9	1	**250 ≤ Ptot < 500**
	4.3	1.7	0.7	0.3	**500 ≤ Ptot < 1000**
	0.7	0.7	0.7	0.3	**1000 ≤ Ptot < 1500**
	0.7	0.7	0.7	0.3	**Ptot ≥ 1500**
Sc	192.6	96.3	28.9	28.9	**Ptot < 250**
	57.8	19.3	3.9	3.9	**250 ≤ Ptot < 500**
	5.1	3.4	1.7	1.7	**500 ≤ Ptot < 1000**
	2.6	1.7	1.7	1.7	**1000 ≤ Ptot < 1500**
	1.7	1.7	1.7	1.7	**Ptot ≥ 1500**
Pb	1.2	1.2	1.2	1.2	**Ptot < 250**
	1.2	1.2	1.2	1.2	**250 ≤ Ptot < 500**
	1.0	1.0	1.0	1.0	**500 ≤ Ptot < 1000**
	1.0	1.0	1.0	1.0	**1000 ≤ Ptot < 1500**
	0.5	0.5	0.5	0.5	**Ptot ≥ 1500**

6.4 Drainage density maps for FAO data and LC model

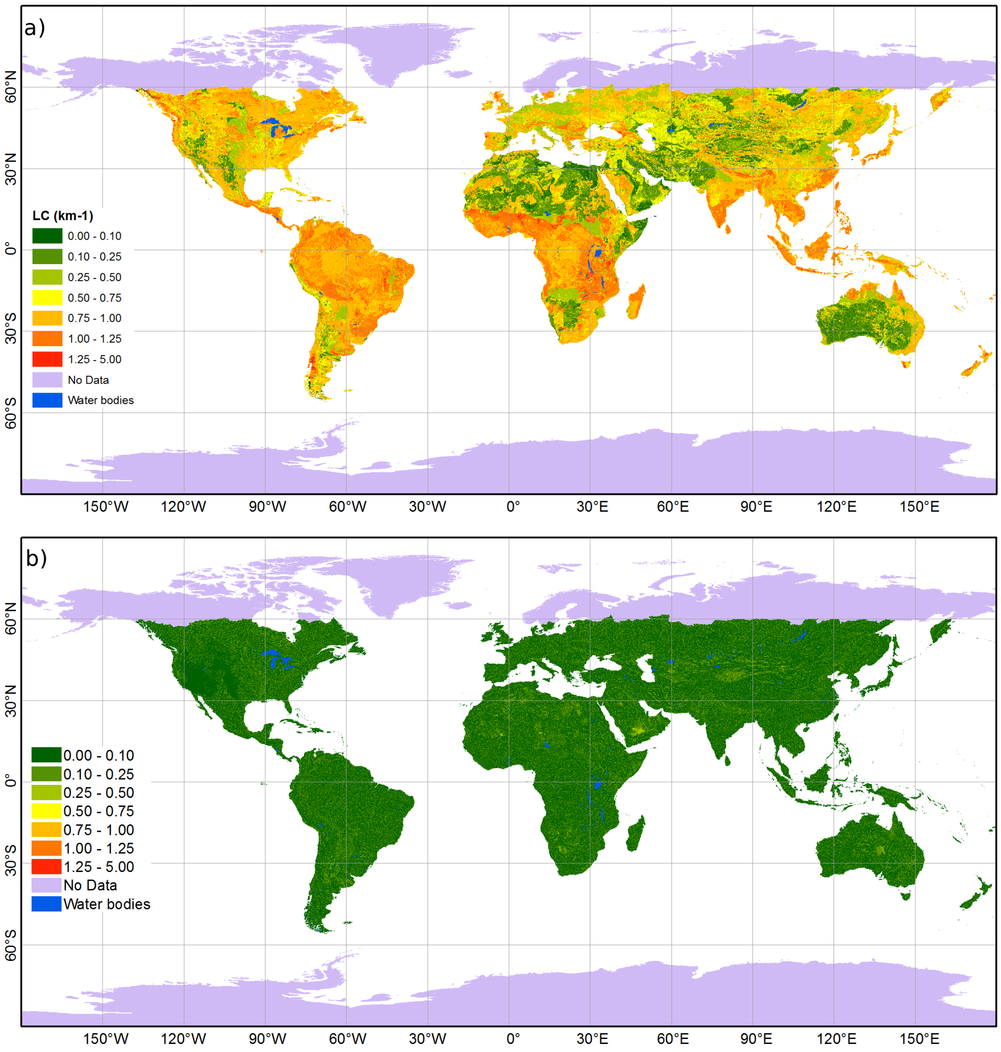

a)
LC (km-1)
0.00 - 0.10
0.10 - 0.25
0.25 - 0.50
0.50 - 0.75
0.75 - 1.00
1.00 - 1.25
1.25 - 5.00
No Data
Water bodies
60°N
30°N
0°
30°S
60°S
150°W
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
b)
0.00 - 0.10
0.10 - 0.25
0.25 - 0.50
0.50 - 0.75
0.75 - 1.00
1.00 - 1.25
1.25 - 5.00
No Data
Water bodies

Figure 6.1 – Drainage density (km^{-1}) for a) LC model and b) FAO data.

6.5 Descriptive statistics of τ

Table 6.1 – Descriptive statistics for the evaluated τ (in years). Gray rows represent results at 0.5°, and white rows at 7.5'.

τ	Min	1st. Q.	Median	Mean	3rd. Q.	Max
τ_{ini}	1	326	850	23110	6880	19880000
τ_{ini}	0	231	786	72070	13330	178500000
τ_{G1}	1	714	2893	70960	21420	99420000
τ_{G1}	0	468	2069	206800	30150	585500000
τ_{G2}	0	332	1217	30190	9338	39770000
τ_{G2}	0	220	902	88900	13850	234200000
τ_{WT}	0	190	539	15820	3803	5487000
τ_{WT}	0	120	420	34370	5700	125900000
τ_{TP}	0	124	416	13270	2494	3791000
τ_{TP}	0	58	272	23460	2743	53690000
τ_{LCS}	0	17	88	216	559	3008000
τ_{LCS}	0	11	71	5381	861	172600000
τ_{ne}	0	14	65	517	384	267600
τ_{ne}	0	9	54	931	485	19990000
τ_{U}	0	5	17	128	62	117300
τ_{U}	0	3	13	235	65	2668000
τ_{UG}	0	2	7	32	21	21692
τ_{UG}	0	2	6	75	24	1255000

6.6 Statistical coefficients of TWS anomalies

Table 6.2 – Coefficient of determination (R^2) and Nash-Sutcliffe coefficient (Nash) for TWS anomalies (mm) of the evaluated watersheds.

Basin	τ_{ORC}		$\tau_{UG/10}$		τ_{UG}		τ_{ne}	
	R^2	Nash	R^2	Nash	R^2	Nash	R^2	Nash
Amazon	0.99	0.98	0.97	0.96	0.93	0.91	0.88	0.68
Congo	0.95	0.87	0.95	0.89	0.94	0.89	0.92	0.84
Mississippi	0.97	0.96	0.98	0.97	0.99	0.98	0.99	0.95
Parana	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.92
Yenisei	0.94	0.78	0.92	0.68	0.90	0.56	0.77	0.33
Ob	0.94	0.89	0.94	0.84	0.93	0.77	0.84	0.57
Lena	0.63	0.36	0.60	0.31	0.57	0.26	0.45	0.10
Niger	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.95
Zambezi	0.99	0.93	0.99	0.93	0.99	0.92	0.99	0.94
Yangtze	0.94	0.92	0.92	0.89	0.89	0.84	0.85	0.65
Mackenzie	0.95	0.92	0.94	0.89	0.93	0.86	0.85	0.76
Brahmaputra	0.99	0.96	0.99	0.95	0.99	0.93	0.87	0.84
Orinoco	0.99	0.97	0.97	0.95	0.93	0.91	0.79	0.79
Orange	0.65	0.58	0.64	0.57	0.64	0.57	0.64	0.61
Yellow	0.88	0.87	0.87	0.87	0.87	0.87	0.86	0.85
Yukon	0.93	-0.31	0.92	-0.48	0.91	-0.69	0.85	-0.49
Danube	0.89	0.80	0.91	0.83	0.93	0.85	0.95	0.74
Mekong	0.99	0.99	0.99	0.98	0.96	0.96	0.85	0.82
Tocantins	0.96	0.94	0.99	0.97	0.99	0.97	0.96	0.81

Résumé

La constante de temps du débit de base (τ) représente le temps moyen pour que l'eau souterraine arrive à la rivière depuis la zone de recharge dans un bassin donné. C'est un élément clé pour simuler le débit de base dans les modèles simples des eaux souterraines, tels qu'ORCHIDEE. τ a été estimée à l'échelle globale à partir d'une solution de l'équation de Boussinesq pour les aquifères libres en pente. τ dépend de la porosité efficace, de la transmissivité, de la pente de l'aquifère et de la densité de drainage (δ). Calculées à partir de bases de données globales, les valeurs de τ sont surestimées par rapport à celles obtenues par analyse des courbes de récession. Une analyse de sensibilité a montré que la transmissivité et δ sont les principales sources d'incertitude de τ. L'extraction d'un nouveau réseau de drainage, qui dépend de la lithologie, du climat, de la pente et des δ observées, a permis d'obtenir des δ conformes aux valeurs observées aux échelles régionales et à la variabilité spatiale. L'utilisation de ces nouvelles δ et la combinaison de deux jeux de données de conductivité hydraulique pour le sol et l'aquifère a réduit τ de deux ordres de grandeur, mais les valeurs calculées restent surestimées. L'utilisation de τ dans le modèle de surface ORCHIDEE a montré une forte sensibilité du débit simulé à l'augmentation de τ, qui dégrade les débits simulés par rapport aux observations. Cette méthodologie nécessite des valeurs plus adaptées de transmissivité et porosité efficace par rapport aux jeux de données globaux actuellement disponibles pour obtenir des valeurs de τ plus proches de celles attendues et qui permettent de reproduire les débits observés.

Mots-clefs : constante de temps du débit de base, échelle globale, densité de drainage

Abstract The base flow time constant (τ) represents the mean amount of time the groundwater takes to reach the stream from the recharge zone in a given watershed. τ is a key element to simulate base flow in simple groundwater models as ORCHIDEE. τ was estimated at global scale based on a solution of the Boussinesq equation for unconfined sloping aquifers. τ depends on the effective porosity, transmissivity, aquifer slope, and drainage density (δ). When estimated from global available datasets, τ results are overestimated when compared to recession analysis results. A sensitivity analysis showed that transmissivity and δ are the main uncertainty sources of τ. A river network extraction based on lithology, climate, slope, and observed δ allowed to obtain δ values close to reference data and spatially variable at regional scale. The use of a new δ and the combination of two hydraulic conductivity datasets of soil and aquifer reduced τ of two orders of magnitude, however the values remained overestimated. The use of τ in ORCHIDEE land surface model showed a strong sensitivity of the river discharge buffer effect to τ, which worsen simulated river discharge when compared to observations. This methodology needs more adequate porosity and transmissivity values when compared to global available datasets that will result in close results to observed river discharge.

Keywords: base flow time constant, global scale, drainage density