Identification of a lumped, mass-conserving rainfall-discharge model of the Amazon basin for GRACE data assimilation

Karim Douch[1], Peyman Saemian[1], and Nico Sneeuw[1]

[1]University of Stuttgart

November 23, 2022

Abstract

Previous work based on Gravity Recovery and Climate Experiment (GRACE) data has shown that for certain large river basins like the Amazon, the empirical storage-discharge relationship reveals an underlying dynamics that is approximately linear and time-invariant. This is particularly true for the catchment upstream of the Óbidos stream gauge station on the Amazon river. We build on this observation to put forward, in this case, a simple first-order differential equation that approximates the observed dynamics. The model formulation includes one parameter that can be physically interpreted as an offset determining the total drainable water stored in the catchment, while a second parameter characterizes the typical time constant of the draining of the basin. We determine a value of 1925 km^3 for the average total drainable water stored in the catchment during the period 2004 to 2009 and a draining time constant of 27.4 days. The same approach is also tested over eight smaller catchments of the Amazon to investigate whether or not the storage-discharge relationship is governed by a similar dynamics. Combined with the water mass balance equation, we eventually obtain two coupled linear differential equations which can be easily recast into a discrete state-space representation of the rainfall-storage-discharge dynamics of the considered basin. This set of equations is equivalent to defining an analytical instantaneous unit hydrograph for the whole basin. Besides, the proposed model is particularly suitable for Bayesian filtering and smoothing or the reconstruction of past unobserved states.

Identification of a lumped, mass-conserving rainfall-discharge model of the Amazon basin for GRACE data assimilation

K. Douch[1], P. Saemian[1] and N. Sneeuw[1]

[1]Institute of geodesy Stuttgart, University of Stuttgart

Key Points:

- The storage-discharge relationship of the Amazon basin is modelled by a differential equation with physically interpretable parameters
- The model is calibrated and validated using GRACE data and coupled with the water mass balance equation to yield a rainfall-discharge model
- A similar approach is tested over smaller sub-catchments but does not always yield satisfactory results, indicating more complex dynamics

Corresponding author: Karim Douch , `karim.douch@gis.uni-stuttgart.de`

Abstract

Previous work based on Gravity Recovery and Climate Experiment (GRACE) data has shown that for certain large river basins like the Amazon, the empirical storage-discharge relationship reveals an underlying dynamics that is approximately linear and time-invariant. This is particularly true for the catchment upstream of the Óbidos stream gauge station on the Amazon river. We build on this observation to put forward, in this case, a simple first-order differential equation that approximates the observed dynamics. The model formulation includes one parameter that can be physically interpreted as an offset determining the total drainable water stored in the catchment, while a second parameter characterizes the typical time constant of the draining of the basin. We determine a value of 1925 km^3 for the average total drainable water stored in the catchment during the period 2004 to 2009 and a draining time constant of 27.4 days. The same approach is also tested over eight smaller catchments of the Amazon to investigate whether or not the storage-discharge relationship is governed by a similar dynamics. Combined with the water mass balance equation, we eventually obtain two coupled linear differential equations which can be easily recast into a discrete state-space representation of the rainfall-storage-discharge dynamics of the considered basin. This set of equations is equivalent to defining an analytical instantaneous unit hydrograph for the whole basin. Besides, the proposed model is particularly suitable for Bayesian filtering and smoothing or the reconstruction of past unobserved states.

1 Introduction

The Amazon river and its tributaries drain an area of around $5.9 \cdot 10^6\,\mathrm{km}^2$ representing the largest drainage basin in the world and covering about 35% of the South-American continent. With an average flow of around $6.6\cdot 10^3\,\mathrm{km}^3\,\mathrm{yr}^{-1}$ (Sproles et al., 2015; Dai & Trenberth, 2002; Marengo, 2005), it contributes by about 15 to 20% to the world's total freshwater discharge into the oceans (Clark et al., 01 Aug. 2015; Coe et al., 2016), far beyond any other river. The basin is also home to the world's largest tropical rainforest and receives annually around 2100 mm of precipitation across most of the basin, with inter-annual variations between 1000 and 3000 mm and peak values up to 4000 mm in the northwestern part of the basin (de Paiva et al., 2013). The whole watershed can be seen as a moisture sink since the yearly aggregated precipitation over the whole basin largely outbalances evapotranspiration, the excess water being eventually discharged into the ocean. Yet, it accounts for around 15% of global terrestrial evapotranspiration, part of which is recycled as new precipitation over the basin. As such, the Amazon plays a critical role in the global hydrologic cycle as well as in the global atmospheric circulation and regional climate (Malhi et al., 2008; Coe et al., 2016; Jahfer et al., 2017).

Despite this importance, proper quantification of the Amazon basin hydrologic balance remains partly elusive: different estimates of precipitation, evapotranspiration, water storage, and discharge across the basin can show significant discrepancies and, when combined, can fail to close the water mass budget (see equation (1)) within the error bounds (Lehmann et al., 2021). This is particularly true for the evapotranspiration, whose estimates can significantly differ from one data set to another, both in the spatial and temporal domain (Werth & Avissar, 2004; Chen et al., 2020). For an example of a systematic comparison of different estimates of these components for the Amazon basin, the reader is referred to the article by Chen et al. (2020) and Sneeuw et al. (2014); Lorenz et al. (2014) for other river basins.

Our main goal is to propose an approach to refine the discharge and the aggregated terrestrial water storage estimates for the Amazon basin by combining in a Bayesian framework space gravity observations from the Gravity Recovery and Climate Experiment (GRACE) mission with in situ discharge measurements and estimated precipitation and evapotranspiration across the basin. To this aim, we first need to build a rainfall-discharge model of the Amazon. This is the object of this article.

Modelling the water cycle at regional scales relies essentially upon the determination of four quantities: precipitation (P), evapotranspiration (ET), terrestrial water storage (TWS but referred to as S in the equations), which corresponds to the summation over a vertical column of surface and subsurface water storage, and runoff (Q), which represents the surface and subsurface flow of liquid water. These quantities are related to each other via the water mass balance equation equation (1)

$$\frac{dS}{dt} = P - ET - Q \tag{1}$$

which states the conservation of mass. Besides this equation, a crucial step in the proposed approach is to model the coupled dynamics of both terrestrial water storage and discharge. Depending on the spatial and temporal resolution we seek to achieve, the numerical modelling of a real-world basin can quickly become intractable as the involved physical processes can be multi-scale, nonstationary and controlled by complex, spatially heterogeneous features, whose typical size can be orders of magnitude smaller than the catchment. A faithful representation of these processes would therefore require collecting an immense amount of information about the surface and sub-surface spatially distributed properties over the whole basin, which is not conceivable nowadays or in the near future. Consequently, it should be borne in mind that all hydrologic models are inevitable, at some level of detail, a simplified representation of real-world physical processes (Singh & Woolhiser, 2002; Vrugt et al., 2008; Yilmaz et al., 2010).

When only quantities aggregated over the whole basin are considered, the combination of all the hydrologic and hydraulic processes can sometimes result in an apparent simpler behaviour. In such a case, parsimonious heuristic or conceptual models can reproduce the observed dynamics satisfactorily. We use this observation to model the storage-discharge relationship in the large catchment upstream of the Óbidos gauge (see Figure 1). Óbidos (state of Parà, Brazil) is located in the downstream region of the Amazon river at a section where the river has a single and relatively narrow stem. It is approximately 800 km upstream of the river mouth, making the influence of the Atlantic ocean tides on the flow negligible. Still, with an area approximately 4680 000 km^2, the catchment covers 80% of the whole Amazon basin.

After analysing the empirical function $Q = f(\Delta S)$ for the Óbidos catchment, where ΔS is the monthly TWS anomaly estimated from GRACE data, Riegger and Tourian (2014) concluded on the existence of an underlying linear time-invariant (hereafter abbreviated LTI) dynamic system governing the time evolution of these quantities. The plot of the discharge Q as a function of ΔS is reproduced in Figure 2 with updated and extended time series. As already noticed, the observed dynamics exhibit an annual hysteresis cycle demonstrating a dependence of the base flow, not only on the TWS anomaly, but also on its own past values. Besides, it evolves in an anti-clockwise direction, meaning that for a given value of the TWS anomaly, the corresponding discharge is larger when the TWS is decreasing (mostly during the dry season) than during the wet season. We build on these observations to put forward a simple dynamic model relating Q and ΔS in the next sections. Furthermore, we investigate whether such an approach is generalisable and to what extent similar dynamics can be observed in eight smaller catchments of the Amazon basin (see Figure 1) and be modelled in the same way.

The remainder of the article is organised as follows: in section 2 we describe the different data sets needed to quantify the discharge and the aggregated TWS anomaly for the different gauge stations considered in this study. In the next section, we propose and analyse a linear storage-discharge model that will construct the final lumped rainfall-discharge model and discuss its physical interpretation in detail. We also describe the

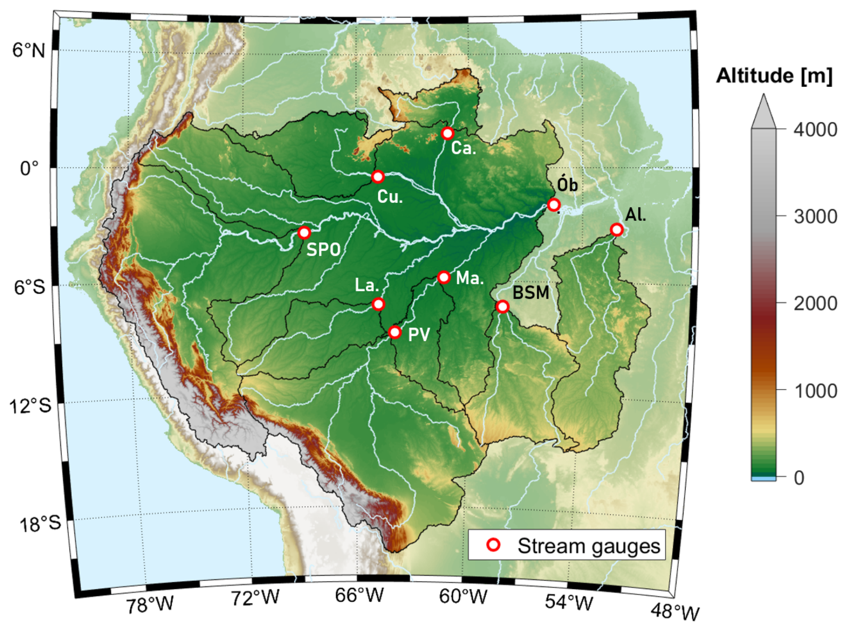

Figure 1. Location of the different stream flow gauges considered in this study and their corresponding upstream catchments (in bright colours). The gauges are Óbidos (Ób.), Caracarai (Ca.), Curicuriari (Cu.), São Paulo de Olivença (SPO), Labrea (La.), Porto Velho (PV), Manicoré (Ma.), Barra de São Manuel (BSM) and Altamira (Al.). Note that the Manicore upstream catchment includes the Porto Velho catchment and the Óbidos upstream catchment encompasses all the other catchments except Altamira and Barra de São Manuel.

method used to calibrate and evaluate the model against observation data. Finally, in section 4, the model fitness for discharge simulation is discussed in light of the calibration results.

2 Data sources and pre-processing

2.1 Discharge records

We use river discharge time series from gauge records provided by the Global Runoff Data Centre (GRDC (*GRDC data archive*, 2021)). Besides the Óbidos gauge, we have selected eight other streamflow gauges located in the Amazon basin fulfilling the following conditions:

- the daily discharge records are available without major data gaps during most of the GRACE mission lifetime, and
- the corresponding upstream catchment has an area sufficiently large to be compatible with the spatial resolution of GRACE gravity field solutions, which is is estimated by Vishwakarma et al. (2018) to be around 63 000 km^2.

The name, location and divide of the upstream catchment of these gauges are provided in Figure 1. The daily data are used to compute the monthly averaged discharge corresponding to the GRACE monthly solutions. A month is discarded if more than 10 data points are missing. As an example, the time series of the discharge at the Óbidos gauge used in Figure 2 is plotted in Figure 3.

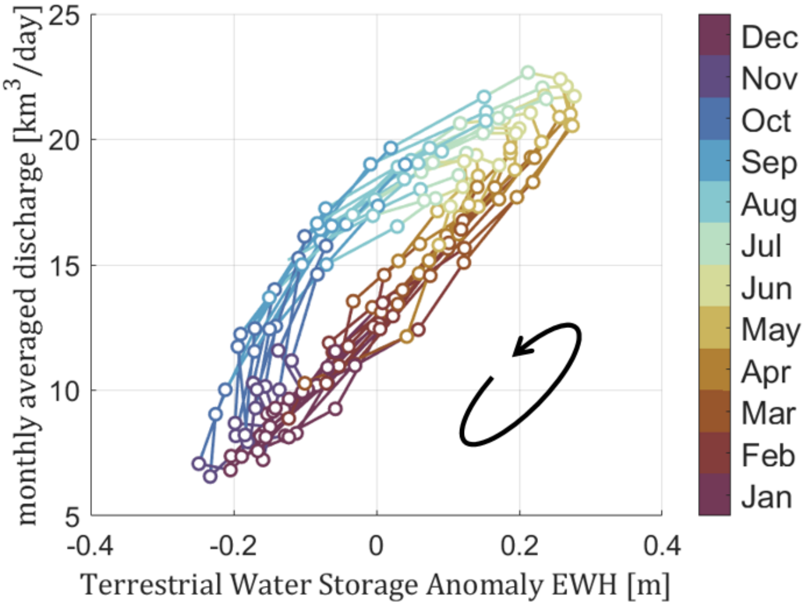

Figure 2. Phase portrait representing the discharge as a function of the aggregated TWS anomaly across the catchment upstream Óbidos (expressed in terms of Equivalent Water Height in meters). The observed dynamics clearly displays a hysteresis cycle in an anti-clockwise direction.

2.2 TWS anomaly from GRACE observations

GRACE was a satellite gravimetry mission consisting of a constellation of two identical satellites launched in March 2002, aiming at mapping the month-to-month Earth's gravity field variations (Tapley et al., 2004, 2019). Until its end in October 2017, GRACE's observation records have been used to estimate 163 monthly solutions of the Earth's gravity field out of 187 months. These monthly models can be inverted to quantify the corresponding mass change on the Earth's surface, which on the continents reflect the TWS redistribution. Yet, it should be borne in mind that the physical quantity related to the TWS estimated with GRACE and GRACE-FO observations is by no mean an absolute quantity but rather an anomaly, that is, the deviation of the current TWS for a given month relative to an unknown reference which is the average total TWS during the period 01.01.2004 to 09.31.2009 (Save et al., 2016). We show in the next section how the proposed methodology can help estimate this unknown constant hereafter denoted S_0. In May 2018, a GRACE Follow-On mission (abbreviated GRACE-FO) was launched to continue the multi-decadal record of the Earth's gravity field variations (Tapley et al., 2019). First analyses show that the quality of released monthly gravity models of GRACE-FO is slightly improved compared to GRACE (Landerer et al., 2020).

Among the various gravity model solutions computed from GRACE data by the three official Science Data System centres, we consider only the latest (RL06 version 2) mascon solutions released by JPL (Wiese et al., 2016) and CSR (Save et al., 2016; Save, n.d.). Mascon solutions offer the possibility to constrain the gravity field solutions with prior information drawn from geophysical models, preventing, in particular, the apparition of north-south stripes degrading the classical unconstrained spherical harmonic solutions (Watkins et al., 2015). In addition, the mascon solutions are less damaged by leakage error and compare better to in situ data (Landerer et al., 2020). Besides the data gap of 11 months between GRACE and GRACE-FO, GRACE data after August 2016

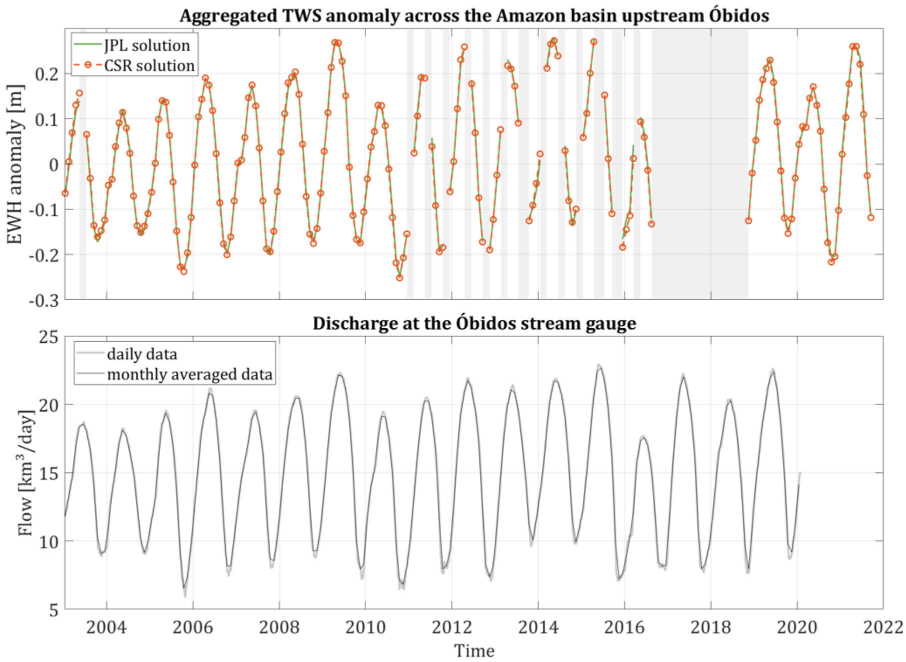

Figure 3. Top: Comparison of the TWS anomaly expressed in terms of Equivalent Water Height (EWH) aggregated over the Óbidos watershed derived from GRACE data and computed from CSR and JPL mascons solutions. The grey patches correspond to data gaps. Bottom: daily (in light grey) and monthly averaged discharge observed at Óbidos from January 2003 to January 2020.

have been purposely discarded since the raw data collected by the two satellites were much more degraded at the end of the mission. As a consequence, only the data from January 2003 to August 2016 are used for the model calibration and validation since a time series with a constant sampling period is needed. Small data gaps during this period are filled in by spline interpolation. Comparing the aggregated TWS anomaly across the Óbidos catchment computed from the CSR and JPL solutions shows no significant difference (see Figure 3). As a consequence, we only use the JPL mascon solution. Based on this dataset and the discharge records, the phase portrait for the eight other flow gauges is computed and plotted in Figure 4.

3 Storage-discharge relationship in the Óbidos Catchment

This section aims to show that for monthly-averaged values, the storage-discharge relationship in the case of the catchment upstream of the Óbidos gauge station on the Amazon river is well approximated by a first-order ordinary differential equation, whose parameters can be estimated from data. For the remainder of this article, all the quantities such as ΔS or Q are considered as monthly averaged, continuous-time variables.

3.1 A dynamic model

Following the investigations of Riegger and Tourian (2014) on global scale storage-discharge relationships for large drainage basins, Tourian et al. (2018) proposed a basic storage-discharge model for the case of the Óbidos upstream catchment and used it to determine the total drainable water storage (TDWS) in the catchment. They define this quantity as "the total stored water that can exit or drain the river basin through

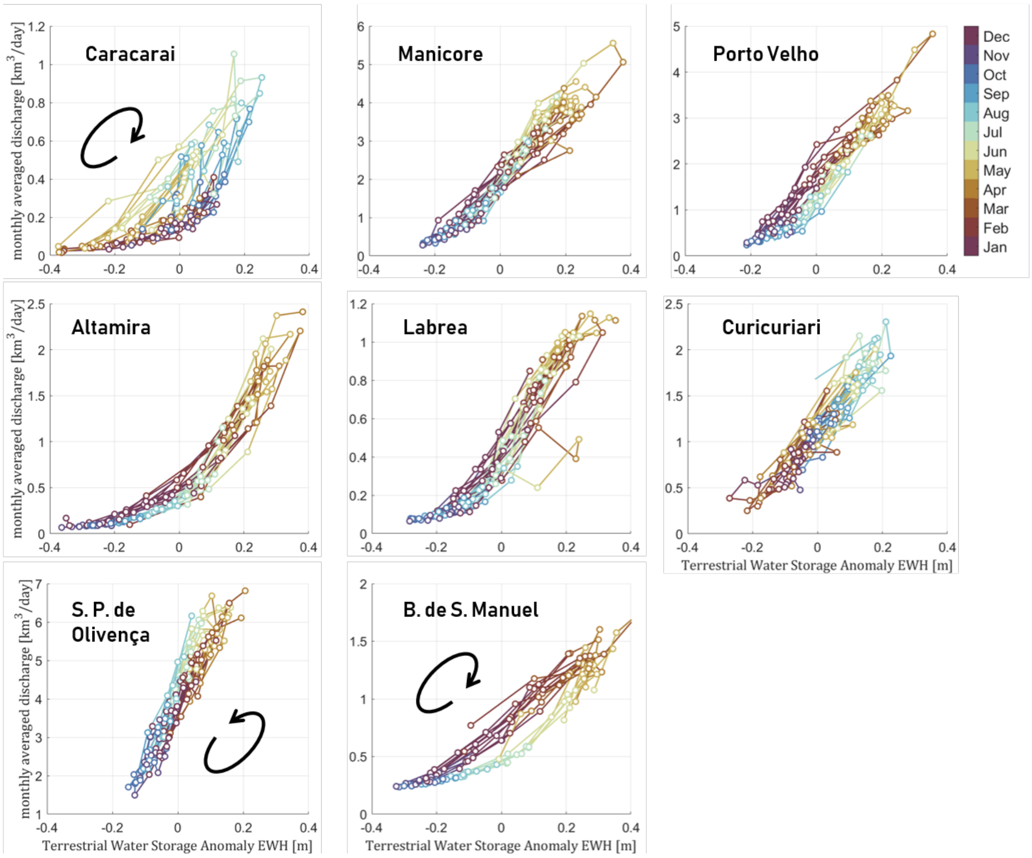

Figure 4. Phase portrait representing the discharge as a function of the aggregated TWS anomaly across the eight different catchments considered besides Óbidos. When a hysteresis cycle is clearly visible and the direction of cycling unambiguous, a black arrow indicates the direction of evolution.

natural hydrologic runoff generation as the time approaches infinity, [..] given no additional inputs". The model has the form

$$Q(t) = \frac{1}{\theta} S(t + \Delta t) = \frac{1}{\theta}(S_0 + \Delta S(t + \Delta t)) \tag{2}$$

where θ is a time constant and Δt is a time shift. This equation assumes that the waveform of the TDWS time series S is identical to the waveform of the base flow Q, except for scaling and a shift in time by Δt. This seems to be an acceptable approximation since the temporal evolution of both quantities is largely dominated by a smooth seasonal variation (see Figure 3).

The core idea of our approach originates from the observation that a hysteresis cycle in an LTI system can rather be approximated by a model as simple as a first-order ordinary differential equation (abbreviated hereafter ODE). We, therefore, propose the following continuous-time ODE to model the relationship between Q and TDWS S_d:

$$\frac{dQ}{dt} + \frac{Q}{\tau} = \omega_{\mathrm{n}}^2 S_d \tag{3}$$

where τ is a time constant while ω_n^2 is a squared frequency, note that τ must be positive in order to keep the system stable. A more detailed physical interpretation of these parameters is given in the next section. This model is totally heuristic in the sense that it is not derived from first principles but corresponds rather to an approximate and parsimonious representation of the dynamic system. Similarly to equation (2), We can adapt equation (3) to the water storage observations derived from GRACE data and decompose the terrestrial water storage as $S_d(t) = S_0 + \Delta S(t)$ where $\Delta S(t)$ is the observed monthly anomaly with respect to an unknown volume of water S_0:

$$\frac{dQ}{dt} + \frac{Q}{\tau} = \omega_{\mathrm{n}}^2(\Delta S + S_0) \tag{4}$$

An implicit yet fundamental assumption made in the previous equation is that the variations of the TWS anomaly observed by GRACE correspond to the variations of the quantity of water available for drainage: $\Delta S = \Delta S_d$. In other words, the TWS anomaly is assumed to only reflect water storage connected in one way or another to the drainage system. As such, no precipitation water remains indefinitely in storage disconnected from the river network.

3.2 Physical interpretation of the model

Provided that equation (4) is a reasonable model of the observed dynamic, one can get a more insightful interpretation of the three unknown parameters τ, ω_n^2 and S_0 by considering, in addition, the water mass balance equation (1). Let us assume that there are no underground outflow or secondary arms of the river draining water out of the catchment. It follows that the spatial integral over the whole catchment of the runoff is equal to the discharge Q at the outlet in Óbidos. We can, therefore, couple equation (4) and the mass balance equation (1) which remains identical for monthly averaged quantities and where the time derivative of S is replaced by the time derivative of ΔS. These two coupled first-order ODEs can be recast into a linear state-space representation:

$$\frac{d}{dt}\begin{pmatrix} Q \\ \Delta S \end{pmatrix} = \underbrace{\begin{pmatrix} -\frac{1}{\tau} & \omega_{\mathrm{n}}^2 \\ -1 & 0 \end{pmatrix}}_{\mathbf{A_c}} \underbrace{\begin{pmatrix} Q \\ \Delta S \end{pmatrix}}_{\mathbf{X}} + \underbrace{\begin{pmatrix} \omega_n^2 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}}_{\mathbf{B_c}} \underbrace{\begin{pmatrix} S_0 \\ P \\ ET \end{pmatrix}}_{\mathbf{u}} \tag{5}$$

where for brevity we note the state vector $\mathbf{X}$, the system matrix $\mathbf{A}_c$, the input matrix $\mathbf{B}_c$ and the input $\mathbf{u}$ as indicated in equation (5). For completeness, we can adjoin an observation equation

$$\mathbf{Y} = \mathbf{C}_c\mathbf{X} \tag{6}$$

describing whether Q, ΔS or both are observed. Note that we append an index $\mathbf{c}$ to all the matrices names to stress that they describe continuous-time dynamics. The continuous-time state-space representation ((5), (6)) constitutes a deterministic and complete description of the dynamics of the basin aggregated state. In particular, the integration of the model equations given an initial condition $\mathbf{X}(t_0)$ and the input time series $\mathbf{u}$ yields the solution:

$$\mathbf{Y}(t) = \mathbf{C_c}e^{\mathbf{A_c}(t-t_0)}\mathbf{X}(t_0) + \mathbf{C_c}\int_{t_0}^{t} e^{\mathbf{A_c}(t-\tau)}\mathbf{B_c}\mathbf{u}(\tau)d\tau \tag{7}$$

Together with a characterization of the observation and process noise, the system ((5), (6)) would form a stochastic model perfectly amenable to Kalman filtering and smoothing and in this way, data assimilation. Yet, the structure of the model must still be validated and the parameters estimated.

When $\mathbf{C_c} = (1\ 0)$, the model equations boil down to a rainfall-discharge model of the basin. Alternatively, one can derive the direct rainfall-discharge equation by taking the time derivative of equation (4) and replacing the time derivative of ΔS by $P - ET - Q$. Reformulated in a canonical form, the equation reads

$$\frac{1}{\omega_{\mathrm{n}}^2}\frac{d^2Q}{dt^2} + \frac{2\xi}{\omega_{\mathrm{n}}}\frac{dQ}{dt} + Q = P - ET \tag{8}$$

where

$$\xi = \frac{1}{2\omega_{\mathrm{n}}\tau}$$

A similar differential equation can be derived for the evolution of ΔS.

Equation (8) corresponds to the well-known linear system described by a second-order ODE, ubiquitous in engineering and a fundamental control theory textbook case. The parameter ξ is unitless and is usually called the *damping ratio*. Here again, the stability of the system described by equation (8) requires ξ to be non-negative, that is, $\omega_{\mathrm{n}} > 0$. The parameter ω_{n} is the *natural frequency* of the system, that is, the frequency at which the system would oscillate if the damping modelled by the term $1/\tau$ was zero (in other words, τ was infinite). Now, if we assume that after time $t = t_0$ $P - ET = 0$ and providing that the initial values $Q(t_0)$ is not null, the solution of equation (8) can be classified into four categories inspired by the mechanical analogue of the spring-mass-damper system

- the "undamped" case corresponding to $\xi = 0$: the solution general form is $S(t) = c_1 \sin(\omega_{\mathrm{n}}t) + s_2 \cos(\omega_{\mathrm{n}}t)$ where c_1 and s_1 are integration constants
- the "underdamped" case for $0 < \xi < 1$: the solution is an exponentially attenuated oscillations of the form $S(t) = e^{-\sigma t}(s_1 \sin(\omega_{\mathrm{d}}t) + c_1 \cos(\omega_{\mathrm{d}}t))$ where $\sigma = \xi\omega_{\mathrm{n}}$ is the attenuation and $\omega_{\mathrm{d}} = \omega_{\mathrm{n}}\sqrt{1-\xi^2}$ is the damped natural frequency
- the "critically damped" case corresponding to $\xi = 1$: the general solution form is $S(t) = (s_1 + c_1 t)e^{-\omega_{\mathrm{n}}t}$
- the "overdamped" case ($\xi > 1$): the solution has the general form $S(t) = c_1 e^{-\frac{t}{T_1}} + c_2 e^{-\frac{t}{T_2}}$ where $1/T_1 = (\xi - \sqrt{\xi^2 - 1})\omega_{\mathrm{n}}$ and $1/T_2 = (\xi + \sqrt{\xi^2 - 1})\omega_{\mathrm{n}}$.

The two first cases must be a priori discarded as we would expect the gravity-driven drainage of a basin where no precipitation takes place to decrease the amount of stored water monotonically, without any oscillation or overshot. We would therefore expect from the calibration results that at least $\xi \geq 1$ that is

$$2\omega_{\mathrm{n}}\tau \leq 1 \tag{9}$$

In this case, the solution converges asymptotically and monotonically from $Q(t_0)$ to zero. Similarly, the storage anomaly tends monotonically from $S_0 + \Delta S(t_0)$ toward zero, validating the interpretation of S_0 in equation (4) as an offset which, added to ΔS, forms the TDWS.

It is worth noticing that in terms of rainfall-discharge relationship, the proposed model is equivalent to a generalized Nash model (J. E. Nash, 1957; Lee & Singh, 1998) consisting of a cascade of 2 linear reservoirs with different storage characteristics $1/\tau_1$ and $1/\tau_2$ as illustrated in Figure 5. The parameters of the suggested model would then correspond to: $\omega_n^2 = \frac{1}{\tau_1\tau_2}$ and $\tau = \frac{\tau_1\tau_2}{\tau_1+\tau_2}$. By construction, the damping ratio $\xi = \frac{\tau_1+\tau_2}{2\sqrt{\tau_1\tau_2}}$ of the Nash model is always larger or equal to 1 in virtue of the inequality of arithmetic and geometric means and equal to 1 when $\tau_1 = \tau_2$. This is to be contrasted with our model, for which there is no guarantee that equation (9) is satisfied unless the calibration is explicitly constrained with this inequality. As such, our model can potentially yield unrealistic output. Yet, if the goal of the model calibration is to minimize the prediction error, it might be relevant not to constrain the estimation of the parameters with such an inequality and give more flexibility to the model to fit the data at the price of lower physical interpretability.

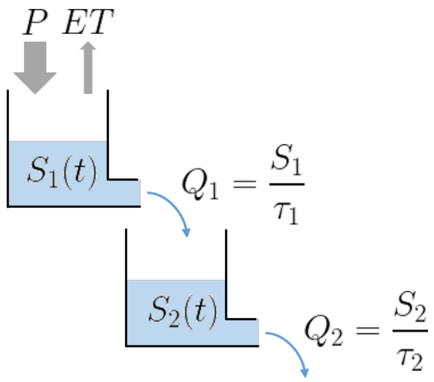

Figure 5. Generalized Nash model consisting of a cascade of 2 linear reservoirs.

Finally, it is noteworthy that solving equation (8) for a Dirac impulse as input ($P = \delta(t)$, $ET = 0$) provides an analytical expression $h(t)$ of the Instantaneous Unit Hydrograph (IUH) (J. E. Nash, 1957) :

$$h(t) = \frac{\omega_n}{2\sqrt{\xi^2 - 1}} \left(e^{-t/T_1} - e^{-t/T_2} \right) \tag{10}$$

where T_1 and T_2 were defined hereinabove. The existence of a "reasonable" IUH for such a catchment as vast as the Óbidos upstream catchment may look surprising at first sight but it is actually a direct consequence of approximating the storage-discharge dynamics as an LTI system. However, this IUH must be considered with caution since a usual assumption when using the unit hydrograph theory for flow prediction is that the precipitations occur uniformly across the whole area, which is obviously not the case for the Amazon basin.

3.3 Model calibration and evaluation

The next step requires to identify the three parameters τ, S_0 and ω_n^2 of the linear time-continuous (hereafter abbreviated CT) model (4) from discrete-time series data. The advantages of estimating directly a CT model are manifold. Besides the fact that most physical systems are by nature CT, one key advantage in our case is that the estimated parameters have a physical meaning, as discussed in the previous section, making easier their comparison to reasonable values for falsification. Conversely, discrete model parameters depend on the sampling time and it is not always clear how they relate to true physical parameters. An extensive discussion on the advantages of direct CT model identification from sampled data can be found in Garnier and Young (2012) and Garnier et al. (2008). In this study, we use the popular *Simplified Refined Instrumental Variable* method for *Continuous-time* system (hereafter abbreviated SRIVC method) as implemented in the CONTSID toolbox for MATLAB™ (Garnier & Gilson, 2018). SRIVC, as its name would suggest, is a simplified version of the RIVC method developed by P. Young and Jakeman (1980), which estimates recursively the parameters of a CT model in differential equation form along with the parameters of a discrete-time autoregressive moving average process associated to the additive coloured noise corrupting the output measurements. In the SRIVC method the additive noise is assumed to be purely white, allowing to ignore the ARMA process identification and simplifying greatly the algorithm. While both RIVC and SRIVC estimators are under mild conditions consistent and asymptotically unbiased (Pan et al., 2020), the SRVIC loses the property of minimum variance estimates (Garnier, 2011). An in depth description of the SRIVC method is beyond the scope of this article and the reader is referred to Garnier (2011), P. C. Young (2002), P. C. Young (2008) and the prior references therein for further details.

We have divided the available time series into two periods: the first, from January 2003 to December 2010, is used for the model estimation, whereas the second, from January 2011 to August 2016, is used for validation. Only 1 data point out of 96 has been interpolated for the calibration period, whereas they are 15 out of 68 for the validation period.

Although the continuous-time model parameters are directly estimated from discrete data, it is necessary to discretize the ODE equation (4) at the same sampling period Δt as the data at hand for simulation purposes. This requires in particular to model the inter-sample behaviour of the input signal $\Delta S(t)$. In our case, it is sufficient to assume a piecewise linear behaviour that is

$$\Delta S(t) = \Delta S(t_n) + \frac{\Delta S(t_n + \Delta t) - \Delta S(t_n)}{\Delta t}(t - t_n)$$

for $t \in [t_n, t_n + \Delta t]$, with $t_n = t_0 + n\Delta t$, $n \in \mathbb{N}$. Under this approximation, the exact discrete state-space representation of the model becomes:

$$Q(k+1) = e^{-\frac{\Delta t}{\tau}} Q(k) + C_0 S_0 + C_1 \Delta S(k) + C_2 \Delta S(k+1) \tag{11}$$

where the coefficients C_0, C_1 and C_2 are given in Appendix A and the time notation t_k is abbreviated k for simplicity.

In order to quantify the simulation accuracy of the models, we use the following indices:

- the root-mean-square error (RMSE), which is simply the square root of the mean of the squares of the deviations of the simulated data with respect to the observed data:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Q_{\text{obs}}(i) - Q_{\text{sim}}(i)\right)^2} \tag{12}$$

  The lower the RMSE, the better the model simulation performance with the limit RMSE $= 0$ in the case where the observed hydrograph is exactly reproduced.
- the Nash-Sutcliffe efficiency (NSE) (J. Nash & Sutcliffe, 1970) is a normalized statistic that characterizes the relative magnitude of the simulation error with respect

to the observed signal variance:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N} \left(Q_{\mathrm{obs}}(i) - Q_{\mathrm{sim}}(i)\right)^2}{\sum_{i=1}^{N} \left(Q_{\mathrm{obs}}(i) - Q_{\mathrm{mean}}\right)^2} \tag{13}$$

NSE varies between $-\infty$ and 1 with a value of 1 corresponding to a perfect reproduction of the observed hydrograph. A negative value indicates that the observed discharge mean value is a better predictor than the model. Conversely, a positive value indicates a better performance of the model. According to Moriasi et al. (2007), a model can be regarded as satisfactory if NSE > 0.5.

Since the discharge of most Amazon sub-basins is dominated by a strong seasonal signal, it is more pertinent to characterize the performance of the simulation by computing a modified version of the NSE called the cyclostationary NSE and expressed as follows:

$$\mathrm{CSNSE} = 1 - \frac{\sum_{i=1}^{N} \left(Q_{\mathrm{obs}}(i) - Q_{\mathrm{sim}}(i)\right)^2}{\sum_{i=1}^{N} \left(Q_{\mathrm{obs}}(i) - Q_{\mathrm{month}}(i)\right)^2} \tag{14}$$

where Q_{month} is the 12-month periodic signal with a cycle consisting for each month of the multi-year mean discharge value of this month. Similarly to the NSE, a positive value of the CSNSE indicates a better ability of the model to capture the annual and interannual variability than the mean annual cycle.

4 Results and validation of the model

4.1 Estimated parameters and evaluation of the models

The estimated parameter values and the simulation evaluation statistics for the 9 sub-basins are summarized in Table 1 and 2, respectively. The simulated discharges are plotted in Figure 6 for both the estimation and validation datasets. We can first notice that, although the condition $\xi \geq 1$ was not enforced in the parameter estimation process, it is verified by all catchments. The simulation statistics are rather satisfactory for most catchments, with a positive CNSE for six of them when tested on the validation dataset. For these cases, it is therefore safe to assert that the storage controls mostly the discharge dynamics and a significant part of this dynamics is captured by the proposed model. This is particularly true for the Óbidos and São Paulo de Olivença catchments for which a CNSE larger than or very close to 0.5 is achieved with the training and validation data. It is notable that they correspond to the two largest catchments analyzed here.

With a negative NSE for both the estimation and validation dataset, the proposed model is clearly not able to capture the discharge dynamics at Caracarai, leading to unrealistic parameter values like for instance for τ. We conjecture that the small size of the corresponding catchment combined with the inevitable leakage error degrade significantly the estimation of the aggregated TWS anomaly derived from GRACE data and thus the model parameters. In the case of Barra de São Manuel, we observe a time delay of the simulated discharge with respect to the ground truth, leading also to poor performance statistics. These poor performances for both catchments could actually have been anticipated from the observation of their respective phase portrait. The solution Q for a

Table 1. Estimated parameter of equation (5) for the different catchments . The minimum offset value S_0^{min} discussed in section 4.2 is also given for comparison.

catchment	τ (days)	ω_n^2 (days^{-2})	S_0 (km^3)	S_0^{min} (km^3)	ξ
Caracarai	544	$2.06 \cdot 10^{-5}$	26.7	56.8	2.0
Barra de São Manuel	17.4	$3.57 \cdot 10^{-4}$	114.5	143.2	1.5
Porto Velho	8.20	$8.90 \cdot 10^{-4}$	208.9	235.2	2.0
Altamira	5.60	$1.20 \cdot 10^{-3}$	106.7	216.0	2.6
Labrea	0.19	$4.88 \cdot 10^{-2}$	53.9	70.3	11.9
Curicuriari	2.61	$5.60 \cdot 10^{-3}$	77.1	87.4	2.6
São Paulo de Olivença	7.65	$2.20 \cdot 10^{-3}$	239.1	192.2	1.4
Óbidos	27.4	$2.75 \cdot 10^{-4}$	1925.0	1739.0	1.1
Manicoré	3.15	$2.40 \cdot 10^{-3}$	268.9	311.8	3.2

Table 2. Simulation evaluation statistics for both the period of estimation (est. data) and of validation (val. data).

catchment	est. data			val. data		
	RMSE (km^3)	NSE	CSNSE	RMSE (km^3)	NSE	CSNSE
Caracarai	0.25	-0.07	-3.11	0.20	-0.04	-1.20
Barra de São Manuel	0.26	0.60	-3.28	0.24	0.68	-1.05
Porto Velho	0.34	0.88	-0.41	0.41	0.86	0.32
Altamira	0.30	0.79	-0.21	0.27	0.81	-0.28
Labrea	0.09	0.94	-0.50	0.15	0.84	0.32
Curicuriari	0.17	0.86	0.48	0.19	0.85	0.33
São Paulo de Olivença	0.35	0.93	0.64	0.52	0.88	0.46
Óbidos	0.72	0.97	0.72	1.10	0.94	0.63
Manicoré	0.28	0.96	0.23	0.35	0.93	0.66

stable model given by Equation (4) is necessarily delayed with respect to the dynamic input ΔS, which results in a counter-clockwise cycling of the phase portrait $Q = f(\Delta S)$. Yet, as noticed in Figure 4, Caracarai and Barra de São Manuel clearly exhibit a clockwise cycling. If we remain within the framework of a lumped storage-discharge relationship modelled by a first order ODE, this leads to the conclusion that the discharge is actually driving the storage and not the inverse like for Óbidos.

4.2 Lower bound of S_0

The total volume of water participating dynamically to the water cycle and temporarily stored in large basins is difficult to quantify from ground measurements and has anyway received little attention (Riegger, 2020). Tourian et al. (2018) obtain a value of 1766km^3 for S_0, which is very similar to our result. This is not surprising since their estimate relies on the identification of the parameters of equation equation (2), which is a actually the solution of equation equation (4) for a sinusoidal ΔS. Besides, we can, with the help of GRACE observations, set a lower bound for the value of S_0 by considering the following argument. To avoid confusion, we note the field associated to an aggregated quantity with the corresponding lower case letter. For instance, $s_\text{d}(\mathbf{r}, t)$ is the field of total drainable water storage defined at any point $\mathbf{r}$ across the catchment and at any time t. Expressed in terms of Equivalent Water Height, the field $s_\text{d}(\mathbf{r}, t)$ is naturally a positive or null quantity (there is no such thing as a negative volume of water). Furthermore, the spatially distributed offset $s_0(\mathbf{r})$, whose aggregated value is S_0, is by definition independent of time so that we can write $s_\text{d}(\mathbf{r}, t) = \Delta s(\mathbf{r}, t) + s_0(\mathbf{r})$ and thus $s_0(\mathbf{r}) \geq -\Delta s(\mathbf{r}, t)$ for any time time t. In particular, this means that the spatially distributed

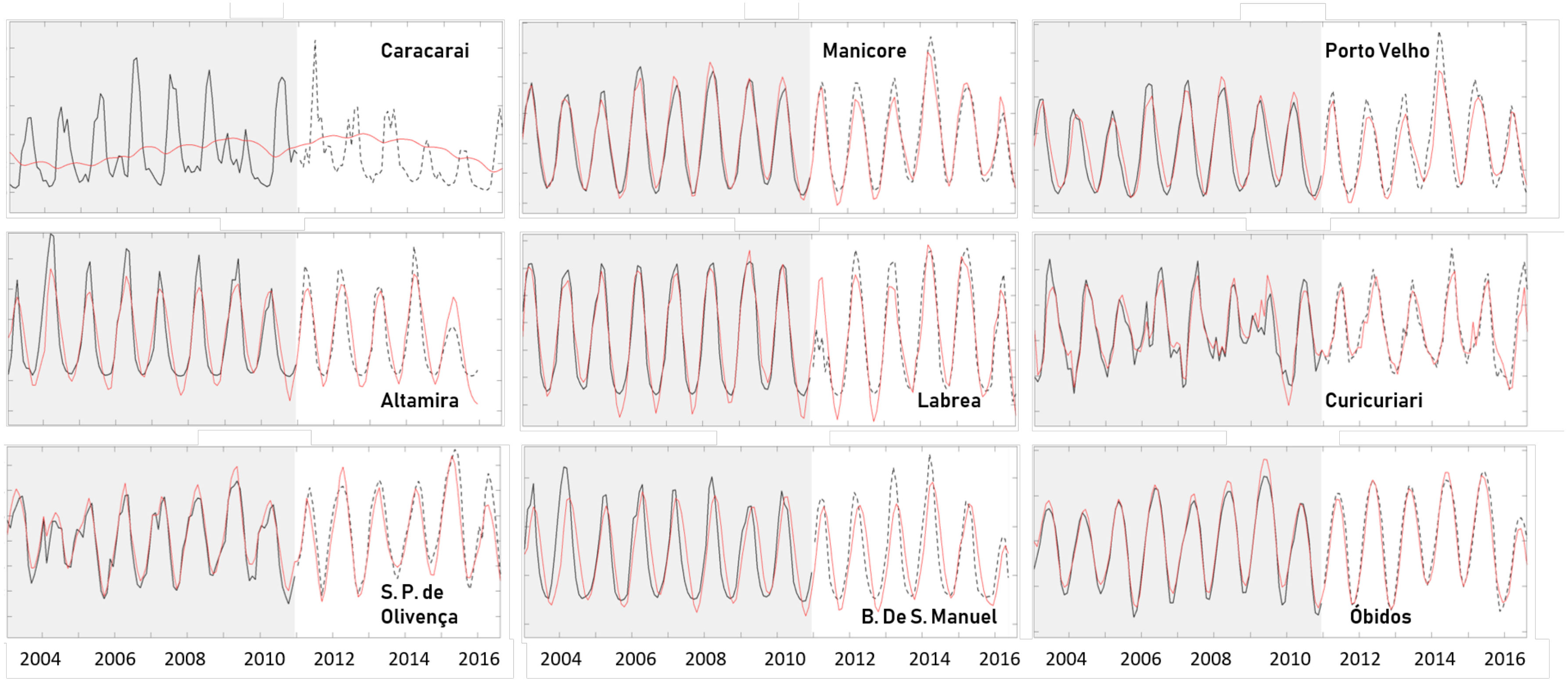

Figure 6. Comparison of the observed discharge (black line) and the simulated one (red line) during both the estimation period (grey patch) and the following validation period.

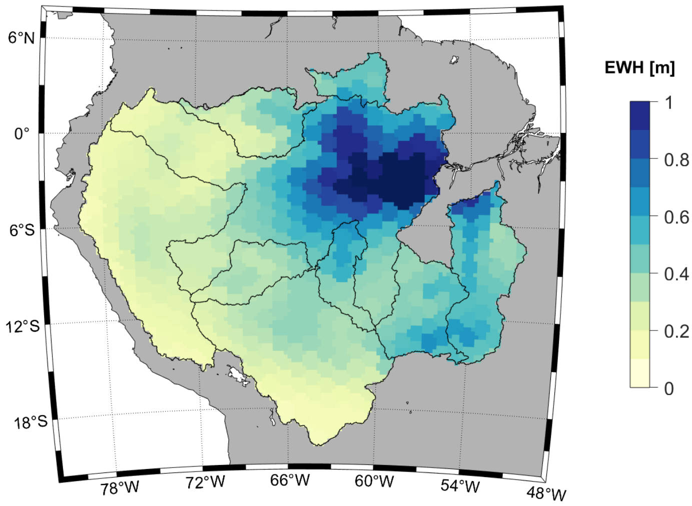

Figure 7. Map of the minimum positive offset to add to the TWS anomaly field estimated from GRACE and GRACE-FO (CSR mascon solution), to ensure a positive equivalent water height during both missions lifetime. The spatial integral of this field corresponds to the integral in equation (15).

offset must be everywhere at least larger than the largest negative variation Δs observed by GRACE and GRACE-FO. A map of this minimum positive offset is plotted in Figure 7. A direct consequence of this result is that S_0 must satisfy

$$S_0 \geq S_0^{\min} = -\iint_{\text{basin}} \min(\Delta S) d\Sigma \tag{15}$$

The latter inequality (15) is only verified by our estimate of S_0 for Óbidos and São Paulo de Olivença confirming again that the proposed lumped model is relevant in these cases and captures correctly the observed dynamics. It should be noticed, however, that the minimum TWS anomaly is not reached everywhere at the same time. The minimum aggregated ΔS observed by GRACE and GRACE-FO over the Óbidos watershed is actually $-1181\,\text{km}^3$ in October 2010. Conversely

5 Discussion and conclusion

In this contribution, we have further developed the idea that the relationship between TWS anomaly and discharge in the Óbidos upstream catchment can be reasonably modelled as an LTI system (Tourian et al., 2018). Assuming that the discharge is primarily driven by the total drainable water storage, we have modelled the storage-discharge dynamics by a first-order ODE, which requires adjusting only three parameters. The data-driven estimation of these parameters has been carried out using the SRIVC method. With an NSE = 0.94 and a CNSE = 0.63 over the validation data, the simulated discharge for Óbidos shows a good agreement with in situ data. Provided the proposed model captures correctly the global dynamics of storage and discharge, a byproduct of the calibration is an estimate of the average total volume of drainable water stored in the Óbidos catchment during the calibration period from January 2003 to December 2010. This volume corresponds to an equivalent water height of 41 cm covering the whole catchment.

Another physically interpretable parameter estimated in the calibration is the time constant of 27.4 days characterizing the exponential decay of the drainage.

By coupling the storage-discharge equation to the water mass conservation equation, we eventually obtain a system of two ODEs that describes the rainfall-discharge dynamics at the basin scale in a consistent manner. As such and despite the large area covered by this catchment, this makes classical hydrology tools such as the IUH still relevant. However, rather than estimating an IUH directly, we advocate the identification of a continuous-time ODE relating discharge to TWS anomaly. The proposed approach offers the advantage to keep the model parameters physically interpretable. Furthermore, it is naturally formulated in a state-space representation that can be exactly discretized and which gives the possibility to apply filtering (resp. smoothing) techniques such as the Kalman filter (resp. smoother) for the optimal estimation of the discharge, TWS anomaly and their respective uncertainty.

The proposed heuristic model relies on a few assumptions that can potentially limit its generalization to other drainage basins:

1. *The TDWS anomaly is equal to the TWS anomaly observed by satellite gravimetry.* This seems to be the case for the Amazon basin where no significant trend in the TWS anomaly field is observed. This may not be the case for other catchments where the variations of TWS could be due for instance to the depletion of groundwater caused by human activities or conversely, the permanent storage of water in man-made reservoirs.
2. *The discharge is driven by the TDWS.* Clearly, this is not the case for the Caracarai and the Barra de São Manuel catchments. For them, and given the data at hand, it is rather correct to say that the TWS anomaly is driven by the discharge dynamics, as the clockwise direction of their respective phase portrait suggests.
3. *The model is time-invariant.* In the proposed model, all the parameters are constant. This may not be true in general (Heerspink et al., 2020). For instance, it has been observed recently that ongoing deforestation and, more generally, changes in land cover and land use alter the partition between evapotranspiration and runoff in favour of the latter (Baidya Roy & Avissar, 2002; Coe et al., 2011, 2017). We should therefore regard the time-invariance of the suggested model as a satisfactory approximation for the period of study rather than the mathematical formulation of inherent stationarity of the observed system.
4. *The observed discharge constitutes the only outflow from the catchment.* This assumption needs to be qualified and quantified. In (Chen et al., 2020) the authors investigated the difference between the in situ discharge observation at the basin outlet and the total runoff estimated as a residual of the water mass budget closure, for which satellite gravity measurements and independent precipitation and evapotranspiration data are combined. While both flow rates estimates are inevitably contaminated with errors, they argue that the latter is more reliable than the former. As a consequence, they hypothesize that the discharge observed at the stream gauge during the wet season is probably underestimated as the water may overflow the riverbanks and surrounding floodplains, creating temporary drainage channels which are not accounted for (Chen et al., 2020; Eom et al., 2017). In addition, they recall that while the stream gauge measures the total surface runoff, the indirect method based on the closure of the water mass balance estimates total runoff, which includes a possible subsurface runoff. In (Chen et al., 2020), the greater yearly accumulated runoff derived from water mass budget closure has been interpreted by the authors as a confirmation of the existence of unobserved groundwater flows to the ocean underneath the Amazon river, as hypothesized by Pimentel and Hamza (2012) following geothermal studies. They estimated this flow rate to be 2% of the observed surface river discharge.

Finally, we have partly omitted an important step of the modelling process, which is the choice of the model structure and thereby the number of adjustable parameters. To promote parsimony, we have prescribed in this article a first-order ODE to represent the TWS-discharge dynamics. However, it may not be the most appropriate order. To go even further, one can drop the linearity approximation: in the case of Óbidos, the proposed model performs in general badly when the discharge reaches its yearly maximum, meaning that it fails to capture the real dynamics at work. A more appropriate model would probably distinguish two different dynamics: a linear one as suggested in this article when TWS is below a certain threshold and a second, non-linear one above this threshold, in which the right-hand side of equation (4) is replaced by a saturation function of $(S_0{+}\Delta S)$. The identification of such a non-linear function will be the object of future work.

Appendix A Exact discretization of the ODE

The general solution of equation (4) between time $t_0 = k\Delta t$ and $t = (k+1)\Delta t$ is given by

$$\begin{aligned} Q((k+1)\Delta t) =& e^{-\frac{\Delta t}{\tau}} Q(k\Delta t) \\ &+ \omega_n^2 \int_{k\Delta t}^{(k+1)\Delta t} e^{-\frac{(k+1)\Delta t - v}{\tau}} (S_0 + \Delta S(v)) dv \end{aligned}$$

where v is a dummy integration variable. If we consider a piecewise linear behaviour of the input $\Delta S(t)$ than the integral in the equation hereinabove reduces to the sum $C_0 S_0 + C_1 \Delta S(k) + C_2 \Delta S(k+1)$ where

$$\begin{aligned} C_0 &= \omega_n^2 \tau (1 - e^{-\frac{\Delta t}{\tau}}) \\ C_1 &= \omega_n^2 \tau (\tau \frac{1 - e^{-\frac{\Delta t}{\tau}}}{\Delta t} - e^{-\frac{\Delta t}{\tau}}) \\ C_2 &= \omega_n^2 \tau (1 - \tau \frac{1 - e^{-\frac{\Delta t}{\tau}}}{\Delta t}) \end{aligned}$$

Appendix B Open Research

The CSR_GRACE/GRACE-FO_RL06_v02 (respectively RL06M.MSCNv02) mascon solutions derived from GRACE and GRACE Follow-On observations by the CSR (respectively JPL) processing centre and used to compute the monthly, basin-aggregated terrestrial water storage anomaly are available at `http://www2.csr.utexas.edu/grace` or via `dx.doi.org/10.15781/cgq9-nh24` (resp. `http://grace.jpl.nasa.gov` or via `dx.doi.org/10.5067/TEMSC-3MJC6`). In both cases, we used the data with all corrections applied.

Daily discharge data at the 9 flow gauges considered in this study along with the boundaries of their corresponding upstream catchment are made available by The Global Runoff Data Centre (GRDC), 56068 Koblenz, Germany via `https://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/riverdischarge_node.html`.

The continuous-time system identification (CONTSID) toolbox version 7.4 used to build a continuous-time dynamic model of the storage-discharge relationship can be downloaded via `http://www.contsid.cran.univ-lorraine.fr/`. The CONTSID toolbox is run with MATLAB™ and requires in addition the MATLAB Control and System Identification toolboxes.

Maps were plotted with MATLAB™ and the mapping package for MATLAB™ M_Map, version 1.4m, from Pawlowicz, R., 2020, available online at `www.eoas.ubc.ca/~rich/map.html`.

Acknowledgments

The authors are grateful to Mohammad Javad Tourian from the Institute of Geodesy, University of Stuttgart, for insightful comments.

References

Baidya Roy, S., & Avissar, R. (2002). Impact of land use/land cover change on regional hydrometeorology in Amazonia. *Journal of Geophysical Research: Atmospheres*, *107*(D20), LBA 4-1-LBA 4-12. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD000266` doi: https://doi.org/10.1029/2000JD000266

Chen, J., Tapley, B., Rodell, M., Seo, K.-W., Wilson, C., Scanlon, B. R., & Pokhrel, Y. (2020). Basin-scale river runoff estimation from grace gravity satellites, climate models, and in situ observations: A case study in the amazon basin. *Water Resources Research*, *56*(10), e2020WR028032. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR028032` (e2020WR028032 2020WR028032) doi: https://doi.org/10.1029/2020WR028032

Clark, E. A., Sheffield, J., van Vliet, M. T. H., Nijssen, B., & Lettenmaier, D. P. (01 Aug. 2015). Continental runoff into the oceans (1950-2008). *Journal of Hydrometeorology*, *16*(4), 1502 - 1520. Retrieved from `https://journals.ametsoc.org/view/journals/hydr/16/4/jhm-d-14-0183_1.xml` doi: 10.1175/JHM-D-14-0183.1

Coe, M. T., Brando, P. M., Deegan, L. A., Macedo, M. N., Neill, C., & Silvério, D. V. (2017). The Forests of the Amazon and Cerrado Moderate Regional Climate and Are the Key to the Future. *Tropical Conservation Science*, *10*, 1940082917720671. Retrieved from `https://doi.org/10.1177/1940082917720671` doi: 10.1177/1940082917720671

Coe, M. T., Latrubesse, E. M., Ferreira, M. E., & Amsler, M. L. (2011). The effects of deforestation and climate variability on the streamflow of the Araguaia River, Brazil. *Biogeochemistry*, *105*(1), 119–131.

Coe, M. T., Macedo, M. N., Brando, P. M., Lefebvre, P., Panday, P., & Silvério, D. (2016). The Hydrology and Energy Balance of the Amazon Basin. In L. Nagy, B. R. Forsberg, & P. Artaxo (Eds.), *Interactions between biosphere, atmosphere and human land use in the amazon basin* (pp. 35–53). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from `https://doi.org/10.1007/978-3-662-49902-3_3` doi: 10.1007/978-3-662-49902-3_3

Dai, A., & Trenberth, K. E. (2002). Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *Journal of hydrometeorology*, *3*(6), 660–687.

de Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., & Bulhões Mendes, C. A. (2013). Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resources Research*, *49*(3), 1226-1243. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/wrcr.20067` doi: https://doi.org/10.1002/wrcr.20067

Eom, J., Seo, K.-W., & Ryu, D. (2017). Estimation of Amazon river discharge based on EOF analysis of GRACE gravity data. *Remote Sensing of Environment*, *191*, 55–66.

Garnier, H. (2011). Data-based continuous-time modelling of dynamic systems. In *2011 international symposium on advanced control of industrial processes (adconip)* (pp. 146–153).

Garnier, H., & Gilson, M. (2018). Contsid: a matlab toolbox for standard and advanced identification of black-box continuous-time models. *IFAC-PapersOnLine*, *51*(15), 688-693. Retrieved from `https://www.sciencedirect`

.com/science/article/pii/S2405896318318755 (18th IFAC Symposium on System Identification SYSID 2018) doi: https://doi.org/10.1016/j.ifacol.2018.09.203

Garnier, H., Wang, L., & Young, P. C. (2008). Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance. In *Identification of continuous-time models from sampled data.* Springer.

Garnier, H., & Young, P. C. (2012). What does continuous-time model identification have to offer? *IFAC Proceedings Volumes*, *45*(16), 810–815.

GRDC data archive [dataset]. (2021). Retrieved from {https://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/riverdischarge_node.html} (The Global Runoff Data Centre, 56068 Koblenz, Germany, accessed: 27.06.2021)

Heerspink, B. P., Kendall, A. D., Coe, M. T., & Hyndman, D. W. (2020). Trends in streamflow, evapotranspiration, and groundwater storage across the Amazon Basin linked to changing precipitation and land cover. *Journal of Hydrology: Regional Studies*, *32*, 100755. Retrieved from https://www.sciencedirect.com/science/article/pii/S2214581820302299 doi: https://doi.org/10.1016/j.ejrh.2020.100755

Jahfer, S., Vinayachandran, P., & Nanjundiah, R. S. (2017). Long-term impact of Amazon river runoff on northern hemispheric climate. *Scientific reports*, *7*(1), 1–9.

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., . . . Yuan, D.-N. (2020). Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance. *Geophysical Research Letters*, *47*(12), e2020GL088306. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL088306 (e2020GL088306 2020GL088306) doi: https://doi.org/10.1029/2020GL088306

Lee, Y. H., & Singh, V. P. (1998). Application of the Kalman filter to the Nash model. *Hydrological Processes*, *12*(5), 755-767. doi: https://doi.org/10.1002/(SICI)1099-1085(19980430)12:5⟨755::AID-HYP623⟩3.0.CO;2-#

Lehmann, F., Vishwakarma, B. D., & Bamber, J. (2021). How well are we able to close the water budget at the global scale? *Hydrology and Earth System Sciences Discussions*, *2021*, 1–53. Retrieved from https://hess.copernicus.org/preprints/hess-2021-279/ doi: 10.5194/hess-2021-279

Lorenz, C., Kunstmann, H., Devaraju, B., Tourian, M. J., Sneeuw, N., & Riegger, J. (2014). Large-scale runoff from landmasses: A global assessment of the closure of the hydrological and atmospheric water balances. *Journal of Hydrometeorology*, *15*(6), 2111 - 2139. Retrieved from https://journals.ametsoc.org/view/journals/hydr/15/6/jhm-d-13-0157_1.xml doi: 10.1175/JHM-D-13-0157.1

Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., & Nobre, C. A. (2008). Climate Change, Deforestation, and the Fate of the Amazon. *Science*, *319*(5860), 169–172. Retrieved from https://science.sciencemag.org/content/319/5860/169 doi: 10.1126/science.1146961

Marengo, J. A. (2005). Characteristics and spatio-temporal variability of the Amazon River Basin Water Budget. *Climate Dynamics*, *24*(1), 11–22.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900.

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282-290. Retrieved from https://www.sciencedirect.com/science/article/pii/0022169470902556 doi: https://doi.org/10.1016/0022-1694(70)90255-6

Nash, J. E. (1957). The form of the instantaneous unit hydrograph. *International Association of Scientific Hydrology, Publ*, *3*, 114–121.

Pan, S., González, R. A., Welsh, J. S., & Rojas, C. R. (2020). Consistency analysis of the simplified refined instrumental variable method for continuous-time systems. *Automatica*, *113*, 108767. Retrieved from https://www.sciencedirect.com/science/article/pii/S0005109819306302 doi: https://doi.org/10.1016/j.automatica.2019.108767

Pimentel, E. T., & Hamza, V. M. (2012). Indications of regional scale groundwater flows in the Amazon Basins: Inferences from results of geothermal studies. *Journal of South American Earth Sciences*, *37*, 214-227. Retrieved from https://www.sciencedirect.com/science/article/pii/S0895981112000375 doi: https://doi.org/10.1016/j.jsames.2012.03.007

Riegger, J. (2020). Quantification of drainable water storage volumes on landmasses and in river networks based on GRACE and river runoff using a cascaded storage approach - first application on the Amazon. *Hydrology and Earth System Sciences*, *24*(3), 1447–1465.

Riegger, J., & Tourian, M. J. (2014). Characterization of runoff-storage relationships by satellite gravimetry and remote sensing. *Water Resources Research*, *50*(4), 3444–3466.

Save, H. (n.d.). *CSR GRACE and GRACE-FO RL06 Mascon Solutions v02.*

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7547–7569.

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. *Journal of hydrologic engineering*, *7*(4), 270–292.

Sneeuw, N., Lorenz, C., Devaraju, B., Tourian, M. J., Riegger, J., Kunstmann, H., & Bárdossy, A. (2014). Estimating runoff using hydro-geodetic approaches. *Surveys in Geophysics*, *35*(6), 1333–1359.

Sproles, E. A., Leibowitz, S. G., Reager, J. T., Wigington Jr, P. J., Famiglietti, J. S., & Patil, S. D. (2015). GRACE storage-runoff hystereses reveal the dynamics of regional watersheds. *Hydrology and Earth System Sciences*, *19*(7), 3253–3272. Retrieved from https://hess.copernicus.org/articles/19/3253/2015/ doi: 10.5194/hess-19-3253-2015

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505.

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., ... others (2019). Contributions of GRACE to understanding climate change. *Nature climate change*, *9*(5), 358–369.

Tourian, M., Reager, J., & Sneeuw, N. (2018). The total drainable water storage of the Amazon river basin: A first estimate using GRACE. *Water Resources Research*, *54*(5), 3290–3312.

Vishwakarma, B. D., Devaraju, B., & Sneeuw, N. (2018). What Is the Spatial Resolution of grace Satellite Products for Hydrology? *Remote Sensing*, *10*(6). Retrieved from https://www.mdpi.com/2072-4292/10/6/852 doi: 10.3390/rs10060852

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12).

Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2648-2671. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JB011547 doi: https://doi.org/10.1002/2014JB011547

Werth, D., & Avissar, R. (2004). The Regional Evapotranspiration of the Amazon. *Journal of Hydrometeorology*, *5*(1), 100 - 109. Retrieved from https://journals.ametsoc.org/view/journals/hydr/5/1/1525-7541_2004_005_0100_treota_2_0_co_2.xml doi: 10.1175/1525-7541(2004)005⟨0100:TREOTA⟩2.0.CO;2

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, *52*(9), 7490-7502. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019344 doi: https://doi.org/10.1002/2016WR019344

Yilmaz, K. K., Vrugt, J. A., Gupta, H. V., & Sorooshian, S. (2010). Model calibration in watershed hydrology. In *Advances in data-based approaches for hydrologic modeling and forecasting* (p. 53-105). World Scientific. Retrieved from https://www.worldscientific.com/doi/abs/10.1142/9789814307987_0003 doi: 10.1142/9789814307987_0003

Young, P., & Jakeman, A. (1980). Refined instrumental variable methods of recursive time-series analysis Part III. Extensions. *International Journal of Control*, *31*(4), 741-764. Retrieved from https://doi.org/10.1080/00207178008961080 doi: 10.1080/00207178008961080

Young, P. C. (2002). Optimal iv identification and estimation of cotinuous-time tf models. *IFAC Proceedings Volumes*, *35*(1), 109-114. Retrieved from https://www.sciencedirect.com/science/article/pii/S1474667015394258 (15th IFAC World Congress) doi: https://doi.org/10.3182/20020721-6-ES-1901.01004

Young, P. C. (2008). The refined instrumental variable method. *Journal Européen des Systemes Automatisés*, *42*(2-3), 149–179.