Dear Dr. Kurtz,

Thank you for your patience in the review process. Reviewer #5 provided additional input that was very helpful to include in the final version. Additionally, to his remarks we modified the following:

1) Figure 2 was extended by FigS4.5 (as it was often references in the text).
2) Table 1 and 2 changed their order.
3) Figure 10 contained an error in the data that was fixed and the overall representation improved.
4) Multiple minor spelling mistakes and minor rephrases to streamline the final version.

Our response to the Reviewer comments are in italic.

Response to Reviewer #5

1)

P1L20-21, P2L40-P3L4 and other lines related to '... additional drainage above flood plain ...' in PCR-GLOBWB-MODFLOW: The 'additional drainage above flood plain' in the PCR-GLOBWB-MODFLOW works (e.g. de Graaf et al., 2015, 2017) was not intended for improving groundwater head simulation performance. Yet, such drainage was introduced to improve/discharge performance of the online coupled PCR-GLOBWB-MODFLOW. In fact, the introduction of the drainage above flood plain was based on the earlier works in Sutanudjaja et al. (2011, 2014). Initially, such drainage was not used in Sutanudjaja et al. (2011), which focused on offline coupling approach of PCR-GLOBWB-MODFLOW. In this offline and one-way coupling approach for modeling spatio-temporal groundwater head dynamics, Sutanudjaja et al. (2011) conceptualized that groundwater discharge/baseflow as merely a function based on groundwater and surface water head differences, via RIV and DRN packages of MODFLOW (McDonald and Harbaugh, 1988; Harbaugh et al., 2000; Harbaugh, 2005). However, as the online two-way coupling approach between PCR-GLOBWB and MODFLOW was established in Sutanudjaja et al. (2014), we realized that flows from RIV and DRN are too slow to satisfy fast/quick-response component of groundwater discharge originating from mountainous regions where many springs tapping groundwater are located higher up in the valleys and feeding tributaries and main rivers. To include such fast groundwater discharge (baseflow) component, it is assumed that groundwater above flood plain is drained based on a linear reservoir concept (for more detailed, see Sutanudjaja et al., 2014 and Sutanudjaja, 2012).

Thank you very much for the clarification and the interesting references! That helped us a lot to understand your model much better. Of course these statements are then wrong in our manuscript and have been removed from the abstract and modified in introduction, and discussion.

Introduction now reads (P.2 L.39 ff):

"However, to achieve plausible discharge performance, they found it necessary to increase drainage from GW to rivers beyond the drainage driven by the hydraulic head difference between GW and river."

Discussion now reads (P. 20 L.18 ff):

"Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to improve the discharge to rivers (Sutanudjaja et al., 2014)."

2)

Related to the aforementioned comment, I am just wondering how the discharge/flow

WaterGAP model will perform when an online two-way coupling/integration between

G3M and WaterGAP is used. I know that this is still outside the scope of your current

study/manuscript, which still focusses on steady-state (and offline approach)

simulation. Yet, could you please speculate about this in the discussion part of yourmanuscript? Do you expect that you have to calibrate your parameter values such as

river conductances (e.g. c_swb and c_riv in Equations 5 and 6) in order to get good

discharge performance? If calibration is required, could you please hypothesize about

its consequence to your groundwater head simulation performance?

That is a very interesting question! We assume that the fully coupled model will help us especially in times of drought to improve streamflow simulation. It is very likely that the conductance will need to be calibrated. Yet we think it is very unclear how this could affect head performance. Future research needs to show how we can achieve a good fit to streamflow as well as to head observations.

Now reads (P. 18 L.21 ff):

"We assume that the fully coupled model will lead to an improved WaterGAP performance during droughts with an increased drop in streamflow due to the now possible switch from gaining to losing conditions. Presumable a calibration of c_{swb} and c_{riv} is necessary to achieve a good discharge performance."

Minor comments

1

P1L20-21: What do you mean by "... externally provided values for GW storage ..."? Please rephrase. GW storages of PCR-GLOBWB-MODFLOW are always based on (internally) simulated groundwater heads.

Thank you very much for the clarification! Of course these statements are then wrong in our manuscript and have been removed from the abstract, introduction, and discussion.

2

P4L18-20: This sentence is not clear for me. Please consider to rephrase. Do you mean that you excluded large mountainous areas in your model simulation? Could you please be more specific about how you defined mountainous areas? It may be helpful for readers if you provide some examples of such mountainous area locations.

No, they were not excluded. This paragraph speculates about the impacts if not the specific saturated thickness approach (criticized by the former reviewers) but a head-based transmissivity was to be used. For clarification the sentence has been revised and now reads (P. 4 L. 10-15):

"Both approaches have proven to be insufficient to simulate head-based transmissivities (unconfined contions) on the global scale. Large mountainous areas would be excluded if unconfined conditions are assumed from the beginning of the solution step, as the head is often far below the deepest model layer, resulting in a no-flow condition and imposing convergence issues to the matrix solver. We choose to simulate both layers with a specific saturated thickness even though the upper layer can be expected to decrease in water level and thus in transmissivity (hydraulic conductivity times saturated depth)."

3

P7L20: "Globally constant but different values ..." This is hard to read for me. Please consider to rephrase.

Now reads (P. 7 L. 1):

"Globally constant values are used for B_{swb} for wetlands, local lakes and global lakes (Table 2)."

4

P9L32: I suggest providing global flux values in annual unit, e.g. m3 year-1 or km3 year-1 (as commonly done in other hydrological studies, such as Döll et al, 2014; Rodell et al., 2012).

Page 10, Figure 3: Please provide values in annual unit, e.g. m3 year-1 or km3 year-1.

Figure 3 is now in m^3 year-1 and values in the text are reported accordingly. Figure 4, Figure S4.10 in mm/year.

5

P11L10-13: Could you please share your hypothesis or reason why the model cannot simulate

losing rivers in Niger? Is it related to the forcing/input error?

Now reads (P. 11 L. 20 ff):

“On the other hand, no losing stretches are simulated along the Niger River and its wetlands and almost none in the North-eastern Brazil even though that losing conditions are known to occur there (Costa et al., 2013; FAO, 1997). This is also true when the minimum elevation for SW bodies is assumed (compare Fig. S4.10) leading to the conclusion that the misrepresentation might be linked to an inadequate representation of the local geology.“

6

P19L1: ... world wide ...

Fixed.

7

P19L24-25: Please give a brief explanation about the method of Morel-Seytoux et al. (2017).

Has been added (P. 20 L. 13-16):

“The simple conductance approach applied in G^3M could possibly be improved by the approach by Morel-Seytoux et al. (2017) who proposes an analytical, and physically based, estimate of the leakance coefficient for coarse scale models based on river and aquifer properties.”

Challenges in developing a global gradient-based groundwater model (G^3M v1.0) for the integration into a global hydrological model

Robert Reinecke[1], Laura Foglia[3], Steffen Mehl[4], Tim Trautmann[1], Denise Cáceres[1], Petra Döll[1,2]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany
[2]Senckenberg Leibniz Biodiversity and Climate Research Centre Frankfurt (SBiK-F), Frankfurt am Main, Germany
[3]Department of Land, Air and Water Resources, University of California, Davis, USA
[4]Department of Civil Engineering, California State University, Chico, USA

Correspondence to: Robert Reinecke (reinecke@em.uni-frankfurt.de)

Abstract. In global hydrological models, groundwater (GW) is typically represented by a bucket-like linear groundwater reservoir. Reservoir models, however, can (1) only simulate GW discharge to surface water (SW) bodies but not recharge from SW to GW, (2) provide no information on the location of the GW table and (3) assume that there is no GW flow among grid cells. This may lead, for example, to an underestimation of groundwater resources in semi-arid areas where GW is often replenished by SW or to an underestimation of evapotranspiration where the GW table is close to the land surface. To overcome these limitations, it is necessary to replace the reservoir model in global hydrological models with a hydraulic head gradient-based GW flow model.

We present G³M, a new global gradient-based GW model with a spatial resolution of 5', which is to be integrated in the 0.5° WaterGAP Global Hydrology Model (WGHM). The newly developed model framework enables in-memory coupling to WGHM while keeping overall runtime relatively low, which allows sensitivity analyses, calibration and data assimilation. This paper presents the G³M concept and model design decisions that are specific to the large grid size required for a global scale model. Model results under steady-state naturalized conditions, i.e. neglecting GW abstractions, are shown. Simulated hydraulic heads show better agreement to observations around the world than model output of de Graaf et al. (2015). Locations of simulated SW recharge to GW are found, as is expected, in dry and mountainous regions but areal extent of SW recharge may be underestimated. Globally, GW discharge to rivers is by far the dominant flow component such that lateral GW flows only become a large fraction of total diffuse and focused recharge in case of losing rivers, some mountainous areas and some areas with very low GW recharge. A strong sensitivity of simulated hydraulic heads to the spatial resolution of the model and the related choice of the water table elevation of surface water bodies was found. We suggest to investigate how global-scale groundwater modelling at 5' spatial resolution can benefit from more highly resolved land surface elevation data.

Kommentiert [RR1]: #5 1) and M.1

1 Introduction

Groundwater (GW) is the source of about 40% of all human water abstractions (Döll et al., 2014) and is also an essential source of water for freshwater biota in rivers, lakes and wetlands. GW strongly affects river flow regimes and supplies the majority of river water during ecologically and economically critical periods with little precipitation. GW storage and flow dynamics have been altered by human GW abstractions as well as climate change and will continue to change in the future (Taylor et al., 2012). Around the globe, GW abstractions have led to lowered water tables and, in some regions, even GW depletion (Döll et al., 2014; Scanlon et al., 2012; Wada et al., 2012; Konikow, 2011). This has resulted in reduced base flows to rivers and wetlands (with negative impacts on water quality and freshwater ecosystems), land subsidence and increased pumping costs (Wada, 2016; Döll et al., 2014; Gleeson et al., 2012; 2016). The strategic importance of GW for global water and food security will probably intensify under climate change as more frequent and intense climate extremes increase variability of SW flows (Taylor et al., 2012). International efforts have been made to promote sustainable GW management and knowledge exchange among countries, e.g., UNESCO's program on International Shared Aquifer Resources Management (ISARM) (http://isarm.org) and the ongoing GW component of the Transboundary Waters Assessment Program (TWAP) (http://www.geftwap.org). To support prioritization for investment among transboundary aquifers as well as identification of strategies for sustainable GW management, information on current conditions and possible trends of the GW systems is required (UNESCO-IHP, IGRAC, WWAP, 2012). In a globalized world, an improved understanding of GW systems and their interaction with SW and soil is needed not only at the local and regional but also at the global scale.

To assess GW at the global scale, global hydrological models (GHMs) are used e.g. (Wada et al., 2012; 2016; Döll et al., 2012; 2014). In particular, they serve to quantify GW recharge (Döll and Fiedler, 2008). Like typical hydrological models at any scale, GHMs simulate GW dynamics by a linear reservoir model. In such a model, the temporal change of GW storage in each grid cell is computed from the balance of prescribed inflows and an outflow that is a linear function of GW storage. Linear reservoir models can only simulate GW discharge to SW bodies but not a reversal of this flow, even though losing streams may provide focused GW recharge that allows the aquifer to support ecosystems alongside the GW flow path (Stonestrom et al., 2007) as well as human GW abstractions. Losing streams typically occur in semi-arid and arid but seasonally also in humid regions. In addition, such linear reservoir models provide no information on the location of the GW table, and assume that GW flow among grid cells is negligible. To simulate the dynamics of water flow between SW bodies and GW in both directions as well as the effect of capillary rise on evapotranspiration, it is necessary to compute lateral GW flows among grid cells as function of hydraulic head gradients and thus the dynamic location of the GW table. To achieve an improved understanding of GW systems at the global scale, and in particular of the interactions of GW with SW and soil, it is therefore necessary to replace the linear GW reservoir model in GHMs by a hydraulic gradient-based GW flow model.

Large-scale gradient-based GW flow models are still rare and mainly available for data-rich regions, e.g. for the Death Valley (Belcher and Sweetkind, 2010) and the Central Valley (Belcher and Sweetkind, 2010; Faunt, 2009; Dogrul et al., 2016) in the USA, but also for large fossil groundwater bodies in arid regions (e.g. the Nubian Aquifer System in North Africa, (Gossel et al., 2004)). However, they are in most cases not integrated within hydrological models that quantify GW recharge based on climate data and provide information on the condition of SW (e.g. streamflow and storage). For North America, Fan et al. (2007) and Miguez-Macho et al. (2007) linked a land surface model with a two-dimensional gradient-based GW model and computed, with a daily time step, GW flow, water table elevation, GW–SW interaction, and capillary rise, using a spatial resolution of 1.25 km. One challenge was the determination of the river conductance that affects the degree of GW–SW interaction. A computationally very expensive integrated simulation of dynamic SW, soil and GW flow using Richards' equation for variably saturated flow was achieved at a spatial resolution of 1 km for the continental US by applying the ParFlow model (Maxwell et al., 2015). In both studies, GW abstractions were not taken into account.

A first simulation of the steady-state GW table for the whole globe at the very high resolution of 30" was presented by Fan et al. (2013) and compared to an extensive compilation of observed hydraulic heads. However, there was no head-based interactions with SW; GW above the land surface was simply discarded. Global GW flow modeling is strongly hampered by data availability, including the geometry of aquifers and aquitards as well as parameters like hydraulic conductivity (de Graaf et al., 2017), and by computational restrictions on spatial resolution leading to conceptual problems, e.g., regarding SW-GW interactions (Morel-Seytoux et al., 2017). Recently, some GW flow models that are in principle applicable for the global scale were developed but were applied only regionally in data-rich regions (Rhine basin: Sutanudjaja et al., 2011; France: Vergnes et al., 2012; 2014). The first global gradient-based GW model that was run for both steady-state (de Graaf et al., 2015) and transient conditions (de Graaf et al., 2017) was driven by GW recharge and SW data of the GHM PCR-GLOBWB (van Beek et al., 2011). However, to achieve plausible discharge performance, they found it necessary to increase drainage from GW to rivers beyond the drainage driven by the hydraulic head difference between GW and river. This additional drainage, which accounts for about 50% of global GW flow into SW, is simulated as a function of GW storage above the floodplain.

Kommentiert [RR2]: #5 1)

In this study, we present the Global Gradient-based Groundwater Model (G^3M) that is to be integrated into the GHM WaterGAP 2 to improve estimation of flows between SW and GW (affecting both streamflow and groundwater recharge and thus water availability for humans and ecosystems) and implement capillary rise (affecting evapotranspiration). Table 1 provides a comparative summary of G³M as well as the global groundwater models of Fan et al. (2013), de Graaf et al. (2015; 2017), and the continental scale model ParFlow (Maxwell et al., 2015).

The objective of this paper is to learn from a steady-state model, a well-established first step in groundwater model development, to (1) understand the basic model behaviour by limiting model complexity and degrees of freedom, and thus (2) providing insights into dominant processes and uncovering potential model-inherent characteristics difficult to observe in a fully coupled transient model. A transient model might obfuscate model inherent trends due to the slow changing nature of groundwater processes e.g. trends towards large over/under-estimation due to wrong parameterisation. A fully coupled model furthermore adds complexity and uncertainty to the model outcome. The presented steady-state model is furthermore used to (3) investigate parameter sensitivity and sensitivity to spatial resolution. In addition, the steady-state solution can be used as (4) initial condition for future fully coupled transient runs.

Model concept and equations as well as applied data and parameter values are presented in section 2. In section 3, we show steady-state results of G^{3}M driven by WGHM data. Simulated hydraulic heads are compared to observations world-wide and to the output of existing large-scale GW models (Table 1). Furthermore, sensitivity to parameters and grid size is shown for the example of New Zealand. Finally, the implications of modeling decisions and grid size are discussed (section 4) and conclusions are drawn (section 5).

2 Model description

2.1 G³M model concept

Although G³M is based on principles of the well-known GW flow modelling software MODFLOW (Harbaugh, 2005), G³M differs in its parameterization from traditional local and regional GW models. These models are generally based on rather detailed information on hydrogeology (including aquifer geometry and properties such as hydraulic conductivity derived from pumping tests), topography, pumping wells, location and shape of SW bodies as well as on observations of hydraulic head in GW and SW. Local observations guide the developer in constructing the model such that local conditions and processes can be properly represented. The lateral extent of individual grid cells of such GW flow models is generally smaller or similar to the depth of the aquifer(s) and the size of the SW bodies that interact with the GW. The global GW flow model G³M, however, covers all continents of the Earth except Greenland and Antarctica. At this scale, information listed above is poor or non-existing, and the lateral extent of grid cells needs to be relatively large due to computational (and data) constraints. We selected a grid cell size of 5' by 5' (approx. 9 km by 9 km at the equator), as this size fits well to WaterGAP and is smaller than the suggested 6' of Krakauer et al. (2014). WaterGAP 3 (Eisner, 2016) has the same cell size, and 36 of such cells fit to into one 0.5° WaterGAP 2 cell. Global climate data are only available for 0.5° grid cells. The landmask of G³M, i.e. location and size of 5' grid cells, is that of WaterGAP 3 end encompasses 2.2 million 5' grid cells on each layer.

Due to the lack of the spatial distribution of hydrogeological properties, we chose to use, in the current version of G³M, two GW layers with a vertical size of 100 m each (Fig. 1). We performed a sensitivity analysis that confirmed the findings of others (de Graaf et al., 2015) that the aquifer thickness has a relatively small impact on the model results. Therefore, selecting a uniform thickness of 100 m (motivated by the assumed depth of validity of the lithology data) (Fig. 1) worldwide for the first layer and also for the second layer is expected to lead to less uncertainties as compared to hydraulic conductivities and the surface water table elevation.

G³M focuses on a plausible simulation of water flows between GW and SW, and we deemed it suitable to have an upper GW layer that interacts with SW and soil (the soil layer of WaterGAP is described in detail in S1) and a lower one in which GW may flow laterally without such interactions. As land surface elevation within each 5' grid cell, with an area of approximately 80 km^2, may vary by more than 200 m (Fig. S4.1), neighbouring cells in G³M may not be adjacent anymore (Fig. 1), in contrast to (regional) GW models with smaller grid cells. This makes G³M a rather conceptual model in which water exchange between groundwater cells is driven by hydraulic head gradients but flow can no longer be conceptualized as occuring through continuous pore space. In addition, due to the coarse spatial scale and the possible large variations of land surface elevations within each grid cell, the upper model layers should not be considered to be aligned with an average land surface elevation. The model layers can be rather thought to be vertically aligned with the elevation of the surface water body table, as this prescribed elevation is, together with the sea level, the only elevation included in the groundwater flow equation (Eq. (1)).

The simulation of aquifers that contain dry cells and/or cells that oscillate between wet and dry states poses great challenges to solving Eq. (1) (Niswonger et al., 2011). G³M-f (the framework code used to implement G³M) implements the traditional wetting approach from Harbaugh (2005) as well as the approach proposed by Niswonger et al. (2011) along with the proposed damping scheme. Both approaches have proven to be insufficient to simulate head-based transmissivities (unconfined contions) on the global scale. Large mountainous areas would be excluded if unconfined conditions are assumed from the beginning of the solution step, as the head is often far below the deepest model layer, resulting in a no-flow condition and imposing convergence issues to the matrix solver. We choose to simulate both layers with a specific saturated thickness even though the upper layer can be expected to decrease in water level and thus in transmissivity (hydraulic conductivity times saturated depth). The large uncertainties regarding hydraulic conductivities (possibly an order of magnitude), further justifies using the computationally more efficient assumption of specified saturated thickness. This approach is consistent with findings that this is accurate for large, complex groundwater models (Sheets et al., 2015). Furthermore, it is consistent with recent presented large scale studies e.g. for the Rhine Meuse basin of Sutanudjaja et al. (2011) (using one confined layer), the Death Valley Regional Flow Model (Belcher, 2004; Faunt et al., 2011), and the global groundwater model of de Graaf et al. (2017) (two layers and partially unconfined conditions are simulated by parametrization of the model input and not by a head-dependant transmissivity).

Kommentiert [RR3]: #5 M.2

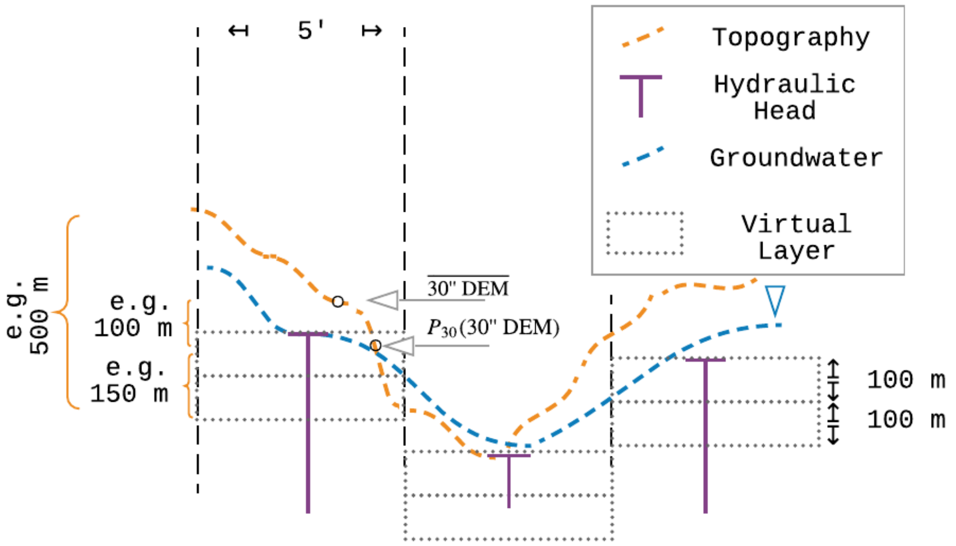

Figure 1 Schematic of G³M's spatial structure, with 5**'** grid cells, hydraulic head per cell, and the conceptual virtual layers (virtual because at this stage only confined conditions are computed). The underlying variability of the topography changes the perception of simulated depth to groundwater depending on what metrics are used to represent it on a coarser resolution. Layers in G³M are of a conceptual nature and describe the saturated flow between locations of head laterally and vertically. The P_{30} is used in the presented steady-state model as SW elevation instead of an average or minimum per grid cell.

Three-dimensional groundwater flow is described by a partial differential equation (approximated in the model implementation by using the finite differences method (Sect. 2.4))

$$\frac{dGWS}{dt} = (\frac{\partial}{\partial x}(K_x \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y}(K_y \frac{\partial h}{\partial y}) + \frac{\partial}{\partial z}(K_z \frac{\partial h}{\partial z}) + W)\, \Delta x \Delta y \Delta z = S_s \frac{\partial h}{\partial t} \Delta x \Delta y \Delta z \quad (1)$$

where $K_{x,y,z}$ is the hydraulic conductivity $[LT^{-1}]$ along the x, y, and z axis between the cells (harmonic mean of grid cell conductivity values), S_s the specific storage $[L^{-1}]$, $\Delta x \Delta y \Delta z$ $[L^3]$ the volume of the cell, and h the hydraulic head $[L]$. In- and out-flows in the groundwater are accounted for as

$$W\, \Delta x \Delta y \Delta z = R_g + Q_{swb} - NA_g - Q_{cr} + Q_{ocean} \quad (2)$$

where Q_{swb} is flow between the SW bodies (rivers, lakes, reservoirs and wetlands) and GW $[L^3T^{-1}]$, Q_{cr} is capillary rise, i.e. the flow from GW to the soil, and Q_{ocean} is the flow between ocean and GW[L^3T^{-1}], representing the boundary condition. In case of Q_{swb} and Q_{ocean}, a positive value represents a flow into the groundwater.

Q_{swb} in Eq. (3) replaces $k_g\ GWS$ and R_{g_swb} in the linear storage equation of WaterGAP (Eq. (S1)), such that losing conditions of all types of SW bodies can be simulated dynamically. It is calculated as a function of the difference between the elevation of the water table in the SW bodies h_{swb} $[L]$ and h_{aq} as

$$Q_{swb} = \begin{cases} c_{swb}(h_{swb} - h_{aq}) & h_{aq} > B_{swb} \\ c_{swb}(h_{swb} - B_{swb}) & h_{aq} \leq B_{swb} \end{cases} \quad (3)$$

where c_{swb} is the conductance $[L^2T^{-1}]$ of the SW body bed (river, lake, reservoir or wetland) and B_{swb} the SW body bottom elevation $[L]$.

Conductance of SW bodies is often a calibration parameter in traditional GW models (Morel-Seytoux et al., 2017). Following Harbaugh (2005), it can be estimated by

$$c_{swb} = \frac{K\, L\, W}{h_{swb} - B_{swb}} \quad (4)$$

where K is hydraulic conductivity, L is length and W is width of the SW body per grid cell. For lakes (including reservoirs) and wetlands, c_{swb} is estimated based on hydraulic conductivity of the aquifer K_{aq} and SW body area (Table 2). For gaining rivers, conductance is quantified individually for each grid cell following an approach proposed by Miguez-Macho et al. (2007). The value of river conductance c_{riv}, according to Miguez-Macho et al. (2007), in a GW flow model needs to be set to such a values that, for steady-state conditions, the river is the sink for all the inflow to the grid cell (GW recharge and inflow from neighbouring cells) that is not transported laterally to neighbouring cells such that

$$c_{riv} = \frac{R_g + Q_{eq_{lateral}}}{h_{eq} - h_{riv}} \qquad h_{aq} > h_{riv} \quad (5)$$

For G^{3}M, we computed the equilibrium head h_{eq} as the 5' average of the 30" steady-state heads calculated by Fan et al. (2013). Using WGHM diffuse GW recharge lateral equilibrium flow $Q_{eq_{lateral}}$ $[L^3T^{-1}]$ is net lateral inflow into the cell computed based on the h_{eq} distribution as well as G^{3}M K_{aq} and cell thickness (Table 2). Elevation of the river water table h_{riv} [L] is to be provided by WGHM. Using a fully dynamic approach, i.e. utilizing the hydraulic head and lateral flows from the current iteration to re-calculate c_{riv} in each iteration towards the steady-state solution, has proven to be too unstable due to its non-linearity affecting convergence. We limit c_{riv} to a maximum of $10^7\ m^2day^{-1}$; this would be approximately the value for a 10 km long and 1 km wide river with a head difference between GW and river of 1 m and hydraulic conductivity of the river bed of 10^{-5} m/s.

If the river recharges the GW (losing river), Eq. (5) cannot be used as the Fan et al. (2013) high-resolution equilibrium model only models groundwater outflows but not inflows from SW bodies. If h_{aq} drops below h_{riv}, Eq. (4) is used to compute c_{riv}, with K equals to K_{aq}.

The flux across the model domain boundary Q_{ocean} is modeled as a head-dependent flow based on a static head boundary.

$$Q_{ocean} = c_{ocean}(h_{ocean} - h_{aq}) \quad (6)$$

where h_{ocean} is the elevation of the ocean water table $[L]$, h_{aq} the hydraulic head in the aquifer $[L]$ and c_{ocean} the conductance of the boundary condition $[L^2T^{-1}]$ (Table 2). We assume that density difference to sea-water is negligible at this scale. Q_{cr} is not yet implemented in G³M.

2.2 The steady-state uncoupled model version

In a first implementation stage, G³M was developed as a steady-state (right-hand side of Eq. (1) is zero) standalone model that represents naturalized conditions (i.e. without taking into account human water use) during 1901-2013. Input data and parameters used are listed in Table 2 and described below.

Gleeson et al. (2014) provided a global subsurface permeability data set from which K_{aq} was computed. The data set was derived by relating permeabilities from a large number of local to regional GW models to the type of hydrolithological units (e.g., “unconsolidated” or “crystalline”). The geometric mean permeability values of nine hydrolithological units were mapped to the high-resolution global lithology map GLiM (Hartmann and Moosdorf, 2012). In continuous permafrost areas, a very low permeability value was assumed by Gleeson et al. (2014). The estimated values represent the shallow surface on the scale of 100 m depth. The unique dataset has three inherent problems when used as input for a GW model: (1) At this scale, important heterogeneities such as discrete fractures or connected zones of high hydraulic conductivity controlling the GW flow are not visible. (2) Jurisdictional boundaries due to different data sources in the global lithological map lead to artifacts. (3) The differentiation between coarse and fine-grained unconsolidated deposits is only available in some regions resulting in $10^{-4}\ m\ s^{-1}$ as hydraulic conductivity for coarse-grained unconsolidated deposits. If the distinction is not available, a rather low value of $10^{-6}\ m\ s^{-1}$ is set for unconsolidated porous media (Fig. S4.3). The original data was gridded to 5' by using an area-weighted average and used as hydraulic conductivity of the upper model layer. For the second layer, hydraulic conductivity of the first layer is reduced assuming that conductivity decreases exponentially with depth. Based on the e-folding factor f used by Fan et al. (2013) (a calibrated parameter based on terrain slope), conductivity of the lower layer is calculated by multiplying the upper layer value by $\exp(-50\ m\ f^{-1})^{-1}$ (Fan et al., 2007).

Mean annual GW recharge computed by WaterGAP 2.2c for the period 1901-2013 is used as input (Fig. S4.4), while no net abstraction from GW was taken into account. It would not be meaningful to try to derive a steady-state solution under existing net groundwater abstractions that in some regions cause GW depletion with continuously dropping water tables. The 0.5° data of WaterGAP was equally distributed to the pertaining cells. Regarding the ocean boundary condition, h_{ocean} is set to 0 m and c_{ocean} to $10\ m^2\ day^{-1}$ (Table 2), reflecting a global average conductance based on hydraulic conductivity and lateral surface area.

It is assumed that there is exchange of water between GW and one river stretch in each 5' grid cell, and in addition where lakes and wetlands exist according to WaterGAP 3, which provides, for each grid cell, the area of “local” and ”global” lakes and wetlands. In WaterGAP, “local” SW bodies are only recharged by runoff produced within the grid cell, while ”global” SW bodies also obtain inflow from the upstream cell. In an uncoupled model, it is difficult to prescribe the, in reality temporal variable, area of lakes and wetlands that affect the flow exchange between SW body and GW. Maps generally show the maximum spatial extent of SW bodies. This maximum extent is seldom reached, in particular in case of wetlands in dry areas. For global wetlands (wetlands greater than one 5' cell), it is therefore assumed in this model version that only 80% of their maximum extent is reached. In the transient model SW body areas change over time. A further difficulty in an uncoupled model run is that the water table elevation of SW bodies does not react to the GW-SW exchange flows Q_{swb} and that water supply from SW is not limited by availability. A losing river may in reality dry out and therefore cease to lose any more water. For rivers B_{swb} is set to $h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994), where $Q_{bankfull}$ is the bankfull river discharge in the 5' grid cell (Verzano et al., 2012). Globally constant values are used for B_{swb} for wetlands, local lakes and global lakes (Table 2).

Kommentiert [RR4]: #5 M.3

For the steady-state model, all SW bodies in a grid cell are assumed to have the same head i.e $h_{riv} = h_{swb}$. We found that for both gaining and losing conditions, Q_{swb} and thus computed hydraulic heads are highly sensitive to h_{swb}. The overall best agreement with the hydraulic head observations of Fan et al. (2013) was achieved if h_{swb} (Eq. (3), (4) and (5)) was set to the 30th percentile (P_{30}) of the 30" land surface elevation values of Fan et al. (2013) per 5' cell, e.g. the 30" elevation that is exceeded by 70% of the 100 30" elevation values within one 5' cell. To decrease convergence time we used h_{eq} derived from the high-resolution steady-state hydraulic head distribution of Fan et al. (2013) as initial guess of h_{aq}. In each outer iteration (Sect. 2.4) gaining and losing conditions may change depending on the current head solution.

2.3 Integration into WGHM

We intend to integrate G³M into WaterGAP 2, i.e. the 0.5° version of WATERGAP (for details see S1), to keep computation time low enough for performing sensitivity analyses and ensemble-based data assimilation and calibration, instead of integrating it into WaterGAP 3 (Eisner, 2016), which has the same spatial resolution as G³M. However, data from WaterGAP 3 were used to set up G³M. Location and area of the 5' grid cells of G³M are the same as in the landmask of WaterGAP 3. In addition, the percentage of the 5' grid cell area that is covered by lakes (including reservoirs) and by wetlands, based on Lehner and Döll (2004), is taken from WaterGAP 3, as well as the length and width of the main river within each 5' grid cell as (Table 2).

2.4 Model implementation

G³M is implemented using a newly developed open-source model framework G³M-f (Reinecke, 2018). The main motivation to develop a new model framework is the efficient in-memory coupling to the GHM and flexible adaptation to the specific requirements of global-scale modelling. Written in C++14, the framework allows the implementation of global and regional groundwater models alike while providing an extensible purely object-oriented model environment. It is primarily targeted as extension to WaterGAP but allows an in-memory coupling to any GHM or can be used as a standalone groundwater model. It provides a unit-tested (Dustin, 2006) environment offering different modules that can couple results in-memory to a different model or write out data flows to different file formats. G³M-f has the following advantages over using an established GW modelling software such as MODFLOW. G³M-f enables an improved coupling capability. Unlike MODFLOW it provides a clear development interface to the programmer coupling a model to G³M-f. It can be easily compiled as a library, and provides a clearly separated logic between computation and data read-in/write-out). It is written in the same language as the target GHM enabling a straight-forward in-memory access to arrays without the need to write data to disk, required when coupling with MODFLOW (a very expensive operation even if that disk is a RAM-disk). Even though it is possible to call FORTRAN functions from C++, it is very complicated to pass file pointers properly, as the I/O implementation of both languages differ substantially and it is widely considered bad practice to handle I/O in two different languages at once. As MODFLOW was never designed to be coupled/integrated to/into other models, it is not possible to separate the I/O logic fully from the computational logic without substantial code changes that are hard to test. To this end, G³M-f provides a highly modularized framework that is written with extensibility as design goal while implementing all required groundwater mechanisms.

Eq. (1) is reformulated as finite-difference equation and solved using a conjugate gradient approach and an Incomplete LUT preconditioner (Saad, 1994). In order to keep the memory footprint low, the conjugate gradient method makes use of the sparse matrix. Furthermore, it solves the equations in parallel (preconditioner currently non-parallel). As internal numerical library G³M uses Eigen3 (eigen.tuxfamily.org). G³M can compute the presented steady-state solution (with the right-hand side of Eq. (1) being zero and the heads of Fan et al. (2013) as initial guess (Table 1,2)) on a commodity computer with four computational cores and a standard SSD in about 30 minutes while occupying 6 GB of RAM.

Similar to MODFLOW, G³M-f solves Eq. (1) in two nested loops using a Picard iteration (Mehl, 2006): (1) the outer iteration checks the head and residual convergence criterion (if the maximum head change between iterations is below a given value in three consecutive iterations and/or the norm of the residual vector of the conjugate gradient (Harbaugh, 2005; Niswonger et al., 2011) is below a given value). It adjusts head-dependant values e.g. from gaining to losing conditions and updates the system of linear equations if flows are no longer head dependent. (2) The inner loop primarily consists of the conjugate gradient solver, which runs for a number of iterations defined by the user or until the residual convergence criterion is reached (Table 2), solving the current system of linear equations.

Because switching between Eq. (4) and Eq. (5), which occurs if e.g. h_{aq} drops below h_{riv} from one iteration to the next causes an abrupt change of c_{riv} inducing a nonlinearity that affects convergence we introduced an ϵ = 1 m interval around h_{riv} and interpolate c_{riv} between the two Eq. (4) and (5) by a cubic hermite spline polynomial over that interval. This allows for a smoother transition between both states, reducing the changes in the solution if a river is in a gaining condition in one iteration and in a losing condition in the next or vice versa.

Different from Vergnes et al. (2014), G³M's computations are not based on spherical coordinates directly but on an irregular grid of quadratic cells of different size depending on the latitude. Cell sizes are provided by WaterGAP3 and are derived from their spherical coordinates maintaining their correct area and centre location. The model code will be adapted in the future to account for the different length in x and y direction per cell correctly.

3 Results

3.1 Global hydraulic head and water table depth distribution under natural steady-state conditions

As expected, the computed global distribution of steady-state hydraulic head (in the upper model layer) under natural conditions (Fig. 2a) follows largely the land surface elevation (Fig. S4.2), albeit with a lower range and locally different ratios between the hydraulic head and land surface gradients (Fig. S4.6). Water table depth (WTD), i.e. the distance between the groundwater table and the land surface, can be computed by subtracting the hydraulic head computed by G^{3}M for the upper layer of each 5' grid cell from the arithmetic mean of the land surface elevations of the 100 30" grid cells within each 5' cells (Fig. S4.2). The global map of steady-state WTD (Fig. 2b) clearly resembles the map of differences between surface elevation and P_{30} , the assumed water level of SW bodies h_{swb}, shown in Fig. S4.1**,** which indicates that simulated WTD is strongly governed by the assumed water level in SW bodies.

Deep GW, i.e. a large WTD, occurs mainly in mountainous regions (Fig. 2b). These high values of WTD are mainly a reflection of the steep relief in these areas as quantified either by the differences of mean land surface elevations between neighbouring grid cells (Fig. S4.7) or the difference between mean land surface elevation and P_{30}, the 30th percentile of the 30" land surface elevations (Fig. S4.1). When computed hydraulic head is subtracted not from average land surface elevation but from P_{30}, the assumed water table elevation of SW bodies, the resulting map shows that the groundwater table is mostly above P_{30}, in both flat and steep terrain (Fig. 2c). Thus, high WTD values at the 5' resolution do not indicate deep unsaturated zones and losing rivers but just high land surface elevation variations within a grid cell. In steep terrain, 5' water tables are higher above water level in the surface water bodies than in flat terrain (Fig. 2c). Deep GW tables that are not only far below the mean land surface elevation but also below the water table of surface water bodies are simulated to occur in some (steep or flat) desert area with very low GW recharge. Negative WTD only occurs in places were the P_{30} is above the mean surface elevation e.g. parts of the Netherlands (Fig. 2b). Less than 10 cells experience WTD smaller than -10 m, which is very likely due to a not fully converged head solution.

In 2.1 % of all cells, GW head is simulated to be above the average land surface elevation, by more than 1 m in 0.3 % and by more than 100 m in 0.004 % of the cells. The shallow water table in large parts of the Sahara is caused by losing rivers (and some wetlands) that cannot run dry in the model, causing an overestimation of the GW table (section 2.2). Please note that the computed steady-state WTD certainly underestimates the steady WTD in GW depletion areas such as the High Plains Aquifer and the Central Valley in the USA (section S2), North-western India, North China Plain and parts of Saudi Arabia and Iran (Döll et al., 2014) as groundwater withdrawals are not taken into account in the presented steady-state simulation of G^3M.

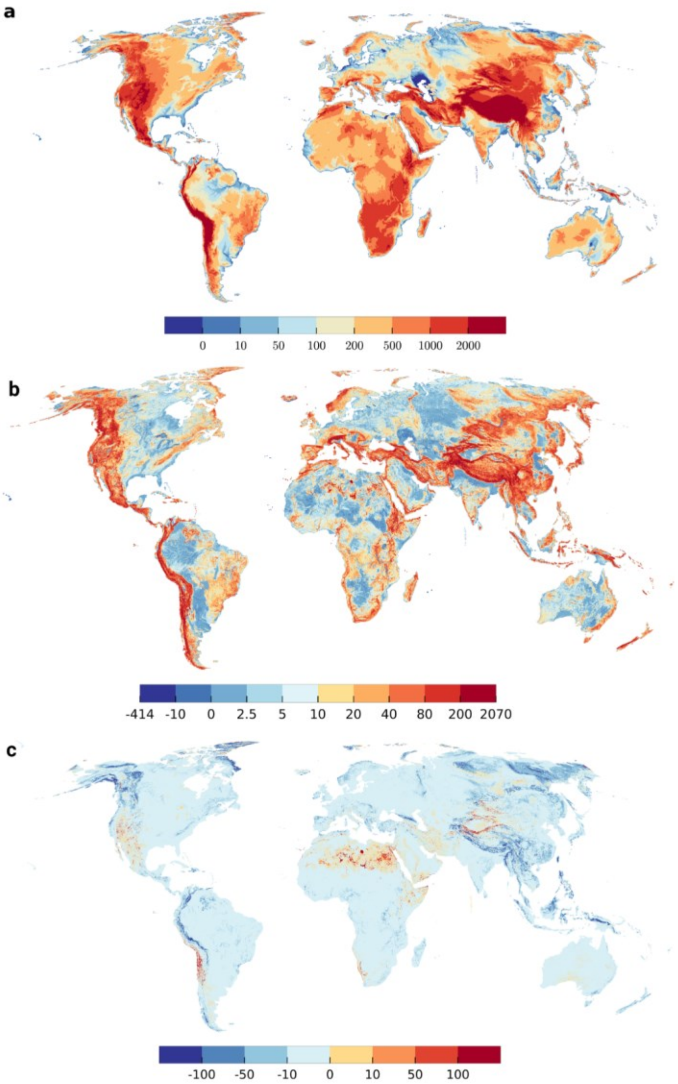

Figure 2 (a) Simulated steady-state hydraulic head of groundwater above sea level [***m***]. Maximum value 6375 m, minimum -414 m (Extremes included in dark blue and dark red). (b) Water table depth [***m***]. (c) Difference between 30th percentile of the 30'' land surface elevation per 5' grid cell (chosen elevation for surface water bodies h_{swb}) and simulated groundwater head [***m***]. Maximum value 1723 m, minimum value -1340 m (Extremes included in dark blue and dark red).

Kommentiert [RR5]: Added Figure from supplement

3.2 Global water budget

Inflows to and outflows from GW of all G^3M grid cells were aggregated according to the compartments ocean, river, lake, wetland, and diffuse GW recharge from soil (Fig. 3). The difference between the global sum of inflows and outflows is less than 10^{-6} %. This small volume balance error indicates the correctness of the numerical solution.

Total diffuse GW recharge, model input from WaterGAP, from soil is $10^4\ km^3\ year^{-1}$ and approximately equal to the simulated flow of GW to rivers (Fig. 3). Rivers are the ubiquitous drainage component of the model, followed by wetlands, lakes and the ocean boundary. According to G^3M, the amount of river water that recharges GW is more than one order of magnitude smaller than GW flow to rivers (Fig. 3). Possibly, flow from SW bodies to GW is overestimated, as outflow from SW is not limited by water availability in the SW, and depending on the hydraulic conductivity, Eqs. (4) and (5) can lead to rather large flows. Inflow from the ocean, which is more than two magnitudes smaller than outflow to ocean, occurs in regions where h_{swb} = P_{30} is below h_{ocean} e.g. the Netherlands. Globally, lakes and wetlands are computed to receive up to $10^3\ km^3\ year^{-1}$ of water from GW, and lose 1-2 orders of magnitude less.

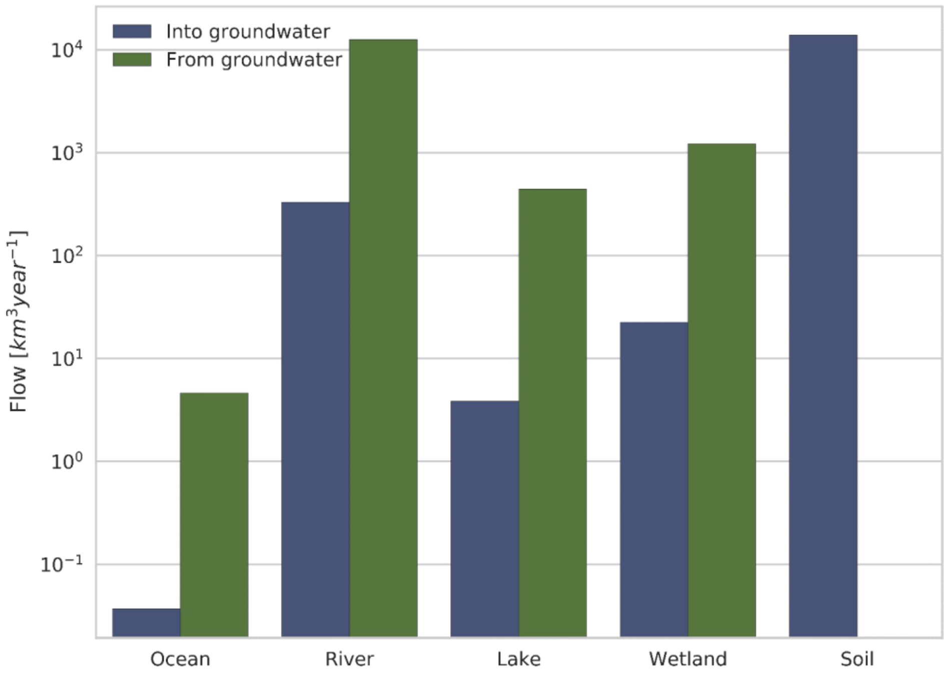

Figure 3 Global sums of flows from different compartments into or from GW at steady state. Flows into the GW are denoted by the color blue, flows out of the GW into the different compartments by green. The compartment soil is the diffuse GW recharge from soil calculated by WaterGAP.

3.3 GW-SW interactions

Figure 4 plots the spatial distribution of simulated flows from and to lakes and wetlands (Fig. 4a) as well as from and to rivers (Fig. 4b). Parallel to the overall budget (Fig. 3), the map reveals the globally large but locally strongly varying influence of lakes and wetlands (Fig. 4a). Rivers with riparian wetlands such as the Amazon River receive comparably small amounts of

GW as most of the GW is drained by the wetland (compare Figs. 4a and 4b). Similarly, areas dominated by wetlands and lakes (e.g. parts of Canada and Scandinavia) show less inflow for rivers (Fig. 4b). In G^3M, all SW bodies (rivers, lakes and wetlands) in a grid cell either loose or gain water. Consistent with negative or positive differences between h_{swb} and h_{aq} (Fig. 2c). 93 % of all grid cells contain gaining rivers, and only 7% losing rivers. Gaining lakes and wetlands are found in 12 % and 11 % of the cells, respectively, whereas only 0.2 % contain a losing lake or wetland.

Gaining rivers, lakes and wetlands with very high absolute Q_{swb} values over 500 mm year^{-1} (averaged over the grid cell area of approximately 80 km^2) can be found in the Amazon and Congo basin as well in Bangladesh and Indonesia, where GW recharge in very high (Fig. S4.4). Values below 1 mm year^{-1} occur in dry and in permafrost areas where groundwater recharge is small.

Losing SW bodies are caused by a combination of low GW recharge from soil (Fig. S4.4) and steep mountainous terrain (Fig. S4.7). While the steep Himalayas receive enough GW recharge to have gaining SW bodies, this is not the case for the much dryer mountain ranges around the Taklamakan desert in Central Asia, or mountainous Iran where SW bodies are losing. In the Sahara, GW recharge is so low that SW bodies are losing even in relatively high terrain.

Rivers lose more than 100 mm year^{-1} in Ethiopia and Somalia, West Asia, Northern Russia, the Rocky Mountains and the Andes whereas lower values can be observed in Australia and in the Sahara. High values of outflow from wetlands and lakes are found in Tibet, the Andes and northern Russia, lower values in the Sahara and Kazakhstan. The river Nile in the Northern Sudan and Egypt is correctly simulated to be a losing river (Fig. 4b), being an allogenic river that is mainly sourced from the upstream humid areas, including the man-made Lake Nasser (Elsawwaf et al., 2014) (Fig. 4a). Furthermore, the following lakes and riparian wetlands are simulated to recharge GW: parts of the Congo River, Lake Victoria, the Ijsselmeer, Lake Ladoga, the Aral Sea, parts of the Mekong Delta, the Great Lakes of North America. On the other hand, no losing stretches are simulated along the Niger River and its wetlands and almost none in the North-eastern Brazil even though that losing conditions are known to occur there (Costa et al., 2013; FAO, 1997). This is also true when the minimum elevation for SW bodies is assumed (compare Fig. S4.10) leading to the conclusion that the misrepresentation might be linked to an inadequate representation of the local geology.

Kommentiert [RR6]: #5 M.5

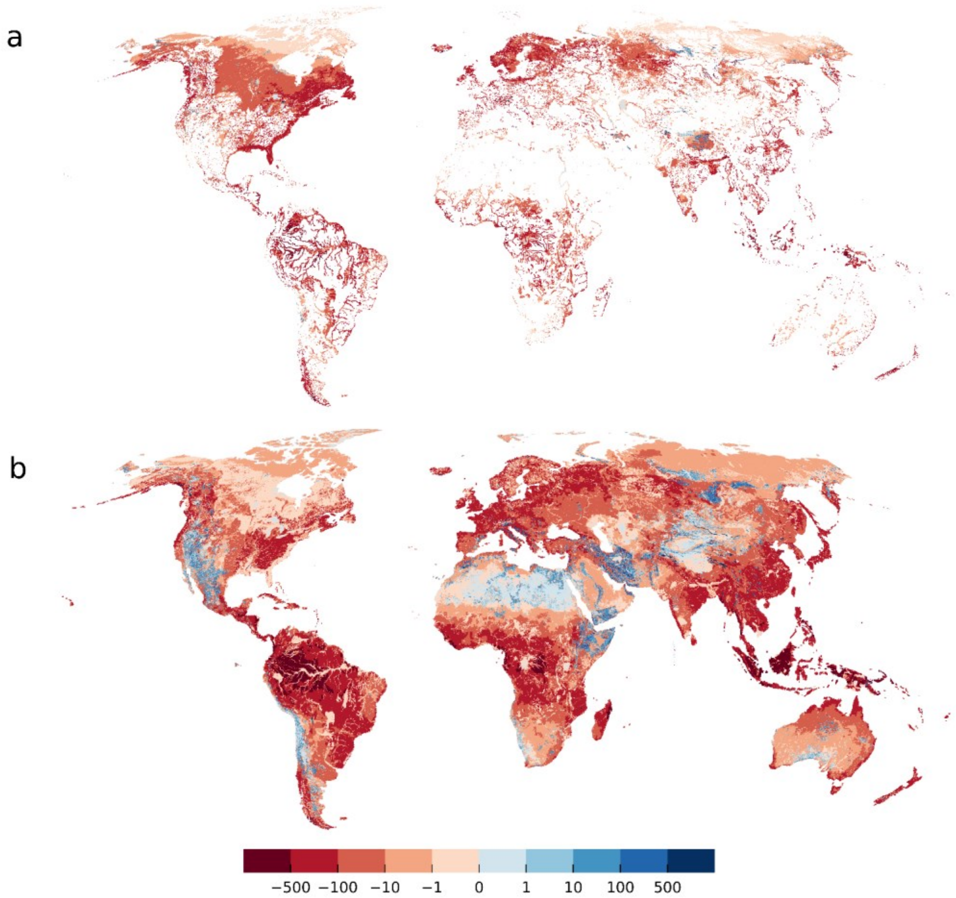

Figure 4 Flow $\boldsymbol{Q_{swb}}$ [$\boldsymbol{mm\ year^{-1}}$] from/to wetlands, lakes (a) and losing/gaining streams (b) with respect to the 5' grid cell area. Gaining surface water bodies are shown in red, surface water bodies recharging the aquifer in blue. Focused aquifer recharge occurs in arid regions, e.g. alongside the river Nile, and in mountainous regions where the average water table is well below the land surface elevation.

Simulated flows between GW and SW depend on assumed conductances for both rivers and lakes/wetlands (Eqs. (3), (4), (5)) shown in Fig. 5. Q_{swb} (Fig. 4) correlates positively with conductance. Conductance for gaining rivers correlates positively with GW recharge (Eq. (5) and Fig. S4.4). High river conductance values are reached in the tropical zone due to a high GW recharge but are capped at a plausible maximum value of $10^7\ m^2 day^{-1}$s in case of a river (section 2.1) (Fig. 5b). Lakes and wetlands, can have larger values of conductance due to their large areas, e.g. in Canada or Florida.

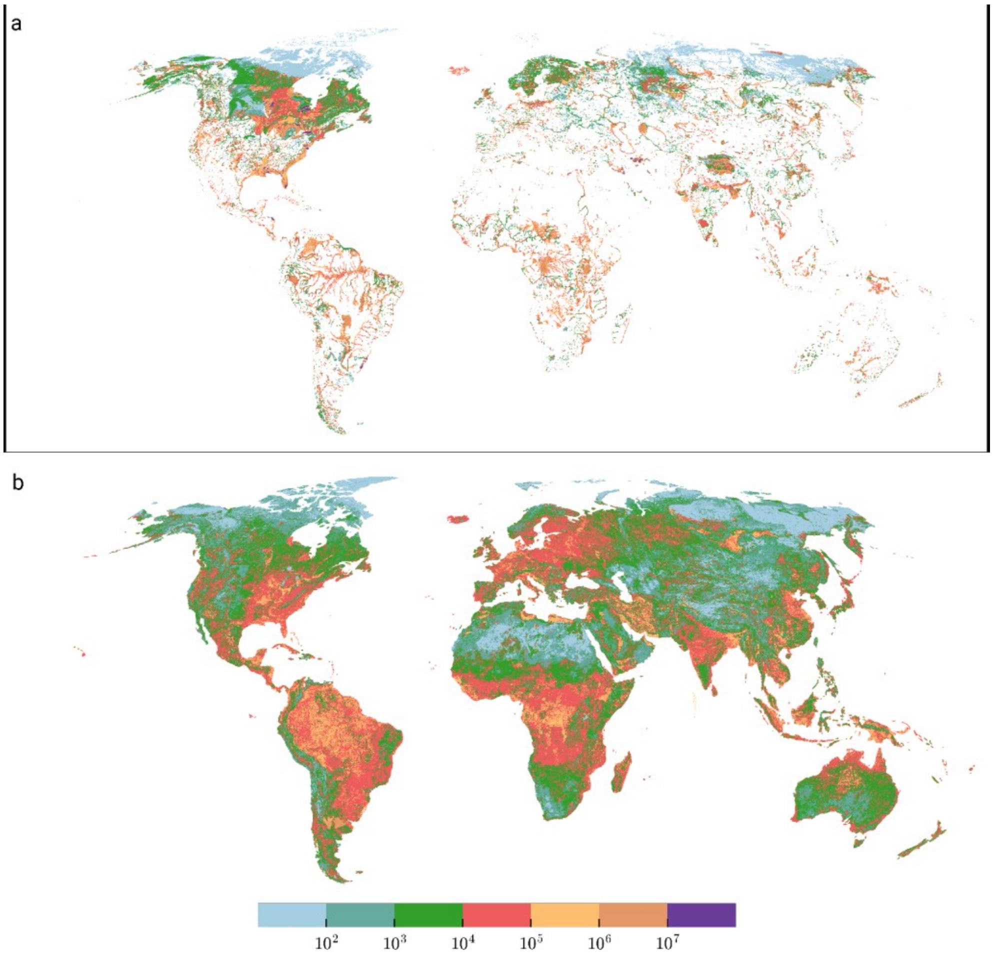

Figure 5 Conductance [$\boldsymbol{m^2 day^{-1}}$] of lakes and wetlands (a) and rivers (b). In regions close to the pole conductance is in general lower due to the influence of the low aquifer conductivity (losing conditions), and relatively small GW recharge due to permafrost conditions (only applies for gaining conditions). Max conductance of wetlands is 10^8 $\boldsymbol{m^2 day^{-1}}$.

3.4 Lateral flows

Figure 6 shows lateral GW flow (between grid cells, summing up over all model layers) in percent of the sum of diffuse GW recharge from soil and GW recharge from SW bodies. The percentage of recharge that is transported through lateral flow to neighbouring cells depends on five main factors: (1) hydraulic conductivity (Fig. S4.3), (2) diffuse GW recharge (Fig. S4.4), (3) losing or gaining SW bodies (Fig. 4), (4) their conductance (Fig. 5) and (5) the head gradients (Fig. 2a).

On large areas of the globe, where GW discharges to SW bodies, the lateral flow percentage is less than 0.5% of the total GW recharge to the grid cell, as most of the GW recharge in a grid cells is simulated to leave the grid cell by discharge to SW bodies. For example, in the permafrost regions, the assumed very low hydraulic conductivity limits the outflow to neighbouring cells of the occurring recharge, leading to these very low percent values. Such values also occur in regions with high SW conductances and rather low hydraulic conductivities, e.g. in the Amazon Basin. Values of more than 5% occur where hydraulic conductivity is high even if the terrain in rather flat, such as in Denmark. Higher values may occur in case of gaining

SW bodies in dry areas like Australia or in the Taklamakan desert. They can also be observed in mountainous regions where large hydraulic gradients can develop. In mountains with gaining surface water bodies, lateral outflows may even exceed GW recharge of the cell. In grid cells where SW bodies recharge the GW, outflow tends to be a large percentage of total GW recharge as there is no outflow from GW other than in lateral direction, and values often exceed 100% (Fig. 6).

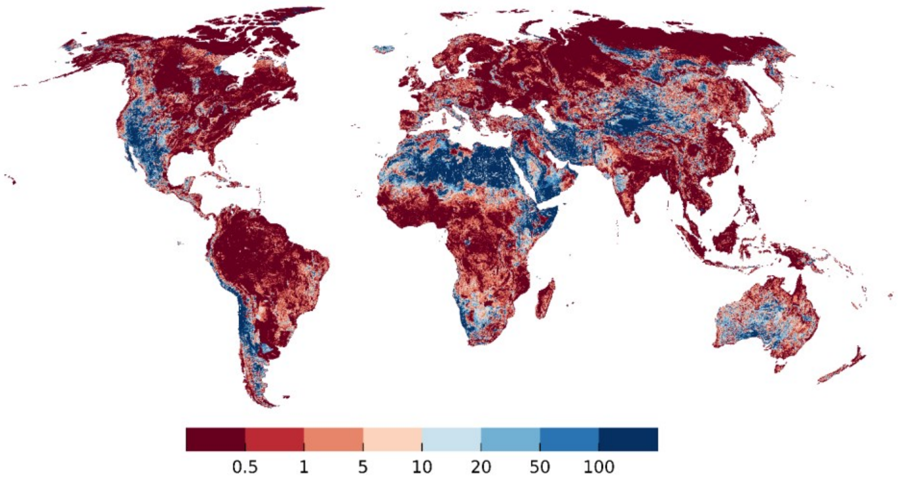

Figure 6 Percentage of GW recharge from soil and surface water inflow that is transferred to neighboring cells through lateral out flow (sum of both layers). Grid cells with zero total GW recharge are shown in white (a few cells in the Sahara and the Andes).

3.5 Comparison to groundwater well observations and the output of two higher-resolution models

Global observations of WTD were assembled by Fan et al. (2007; 2013). We selected only observations with known land surface elevation and removed observations where a comparison to local studies suggested a unit conversion error. This left total of 1,070,402 WTD observations. An "observed head" per 5' model cell was then calculated by first computing hydraulic head of each observation by subtracting WTD from the 5' average of the 30'' land surface elevation and then calculating the arithmetic mean of all observations within the 5' model cell. This resulted in 78,664 grid cells with observations out of a total of 2.2 million G³M top-layer grid cells. Multiple obstacles limit the comparability of observations to simulated values. (1) Observations were recorded at a certain moment in time influenced by seasonal effects and abstraction from GW, whereas the simulated heads represent a natural steady-state condition. (2) Observation locations are biased towards river valleys and productive aquifers. (3) Observations may be located in valleys with shallow local water tables too small to be captured by a coarse resolution of 5'.

Simulated steady-state hydraulic heads in the upper model layer are compared to observations in Fig. 7. Shallow GW is generally better represented by the model than deeper GW. Especially the water table in mountainous areas is underestimated, which may be related to observations in perched aquifers caused by low permeability layers (Fan et al., 2013) that are not represented in G^3M due to lacking information. Because the steady-state model cannot take into account the impact of GW abstraction, the computed WTD values are considerably smaller than currently observed values in GW depletion areas like the Central Valley in California (where once wetlands existed before excessive GW use depleted the aquifer) and the High Plains Aquifer in the Midwest of the USA. Still, the elevation of the GW table in the non-depleted Rhine valley in Germany is overestimated, too. Overestimates in the Netherlands may partially be due to artificial draining. Figure 8a shows the hydraulic head comparison as scatter plot. Overall, the simulation results tend to underestimate observed hydraulic head but much less than the steady-state model presented by de Graaf et al. (2015) (their Fig. 6).

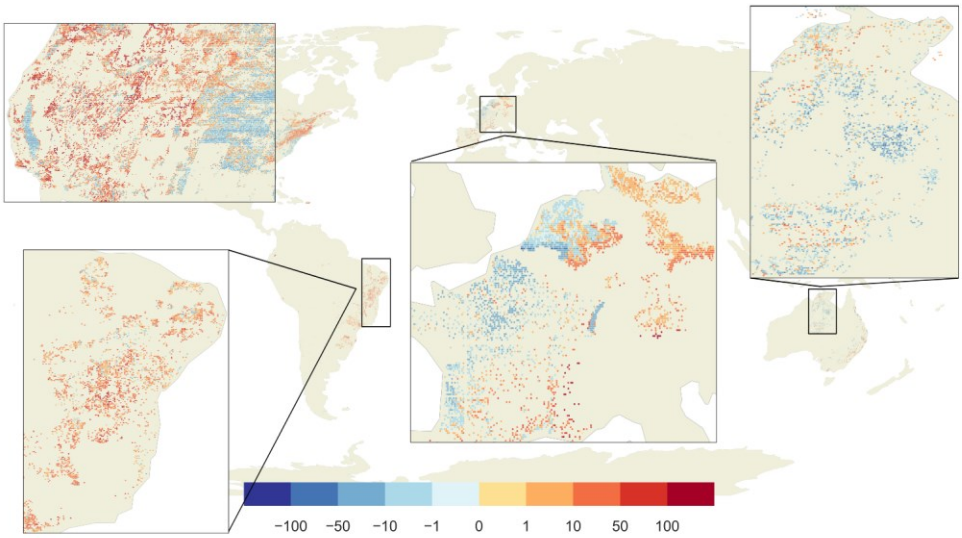

Figure 7 Differences between observed and simulated hydraulic head [$\boldsymbol{m}$]. Red dots show areas where the model simulated deeper GW as observed, blue shallower GW. In the grey areas, no observations are available.

To compare performance of G^{3}M to the steady-state results of two high-resolution model of Fan et al. (2013) and ParFlow (Maxwell et al., 2015) (Table 1), heads in 30" (Fan et al., 2013) and 1 km (ParFlow) grid cells were averaged to the G^{3}M 5' grid cells. The comparison of 5' observations to the 5' average of ParFlow seems to be consistent with the 1 km model comparison in Maxwell et al. (2015) (their Fig. 5), even though over/under -estimates in the original resolution seem to be smoothed out by averaging to 5' (not shown). The heads of Fan et al. (2013) fit better to observations than G^{3}M heads, with less underestimation (Fig. 8b) and a RMSE (Root Mean Square Error) of 26.0 m compared to the 32.4 m RMSE of G^{3}M. The comparison of G^{3}M heads to Fan et al. (2013) values for all 5' grid cells, which are also the initial heads of G^{3}M and the basis to compute river conductances, show that heads computed with the G^{3}M are mostly much lower except in regions with a shallow GW (Fig. 8c), RMSE is 46.7 m. This cannot be attributed to the 100 times lower spatial resolution per se but to the selection of the 30th percentile of the 30’’ as the SW drainage level. Outliers in the upper half of the scatter plot, with much larger G^{3}M heads than the initial values (Fan et al., 2013), are mainly occurring in steep mountain areas like the Himalayas where the 5’ model is not representing smaller valleys with a lower head. For the continental US, the computationally expensive 1-km integrated hydrological model ParFlow (Maxwell et al., 2015) fits much better to observations than G^{3}M (Figs. 8d, e), with a RMSE of 14.3 m (ParFlow) compared to 34.2 m (G^{3}M). G^{3}M produces a generally lower water table (Fig. 8f), a main reason being that ParFlow assumes an impermeable bedrock at a depth of 100 m below the land surface elevation.

The global map of head comparison (Fig. 7) suggests that G^{3}M performs resonably well in flat areas compared to mountainous regions. This is corroborated by Fig. 9 that shows the difference between observed and simulated hydraulic heads for five land surface elevation categories. It is evident that model performance deteriorates with increasing land surface elevation and positively correlates with variations of land surface elevation within each grid cell (Fig. S4.7).

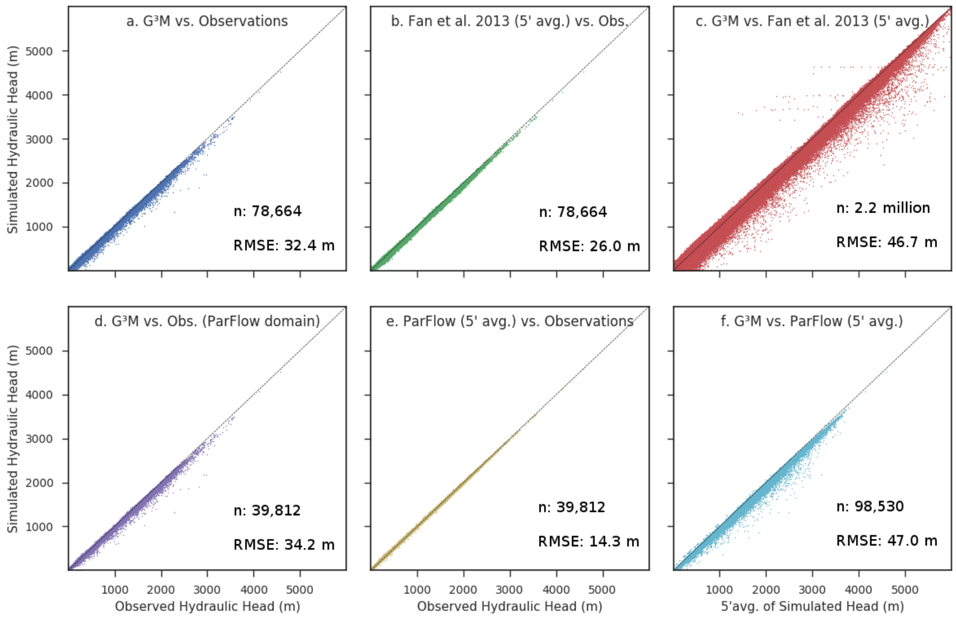

Figure 8 Scatterplots of simulated vs. observed hydraulic head and inter-model comparison of heads. (Upper panel) The steady-state run of G³M vs. observations (a), the 5' average of the equilibrium head of Fan et al. (2013) vs. observations (b) and the avg. equilibrium vs. G³M (c). (Lower panel) The steady-state run of G³M vs. observations only for the ParFlow domain (d), the 5' average of the ParFlow average annual GW table (Maxwell et al., 2015) vs. observations (e) and the steady-state run of G³M vs. 5' average of the ParFlow average annual GW table (f).

Plotting hydraulic head instead of WTD has the disadvantage that the goodness of fit is dominated by the topography as the observed heads are calculated based on the surface elevation of the model. Well observations provide WTD and only sometimes contain complementary data specifying the elevation at which the measurements were taken. Even though hydraulic heads are a direct result of the model and are forcing lateral GW flows, WTD is more relevant for processes like capillary rise. For G³M, there is almost no correlation between WTD observations and simulated values. To our knowledge, no publication on large-scale GW modeling has presented correlations of simulated with observed WTD.

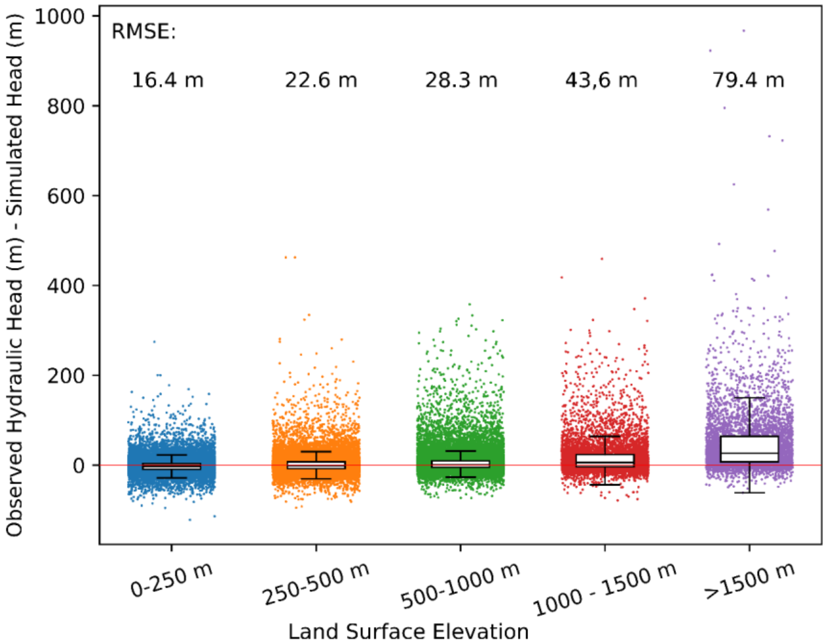

Figure 9 Observed minus simulated hydraulic head for different land surface elevation categories. The whiskers of the boxplots show the interquartile range.

3.6 Testing sensitivity of computed steady-state hydraulic heads to parameter values and spatial resolution

To limit the computational effort for assessing model sensitivity to both parameters and grid size, we selected New Zealand as a representative "small world" that includes a complex topography and the ocean as a clear boundary condition. All inputs and parameters are the same as in the global 5' model.

3.6.1 Parameter sensitivity

To determine to which parameters simulated hydraulic heads are most sensitive to, we used the established sensitivity tool UCODE 2005 (Hill and Tiedeman, 2007) to compute composite scaled sensitivity (CSS) values for seven model parameters (S3). CSS of h_{swb} is orders of magnitude larger than the CSS of the other parameters. This confirms our observations during model development when an appropriate value for had to be found (section 2.2). The second-most important parameter is K_{aq}, the third most important R_g. CSS of the conductance of lakes is one magnitude less than CSS of R_g but as only few cells contain lakes, the CSS value that averages over all grid cells indicates a large sensitivity to c_{Lakes} for grid cells with lakes. Simulated hydraulic heads were found to be rather insensitive to changes in the conductance of rivers, wetlands, and ocean boundary.

3.6.2 Sensitivity to spatial resolution

The extremely high sensitivity of simulated hydraulic heads to the choice of h_{swb} (section 3.6.1) and the better agreement of the continental models with a higher spatial resolution of approx. 30'' (the Fan et al. (2013) model and ParFlow (section 3.5)) motivated us to run G³M for New Zealand with a spatial resolution of 30'', to understand the impact of spatial resolution on simulated hydraulic heads. The 30'' G³M model uses the same input as the 5' model except for the land surface elevation, h_{swb}

and the location of rivers. While the total length and width of the rivers is equal in both models, a river is assumed to exist in all 5' grid cells, the river is concentrated, in the 30" model, to a few grid cells with each 5' grid cell. The river cell locations at 30" are determined based on 30" HydroSHEDS (hydrosheds.org) information on flow accumulation. Starting with the 30" cell with the highest number of upstream cells per 5' cell, a river is added to this 30" cell using the length and information of HydroSHEDS till the size of the river of the 5' model is reached for all 30" cells within a 5' cell. The areal fractions of all other SW bodies from 5' grid data where used for all 30" grid cells within the 5' grid cell. h_{swb} is set to the land surface elevation.

Figure 10 compares the performance of the two model versions. The comparison of simulated hydraulic head to observations for the Canterburry region (Westerhoff et al., 2018) shows that the overall performance of the 30" model is better, with a smaller RMSE of 26.7 m as compared to a RMSE of 53.8 m in case of the original spatial resolution of 5'. The 30" model results in generally lower simulated hydraulic heads leading to a closer fit to the observed values. This is likely caused by the improved estimation of SW body elevation, which generally leads to lower estimates of h_{swb}. On the other hand, overestimates of observed hydraulic heads prevails in the 30" model, even though h_{swb} was set to the land surface elevation, indicating that further investigation is necessary. The underestimates are likely due to large GW abstractions for irrigation in the particular region (Westerhoff et al., 2018).

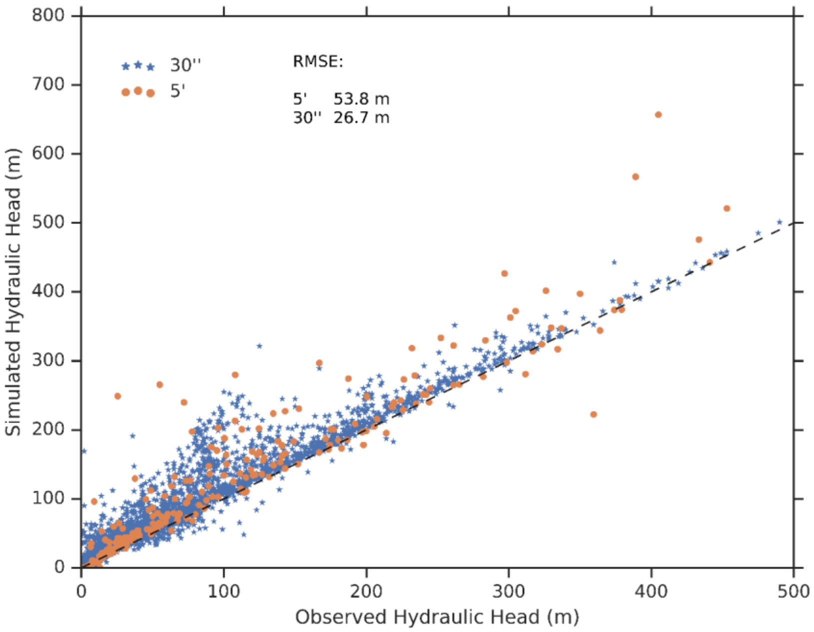

Figure 10 Low (5') vs. high (30'') spatial resolution for the Canterburry region in New Zealand, comparison of observed vs. simulated hydraulic head for both resolutions. The observed head is the geometric mean per 5' and 30" respectively.

Kommentiert [RR7]: Modified

4 Discussion

The objective of global gradient-based groundwater flow modelling with G³M is to better simulate water exchange between SW and GW in the GHM WaterGAP, for example for an improved estimation of GW resources in dry regions of the globe that are augmented by focused recharge from SW bodies. We assume that the fully coupled model will lead to an improved WaterGAP performance during droughts with an increased drop in streamflow due to the now possible switch from gaining to losing conditions. Presumable a calibration of c_{swb} and c_{riv} is necessary to achieve a good discharge performance. The presented steady-state model is a first step in this direction. It helped to understand basic model behaviour, e.g. the sensitivity

Kommentiert [RR8]: #5 2)

to SW body elevation, and the necessary improvement of its parameterization, before moving to the more complex integrated transient model. The reduced runtime of the steady-state model in comparison to a fully integrated transient run supported the investigation of parameter sensitivity and sensitivity to spatial resolution. Additionally, the presented steady-state model can be used in future fully integrated transient runs as initial condition.

A major challenge for simulating GW-SW interactions (but also capillary rise) at the global scale is the large size of grid cells that is required due to computational constraints. Within the 5' grid cells, land surface elevation at the scale of 30" very often varies by more than 20 m, and often by 200 m and more (Fig. S4.1), while the vertical position of the cell and the hydraulic head are approximated in the model by just one value. The question is whether head-dependent flows between grid cells, between GW and SW and from GW to soil (capillary rise) can be simulated successfully at the global scale, i.e. whether an improved quantification of these flows as compared to the simple linear reservoir model currently used in most GHMs can be achieved by this approach. This question cannot be answered before a dynamic coupling of G³M with a global hydrological model has been achieved but one may speculate that some innovative approach to take into account the elevation variations within the grid cells is needed.

It is difficult to assess quality of the presented steady-state G³M results. Model performance is assessment is hindered by data availability and the coarse model resolution. (1) To our knowledge the data collection of depth to groundwater by Fan et al. (2013) is unique. However, they do not represent steady-state values. Apart from depth to groundwater observations, hardly any relevant data is available at the global scale. Especially exchange between surface water and groundwater is difficult to measure even at the local scale. Therefore, we compared G³M results with the results from other large-scale models. Comparison to the results of catchment-scale groundwater flow models is planned for transient runs that will be possible after integration into WaterGAP. (2) Scale differences make the comparison to point observations of depth to groundwater difficult. Often, observations are biased towards alluvial aquifers in valleys. The calculated hydraulic head of the grid cell may represent the average groundwater level per grid cell correctly but can be still far off the local observations of depth to groundwater. As the current model only represents an uncalibrated natural steady-state, a comparison to observations only provides a first indicator where the model and the performance measurements needs to be improved as we move to a fully transient model.

The presented development of the uncoupled steady-state global GW flow model enabled us to better understand how the spatial hydraulic head pattern relates to the fundamental drivers topography, climate and geology (Fan et al., 2007) and how the interaction to SW bodies governs the global head distribution. Simulated depth to groundwater is particularly affected by the assumed hydraulic head in SW bodies, the major GW drainage component in the model. As rivers represent a natural occurring drainage at the lowest point in a given topography, one would assume that the minimum elevation 30" land surface elevation per 5' grid cell is a reasonable choice. Experiments have shown that this will induce a head distribution well below the average 5' elevation that is much below observations of Fan et al. (2013). We also tested setting h_{swb} to the average elevation of all "blue" cells (with a WTD of less than 0.25 m) of the steady-state 30" water table results of Fan et al. (2013) that indicate the locations were GW discharges to the surface or SW bodies. This leads to an overall underestimation of the observed hydraulic heads (Fig. S4.9) as the assumed SW elevation is too low. Furthermore, it leads to an increase in losing SW bodies (comp. Fig. S4.10 with Fig. 4). However, it is difficult to judge whether this improves the simulation. More stretches of the Nile and its adjacent wetlands and also of the Niger wetlands and rivers in North-eastern Brazil are losing in case of lower h_{swb}, which appears to reasonable. Additionally, choosing the average as SW elevation provides on the on hand a better fit to observations (Fig. S4.9 right) but leads to a world- wide flooding (Fig. S4.9) and a much longer convergence time due to an increased oscillation between gaining and losing conditions.

The problem is very likely one of scale. All three models (Fan et al., ParFlow, and G³M 30") (Table 1) high-resolution models, even the simple one of Fan et al. (2013), fit better to observations than the 5' model G³M (Fig. 8,10). In case of high resolution, there are a number of grid cells at an elevation above the average 5' land surface elevation, leading to higher hydraulic heads in parts of the 5' area that drain towards the SW body in a lower 30" grid cell. In case of the low spatial resolution of 5' in which h_{swb} is set to the elevation of the fine-resolution drainage cell, the 5' hydraulic head is rather close to this (low) elevation (Fig. S4.8 center), resulting in an underestimation of hydraulic head and thus an overestimation of WTD. While it is plausible and necessary to assume that there is SW-GW interaction within each of in the approximately 80 km^2, this is not the case for the two orders of magnitude smaller 30" grid cells. Thus, with the high resolution, heads are not strongly controlled everywhere by the head in SW bodies. Selecting the 30th percentile of the 30" land surface elevation as h_{swb} was found, by trial-and-error, to lead to a hydraulic head distribution that fits reasonably well to observed head. It avoids that the simulated GW table drops too low while avoiding the excessive flooding that occurs if h_{swb} is set to the average of 30" land surface elevations, i.e. the 5' land surface elevation (Fig. S4.9).

The constraint that the selected h_{swb} value puts on simulated hydraulic heads is also linked to the conductance of the SW bodies. A higher conductance will lead to aquifer heads closer to h_{swb}. If the hydraulic head drops below the bottom level of the SW body, the hydraulic gradient is assumed to become 1 and the SW body recharges the GW with a rate of K_{aq} per unit SW body area. In case of a K_{aq} value of 10^{-5} m s^{-1}, the SW body would lose approximately 1 m of water each day. Further investigations are needed regarding the appropriate choice of SW body elevation and conductance. The simple conductance approach applied in G^3M could possibly be improved by the approach by Morel-Seytoux et al. (2017) who proposes an analytical, and physically based, estimate of the leakance coefficient for coarse scale models based on river and aquifer properties.

Kommentiert [RR9]: #5 M.7

De Graaf et al. (2015) set their SW head (h_{swb}) to the mean land surface elevation (Table 1) of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge as compared to P_{30} in the 5' G^3M. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to improve the discharge to rivers (Sutanudjaja et al., 2014). On the other hand, the additional drainage leads to drainage of water even if the hydraulic head is below the SW elevation, which might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity to SW body elevation.

Kommentiert [RR10]: #5 1)

Due to the spatial resolution and lack of data G^3M does not capture the actual variability of topography, aquifer depth (Richey et al., 2015) and (vertical) heterogeneity of subsurface properties. The lack of information about the three-dimensional distribution of hydraulic conductivity is expected to negatively impact the quality of simulated GW flow. For example, the lateral conductivity and connectivity of groundwater along thousands of kms from e.g. the Rocky Mountains in the Central USA to the coast as well as the vertical connectivity is likely to be overestimated by G^3M, as vertical faults and interspersed aquitards are not represented; this is expected to lead to an underestimation of hydraulic head in those mountainous areas.

5 Conclusions

We have presented the concept and first results of the new global gradient-based 5' GW flow model G^3M that is to be integrated into the 0.5° GHM WaterGAP. The uncoupled steady-state model has provided important insights into challenges of global GW flow modelling mainly related to the necessarily large grid cells size (5' by 5'). In addition, first global maps of SW-GW interactions were generated. Simulated heads were found to be strongly impacted by assumptions regarding the interaction with SW bodies, in particular the selected elevation of the SW table. We have demonstrated that simulated G^3M hydraulic heads fit better to observed heads than the heads of the comparable steady-state GW model of de Graaf et al. (2015), without requiring additional drainage. Furthermore, we provided insights into how the choice of surface water body elevation h_{swb} affects model outcome. In a next step, approaches for utilizing high-resolution topographic data to improve the selection of h_{swb} will be investigated.

The presented results are the first step towards a fully coupled model in which SW heads are jointly computed, also taking into account the impact of SW and GW abstraction. Especially the interaction with SW bodies that can run dry will make the G^3M behavior more realistic. The fully coupled model will simulate transient behaviour reflecting climate variability and change. Simulated hydraulic head dynamics will be compared to observed head time series as well as to the output of large-scale regional models, while total water storage variations will be compared to GRACE satellite data. However, it will be challenging to judge the dynamics of GW and the quality of simulated GW-SW interactions due to a scarcity of observations.

6 Code and data availability

The model-framework code is available at globalgroundwatermodel.org or at DOI: 10.5281/zenodo.1175540 with a description on how to compile and run a basic GW model. The code is available under the GNU General Public License 3. Model output is available at DOI: 10.5281/zenodo.1315471.

Author contribution

RR led conceptualization, formal analysis, methodology, software, visualization, and writing of the original draft. LF and SM supported review and editing as well as the development of the methodology. TT, CD supported review. PD supervised the work of RR and made suggestions regarding analysis, structure, and wording of the text and design of tables and figures.

Kommentiert [RR11]: Added

Acknowledgments

Part of this study was funded by a Friedrich-Ebert foundation Ph.D. fellowship. We are very grateful to Ying Fan and Gonzalo Miguez-Macho for fruitful discussions and data provisioning. We thank the reviewers for their thoughtful comments that helped to improve the manuscript. Furthermore, we like to thank Alexander Wachholz for contributing to the New Zealand figures and Rogier Westerhof for providing the observations for New Zealand.

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrological Sciences Journal, 48, 317–337, doi:10.1623/hysj.48.3.317.45290, 2003.

Allen, P. M., Arnold, J. C., and Byars, B. W.: Downstream channel geometry for use in planning-level models, J Am Water Resources Assoc, 30, 663–671, doi:10.1111/j.1752-1688.1994.tb03321.x, 1994.

Belcher, W. R. and Sweetkind, D. S.: Death Valley regional groundwater flow system, Nevada and California-Hydrogeologic framework and transient groundwater flow model, U.S. Geological Survey Professional Paper, 398, 2010.

Belcher, W.R. (E.): Death Valley Regional Ground-water Flow System, Nevada and California-Hydrogeologic Framework and Transient Ground-water Flow Model, U.S. Geological Survey Professional Paper, pp. 408, 2004.

Costa, A. C., Foerster, S., de Araújo, J. C., and Bronstert, A.: Analysis of channel transmission losses in a dryland river reach in north-eastern Brazil using streamflow series, groundwater level series and multi-temporal satellite data, Hydrol. Process., 27, 1046–1060, doi:10.1002/hyp.9243, 2013.

de Graaf, I. E.M. de, Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, Hydrol. Earth Syst. Sci., 19, 823–837, doi:10.5194/hess-19-823-2015, 2015.

de Graaf, I. E.M. de, van Beek, R. L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.P.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, Advances in Water Resources, 102, 53–67, doi:10.1016/j.advwatres.2017.01.011, 2017.

Dogrul, E., Brush, C., and Kadir, T.: Groundwater Modeling in Support of Water Resources Management and Planning under Complex Climate, Regulatory, and Economic Stresses, Water, 8, 592, doi:10.3390/w8120592, 2016.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12, 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B.R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, Journal of Geodynamics, 59-60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, Journal of Hydrology, 270, 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, doi:10.1002/2014WR015595, 2014.

Dustin, E.: Effective software testing: 50 specific ways to improve your testing, 5th printing, Addison-Wesley, Boston, Mass. [etc.], XV, 271 s. ;, 2006.

Eisner, S.: Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients, University of Kassel, Kassel, Germany, 2016.

Elsawwaf, M., Feyen, J., Batelaan, O., and Bakr, M.: Groundwater-surface water interaction in Lake Nasser, Southern Egypt, Hydrol. Process., 28, 414–430, doi:10.1002/hyp.9563, 2014.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science (New York, N.Y.), 339, 940–943, doi:10.1126/science.1229881, 2013.

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, J. Geophys. Res., 112, doi:10.1029/2006JD008111, 2007.

FAO: Irrigation potential in Africa: http://www.fao.org/docrep/w4347e/w4347e0i.htm, 1997, last access: 1 May 2018.

Faunt, C. C. (Ed.): Groundwater Availability of the Central Valley Aquifer, California, U.S. Geological Survey Professional Paper, 1776, 225 pp., 2009.

Faunt, C. C., Provost, A. M., Hill, M. C., and Belcher, W. R.: Comment on "An unconfined groundwater model of the Death Valley Regional Flow System and a comparison to its confined predecessor" by R.W.H. Carroll, G.M. Pohll and R.L. Hershey [Journal of Hydrology 373/3–4, pp. 316–328], Journal of Hydrology, 397, 306–309, doi:10.1016/j.jhydrol.2010.11.038, 2011.

Faunt, C. C., Sneed, M., Traum, J., and Brandt, J. T.: Water availability and land subsidence in the Central Valley, California, USA, Hydrogeol J, 24, 675–684, doi:10.1007/s10040-015-1339-x, 2016.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nature Geosci, 9, 161–167, doi:10.1038/ngeo2590, 2016.

Gleeson, T., Moosdorf, N., Hartmann, J., and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898, doi:10.1002/2014GL059856, 2014.

Gleeson, T., Wada, Y., Bierkens, M. F. P., and van Beek, L. P. H.: Water balance of global aquifers revealed by groundwater footprint, Nature, 488, 197–200, doi:10.1038/nature11295, 2012.

Gossel, W., Ebraheem, A. M., and Wycisk, P.: A very large scale GIS-based groundwater flow model for the Nubian sandstone aquifer in Eastern Sahara (Egypt, northern Sudan and eastern Libya), Hydrogeology Journal, 12, 698–713, doi:10.1007/s10040-004-0379-4, 2004.

Harbaugh, A. W.: MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process, US Department of the Interior, US Geological Survey Reston, 2005.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophysics, Geosystems, 13, doi:10.1029/2012GC004370, 2012.

Hill, M. C. and Tiedeman, C. R.: Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty, Wiley-Interscience, Hoboken, N.J., 2007.

Konikow, L. F.: Contribution of global groundwater depletion since 1900 to sea-level rise, Geophys. Res. Lett, 38, doi:10.1029/2011GL048604, 2011.

Krakauer, N. Y., Li, H., and Fan, Y.: Groundwater flow across spatial scales: importance for climate modeling, Environ. Res. Lett., 9, 34003, doi:10.1088/1748-9326/9/3/034003, 2014.

Lange, S.: EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), 2016.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, 296, 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Maxwell, R. M., Condon, L. E., and Kollet, S. J.: 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, 923–937, doi:10.5194/gmd-8-923-2015, 2015.

Mehl, S.: Use of Picard and Newton iteration for solving nonlinear ground water flow equations, Ground Water, 44, 583–594, doi:10.1111/j.1745-6584.2006.00207.x, 2006.

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, J. Geophys. Res., 112, D13108, doi:10.1029/2006JD008112, 2007.

Morel-Seytoux, H. J., Miller, C. D., Miracapillo, C., and Mehl, S.: River Seepage Conductance in Large-Scale Regional Studies, Ground Water, 55, 399–407, doi:10.1111/gwat.12491, 2017.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Naff, Richard L., and Edward R. Banta: The US Geological Survey modular ground-water model-PCGN: a preconditioned conjugate gradient solver with improved nonlinear control, Geological Survey (US), 1331, 2008.

Niswonger, R. G., Panday, S., and Ibaraki, M.: MODFLOW-NWT, a Newton formulation for MODFLOW-2005, US Geological Survey Techniques and Methods, 6, 2011.

Poeter, E. P., Hill, M. C., Lu, D., Tiedeman, C., and Mehl, S. W.: UCODE 2014, with new capabilities to define parameters unique to predictions, calculate w eights using simulated values, estimate parameters w ith SVD, evaluate uncertainty with MCMC, and more, Integrated Groundwater Modeling Center IGWMC, of the Colorado School of Mines, 2014.

Reinecke, R.: G³M-f a global gradient-based groundwater modelling framwork, JOSS, 3, 548, doi:10.21105/joss.00548, 2018.

Richey, A. S., Thomas, B. F., Lo, M.-H., Famiglietti, J. S., Swenson, S., and Rodell, M.: Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework, Water Resour. Res, 51, 5198–5216, doi:10.1002/2015WR017351, 2015.

Saad, Y.: ILUT: A dual threshold incomplete LU factorization, Numer. Linear Algebra Appl., 1, 387–402, doi:10.1002/nla.1680010405, 1994.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, PNAS, 109, 9320–9325, doi:10.1073/pnas.1200311109, 2012.

Sheets, R. A., Hill, M. C., Haitjema, H. M., Provost, A. M., and Masterson, J. P.: Simulation of water-table aquifers using specified saturated thickness, Ground Water, 53, 151–157, doi:10.1111/gwat.12164, 2015.

Stonestrom, D. A., Constantz, J., Ferre, T. P.A., and Leake, S. A.: Ground-water recharge in the arid and semiarid southwestern United States, U.S. Geological Survey Professional Paper 1703, 414 p, 2007.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin, Hydrol. Earth Syst. Sci., 15, 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, Water Resour. Res., 50, 687–705, doi:10.1002/2013WR013807, 2014.

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I., and Treidel, H.: Ground water and climate change, Nature Climate change, 3, 322–329, doi:10.1038/nclimate1744, 2012.

UNESCO-IHP, IGRAC, WWAP: GEF-TWAP Methodology Transboundary Aquifers.: http://isarm.org/sites/default/files/resources/files/TWAP%20Methodology%20Groundwater%20Component%20%28Revised%20Aug%202012%29.pdf, 2012, last access: 9 October 2017.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, doi:10.1029/2010WR009791, 2011.

Vergnes, J.-P., Decharme, B., Alkama, R., Martin, E., Habets, F., and Douville, H.: A Simple Groundwater Scheme for Hydrological and Climate Applications: Description and Offline Evaluation over France, J. Hydrometeor, 13, 1149–1171, doi:10.1175/JHM-D-11-0149.1, 2012.

Vergnes, J.-P., Decharme, B., and Habets, F.: Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model, J. Geophys. Res. Atmos., 119, 11, doi:10.1002/2014JD021573, 2014.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F., and Alcamo, J.: Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe, Journal of Hydrology, 424-425, 238–251, doi:10.1016/j.jhydrol.2012.01.005, 2012.

Wada, Y.: Modeling Groundwater Depletion at Regional and Global Scales: Present State and Future Prospects, Surv Geophys, 37, 419–451, doi:10.1007/s10712-015-9347-x, 2016.

Wada, Y., van Beek, L. P. H., and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: A global assessment, Water Resour. Res., 48, doi:10.1029/2011WR010562, 2012.

Westerhoff, R., White, P., and Miguez-Macho, G.: Application of an improved global-scale groundwater model for water table estimation across New Zealand, Hydrol. Earth Syst. Sci, 22, 6449–6472, doi:10.5194/hess-22-6449-2018, 2018.

Table 1 Comparison of global- and continental-scale groundwater models

Aspect	G³M	de Graaf et al. (2015; 2017)	Fan et al. (2013)	ParFlow
Extent	Global	Global	Global	Continental USA
Resolution	5'	6'	30"	1 km
Software	G³M-f	MODFLOW	Unnamed	ParFlow
Computational expense	Medium	Medium	High	Very high
Flow representation	3D saturated	3D saturated	2D saturated	3D saturated/unsaturated
Time scale	Steady-state/(transient)	Steady-state/transient	Steady-state	Steady-state
Vertical layers	2	2	1	5
Full coupling possible	Yes	No (Conceptual issue)	No	Yes (already coupled)
In-memory coupling	Yes	No	N/A	Yes
Constant saturated thickness	Yes	Yes	No	No
Impermeable bottom	No	No	No	Yes
Surface water body location	In every cell	In almost every cell	No surface water	Created during simulation
Surface water body elevation	P_{30} of 30" DEM	Avg. of 30" DEM	N/A (outflow if WTD < 0.25 m)	N/A
Deviation from observations	Large	Very large	Medium	Medium

Table 2 Model parameter values, input data sources and other information about the steady-state simulation.

Parameter	Symbol	Units	Description	Eq. No.
Landmask	-	-	Location and area of 2161074 cells at 5' resolution based on WaterGAP (Eisner, 2016))	-
GW recharge	R_g	L^3T^{-1}	Mean annual diffuse GW recharge 1901–2013 of WaterGAP 2.2c (Müller Schmied et al., 2014) forced with EWEMBI (Lange, 2016), spatial resolution 0.5° (Fig. S4.4)	2,5,S1
Hydraulic conductivity	K_{aq}	LT^{-1}	Derived from Gleeson et al., 2014 (Fig. S4.3)	1,3
Hydraulic head	$h_{(aq)}$	L	Head of the aquifer in a computational cell, initial estimate based on 5' average of 30" head of Fan et al. (2013)	1,6,5
Ocean boundary conductivity	c_{ocean}	L^2T^{-1}	$10\ m^2\ day^{-1}$	2,6
Ocean boundary head	h_{ocean}	L	Global mean sea-level of 0 m	6
SW head	h_{swb}	L	30 % quantile (P_{30}) of 30" land surface elevation of Fan et al. (2013) per 5' grid cell	3
SW bottom elevation	B_{swb}	L	2 m (wetlands), 10 m (local lakes), 100 m (global lakes) below P_{30}	4
Area of global and local lakes and global and local wetlands	WL	L^2	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016),	4
Length of the river	L	L	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016)	4
Width of the river	W	L	Per 5' grid cell, based d on WaterGAP 3 (Eisner, 2016)	4
River head	h_{riv}	L	h_{swb}	4,5
River bottom elevation	B_{riv}	L	$h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994)	5
Equilibrium hydraulic head	h_{eq}	L	Steady-state hydraulic head of Fan et al. (2013) (averaged to 5' from original spatial resolution of 30")	5
Layers	-	-	2 confined, 100 m thick each	-
Land surface elevation	-	L	5' average of 30" digital elevation map of Fan et al. (2013) (Fig. S4.2)	-
E-folding factor	-	-	Applied only to lower layer for 150 m depth, based on area-weighted average of Fan et al. (2013)	-
Timestep	t	T	Daily timestep	-
Head convergence criterion (outer loop)	-	L	Max head change globally < 10 m in three consecutive iterations	-
Residual convergence criterion (inner loop)	-	-	$\lvert\lvert$conjugate gradient residuals$\rvert\rvert_{inf} < 10^{-100}$	
Maximum number of inner iterations	-	-	Maximum 50 inner iterations between outer Picard iterations (Naff, Richard L., and Edward R. Banta, 2008)	-

Supplement

1 Coupling to WGHM

With a spatial resolution of 0.5° by 0.5° (approximately 55 km by 55 km at the equator), the WaterGAP 2 model (Alcamo et al., 2003) computes human water use in five sectors and the resulting net abstractions from GW and SW for all land areas of the globe excluding Antarctica. These net abstractions are then taken from the respective water storages in the WaterGAP Global Hydrology Model (WGHM) (Müller Schmied et al., 2014; Döll et al., 2003; 2012; 2014). With daily time steps, WGHM simulates flows among the water storage compartments canopy, snow, soil, GW, lakes, man-made reservoirs, wetlands and rivers. As in other GHMs, the dynamic of GW storage (GWS) is represented in WGHM by a linear GW reservoir model, i.e.

$$\frac{dGWS}{dt} = R_g + R_{g\_swb} - NA_g - k_g\, GWS \qquad \text{(S1)}$$

where R_g $[L^3T^{-1}]$ is diffuse GW recharge from soil, R_{g_swb} $[L^3T^{-1}]$ GW recharge from lakes, reservoirs and wetlands (only in arid and semiarid regions, with a global constant value per SW body area), NA_g $[L^3T^{-1}]$ net GW abstraction. The product $k_g\, GWS$ quantifies GW discharge to SW bodies as a function of GWS and the GW discharge coefficient k_g (Döll et al., 2014). G³M is to replace this linear reservoir model in WGHM. Capillary rise is not included in the presented steady-state simulation, as simulation of capillary rise requires information of soil moisture that is only available when G³M is fully integrated into WGHM.

G³M will be integrated into WGHM by exchanging information on (1) R_{g_swb} and NA_g, (2) soil water content, (3) Q_{cr}, (4) h_{swb}, and (5) Q_{swb}. Figure S1.1 indicates the direction of the information flows. Water flows from the 0.5° cells of WGHM are distributed equally to all 5' G³M grid cells inside a 0.5° cell. Flows transferred from the 5' cells of G³M to WGHM are aggregated. GW recharge and net abstraction from GW together with SW tables are the main drivers of the GW model that will be provided dynamically by WGHM. GW-SW flow volumes computed by G³M will be aggregated and added or subtracted from the SW body volumes in WGHM, and SW body heads will be recalculated. WGHM soil water content together with G³M WTD will be used to calculate capillary rise and thus a change of soil water content. WaterGAP includes a one layer soil water storage compartment characterized by land cover specific rooting depth, maximum storage capacity and soil texture (Döll et al., 2014). The water content in the soil storage is increased by incoming precipitation and decreased by evapotranspiration and runoff generation (Döll et al., 2014). Capillary rise is not yet implemented in G³M, and SW heads are currently based on land surface elevation.

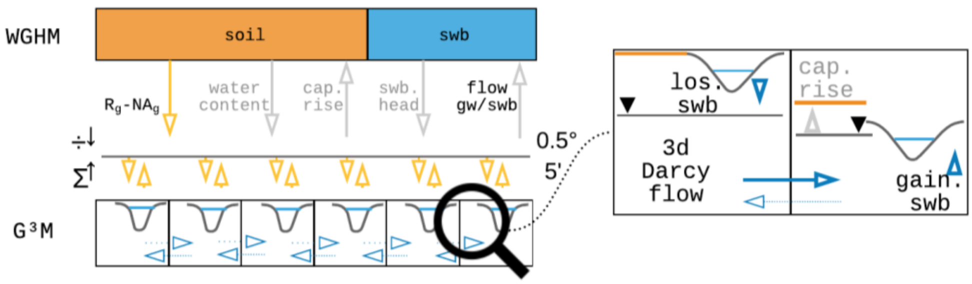

Figure S1.1 Conceptual view of the coupling between WGHM and G³M. WGHM provides calculated GW recharge (R_g) (Döll and Fiedler, 2008) and if the human impact is considered, net abstraction from GW (NA_g) (Döll et al., 2012). G³M spreads this input equally to all 5' grid cells inside a 0.5° cell and calculates hydraulic head and interactions with SW bodies (swb) as well as capillary rise (cap. rise) at the 5' resolution. Grey arrows show information flow that is not yet implemented.

2 Case study Central Valley

To evaluate G[3]M further, its results were analysed for to a well-studied region, the Central Valley in California, USA. The Central Valley is one of the most productive agricultural regions of the world and heavily relies on GW pumpage to meet irrigation demands (Faunt et al., 2016). GW pumping in the valley increased rapidly in the 1960s (Faunt, 2009). Figure S2.1 shows simulated WTD for the Central Valley, the coast and the neighboring Sierra Nevada mountainside as well as parts of the Great Basin. The WTD table represents natural conditions without any pumping and is rather small. It roughly resembles the WTD assumed in the Central Valley Hydrological Model (CVHM) as initial condition, representing a natural state (Faunt, 2009) (Fig. S2.1b). G³M correctly computes the shallow conditions with groundwater above the surface in the north, partially in the south of the valley and decreasing towards the Sierra Nevada. The difference in the extend of flooded area could be due to large wetlands areas still present in the early 60s which are not represented in this extent in the data used by G³M. Beyond the CVHM domain, WTD in mountainous regions is probably overestimated by G³M. The elevation of neighboring cells may differ up to a 1000 meter resulting in a large gradient (Fig. S4.5b and S4.5e).

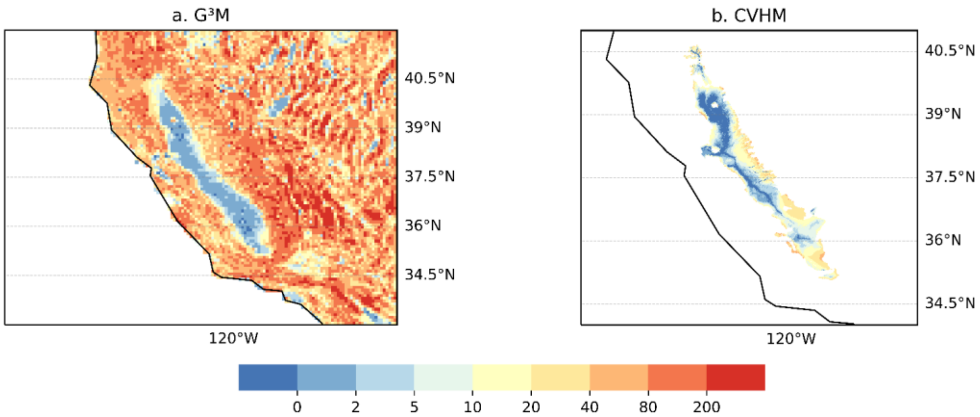

Figure S2.1 Plots of WTD [m] as calculated by G³M for the Central Valley and the Great Basin (a), and as used by CVHM as the natural state and starting condition (Faunt, 2009) (b).

3 Sensitivity Analysis

Sensitivities are calculated using forward differences (Poeter et al., 2014).

$$\frac{\Delta y_i'}{\Delta b_j} = \frac{y_i'(b_j + \Delta b_j - y_i'(b_j)}{\Delta b j_j} \quad \text{(S2)}$$

where y_i' is the simulated hydraulic head at position i from ND number of cells and b_j the perturbed parameter, here a multiplier for grid specific values shown in Table S1, in a vector of all parameter b of length j. Based on these values the composite scaled sensitivity is computed as

$$CSS_j = \sqrt{\sum_{i=1}^{ND} \frac{\Delta y_i'}{\Delta b_j} ND^{-1}} \quad \text{(S3)}$$

The result of the CSS is in units of meters. The higher the CSS, the more sensitive are the computed hydraulic heads to the parameter (Table S1).

Table S1 Ranges of parameter multipliers used in the local sensitivity analysis and their resulting composite scaled sensitivity values. The multiplier for the wetlands applies to global and local wetlands.

Parameter	$\Delta \boldsymbol{b}$	**Composite Scaled Sensitivity [m]**
h_{swb}	0.01	39132.1
K_{aq}	0.01	76.8
R_g	0.1	39.8
c_{Lakes}	0.1	3.2
$c_{Wetlands}$	0.1	0.014
c_{riv}	0.1	0.013
c_{ocean}	0.1	0.013

4 Additional Figures

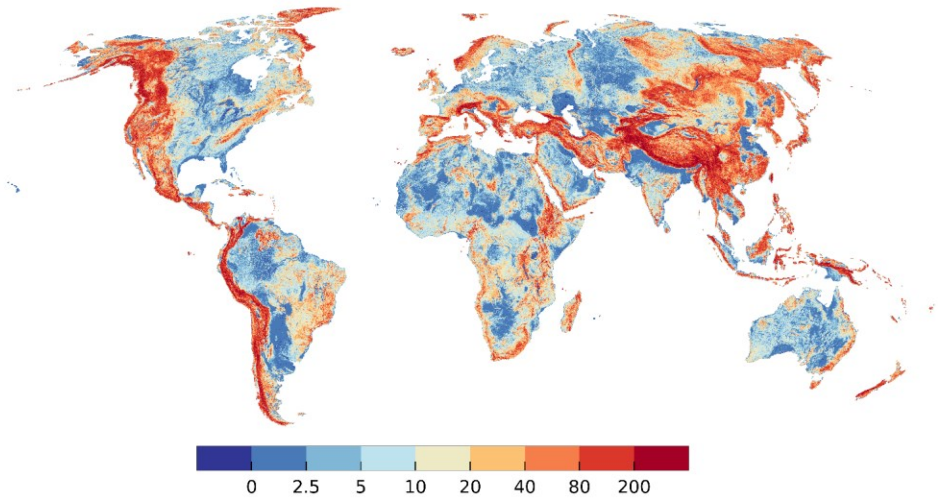

Figure S4.1 Difference [***m***] between 5' average of 30'' land surface elevation and P_{30} elevation. Maximum value 1365 m**.**

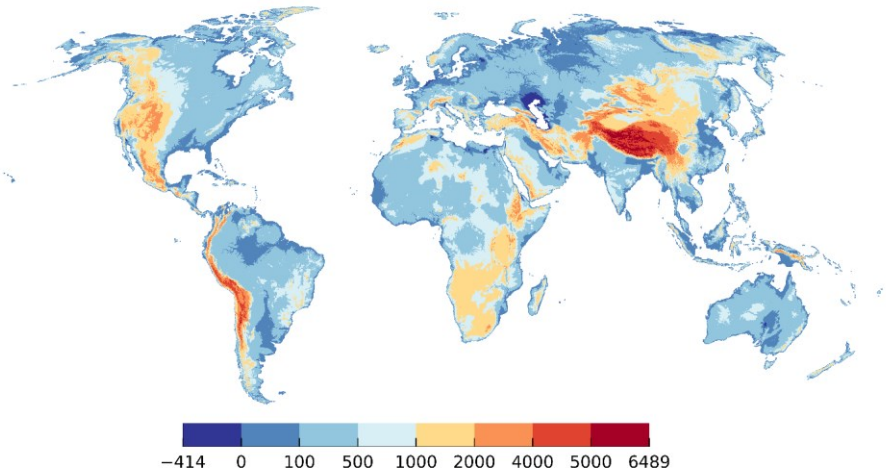

Figure S4.2 Land surface elevation [***m***] used in G^3M: 5' average of 30'' land surface elevation used in Fan et al. (2013).

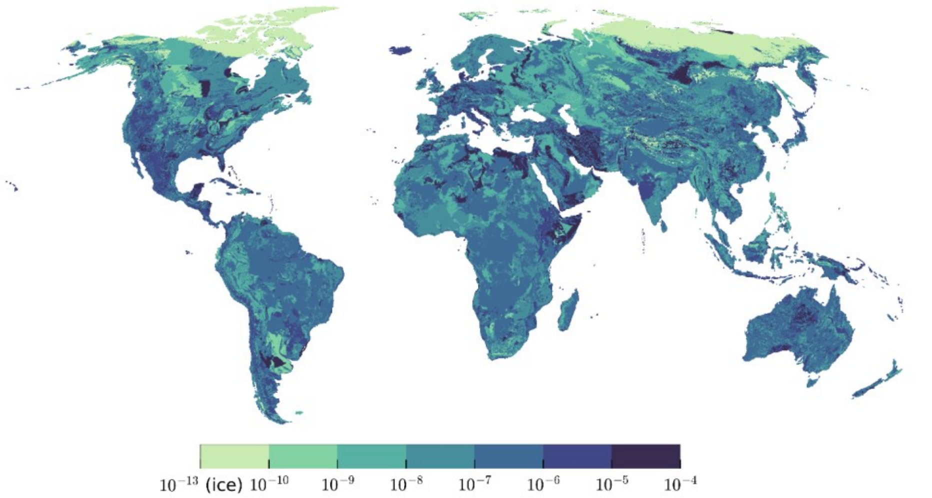

Figure S4.3 Hydraulic conductivity [$\boldsymbol{ms^{-1}}$] derived from Gleeson et al. (2014) by scaling it with the geometric mean to 5'. Very low values in the northern hemisphere are due to permafrost conditions.

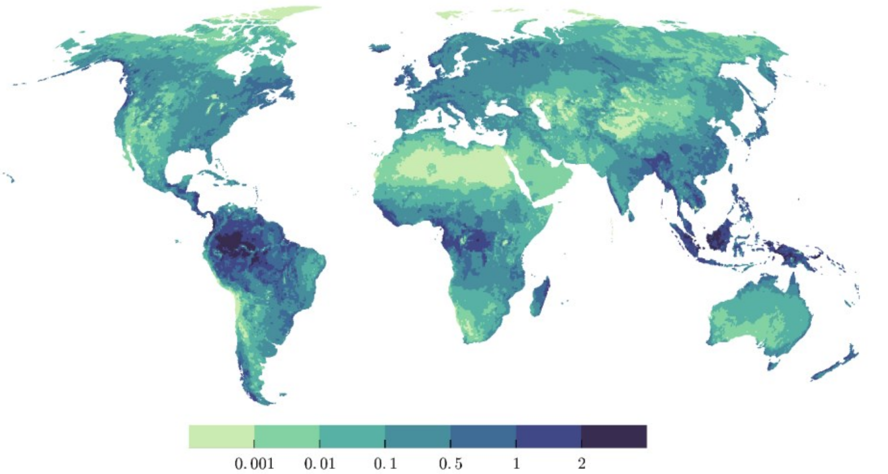

Figure S4.4 Mean annual groundwater recharge [$\boldsymbol{mm\ day^{-1}}$] between 1901-2013, from WaterGAP 2.2c.

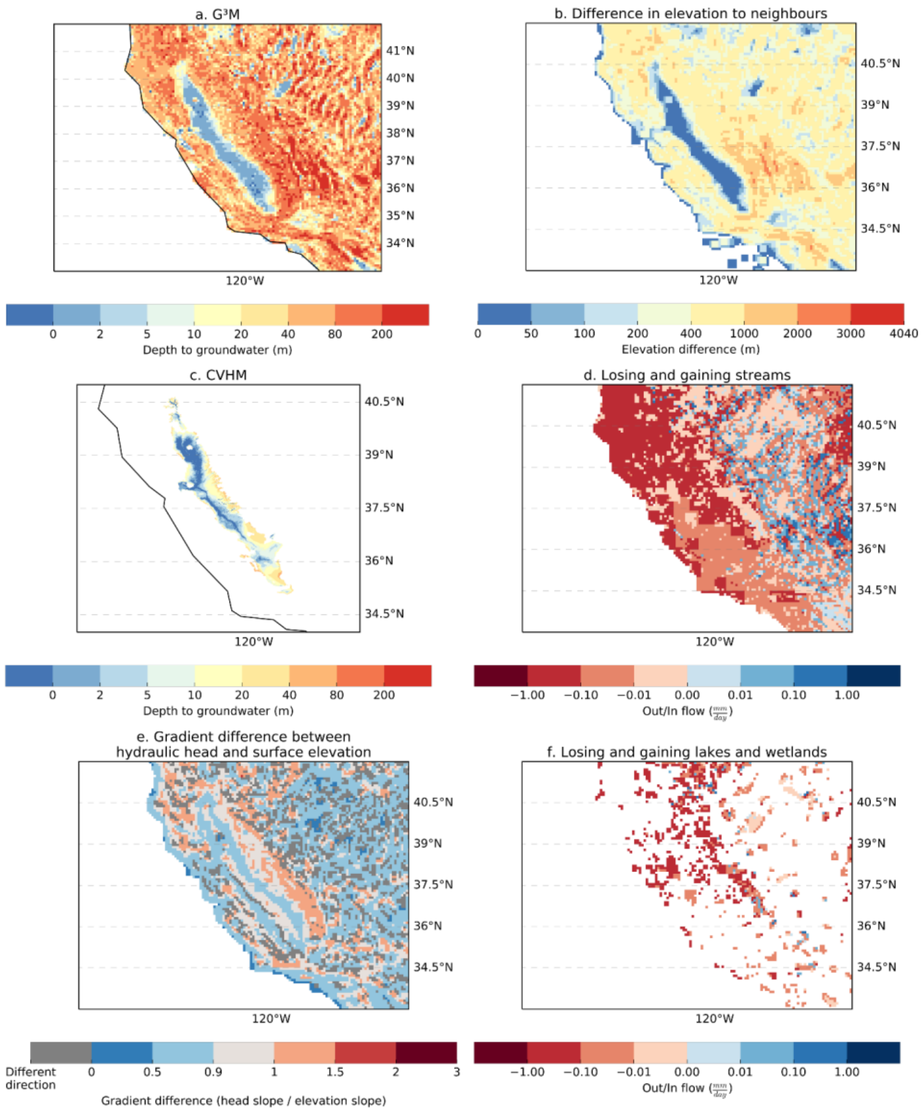

Figure S4.5 Plots of WTD as calculated by G³M (a), difference in surface elevation to neighbouring cells (b), WTD as used by the CVHM as the natural state and starting condition (Faunt, 2009) (c), losing and gaining streams as calculated by G³M (d), difference in gradient of hydraulic head and surface elevation (e), losing and gaining lakes and wetlands as calculated by G³M for the Central Valley and the Great Basin.

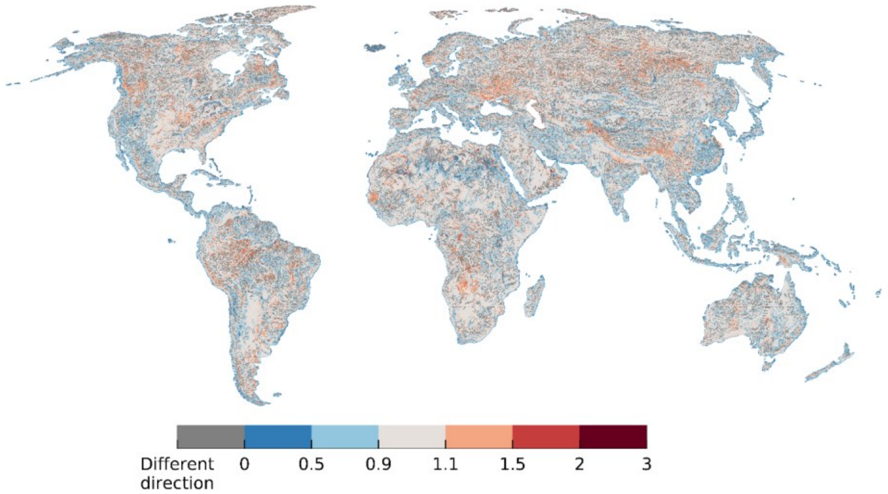

Figure S4.6 Ratio of hydraulic head gradient to 5' mean surface elevation gradient, only computed if the difference in direction of the gradient was smaller than 45°.

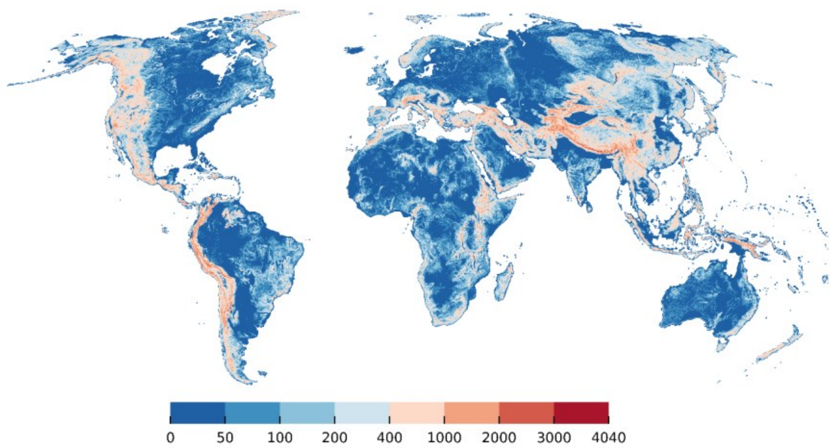

Figure S4.7 Land surface elevation difference between 30'' mean land surface elevation in 5' grid cell and mean elevation of neighboring 5' cells [**m**].

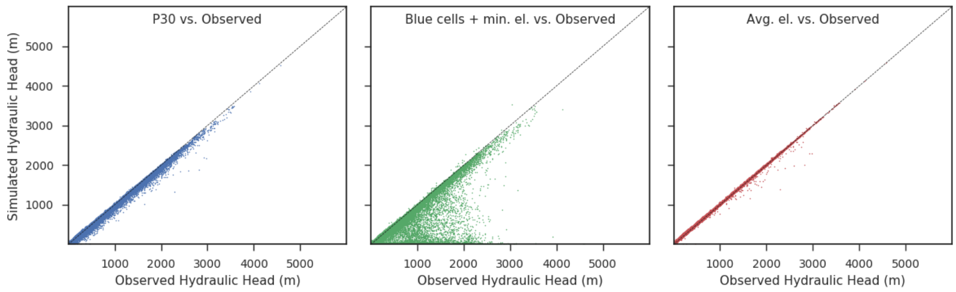

Figure S4.8 Comparison between three alternatives for setting $\boldsymbol{h_{swb}}$. Left to right: Fit of simulated hydraulic heads observations if $\boldsymbol{h_{swb}}$ is set (1) to the 30th percentile of the 30'' land surface elevations (standard model) , (2) alternatively to the average elevation of all “blue” cells of the 30'' water table results of Fan et al. (2013) or (3) is set to the average of the 30'' land surface elevations. A blue cell has a WTD of less than 0.25 m and indicates GW discharge to the surface. If no “blue” cell exists in the 5' cell, the minimum elevation of the 30'' land surface elevation values within the cell was used.

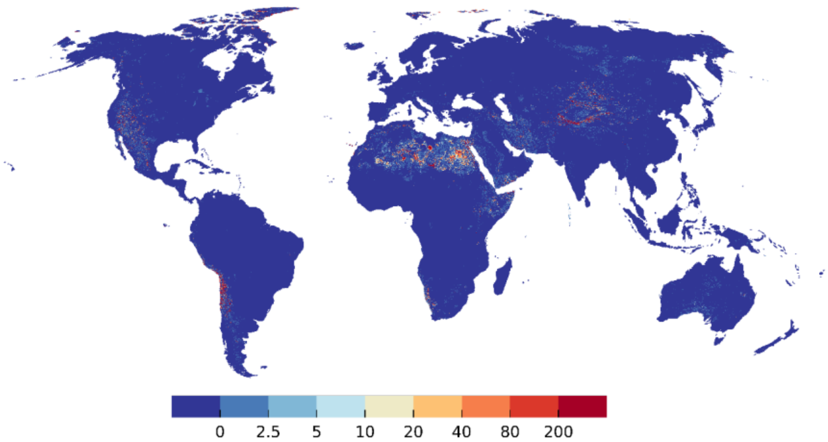

Fig. S4.9 Depth to groundwater [$\boldsymbol{m}$] for SW body elevation $\boldsymbol{h_{swb}}$ at average of 30'' land surface elevations.

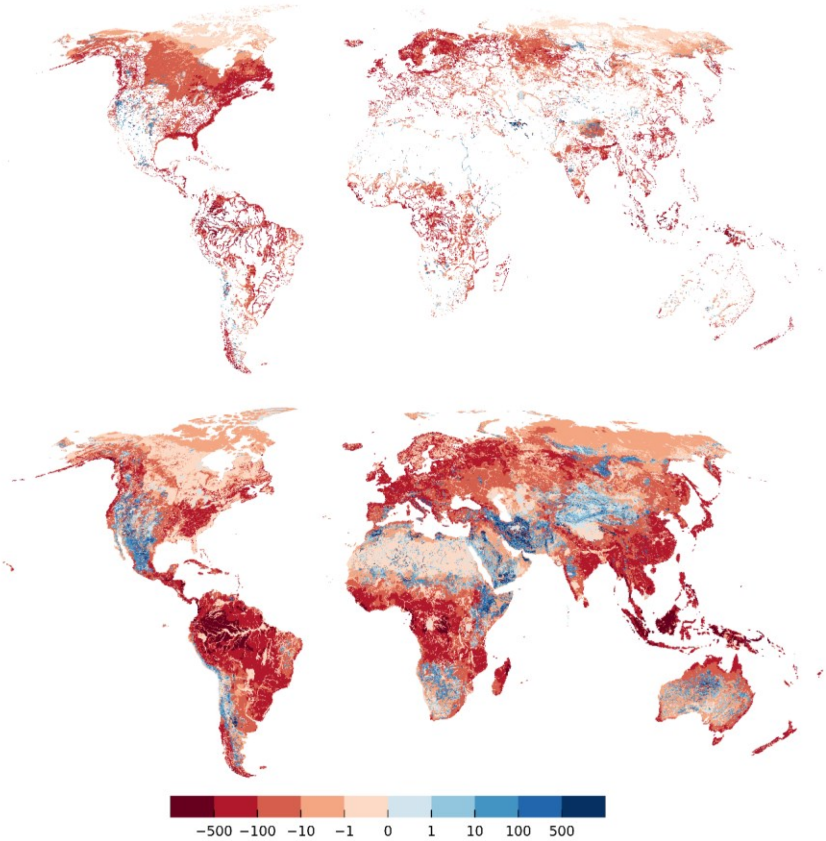

Figure S4.10 Gaining and losing rivers (lower panel) and wetlands and lakes (upper panel) as flow into/out the GW [$\boldsymbol{mm\ year^{-1}}$] if $\boldsymbol{h_{swb}}$ is set to average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) (right). A blue cell is defined as a depth to groundwater of less than 0.25 m. If no "blue" cell exist in the 5' cell, the minimum elevation of the 30'' land surface elevation values is used. Red denotes gaining SW bodies.

Dear Dr. Kurtz,

Thank you for your remarks and the possibility to revise the manuscript again.
In the following, we are presenting additional changes to the manuscript to address the issues raised by the reviewers, as indicated in your letter.
Our comments to the reviewer remarks (numbered) are provided in italics and are marked accordingly in the manuscript.

Reviewer #1

1.1

[..]However, authors have not fully addressed reviewers' and editor's comments, and the manuscript still has some deficiencies that need to be addressed. For example, response to comments 3 and 5 of the Editor is still not convincing after revising the manuscript.

Comment 3 (of the Editor):
At the current stage, it is not clear what was learned from the modeling exercise and what are the advantages and differences compared to already existing model set-ups and software packages. As advertised in the manuscript, a main purpose of this study is to use a steady state groundwater model to gain first insights into the credibility of the model set-up for future coupled transient simulations. However, this is not streamlined at the moment, i.e. it is not clear what experience you gained from the steady state simulations to move forward to transient coupled simulations. The problems outlined by the referees regarding model set-up and model verification will prevail for transient coupled simulations and therefore need to be properly addressed at the current stage. It needs to be pointed out more clearly, how this study serves as a basis for the development of the desired coupled WaterGAP model.

Comment 5 (of the Editor):
The manuscript quite often refers to implementation that are planned in the future (i.e. coupling to WaterGAP, transient simulations). While it is ok to outline these future plans in order to justify the proposed modelling approach for the current global steady-state model, the frequency of references to future work is often misleading (as also acknowledged by referee 2), because it is often not clear which kind of feature is really implemented and which one is intended to be implemented. Therefore, I would ask you to limit references to future work to the necessary minimum.

The abstract and introduction were changed to focus the paper on the steady-state outcome and streamline it with the presented results

Section 2.2 was exchanged with section 2.3 to streamline the paper further. Major parts of the new section 2.3 were moved to the supplement to preserve the information for the interested reader while focusing the paper on the actual presented model.
The new section 2.3 now reads:

In this initial effort, we intend to integrate G³M into WaterGAP 2, i.e. the 0.5° version of WGHM (for details see S1) to keep computation time low enough for performing sensitivity analyses and ensemble-based data assimilation and calibration, instead of integrating it into WaterGAP 3 (Eisner, 2016), which has the same spatial resolution as G³M. However, data from WaterGAP 3 were used to set up G³M. Location and area of the 5' grid cells of G³M are the same as in the landmask of WaterGAP 3. In addition, the percentage of the 5' grid cell area that is covered by lakes (including reservoirs) and by wetlands, based on Lehner and Döll (2004), is taken from WaterGAP 3, as well as the length and width of the main river within each 5' grid cell as estimated by WaterGAP 3 (Table 1).

1.2

I concur with the editor's comment (3) that it is still not clear what is learned from this modeling exercise and comparison with existing methods are required to show the gradient based approach indeed is superior to the existing approaches.

See 1.1. A comparison to existing approaches would mean a comparison to the linear storage approach of the global hydrological model as there are no other existing global approaches to this problem (apart from solving the computationally even more complex Richards equation). This is not possible at this stage because the linear storage describes only a transient groundwater storage not a steady-state equilibrium. This is something that is planned for the future transient integration into the GHM.

1.3

Evaluation of the model performance against existing observations and model simulations such as CVHM and ParFlow models are qualitative. A more robust quantitative analysis using various statistical measures are required to show where the new approach works and where it does not.

Section 3.5 was extended with a Root Mean Squared Errors for plots shown in Fig. 8 and are discussed accordingly (Lines 14-23 P. 14). Furthermore, a new Fig. 9 was added (page 16) that shows observed minus simulated head for different land surface elevation categories of the model including the IQR and Mean as Boxplot. Additionally, this figure shows the RMSE.

1.4

Authors indicate lack of enough observations to validate their modeling approach. Perhaps, they can run the model for a higher resolution and evaluate the impact of coarse simulation on simulated hydrologic fluxes. This could provide valuable insights regarding model development and implementation for the coupling approach as well.

We extended the study with a spatial scale sensitivity analysis of the model for the region of New Zealand and show 5 arcmin. resolution results in comparison with a model on 30 arcsec and compare results to local observations. It adds evidence to the hypothesis that a more elaborate estimation of the surface water body elevation can improve the 5 arcmin results.
See new section 3.6.

1.5

The main contribution of this paper is not entirely clear. Authors have highlighted a number of improvements that the use of a gradient-based groundwater model could have on simulating global hydrological processes but many of those improvements have not been made in this version of the manuscript including the coupling with the global hydrologic model. As pointed out by the reviewers, I suggest authors to focus this paper on the sensitivity analysis of model parameterization and conceptual formulations. This can certainly help with the development of the transient model as well as the coupling method.

See 1.1. Additionally, we added a local parameter sensitivity analysis of New Zealand (section 3.6). These results are then further reflected in the discussion (Line 23 Page 17, Line 9,18 Page 19).
As this paper is an initial model development paper and is already large as it is, we saw fit to submit a separate paper to HESS with a full global sensitivity analysis of the model.

1.6

Could you please further explain how lateral connectivity between neighboring groundwater cells are calculated?

The lateral connection of cells is based on the cell location which is determined by the landmask which is adapted from WaterGAP3. If the question relates to the lateral flow and hydraulic conductivity between cells, this is described in section 2.1. The hydraulic conductivity is calculated as the harmonic mean between neighboring cells. The calculated value is then used to calculate the hydraulic heads based on the in and out fluxes to the cell.

Further information can also be found in the MODFLOW documentation by Harbaugh (2005) also cited in this section.

Additional Remarks

A1.1

Page 1 – L18: “as simulation of unsaturated flow and SW body elevation” is not clear. Please describe.

The abstract was fully revised see #2.1

A1.2

Page 1 - L31-34 : It is not clear.

The abstract was fully revised see #2.1

A1.3

Page 2- L5: replace recharge from soil with “precipitation”

Precipitation is an input of the global hydrological model. Recharge from soil is the actual amount of water that infiltrates the groundwater taking into account e.g. evapotranspiration and surface runoff. Thus, recharge is only a certain percentage of precipitation.

A1.4

Page 2- L27: replace “This flow direction” with “losing streams”

Now reads (Line 14, Page 2): `Losing streams typically occur in semi-arid and arid but seasonally also in humid regions. In addition, such linear reservoir models provide no information on the location of the GW table and assume that GW flow among grid cells is negligible.`

A1.5

Page 3- L16-17: Not clear. Please explain

Revised and now reads (Line 44 ff., Page 2):
`This additional drainage, which accounts for about 50% of global GW flow into SW, is simulated as a function of GW storage above the floodplain. The values needed to compute this additional artificial drainage are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017)) – the model component that the gradient-based model was intended to replace in the first place. This prevents a full integration of the global GW flow model of de Graaf et al. (2017) into a GHM, as then, the linear GW reservoir model would be replaced by the GW flow model.`

A1.6

Figure 1. Add further details to the caption. Describe P30 of 30” DEM and why average of 30” DEM is compared with P30?

Now reads (Description Fig 1.): The P_{30} is used in the presented steady-state model as SW elevation instead of an average or minimum per grid cell.

A1.7

Page 4- Final paragraph needs revision to improve consistency with the rest of the text.

It now reads (Line 14 ff., Page4):
The simulation of aquifers that contain dry cells and/or cells that oscillate between wet and dry states poses great challenges to the solving of Eq. (1) (Niswonger et al. 2011). G³M-f (the framework code used to implement G³M) implements the traditional wetting approach from Harbaugh (2005) as well as the approach proposed by Niswonger et al. (2011) along with the proposed damping scheme.

A1.8

Page 6 – L6: replace "loosing" with losing

Now consistently spelled in manuscript.

A1.9

Page 6 – L28: Replace equation 5 with 6

Has been corrected.

A1.10

Section 2.2. If the coupling is not done yet, why it is discussed in this paper? You could devote this section to assess the impacts of simplifications/improvements you made compared to the linear reservoir GW model. Or this section should come last (future work) if you plan to include the coupling approach in this paper.

See 1.1. this section has been moved to the supplement and was shortened. For the linear GW model see the answer to 1.2.

A1.11

Page 8 – L4: Could you please justify the choice of ocean conductance value.

Now reads(Line 9, Page 7): [..]to $10\, m^2\, day^{-1}$ (Table 1), reflecting a global average conductance based on hydraulic conductivity and lateral surface area.

A1.12

Section 3.1. Why P30 has been considered as the surface elevation of surface water bodies in the model?

This is discussed in section 2.2 (former 2.3) (Line 24, Page 7):
"For the steady-state model, river elevation h_{riv} is set in each grid cell to the same elevation as all other SW bodies, h_{swb}. We found that for both gaining and losing conditions, Q_{swb} and thus computed hydraulic heads are highly sensitive to h_{swb}. The overall best agreement with the hydraulic head observations of Fan et al. (2013) was achieved if h_{swb} (Eq. 4, 5 and 6) was set to the 30th percentile (P_{30}) of the 30'' land surface elevation values of Fan et al. (2013) per 5' cell, e.g. the 30'' elevation that is exceeded by 70% of the thousand 30'' elevation values within one 5' cell. To decrease convergence time we used h_{eq} derived from the high-resolution steady-state hydraulic head distribution of Fan et al. (2013) as initial guess."

A1.13

Section 3.2. L18 – "the amount of river water that recharges GW is only about a 40th of the drainage to GW," not clear

Now reads(Line 33, page 9):
According to G³M, the amount of river water that recharges GW is more than one order of magnitude less than the drainage of GW, and the relative recharge to GW from lakes and wetlands is even smaller (Fig. 3).

A1.14

Section 3.2 – L20 – Why outflow from SW body is not limited by water availability? River stage in equation 5 should control this.

This is not possible in a steady-state model (will be possible with the transient model) as explained in section 2.2 (former 2.3):
"A further difficulty in an uncoupled model run is that the water table elevation of SW bodies does not react to the GW-SW exchange flows Q_{swb} *and that water supply from SW is not limited by availability. A losing river may in reality dry out due to loss to GW and therefore cease to lose any more water."*

A1.15

Section 3.2- Could you verify estimated global water budget against other existing global models?

We would like to but the global budget of the only other existing model is not available. As we move to the transient model we will compare the transient budget of the gradient-based approach with the linear storage approach.

A1.16

Section 3.3 – No attempt has been made to verify SW-GW exchange rate. This is an important contribution of this work and model validation is required. Perhaps, use ParFlow simulations over CONUS to check these fluxes at steady-state.

Fluxes between groundwater and surface water are very complicated to validate. Even at the regional scale riverbed conductance is a calibration parameter as fluxes between groundwater and surface water can change on a very small scale and are challenging to measure. Measurements are furthermore only available for very small fractions of local streams which cannot be interpolated as comparison to 9km by 9km gridcells. ParFlow creates rivers naturally without predefining them in their model. Furthermore, as far as the authors are aware no exchange rate data is available for ParFlow at this point.
As we move to the transient implementation we certainly should consider to revisit this issue and compare the computed fluxes to existing regional models, if possible.

A.1.17

Section 3.6. Comparison with the Central Valley hydrologic model is purely qualitative. It does not seem this qualitative comparison adds any value to assess G3M model performance.

This section has been moved to the supplement (now section S2) for the interested reader and was replaced by the New Zealand study (new section 3.6).

Reviewer #2

2.1

[..]Any comment or observation on the fully-coupled model therefore remains purely speculative, since the work has yet to be done. I recommend to change the focus to what is actually presented in the paper, which is the steady-state uncoupled application. It would greatly clarify the paper. Section 2.2 should therefore be deleted. The long-term objective could be mentioned in the discussion, as a perspective for future work. It would then be quite acceptable for the authors to justify the choice of using G3M instead of MODFLOW (as raised by reviewers), for tighter integration with WaterGap2.

See response to 1.1. Section 2.2 has been shortened to a small paragraph and exchanged with section 2.3. The discussion now reflects the long-term perspective as suggested (Line 20 ff., Page 17):

`The objective of global gradient-based groundwater flow modeling with G³M is to better simulate water exchange between SW and GW in the GHM WaterGAP, for example for improved estimation of GW resources in dry regions of the globe that are augmented by focused recharge from SW bodies. The presented steady-state model is the first step in this direction:`

2.2

I agree with the previous reviewers that the lessons learned from the steady-state application are not clear, compared to previous work. Since this paper does not present any results from a coupled model, arguments about coupling strategy as outcome are not valid. A future paper that actually presents results with the coupled model could address this coupling and assess challenges and issues. The other novel elements mentioned relate to scale challenges and equation solved but they are not unique to G3M. The other gradient-based GW flow models mentioned in the paper also face similar issues. The main outcomes of this steady-state application should be much more clearly highlighted and justified, in the context of previous studies.

See also 1.1.,1.2 and 2.1. The introduction has been changed to reflect that together with the shortened section (new)2.3. Additionally, the discussion has been adapted to reflect that and now reads in Line 20 ff., Page 17:

`The objective of global gradient-based groundwater flow modeling with G³M is to better simulate water exchange between SW and GW in the GHM WaterGAP, for example for improved estimation of GW resources in dry regions of the globe that are augmented by focused recharge from SW bodies. The presented steady-state model is the first step in this direction: (1) It aligns with established GW model development practices that helped (2) to understand basic model behavior e.g. the sensitivity to SW body elevation, and the necessary improvement of its parameterization, before moving to the more complex integrated transient model. The reduced runtime of the steady-state model in comparison to a fully integrated transient run (3) supported the investigation of parameter sensitivity and sensitivity to spatial resolution. Additionally, (4) the presented steady-state model can be used in future fully integrated transient runs as initial condition.`

2.3

The previous reviewers raised concerns about using the fully-saturated groundwater flow equation, as opposed to an unconfined flow equation, for the upper layer. The justification given in response to these concerns are rather confusing and difficult to understand. I think that a short description of the way WCHM treats the soil "compartment" is needed and could help justify the approach. Right now, the reader has to guess what "soil" refers to.

As the Editor and all reviewers asked to reduce the description of coupling to WaterGAP a more detailed description of the soil compartment of WaterGAP was added to S1 and is referred to for the interested read in Line 8 Page 4:

`WaterGAP includes a one layer soil water storage compartment characterized by land cover specific rooting depth, maximum storage capacity and soil texture (Döll et al. 2014). The water content in the soil storage is increased by incoming precipitation and decreased by evapotranspiration and runoff generation (Döll et al. 2014).`

2.4

The paper is, in general, not clearly written. The style is often unnecessarily complicated, with very long sentences and repetition. The paper also does not focus on the essential and there is a lot of unnecessary detail. On the other hand, some important information is not presented clearly, such as a short description of how soils are treated in WCHM. Another example is Line 5 on page 8 that states : "It is assumed that there is exchange of water between GW and one river stretch in each 5' grid cell". If I understand correctly, there is a river for every top cell in the model. If it is the case, it is an important assumption and should be stated much more clearly and earlier, and not "buried" on page 8.

The length of multiple sentences has been shortened wherever possible and repetitions decreased (see also following detailed comments).

For soil see 2.3 above.

The fact that each cell has a river is stated clearly in Table 2 and is compared to other models. The authors cannot agree with the observation that it is "buried" it appears clearly stated in section 2.2 which describes the steady-state assumptions.

2.5

The terminology used is also very confusing for a reader with a hydrogeology background. For example, groundwater is used to refer to both the contained (the "groundwater" or subsurface compartment) and the content (groundwater that flows according to the governing equation). The paper also uses the term "drainage" to represent fluid exchange between the various compartments (subsurface, rivers, soil, wetlands, etc.). The use of drainage is extremely confusing and a better terminology would greatly clarify the paper. Also, hydrogeologists do not use hydraulic head for surface water or surface water table.

The use of drainage has been clarified in multiple places (highlighted in the markup document with #2.5).

We disagree. The use of hydraulic head in the context of surface water hydrology is necessary in models that rely on hydraulic flow routing and are based on hydraulic head. Furthermore, it is correct, and even more precise and accurate, in this specific context as the flux between SW and GW cannot be computed solely based on a surface water table and the hydraulic head of the groundwater. To calculate an exchange, it is pivotal to know the pressure gradient between the surface water and the groundwater.

Detailed comments

A2.1

The abstract provides a good example of the writing style. First, it is much too long. The abstract should be short, precise and to the point. It should not try to explain everything, such as the difference between reservoir and gradient-based GW models. It also contains sentences that are either unclear of very complicated. For example, this excerpt from lines 19-20 : "We identify challenges linked to the coarse resolution, which necessitates the deviation from established processes in regional groundwater modeling as simulation of unsaturated flow and SW body elevation". That sentence is both complicated and unclear, and does not inform the reader.

The abstract has been adapted to reflect the streamlined paper and shortened accordingly:
It now reads:

In global hydrological models, groundwater (GW) is typically represented by a bucket-like linear groundwater reservoir. Reservoir models, however, can (1) only simulate GW discharge to surface water (SW) bodies but not recharge from SW to GW, (2) provide no information on the location of the GW table and (3) assume that there is no GW flow among grid cells. This may lead, for example, to an underestimation of groundwater resources in semi-arid areas where GW is often replenished by SW or to an underestimation of evapotranspiration where the GW table is close to the land surface. To overcome these limitations, it is necessary to replace the reservoir model in global hydrological models with a hydraulic head gradient-based GW flow model.

We present G³M, a new global gradient-based GW model with a spatial resolution of 5', which is to be integrated into the 0.5° WaterGAP Global Hydrology Model (WGHM). The newly developed model framework enables in-memory coupling to WGHM while keeping overall runtime relatively low, which allows sensitivity analyses, calibration, and data assimilation. This paper presents the G³M concept and model design decisions that are specific to the large grid size required for a global scale model. In contrast to the GW model of de Graaf et al. (2015; 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G³M, thus enabling full coupling to a GHM. Model results under steady-state naturalized conditions, i.e. neglecting GW abstractions, are shown. Simulated hydraulic heads show better agreement to observations around the world than the model output of de Graaf et al. (2015). Locations of simulated SW recharge to GW are found, as is expected, in dry and mountainous regions but the areal extent of SW recharge may be underestimated. Globally, GW discharge to rivers is by far the dominant flow component such that lateral GW flows only become a large fraction of total diffuse and focused recharge in case of losing rivers, some mountainous areas and some areas with very low GW recharge. Strong sensitivity of simulated hydraulic heads to the spatial resolution of the model and the related choice of the water table elevation of surface water bodies was found. We suggest to investigate how global-scale groundwater modeling at 5' spatial resolution can benefit from more highly resolved land surface elevation data.

A2.2

P2. The first paragraph in the introduction (lines 1-20) presents only generalities and should be deleted. The paper should focus on the model right from the start.

The first paragraph provides a motivation for the presented research and justifies why modeling of global groundwater resources is worth the effort. Without it, it might be unclear to the reader why a GHM with a complex groundwater model is even necessary. However, the paragraph has been revised to be more succinct.

It now reads:

Groundwater (GW) is the source of about 40% of all human water abstractions (Döll et al. 2014) and is also an essential source of water for freshwater biota in rivers, lakes and wetlands. GW strongly affects river flow regimes and supplies the majority of river water during ecologically and economically critical periods with little precipitation. GW storage and flow dynamics have been altered by human GW abstractions as well as climate change and will continue to change in the future (Taylor et al. 2012). Around the globe, GW abstractions have led to lowered water tables and, in some regions, even GW depletion (Döll et al. 2014; Scanlon et al. 2012; Wada et al. 2012; Konikow 2011). This has resulted in reduced base flows to rivers and wetlands (with negative impacts on water quality and freshwater ecosystems), land subsidence and increased pumping costs (Wada 2016; Döll et al. 2014; Gleeson et al. 2012; 2016). The strategic importance of GW for global water and food security will probably intensify under climate change as more frequent and intense climate extremes increase variability of SW flows (Taylor et al. 2012). International efforts have been made to promote sustainable GW management and knowledge exchange among countries, e.g., UNESCO's program on International Shared Aquifer Resources Management (ISARM) (http://isarm.org) and the ongoing GW component of the Transboundary Waters Assessment Program (TWAP) (http://www.geftwap.org). To support prioritization for investment among transboundary aquifers as well as identification of strategies for sustainable GW management, information on current conditions and possible trends of the GW systems is required (UNESCO-IHP, IGRAC, WWAP (2012) 2012). In a globalized world, an improved understanding of GW systems and their interaction with SW and soil is needed not only at the local and regional but also at the global scale.

A2.3

P2. Line 34 : what are "macro-scale models"?

Now reads (Line 21, Page 2) : large-scale models.

A2.4

P2, Lines 37-38 : what is "the condition of SW"? Be more specific.

Now reads (Line 24, Page 2):
However, they are in most cases not integrated within hydrological models that quantify GW recharge based on climate data and provide information on the condition of SW (e.g. streamflow and storage).

A2.5

P2. Lines 38-39 : the excerpt "Miguez-Macho et al. (2007) linked a land surface model with a two-dimensional gradient-based GW model and computed, with a daily time step, gradient-based GW flow" is one example of unnecessary repetition that does not help the reader. There is no need to repeat that Miguez-Macho et al. used their gradient-based model to compute gradient-based GW flow. Another example of repetition is on page 3, lines 18-19 : "In this study, we present the Global Gradient-based Groundwater Model (G3M) that is to be integrated into the GHM WaterGAP 2" and just a bit further, line 29 is : "G3M is to replace this linear reservoir model in WGHM". Actually, that last repetition is even more confusing because GHM WaterGAP 2 and WGHM are not even the same model. I have noted several such repetitions that I will not list but that the authors should identify and eliminate.

The sentence has been changed to (Line 25 ff., Page 2):
For North America, Fan et al. (2007) and Miguez-Macho et al. (2007) linked a land surface model with a two-dimensional gradient-based GW model and computed, with a daily time step, GW flow, water table elevation, GW–SW interaction, and capillary rise, using a spatial resolution of 1.25 km

The second mention of the integration was deleted.

To clarify WaterGAP and WGHM the following sentence was added (Line 5, Page 3):
In this study, we present the Global Gradient-based Groundwater Model (G^3M) that is to be integrated into the GHM WaterGAP 2 (in the following we refer to WGHM (WaterGAP Global Hydrology Model), which is part of the GHM WaterGAP) to improve[..]

A2.6

P3. Line 3 : "GW above the land surface". Check terminology for more clarity. GW above land surface is no longer GW. Should probably write instead something like : groundwater exfiltration.

The calculated GW head above the land surface elevation does not represent a process like groundwater exfiltration. Such a process is not implemented in the discussed model. GW that is computed to be above *the assumed land surface elevation is simply discarded (! To clarify this statement refers to the Model of Fan and Miguez-Macho not the presented model.).*

A2.7

P3. Line 12 : The difference with the de Graaf et al paper is that they added "an additional drainage flus to GW drainage". That explanation is given several times in the paper but it is never clearly described and it is difficult to understand what de Graaf et al. did exactly.

The introduction states that "This additional drainage, which accounts for about 50% of global GW drainage, is simulated as a function of GW storage above the floodplain, the values of which are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017))[..]" An interested reader can find a detailed description in the original paper of de Graaf and in the according equation.

A2.8

P3, Line 30 : The paper mentions the G3M model and now the G3M-f framework. First, what is the difference between G3M and G3M-f? Second, is it relevant to mention anything else that the model, G3M, since the application here is totally decoupled from a global hydrologic model?

A framework is an established term that e.g. Wikipedia defines as: "In computer programming, a software framework is an abstraction in which software providing generic functionality can be selectively changed by additional user-written code, thus providing application-specific software. A software framework provides a standard way to build and deploy applications. A software framework is a universal, reusable software environment that provides particular functionality as part of a larger software platform to facilitate development of software applications, products and solutions."
This is explained in section 2.4. The goal of G^3M-f was to develop a framework that can also be used to build regional models and be reused by the community to also build GW models for the integration into other models easily (because MODFLOW does not offer this capability) rather than only developing a code that can only be used with WaterGAP.
The reference in the introduction has been removed to avoid confusion.

A2.9

P3, Lines 33-34 : The formulation "We want to find out whether we can use gradient-based groundwater modelling at the global scale, …, to improve estimation of flows between SW and GW … and capillary rise" is rather surprising. Isn't the working hypothesis that gradient-based models will improve simulations?

That is the base hypothesis yes, but this is only true if the assumption above holds true and we aim at providing a demonstration of this approach, keeping in mind that this is the first example available. To clarify it now reads (Line 9 ff.,Page 3):

```
      The objective of this paper is to learn from a steady-state model, a well-
established first step in groundwater model development, to (1) understand the
basic model behaviour by limiting model complexity and degrees of freedom, and
thus (2) providing insights into dominant processes and uncovering potential
model-inherent characteristics difficult to observe in a fully coupled transient
model. A transient model might obfuscate model inherent trends due to the slow
changing nature of groundwater processes e.g. trends towards large over/under-
estimation due to wrong parameterisation. A fully coupled model furthermore adds
complexity and uncertainty to the model outcome. The presented steady-state model
is furthermore used to (3) investigate parameter sensitivity and sensitivity to
spatial resolution. In addition, the steady-state solution can be used as (4)
initial condition for future fully coupled transient runs.
```

A2.10

P3, lines 36-40 : This is an example of a very complicated and unclear sentence : "Steady-state simulations are a well-established first step in groundwater model development to understand the basic model behaviour limiting model complexity and degrees of freedom, thus providing insights into dominant processes and uncovering possible model-inherent characteristics impossible to observe in a fully coupled transient model." For example, what is "the basic model behaviour limiting model complexity and degrees of freedom"? What are the "model-inherent characteristics" that can't be observed in a fully coupled transient model?

Steady-state models can be easily controlled and run quickly. We are not saying that they are superior to fully transient models, but they are a first step to understand the system and provide information that can be critical for developing an efficient transient model which is expected to require a much longer execution time.

Furthermore, in a transient GW model a run of 100 years might contain slight trends that lead to an ever increasing GW head in a specific region. Due to the slow nature of GW this might not be visible to the model developer. A steady-state model on the other hand would possibly show a clear overestimate/flooding due to e.g. the wrong parameterization.
A comma has been added to make clear that the model behavior is not limiting the complexity. Furthermore, the following sentence has been extended by an example to clarify what trends might not be visible in a transient model.
See A2.9

A2.11

P3, lines 40-41: I don't know what is meant by "A transient model might obfuscate model inherent trends due to the slow changing nature of groundwater processes."

See A2.10.

A2.12

P3, line 41 : The following statement is quite bold and I am not sure that I agree : "A fully coupled model furthermore adds complexity and uncertainty to the model outcome". If it is the case, what do the authors want to develop a fully-coupled model if its outcome will be more uncertain?

Because the main purpose of the model is not to build a standalone GW model but to replace the current GW storage model in WaterGAP. For that we need a fully coupled transient model, but the preliminary steady state model will inform us on the key physical processes and it is more controllable as mentioned above.

A2.13

P4. Section 2 on Model description. I suggest to reorganize that section because it is not clear. I suggest to start by presenting the governing equations (equations 2 to 7) and then present the global-scale components. All simplifying hypothesis should be clearly stated. The exact input data originating from the global hydrologic model should also be clearly presented.

We do not agree with this comment. The governing equations are a consequence of the conceptual nature of the global model and can be misinterpreted without the global-scale components. The data is clearly stated at the beginning of the section and further explained in Table 1.
Additionally, see 1.1.

A2.14

P4, lines 10-11 : "G³M differs from traditional local and regional GW models". Is it really the case? I think that the main difference is the scale of application and the use of WCHM output as input.

Yes, it differs in the exceptional spatial resolution of the grid and the necessary assumptions because of the lack of global data. This is further elaborated in the following sentences (Line 25 ff., Page 3):
"[..]These models are generally based on rather detailed information on hydrogeology (including aquifer geometry and properties such as hydraulic conductivity derived from pumping tests), topography, pumping wells, location and shape of SW bodies as well as on observations of hydraulic head in GW and SW. Local observations guide the developer in constructing the model such that local conditions and processes can be

properly represented. The lateral extent of individual grid cells of such GW flow models is generally smaller or similar to the depth of the aquifer(s) and the size of the SW bodies that interact with the GW. The global GW flow model G³M, however, covers all continents of the Earth except Greenland and Antarctica. At this scale, information listed above is poor or non-existing, and the lateral extent of grid cells needs to be relatively large due to computational (and data) constraints."

A2.15

P4, line 17 : "At this scale, information listed above is poor or non-existing". It should be reworded. The information contained for smaller-scale (as mentioned just above) is still available at the large (global) scale. You probably mean something else.

No, it is unavailable because data available for a specific basin is unusable at the global scale without additional processing and assumptions. Furthermore, it might not even be possible to use this information after processing because it may not be possible to reasonably interpolate the data to an appropriate (global) grid-scale resolution.
We assume that the reader is able to understand that the information does not disappear but that it is very challenging to compile global datasets from local information.

A2.16

P4, line 23 : Not clear what is meant by : Due to the lack of the distribution of hydrogeological properties.

Properties like hydraulic conductivity change vertically and laterally and can be very heterogeneous for a given region. The spatial distribution of these properties is not available on the global scale (or even the regional scale).
We added "spatial distribution" (Line 37, Page 3).

A2.17

P4, line 33 : I assume that "groundwater boxes" are actually "groundwater cells". If it's the case, then "cells" should be only consistently.

Now reads cells.

A2.18

P5. Figure 1 could be improved because it is not clear what is shown exactly. Also, what are "virtual layers"?

It is unclear to us what needs improvement. The explanation of virtual layers was added to the figure 1:
`[..], and the conceptual virtual layers (virtual because at this stage only confined conditions are computed)`

A2.19

P5. Equation 2 is not the correct partial differential equation (PDE). The cell volumes (delta_x, delta_y, delta_z) only appear when the PDE is integrated over a 3D cell. Also, writing that the partial differential equation is "a function of hydraulic head gradients" is not rigorously correct. The PDE is derived from applying mass conservation to a representative elementary volume, where groundwater flow is described by its mass flux. The hydraulic head gradients appear because the mass flux is expressed with Darcy's Law.

We clarified that the shown equation is approximated by using the finite element method. Furthermore, the sentence has been revised to be more correct.
It now reads(Line 6, Page 5):
`Three-dimensional groundwater flow is described by a partial differential equation (approximated in the model implementation by using the finite elements method) [equation] where [..]`

A2.20

P5, lines 20-21 : It is confusing to write that "Inflows in the groundwater are accounted for as…" because the equation is for both inflow and outflow.

It now reads In- and outflows.

A2.21

P7, line 35 : in the exponential, what is m? what is the value of f?

m is the unit meter as in the rest of the study. f is described in the sentence before as the e-folding factor

A2.22

P8, line 4 : the value of c_ocean is set to 100 m2/day. It appears to be several orders of magnitude greater than other conductances. Based on equation (4), I suspect that this large value is similar to specifying a first-type boundary condition for all cells located on the ocean boundaries. Is it the case?

No it is set to 10 m^2/day. Relative to the other conductances (see Figure 6) it is relatively small.

A2.23

P9, lines 11-15. The paragraph is not clear.

Now reads(Line 26, Page 8):

Similar to MODFLOW, G^3M-f solves Eq. (1) in two nested loops using a Picard iteration (Mehl, 2006): (1) the outer iteration checks the head and residual convergence criterion (if the maximum head change is below a given value and/or the residual norm is below a given value) and adjusts whether external flows have changed into a different state e.g. from gaining to losing conditions and updates the system of linear equations if flows are no longer head dependant. (2) The inner loop primarily consists of the conjugate gradient solver, which runs for a number of iterations defined by the user or until the residual convergence criterion is reached (Table 1), solving the current system of linear equations.

A2.24

P12, line 15 : "High conductance values are reached in the tropical zone due to a higher GW recharge". Is it really the case and not the opposite, i.e. because the conductance is large, groundwater recharge is larger?

Yes, groundwater recharge is an input to the GW model and is used to compute the shown conductance for gaining rivers (see also section 2.1).

A2.25

P14, line 11 : what is meant by : comparison to local studies suggested a unit conversion error?

The dataset of Fan provides the measurements of depth to groundwater in meter. By carefully comparing the values to local studies it seems that some of them haven't been properly converted and appear to be originally in feet. Because the original unit cannot be determined they were discarded.

Now reads (Line 15, Page 13):

We selected only observations with known land surface elevation and removed observations where a comparison to local studies suggested a unit conversion error. This left total of 1,070,402 depth to GW observations.

A2.26

P15, lines 18-19 : I don't understand the sentence : Plotting hydraulic head instead of depth to GW has the disadvantage that the goodness of fit is dominated by the topography as the observed heads are calculated based on the surface elevation of the model.

Well observations always measure water table depth not a hydraulic head. For some measurements a surface elevation is available but for many it is not. To make a consistent comparison the surface elevation of the model is used. The values of the surface elevation are much bigger than the water table depth and thus "smooth" out small variations. Thus, the comparison of heads is driven more by topography as by the actual water table depth.
This is clarified with an additional sentence(Line 9, Page 15):
Well observations always measure water table depth and only sometimes contain complementary data specifying the elevation at which the measurements were taken.

A2.27

P17, lines 29-30. I don't understand the reference to "model decision". What does it mean?

Now reads (Line 26, Page 18):
The presented comparison to other large-scale models is based on the assumption that same model deficiencies e.g. in available data and scale issues can uncover differences in model decisions e.g. used equations and spatial resolution.

A2.28

Table 2 : It is incorrect to write that the first 3 models solve the 3D Darcy equation. They solve a 3D mass conservation equation where the fluid flux is expressed with Darcy's law. It is also the case for ParFlow, which uses Darcy's Law to represent fluid fluxes in Richards' equation.

Table 2 was changed to indicate that the flow representation is either 2D or 3D and either saturated or unsaturated.

It now reads:

Aspect	**G³M**	**de Graaf et al. (2015; 2017)**	**Fan et al. (2013)**	**ParFlow**
Extent	Global	Global	Global	Continental USA
Resolution	5'	6'	30''	1 km
Software	G³M-f	MODFLOW	Unnamed	ParFlow
Computational expense	Medium	Medium	High	Very high
Flow representation	3D saturated	3D saturated	2D saturated	3D saturated/unsaturated
Time scale	Steady-state/(transient)	Steady-state/transient	Steady-state	Steady-state
Vertical layers	2	2	1	5
Full coupling possible	Yes	No (Conceptual issue)	No	Yes (already coupled)
In-memory coupling	Yes	No	N/A	Yes
Constant saturated thickness	Yes	Yes	No	No
Impermeable bottom	No	No	No	Yes
Surface water body location	In every cell	In almost every cell	No surface water	Created during simulation

Surface water body elevation	P_{30} of 30'' DEM	Avg. of 30'' DEM	N/A (outflow if depth to GW < 0.25 m)	N/A
Deviation from observations	Large	Very large	Medium	Medium

Challenges in developing a global gradient-based groundwater model (G^{3}M v1.0) for the integration into a global hydrological model

Robert Reinecke[1], Laura Foglia[3], Steffen Mehl[4], Tim Trautmann[1], Denise Cáceres[1], Petra Döll[1,2]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany
[2]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany
[3]Department of Land, Air and Water Resources, University of California, Davis, USA
[4]Department of Civil Engineering, California State University, Chico, USA

Correspondence to: Robert Reinecke (reinecke@em.uni-frankfurt.de)

Abstract. In global hydrological models, groundwater (GW) is typically represented by a bucket-like linear groundwater reservoir. Reservoir models, however, can (1) only simulate GW discharge to surface water (SW) bodies but not recharge from SW to GW, (2) provide no information on the location of the GW table and (3) assume that there is no GW flow among grid cells. This may lead, for example, to an underestimation of groundwater resources in semi-arid areas where GW is often replenished by SW or to an underestimation of evapotranspiration where the GW table is close to the land surface. To overcome these limitations, it is necessary to replace the reservoir model in global hydrological models with a hydraulic head gradient-based GW flow model.

We present G^{3}M, a new global gradient-based GW model with a spatial resolution of 5', which is to be integrated into the 0.5° WaterGAP Global Hydrology Model (WGHM). The newly developed model framework enables in-memory coupling to WGHM while keeping overall runtime relatively low, which allows sensitivity analyses, calibration, and data assimilation. This paper presents the G^{3}M concept and model design decisions that are specific to the large grid size required for a global scale model. In contrast to the GW model of de Graaf et al. (2015; 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G^{3}M, thus enabling full coupling to a GHM. Model results under steady-state naturalized conditions, i.e. neglecting GW abstractions, are shown. Simulated hydraulic heads show better agreement to observations around the world than the model output of de Graaf et al. (2015). Locations of simulated SW recharge to GW are found, as is expected, in dry and mountainous regions but the areal extent of SW recharge may be underestimated. Globally, GW discharge to rivers is by far the dominant flow component such that lateral GW flows only become a large fraction of total diffuse and focused recharge in case of losing rivers, some mountainous areas and some areas with very low GW recharge. Strong sensitivity of simulated hydraulic heads to the spatial resolution of the model and the related choice of the water table elevation of surface water bodies was found. We suggest to investigate how global-scale groundwater modeling at 5' spatial resolution can benefit from more highly resolved land surface elevation data.

Commented [o1]: #2.1,A1.1,A1.2

1 Introduction

Groundwater (GW) is the source of about 40% of all human water abstractions (Döll et al., 2014) and is also an essential source of water for freshwater biota in rivers, lakes, and wetlands. GW strongly affects river flow regimes and supplies the majority of river water during ecologically and economically critical periods with little precipitation. GW storage and flow dynamics have been altered by human GW abstractions as well as climate change and will continue to change in the future (Taylor et al., 2012). Around the globe, GW abstractions have led to lowered water tables and, in some regions, even GW depletion (Döll et al., 2014; Scanlon et al., 2012; Wada et al., 2012; Konikow, 2011). This has resulted in reduced base flows to rivers and wetlands (with negative impacts on water quality and freshwater ecosystems), land subsidence and increased pumping costs (Wada, 2016; Döll et al., 2014; Gleeson et al., 2012; 2016). The strategic importance of GW for global water and food security will probably intensify under climate change as more frequent and intense climate extremes increase the variability of SW flows (Taylor et al., 2012). International efforts have been made to promote sustainable GW management and knowledge exchange among countries, e.g., UNESCO's program on International Shared Aquifer Resources Management (ISARM) (http://isarm.org) and the ongoing GW component of the Transboundary Waters Assessment Program (TWAP) (http://www.geftwap.org). To support prioritization for investment among transboundary aquifers as well as identification of strategies for sustainable GW management, information on current conditions and possible trends of the GW systems is required (UNESCO-IHP, IGRAC, WWAP, 2012). In a globalized world, an improved understanding of GW systems and their interaction with SW and soil is needed not only at the local and regional but also at the global scale.

Commented [RR2]: #A2.2

To assess GW at the global scale, global hydrological models (GHMs) are used e.g. (Wada et al., 2012; 2016; Döll et al., 2012; 2014). In particular, they serve to quantify GW recharge (Döll and Fiedler, 2008). Like typical hydrological models at any scale, GHMs simulate GW dynamics by a linear reservoir model. In such a model, the temporal change of GW storage in each grid cell is computed from the balance of prescribed inflows and an outflow that is a linear function of GW storage. Linear reservoir models can only simulate GW discharge to SW bodies but not a reversal of this flow, even though losing streams may provide focused GW recharge that allows the aquifer to support ecosystems alongside the GW flow path (Stonestrom et al., 2007) as well as human GW abstractions. Losing streams typically occur in semi-arid and arid but seasonally also in humid regions. In addition, such linear reservoir models provide no information on the location of the GW table and assume that GW flow among grid cells is negligible. To simulate the dynamics of water flow between SW bodies and GW in both directions as well as the effect of capillary rise on evapotranspiration, it is necessary to compute lateral GW flows among grid cells as a function of hydraulic head gradients and thus the dynamic location of the GW table. To achieve an improved understanding of GW systems at the global scale, and in particular of the interactions of GW with SW and soil, it is therefore necessary to replace the linear GW reservoir model in GHMs by a hydraulic gradient-based GW flow model.

Commented [o3]: #1.4

Large-scale gradient-based GW flow models are still rare and mainly available for data-rich regions, e.g. for the Death Valley (Belcher and Sweetkind, 2010) and the Central Valley (Belcher and Sweetkind, 2010; Faunt, 2009; Dogrul et al., 2016) in the USA, but also for large fossil groundwater bodies in arid regions (e.g. the Nubian Aquifer System in North Africa, (Gossel et al., 2004)). However, they are in most cases not integrated within hydrological models that quantify GW recharge based on climate data and provide information on the condition of SW (e.g. streamflow and storage). For North America, Fan et al. (2007) and Miguez-Macho et al. (2007) linked a land surface model with a two-dimensional gradient-based GW model and computed, with a daily time step, GW flow, water table elevation, GW–SW interaction, and capillary rise, using a spatial resolution of 1.25 km. One challenge was the determination of the river conductance that affects the degree of GW–SW interaction. A computationally very expensive integrated simulation of dynamic SW, soil and GW flow using Richards' equation for variably saturated flow was achieved at a spatial resolution of 1 km for the continental US by applying the ParFlow model (Maxwell et al., 2015). In both studies, GW abstractions were not taken into account.

Commented [RR4]: #A2.3

Commented [RR5]: #A2.4

Commented [o6]: #A2.5

A first simulation of the steady-state GW table for the whole globe at the very high resolution of 30" was presented by Fan et al. (2013) and compared to an extensive compilation of observed hydraulic heads. However, there were no head-based interactions with SW; GW above the land surface was simply discarded. Global GW flow modeling is strongly hampered by data availability, including the geometry of aquifers and aquitards as well as parameters like hydraulic conductivity (de Graaf et al., 2017), and by computational restrictions on spatial resolution leading to conceptual problems, e.g., regarding SW-GW interactions (Morel-Seytoux et al., 2017). Recently, some GW flow models that are in principle applicable for the global scale were developed but were applied only regionally in data-rich regions (Rhine basin: Sutanudjaja et al., 2011; France: Vergnes et al., 2012; 2014). The first global gradient-based GW model that was run for both steady-state (de Graaf et al., 2015) and transient conditions (de Graaf et al., 2017) was driven by GW recharge and SW data of the GHM PCR-GLOBWB (van Beek et al., 2011). However, there is not yet a two-way coupling of a GW flow model and a GHM. This may be due to the way de Graaf et al. (2015; 2017) modeled river-GW interaction. To achieve plausible hydraulic head results, they found it necessary to add a drainage flux additionally to the GW flow driven by the hydraulic head difference between GW and river. This additional drainage, which accounts for about 50% of global GW flow into SW, is simulated as a function of GW storage

Commented [RR7]: #2.5

Commented [RR8]: #2.5

above the floodplain. The values needed to compute this additional artificial drainage are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017)) – the model component that the gradient-based model was intended to replace in the first place. This prevents a full integration of the global GW flow model of de Graaf et al. (2017) into a GHM, as then, the linear GW reservoir model would be replaced by the GW flow model.

In this study, we present the Global Gradient-based Groundwater Model (G^3M) that is to be integrated into the GHM WaterGAP 2 (in the following we refer to WGHM (WaterGAP Global Hydrology Model), which is part of the GHM WaterGAP) to improve estimation of flows between SW and GW (affecting both streamflow and groundwater recharge and thus water availability for humans and ecosystems) and implement capillary rise (affecting evapotranspiration).

The objective of this paper is to learn from a steady-state model, a well-established first step in groundwater model development, to (1) understand the basic model behavior by limiting model complexity and degrees of freedom, and thus (2) providing insights into dominant processes and uncovering potential model-inherent characteristics difficult to observe in a fully coupled transient model. A transient model might obfuscate model inherent trends due to the slow changing nature of groundwater processes e.g. trends towards large over/under-estimation due to wrong parameterization. A fully coupled model furthermore adds complexity and uncertainty to the model outcome. The presented steady-state model is furthermore used to (3) investigate parameter sensitivity and sensitivity to spatial resolution. In addition, the steady-state solution can be used as (4) initial condition for future fully coupled transient runs.

Model concept and equations, as well as applied data and parameter values, are presented in section 2. In section 3, we show steady-state results of G^3M driven by WGHM data. Simulated hydraulic heads are compared to observations worldwide and to the output of established (regional) models. Furthermore, sensitivity to parameters and grid size is shown for the example of New Zealand. Finally, the implications of modeling decisions and grid size are discussed (section 4) and conclusions are drawn (section 5).

Commented [RR9]: #A.1.5

Commented [o10]: #A2.5

Commented [RR11]: #A2.10,11

Commented [o12]: #A2.9

2 Model description

Commented [RR13]: #1.2

2.1 G^3M model concept

Although G^3M is based on principles of the well-known GW flow modeling software MODFLOW (Harbaugh, 2005), G^3M differs in its parameterization from traditional local and regional GW models. These models are generally based on rather detailed information on hydrogeology (including aquifer geometry and properties such as hydraulic conductivity derived from pumping tests), topography, pumping wells, location and shape of SW bodies as well as on observations of hydraulic head in GW and SW. Local observations guide the developer in constructing the model such that local conditions and processes can be properly represented. The lateral extent of individual grid cells of such GW flow models is generally smaller or similar to the depth of the aquifer(s) and the size of the SW bodies that interact with the GW. The global GW flow model G^3M, however, covers all continents of the Earth except Greenland and Antarctica. At this scale, the information listed above is poor or non-existing, and the lateral extent of grid cells needs to be relatively large due to computational (and data) constraints. We selected a grid cell size of 5' by 5' (approx. 9 km by 9 km at the equator), as this size fits well to WaterGAP and is smaller than the suggested 6' of Krakauer et al. (2014). WaterGAP 3 (Eisner, 2016) has the same cell size, and 36 of such cells fit into one 0.5° WaterGAP 2 cell. Global climate data are only available for 0.5° grid cells. The landmask of G^3M, i.e. location and size of 5' grid cells, is that of WaterGAP 3.

Due to the lack of the spatial distribution of hydrogeological properties, we chose to use, in the current version of G^3M, two GW layers with a vertical size of 100 m each (Fig. 1). We performed a sensitivity analysis that confirmed the findings of others (de Graaf et al., 2015) that the aquifer thickness has a relatively small impact on the model results. Therefore, selecting a uniform thickness of 100 m (motivated by the assumed depth of validity of the lithology data) (Fig. 1) worldwide for the first

Commented [RR14]: #A2.16

layer and also for the second layer is expected to lead to fewer uncertainties as compared to hydraulic conductivities and the surface water table elevation.

G³M focuses on a plausible simulation of water flows between GW and SW, and we deemed it suitable to have an upper GW layer that interacts with SW and soil (the soil layer of WaterGAP is described in detail in S1) and a lower one in which GW may flow laterally without such interactions. As land surface elevation within each 5' grid cell, with an area of approximately 80 km^2, may vary by more than 200 m (Fig. S4.1), neighboring cells in G³M may not be adjacent anymore (Fig. 1), in contrast to (regional) GW models with smaller grid cells. This makes G³M a rather conceptual model in which water exchange between groundwater cells is driven by hydraulic head gradients but flow can no longer be conceptualized as occurring through continuous pore space. In addition, due to the coarse spatial scale and the possible large variations of land surface elevations within each grid cell, the upper model layers should not be considered to be aligned with an average land surface elevation. The model layers can be rather thought to be vertically aligned with the elevation of the surface water body table, as this prescribed elevation is, together with the sea level, the only elevation included in the groundwater flow equation (Eq. (1)).

The simulation of aquifers that contain dry cells and/or cells that oscillate between wet and dry states poses great challenges to solving Eq. (1) (Niswonger et al., 2011). G³M-f (the framework code used to implement G³M) implements the traditional wetting approach from Harbaugh (2005) as well as the approach proposed by Niswonger et al. (2011) along with the proposed damping scheme. Both approaches have proven to be insufficient to simulate head-based transmissivities (unconfined conditions) on the global scale. Large mountainous areas would be excluded from the beginning of the solution step, as the head is often far below the deepest model layer, resulting in a no-flow condition and imposing convergence issues to the matrix solver. We choose to simulate both layers with a specific saturated thickness even though the upper layer can be expected to decrease in depth and thus in transmissivity (hydraulic conductivity times saturated depth). The large uncertainties regarding hydraulic conductivities (possibly an order of magnitude), further justify using the computationally more efficient assumption of specified saturated thickness. This approach is consistent with findings that this is accurate for large, complex groundwater models (Sheets et al., 2015). Furthermore, it is consistent with recent presented large scale studies e.g. for the Rhine Meuse basin of Sutanudjaja et al. (2011) (using one confined layer), the Death Valley Regional Flow Model (Belcher, 2004; Faunt et al., 2011), and the global groundwater model of de Graaf et al. (2017) (two layers and partially unconfined conditions are simulated by parametrization of the model input and not by a head-dependant transmissivity).

Commented [RR15]: #2.3

Commented [RR16]: #A2.17

Commented [RR17]: #A1.7

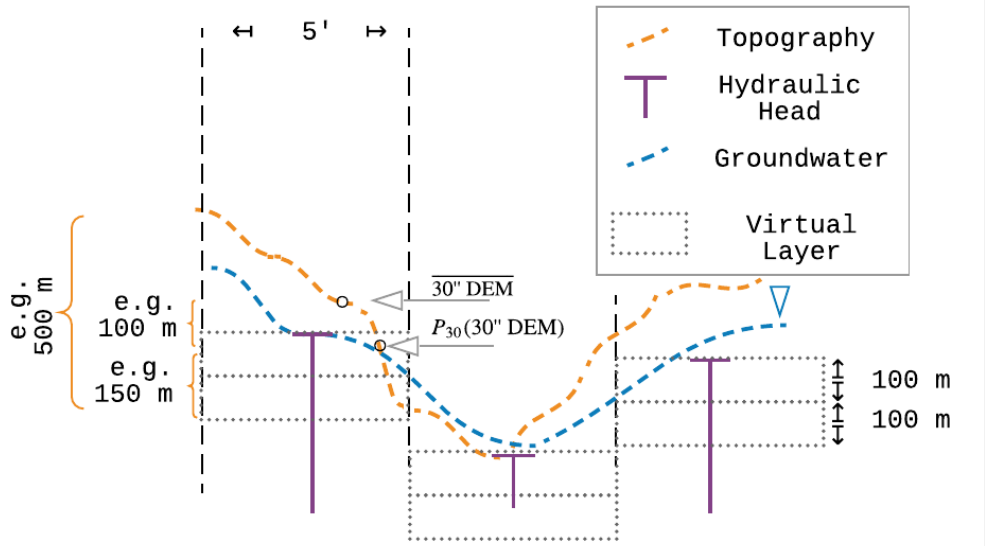

Figure 1 Schematic of G³M's spatial structure, with 5' grid cells, hydraulic head per cell, and the conceptual virtual layers (virtual because at this stage only confined conditions are computed). The underlying variability of the topography changes the perception of simulated depth to groundwater depending on what metrics are used to represent it on a coarser resolution. Layers in G³M are of a conceptual nature and describe the saturated flow between locations of head laterally and vertically. The P_{30} is used in the presented steady-state model as SW elevation instead of an average or minimum per grid cell.

Three-dimensional groundwater flow is described by a partial differential equation (approximated in the model implementation by using the finite elements method)

$$\frac{dGWS}{dt} = \left(\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) + W\right) \Delta x \Delta y \Delta z = S_s \frac{\partial h}{\partial t} \Delta x \Delta y \Delta z \tag{1}$$

where $K_{x,y,z}$ is the hydraulic conductivity $[LT^{-1}]$ along the x, y, and z axis between the cells (harmonic mean of grid cell conductivity values), S_s the specific storage $[L^{-1}]$, $\Delta x \Delta y \Delta z$ $[L^3]$ the volume of the cell, and h the hydraulic head $[L]$. In- and out-flows in the groundwater are accounted for as

$$W \, \Delta x \Delta y \Delta z = R_g + Q_{swb} - NA_g - Q_{cr} + Q_{ocean} \tag{2}$$

where Q_{swb} is flow between the SW bodies (rivers, lakes, reservoirs, and wetlands) and GW $[L^3T^{-1}]$, Q_{cr} is capillary rise, i.e. the flow from GW to the soil, and Q_{ocean} is the flow between ocean and GW$[L^3T^{-1}]$, representing the boundary condition. In case of Q_{swb} and Q_{ocean}, a positive value represents a flow into the groundwater. R_g $[L^3T^{-1}]$ is diffuse GW recharge from soil and NA_g $[L^3T^{-1}]$ net GW abstraction (both calculated by WaterGAP see Eq. (S1)).

The flux across the model domain boundary Q_{ocean} is modeled as a head-dependent flow based on a static head boundary.

$$Q_{ocean} = c_{ocean}\left(h_{ocean} - h_{aq}\right) \tag{3}$$

Here h_{ocean} is the elevation of the ocean water table $[L]$, h_{aq} the hydraulic head in the aquifer $[L]$ and c_{ocean} the conductance of the boundary condition $[L^2T^{-1}]$ (Table 1). We assume that density difference to sea-water is negligible at this scale. Q_{cr} is not yet implemented in G³M.

Q_{swb} in Eq. (4) replaces k_g GWS and R_{g_swb} in the linear storage equation of WaterGAP (Eq. (S1)), such that losing conditions of all types of SW bodies can be simulated dynamically. It is calculated as a function of the difference between the elevation of the water table in the SW bodies h_{swb} $[L]$ and h_{aq} as

Commented [RR18]: #A2.18

Commented [RR19]: #A1.6

Commented [RR20]: #A2.19

Commented [RR21]: #A2.20

$$Q_{swb} = \begin{cases} c_{swb}(h_{swb} - h_{aq}) & h_{aq} > B_{swb} \\ c_{swb}(h_{swb} - B_{swb}) & h_{aq} \le B_{swb} \end{cases} \tag{4}$$

where c_{swb} is the conductance $[L^2T^{-1}]$ of the SW body bed (river, lake, reservoir or wetland) and B_{swb} the SW body bottom elevation $[L]$.

The conductance of SW bodies is often a calibration parameter in traditional GW models (Morel-Seytoux et al., 2017). Following Harbaugh (2005), it can be estimated by

$$c_{swb} = \frac{K\,L\,W}{h_{swb} - B_{swb}} \tag{5}$$

where K is hydraulic conductivity, L is length and W is the width of the SW body per grid cell. For lakes (including reservoirs) and wetlands, c_{swb} is estimated based on hydraulic conductivity of the aquifer K_{aq} and SW body area (Table 1). For gaining rivers, conductance is quantified individually for each grid cell following an approach proposed by Miguez-Macho et al. (2007). The value of river conductance c_{riv}, according to Miguez-Macho et al. (2007), in a GW flow model needs to be set to such a values that, for steady-state conditions, the river is the sink for all the inflow to the grid cell (GW recharge and inflow from neighbouring cells) that is not transported laterally to neighbouring cells such that

$$c_{riv} = \frac{R_g + Q_{eq_{lateral}}}{h_{eq} - h_{riv}} \qquad h_{aq} > h_{riv} \tag{6}$$

For G[3]M, we computed the equilibrium head h_{eq} as the 5' average of the 30'' steady-state heads calculated by Fan et al. (2013). Using WGHM diffuse GW recharge lateral equilibrium flow $Q_{eq_{lateral}}$ $[L^3T^{-1}]$ is net lateral inflow into the cell computed based on the h_{eq} distribution as well as G[3]M K_{aq} and cell thickness (Table 1). Elevation of the river water table h_{riv} [L] is to be provided by WGHM. Using a fully dynamic approach, i.e. utilizing the hydraulic head and lateral flows from the current iteration to re-calculate c_{riv} in each iteration towards the steady-state solution, has proven to be too unstable due to its non-linearity affecting convergence. We limit c_{riv} to a maximum of $10^7\ m^2 day^{-1}$; this would be approximately the value for a 10 km long and 1 km wide river with a head difference between GW and river of 1 m and hydraulic conductivity of the river bed of 10^{-5} m/s.

If the river recharges the GW (losing river), Eq. (6) cannot be used as the Fan et al. (2013) high-resolution equilibrium model only models groundwater outflows but not inflows from SW bodies. If h_{aq} drops below h_{riv}, Eq. (5) is used to compute c_{riv}, with K equals to K_{aq}.

2.2 The steady-state uncoupled model version

In a first implementation stage, G³M was developed as a steady-state (right-hand side of Eq. (1) is zero) standalone model that represents naturalized conditions (i.e. without taking into account human water use) during 1901-2013. Input data and parameters used are listed in Table 1 and described below.

Gleeson et al. (2014) provided a global subsurface permeability data set from which K_{aq} was computed. The data set was derived by relating permeabilities from a large number of local to regional GW models to the type of hydrolithological units (e.g., "unconsolidated" or "crystalline"). The geometric mean permeability values of nine hydrolithological units were mapped to the high-resolution global lithology map GLiM (Hartmann and Moosdorf, 2012). In continuous permafrost areas, a very low permeability value was assumed by Gleeson et al. (2014). The estimated values represent the shallow surface on the scale of 100 m depth. The unique dataset has three inherent problems when used as input for a GW model: (1) At this scale, important heterogeneities such as discrete fractures or connected zones of high hydraulic conductivity controlling the GW flow are not visible. (2) Jurisdictional boundaries due to different data sources in the global lithological map lead to artifacts. (3) The differentiation between coarse and fine-grained unconsolidated deposits is only available in some regions resulting in $10^{-4}\ m\ s^{-1}$ as hydraulic conductivity for coarse-grained unconsolidated deposits. If the distinction is not available, a rather low value of $10^{-6}\ m\ s^{-1}$ is set for unconsolidated porous media (Fig. S4.3). The original data was gridded to 5' by using an area-weighted average and used as hydraulic conductivity of the upper model layer. For the second layer, hydraulic conductivity of the first layer is reduced assuming that conductivity decreases exponentially with depth. Based on the e-folding factor f used by Fan et al. (2013) (a calibrated parameter based on terrain slope), the conductivity of the lower layer is calculated by multiplying the upper layer value by $\exp(-50\ m\ f^{-1})^{-1}$ (Fan et al., 2007).

Mean annual GW recharge computed by WaterGAP 2.2c for the period 1901-2013 is used as input (Fig. S4.4), while no net abstraction from GW was taken into account. It would not be meaningful to try to derive a steady-state solution under existing net groundwater abstractions that in some regions cause GW depletion with continuously dropping water tables. The 0.5° data of WaterGAP was equally distributed to the according 5' grid. Regarding the ocean boundary condition, h_{ocean} is set to 0 m and c_{ocean} to $10\ m^2\ day^{-1}$ (Table 1), reflecting a global average conductance based on hydraulic conductivity and lateral surface area.

Commented [RR22]: #A1.11

It is assumed that there is an exchange of water between GW and one river stretch in each 5' grid cell, and in addition where lakes and wetlands exist according to WaterGAP 3, which provides, for each grid cell, the area of "local" and "global" lakes and wetlands. In WaterGAP, "local" SW bodies are only recharged by runoff produced within the grid cell, while "global" SW bodies also obtain inflow from the upstream cell. In an uncoupled model, it is difficult to prescribe the area of lakes and wetlands that affect the flow exchange between SW body and GW. Maps generally show the maximum spatial extent of SW bodies. This maximum extent is seldom reached, in particular in case of wetlands in dry areas. For global wetlands, it is therefore assumed in this model version that only 80% of their maximum extent is reached. In the transient model, SW extends will be changed over time. A further difficulty in an uncoupled model run is that the water table elevation of SW bodies does not react to the GW-SW exchange flows Q_{swb} and that water supply from SW is not limited by availability. A losing river may in reality dry out due to loss to GW and therefore cease to lose any more water. For rivers B_{swb} is equal to $h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994), where $Q_{bankfull}$ is the bankfull river discharge in the 5' grid cell (Verzano et al., 2012). Globally constant but different values are used for B_{swb} in case of wetlands, local lakes and global lakes (Table 1).

For the steady-state model, river elevation h_{riv} is set in each grid cell to the same elevation as all other SW bodies, h_{swb}. We found that for both gaining and losing conditions, Q_{swb} and thus computed hydraulic heads are highly sensitive to h_{swb}. The overall best agreement with the hydraulic head observations of Fan et al. (2013) was achieved if h_{swb} (Eq. (4), (5) and (6)) was set to the 30[th] percentile (P_{30}) of the 30" land surface elevation values of Fan et al. (2013) per 5' cell, e.g. the 30" elevation that is exceeded by 70% of the thousand 30" elevation values within one 5' cell. To decrease convergence time we used h_{eq} derived from the high-resolution steady-state hydraulic head distribution of Fan et al. (2013) as initial guess.

2.3 Integration into WGHM

In this initial effort, we intend to integrate G³M into WaterGAP 2, i.e. the 0.5° version of WGHM (for details see S1) to keep computation time low enough for performing sensitivity analyses and ensemble-based data assimilation and calibration, instead of integrating it into WaterGAP 3 (Eisner, 2016), which has the same spatial resolution as G³M. However, data from WaterGAP 3 were used to set up G³M. Location and area of the 5' grid cells of G³M are the same as in the landmask of WaterGAP 3. In addition, the percentage of the 5' grid cell area that is covered by lakes (including reservoirs) and by wetlands, based on Lehner and Döll (2004), is taken from WaterGAP 3, as well as the length and width of the main river within each 5' grid cell as estimated by WaterGAP 3 (Table 1).

Commented [RR23]: #1.1, A1.10

2.4 Model implementation

G³M is implemented using a newly developed open-source model framework G³M-f (Reinecke, 2018). The main motivation to develop a new model framework is the efficient in-memory coupling to the GHM and flexible adaptation to the specific requirements of global-scale modeling. Written in C++14, the framework allows the implementation of global and regional groundwater models alike while providing an extensible purely object-oriented model environment. It is primarily targeted as an extension to WGHM but allows an in-memory coupling to any GHM or can be used as a standalone groundwater model. It provides a unit-tested environment offering different modules that can couple results in-memory to a different model or write out data flows to different file formats. G³M-f has the following advantages over using an established GW modeling software such as MODFLOW. G³M-f enables an improved coupling capability: (1) as it is intended to be used as a library-like module (unlike MODFLOW it provides a clear development interface to the programmer coupling a model to G³M-f, can be easily compiled as a library, and provides a clearly separated logic between computation and data read-in/write-out), (2) is written in the same language as the target GHM enabling a straight-forward in-memory access to arrays without the need to write data to disk, required by other global models, (a very expensive operation even if that disk is a RAM-disk). Even though it is possible to call FORTRAN functions from C++, it is very complicated to pass file pointers properly, as the I/O implementation of both languages differ substantially and it is widely considered bad practice to handle I/O in two different languages at once. As MODFLOW was never designed to be coupled/integrated to/into other models, it is not possible to separate the I/O logic fully from the computational logic without substantial code changes that are hard to test. To this end, G³M-f provides a highly modularized framework that is written with extensibility as a design goal while implementing all required groundwater mechanisms.

As internal numerical library, it uses Eigen3 (eigen.tuxfamily.org). Different from Vergnes et al. (2014), G³M's computations are not based on spherical coordinates directly but on an irregular grid of rectangular cells. Cell sizes are provided by WaterGAP3 and are derived from their spherical coordinates maintaining their correct area and center location. The model code will be adapted in the future to account for the different length in x and y-direction per cell correctly.

Eq. (1) is reformulated as a finite-difference equation and solved using a conjugate gradient approach and an Incomplete LUT preconditioner (Saad, 1994). In order to keep the memory footprint low, the conjugate gradient method makes use of the sparse matrix. Furthermore, it solves the equations in parallel (preconditioner currently non-parallel). G³M can compute the presented steady-state solution (with the right-hand side of Eq. (1) being zero and the heads of Fan et al. (2013) as initial guess, Table 1) on a commodity computer with four computational cores and a standard SSD in about 30 minutes while occupying 6 GB of RAM.

Similar to MODFLOW, G³M-f solves Eq. (1) in two nested loops using a Picard iteration (Mehl, 2006): (1) the outer iteration checks the head and residual convergence criterion (if the maximum head change is below a given value and/or the residual norm is below a given value) and adjusts whether external flows have changed into a different state e.g. from gaining to losing conditions and updates the system of linear equations if flows are no longer head dependant. (2) The inner loop primarily consists of the conjugate gradient solver, which runs for a number of iterations defined by the user or until the residual convergence criterion is reached (Table 1), solving the current system of linear equations.

Commented [RR24]: #A2.23

Because of the switch between Eq. (5) and Eq. (6) that occurs if e.g. h_{aq}drops below h_{riv} from one iteration to the next causes an abrupt change of c_{riv} inducing a nonlinearity that affects convergence we introduced an ϵ = 1 m interval around h_{riv} and interpolate c_{riv} by a cubic hermite spline polynomial over that interval. This allows for a smoother transition between both states, reducing the changes in the solution if a river is in a gaining condition in one iteration and in a losing condition in the next and vice versa.

3 Results

3.1 Global hydraulic head and depth to GW distribution under natural steady-state conditions

As expected, the computed global distribution of steady-state hydraulic head (in the upper model layer) under natural conditions (Fig. 2a) follows largely the land surface elevation (Fig. S4.2), albeit with a lower range and locally different ratios between the hydraulic head and land surface gradients (Fig. S4.7). Depth to GW can be computed by subtracting the hydraulic head computed by G^3M for the upper layer of each 5' grid cell from the arithmetic mean of the land surface elevations of the 100 30" grid cells within each 5' cells (Fig. S4.2). The global map of steady-state depth to GW (Fig. S4.5) clearly resembles the map of differences between surface elevation and P_{30}, the assumed water level of SW bodies h_{swb}, shown in Fig. S4.1, which indicates that simulated depth to GW is strongly governed by the assumed water level in SW bodies.

Deep GW, i.e. a large depth to GW, occurs mainly in mountainous regions (Fig. S4.5). These high values of depth to GW are mainly a reflection of the steep relief in these areas as quantified either by the differences of mean land surface elevations between neighbouring grid cells (Fig. S4.8) or the difference between mean land surface elevation and P_{30}, the 30th percentile of the 30" land surface elevations (Fig. S4.1). When the computed hydraulic head is subtracted not from average land surface elevation but from P_{30}, the assumed water table elevation of SW bodies, the resulting map shows that the groundwater table is mostly above P_{30}, in both flat and steep terrain (Fig. 2b). Thus, high depth to GW values at the 5' resolution does not indicate deep unsaturated zones and losing rivers but just high land surface elevation variations within a grid cell. In steep terrain, 5' water tables are higher above the water level in the surface water bodies than in flat terrain (Fig. 2b). Deep GW tables that are not only far below the mean land surface elevation but also below the water table of surface water bodies are simulated to occur in some (steep or flat) desert area with very low GW recharge.

In 2.1 % of all cells, GW head is simulated to be above the average land surface elevation, by more than 1 m in 0.3 % and by more than 100 m in 0.004 % of the cells. The shallow water table in large parts of the Sahara is caused by losing rivers (and some wetlands) that cannot run dry in the model, causing an overestimation of the GW table (section 2.2). Please note that the computed steady-state depth to GW certainly underestimates the steady depth to GW in GW depletion areas such as the High Plains Aquifer and the Central Valley in the USA (see S2), North-western India, North China Plain and parts of Saudi Arabia and Iran (Döll et al., 2014) as groundwater withdrawals are not taken into account in the presented steady-state simulation of G^3M.

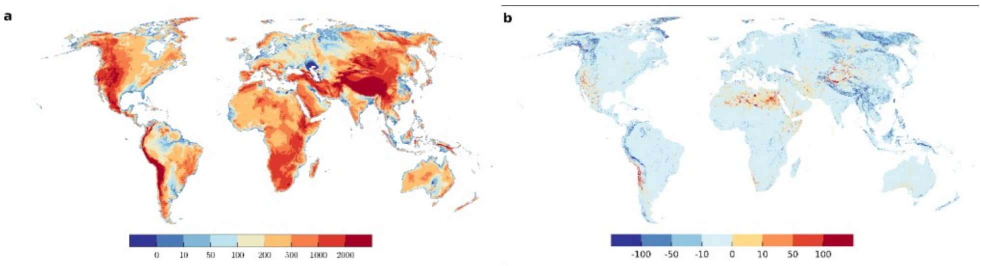

Figure 2 (a) Simulated equilibrium hydraulic head [$\boldsymbol{m}$]. Maximum value 6375 m, minimum value -414 m (Extremes included in dark blue and dark red). (b) Difference between 30th percentile of the 30" land surface elevation per 5' grid cell (chosen elevation for surface water bodies) and simulated equilibrium hydraulic head. Maximum value 1723 m, minimum value -1340 m (Extremes included in dark blue and dark red).

3.2 Global water budget

Inflows to and outflows from GW of all G^3M grid cells were aggregated according to the compartments ocean, river, lake, wetland, and diffuse GW recharge from soil (Fig. 3). The difference between the global sum of inflows and outflows is less than 10^{-6} %. This small volume balance error indicates the correctness of the numerical solution.

Total diffuse GW recharge from soil is $3.9\ 10^{10}\ m^3\ day^{-1}$ and approximately equal to the flow of GW to rivers. Rivers are the ubiquitous drainage component of the model, followed by wetlands, lakes and the ocean boundary. According to G^3M, the amount of river water that recharges GW is more than one order of magnitude less than the drainage of GW, and the relative recharge to GW from lakes and wetlands is even smaller (Fig. 3). Possibly, flow from SW to GW is overestimated, as outflow from SW is not limited by water availability in the SW, and depending on the hydraulic conductivity, Eqs. (5) and

Commented [RR25]: #2.5

Commented [o26]: #A1.13

(6) can lead to rather large flows. Inflow from the ocean, which is more than two magnitudes smaller than outflow to the ocean, occurs in regions where h_{swb} = P_{30} is below h_{ocean}.

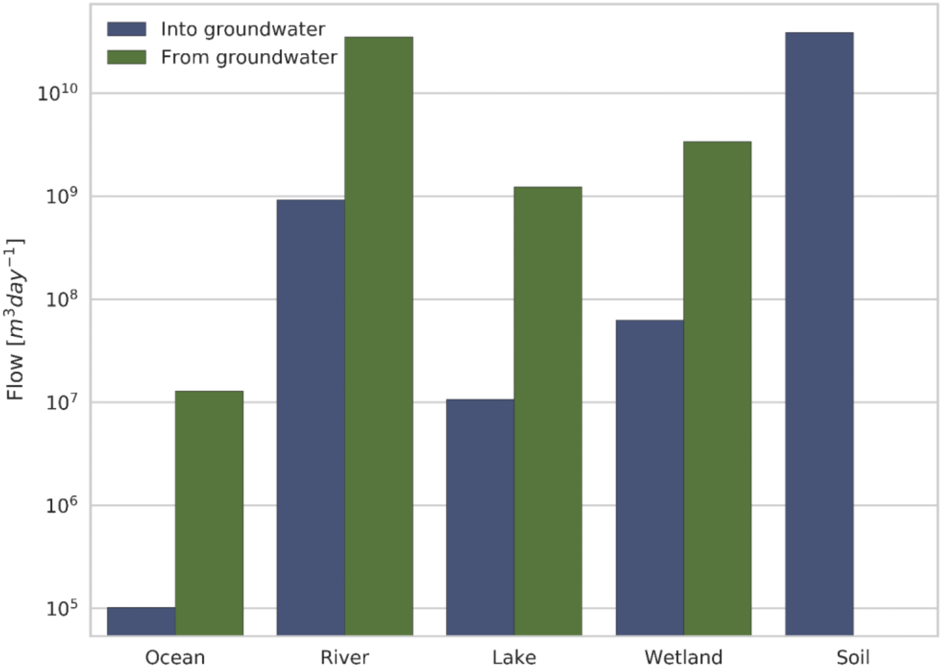

Figure 3 Global sums of flows from different compartments into or from GW at steady state. Flows into the GW are denoted by the color blue, flows out of the GW into the different compartments by green. The compartment soil is the diffuse GW recharge from soil calculated by WaterGAP.

3.3 GW-SW interactions

Figure 4 plots the spatial distribution of simulated flows from and to lakes and wetlands (Fig. 4a) as well as from and to rivers (Fig. 4b). It reveals strong interaction between GW and SW bodies that is dominated by GW discharging to SW bodies. Parallel to the overall budget (Fig. 3), the map reveals the globally large but locally strongly varying influence of lakes and wetlands (Fig. 4a). Rivers with riparian wetlands such as the Amazon River receive comparably small amounts of GW as most of the GW is drained by the wetland (compare Figs. 4a and 4b). Similarly, areas dominated by wetlands and lakes (e.g. parts of Canada and Scandinavia) show less inflow for rivers (Fig. 4b). 93 % of all grid cells contain gaining rivers, and only 7% losing rivers. Gaining lakes and wetlands are found in 12 % and 11 % of the cells, respectively, whereas only 0.2 % contain a losing lake or wetland. In G^3M, all SW bodies (rivers, lakes, and wetlands) in a grid cell either loose or gain water.

Gaining rivers, lakes, and wetlands with very high absolute Q_{swb} values over 1 mm day^{-1} (averaged over the grid cell area of approximately 80 km^2) can be found in the Amazon, Congo, Bangladesh, and Indonesia, where GW recharge in very high (Fig. S3.4). Values below 0.01 mm day^{-1} are present in dry and in permafrost areas where groundwater recharge is small.

Losing SW bodies occur in the model under two conditions, in mountainous regions where depth to GW is high and in arid and semi-arid climates with low diffuse groundwater recharge. Without focused GW recharge, the GW table would drop to even further in the mountains and is necessary to counteract the large hydraulic gradients caused by the large topographic gradients. Rivers lose more than 1 $\mathrm{mm\ day^{-1}}$ in Ethiopia and Somalia, West Asia, Northern Russia, the Rocky Mountains, and the Andes whereas lower values can be observed in Australia and in the Sahara. High values of outflow from wetlands and lakes are found in Tibet, the Andes and northern Russia, lower values in the Sahara and Kazakhstan. The river Nile in the Northern Sudan and Egypt is correctly simulated to be a losing river (Fig. 4b), being an allogenic river that is mainly sourced from the upstream humid areas, including the man-made Lake Nasser (Elsawwaf et al., 2014) (Fig. 4a). Furthermore, the following lakes and riparian wetlands are simulated to recharge GW: parts of the Congo River, Lake Victoria, the Ijsselmeer, Lake Ladoga, the Aral Sea, parts of the Mekong Delta, the Great Lakes of North America. On the other hand, no losing stretches are visible at the Niger River and its wetlands and almost none in the North-eastern Brazil region even though that losing conditions are known to occur there (Costa et al., 2013; FAO, 1997).

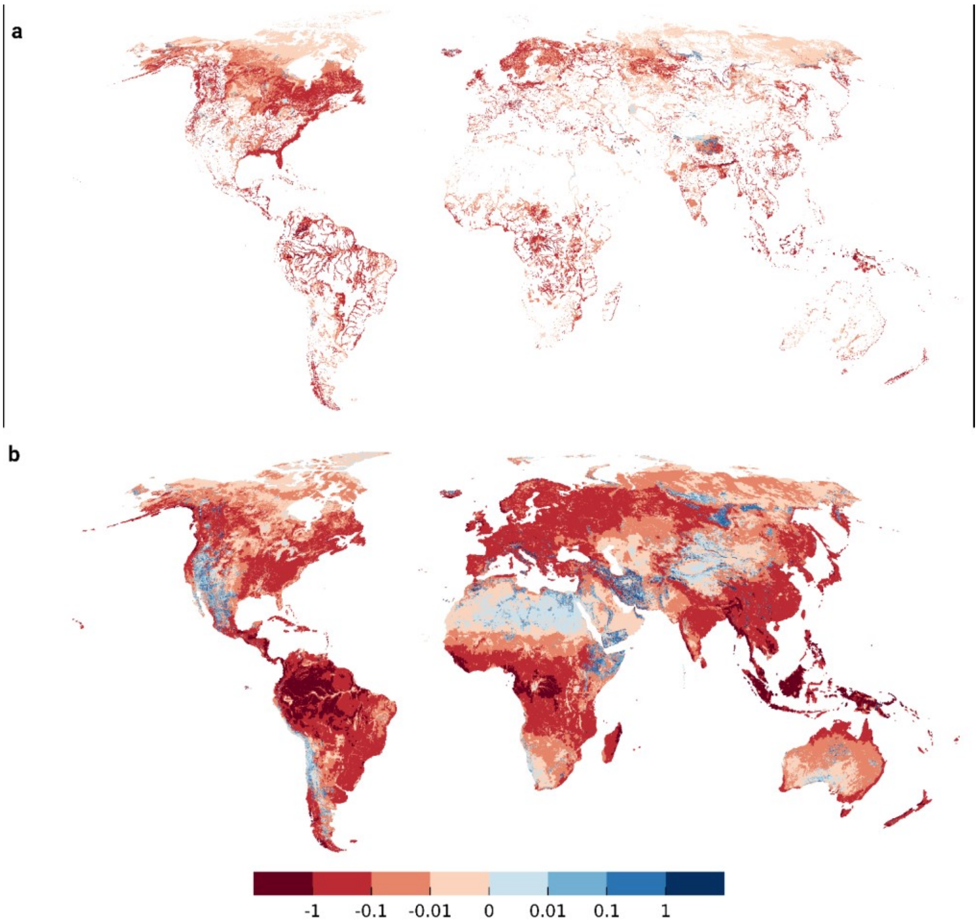

Figure 4 Flow $\boldsymbol{Q_{swb}}$ $[\boldsymbol{mm\ day^{-1}}]$ from/to wetlands, lakes (a) and losing/gaining streams (b) with respect to the 5' grid cell area. Gaining surface water bodies are shown in red, surface water bodies recharging the aquifer in blue. Focused aquifer recharge occurs in arid regions, e.g. alongside the river Nile, and in mountainous regions where the average water table is well below the land surface elevation.

Simulated flows between GW and SW depend on assumed conductances for both rivers and lakes/wetlands (Eqs. (4), (5), (6)) shown in Fig. 5. Q_{swb} (Fig. 4) correlates positively with conductance. Conductance for gaining rivers correlates positively with GW recharge (Eq. (6) and Fig. S4.4). High conductance values are reached in the tropical zone due to a higher GW recharge but are capped at a plausible maximum value of $10^7\ m^2day^{-1}$s in case of a river (section 2.1) (Fig. 5b), while lakes and wetlands, with a larger area, can reach larger values, e.g. in Canada or Florida.

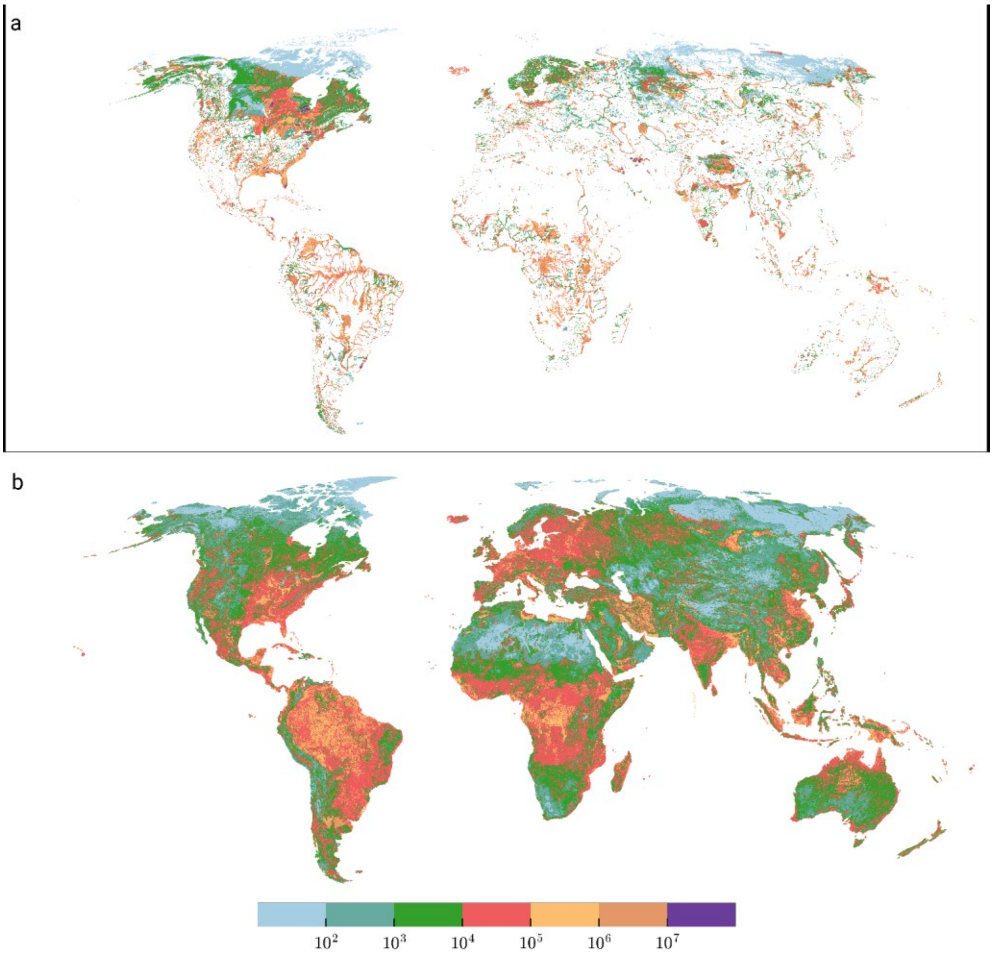

Figure 5 Conductance [m^2day^{-1}] of lakes and wetlands (a) and rivers (b). In regions close to the pole conductance is in general lower due to the influence of the low aquifer conductivity (losing conditions), and relatively small GW recharge due to permafrost conditions (only applies for gaining conditions). Max conductance of wetlands is $10^8\ m^2day^{-1}$.

3.4 Lateral flows

Figure 6 shows lateral outflow from both model layers in percent of the sum of diffuse GW recharge from soil and GW recharge from SW bodies. The percentage of recharge that is transported through lateral flow to neighboring cells depends on 5 main factors: (1) hydraulic conductivity (Fig. S4.3), (2) diffuse GW recharge (Fig. S4.4), (3) losing or gaining SW bodies (Fig. 4), (4) their conductance (Fig. 5) and (5) the head gradients (Fig. 2a).

In large areas of the globe, where GW discharges to SW, the lateral flow percentage is less than 0.5% of the total GW recharge to the grid cell, as most of the GW recharge is simulated to leave the cell by discharge to SW. For example, in the permafrost regions, very low hydraulic conductivity limits the outflow to neighboring cells of the occurring recharge, leading to these very low percent values. Such values also occur in regions with high SW conductances and rather low hydraulic conductivity, e.g. in the Amazon Basin. Values of more than 5% occur where hydraulic conductivity is high even if the terrain is rather flat, such as in Denmark. Higher values may occur for in case of gaining rivers in dry areas like Australia or in mountainous regions where large hydraulic gradients can develop. In mountains with gaining surface water bodies, lateral outflows may even exceed GW recharge of the cell. In grid cells where SW bodies recharge the GW, outflow tends to be a large percentage of total GW recharge as there is no outflow from GW other than in lateral direction, and values often exceed 100% (Fig. 6).

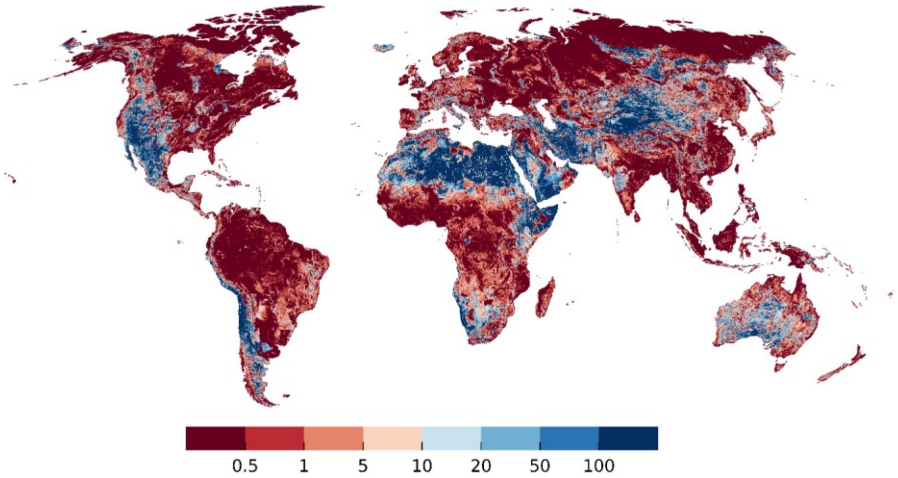

Figure 6 Percentage of GW recharge from soil and surface water inflow that is transferred to neighboring cells through lateral outflow (sum of both layers). Grid cells with zero total GW recharge are shown in white (a few cells in the Sahara and the Andes).

3.5 Comparison to groundwater well observations

Global observations of depth to GW were assembled by Fan et al. (2007; 2013). We selected only observations with known land surface elevation and removed observations where a comparison to local studies suggested a unit conversion error. This left a total of 1,070,402 depth to GW observations. An "observed head" per 5' model cell was then calculated by first computing hydraulic head of each observation by subtracting depth to GW from the 5' average of the 30'' land surface elevation and then calculating the arithmetic mean of all observations within the 5' model cell. Multiple obstacles limit the comparability of observations to simulated values. (1) Observations were recorded at a certain moment in time influenced by seasonal effects and abstraction from GW, whereas the simulated heads represent a natural steady-state condition. (2) Observation locations are biased towards river valleys and productive aquifers. (3) Observations may be located in valleys with shallow local water tables too small to be captured by a coarse resolution of 5'.

Commented [o27]: #2.25

Simulated steady-state hydraulic heads in the upper model layer are compared to observations in Fig. 7. Shallow GW is generally better represented by the model than deeper GW. Especially the water table in mountainous areas is underestimated, which may be related to observations in perched aquifers caused by low permeability layers (Fan et al., 2013) that are not represented in G^3M due to lacking information. Because the steady-state model cannot take into account the impact of GW abstraction, the computed depth to GW values are considerably smaller than currently observed values in GW depletion areas like the Central Valley in California (where once wetlands existed before excessive GW use depleted the aquifer) and the High Plains Aquifer in the Midwest of the USA. Still, the elevation of the GW table in the non-depleted Rhine valley in Germany is overestimated, too. Figure 8a shows the hydraulic head comparison as scatter plot. Overall, the simulation results tend to underestimate observed hydraulic head but much less than the steady-state model presented by de Graaf et al. (2015).

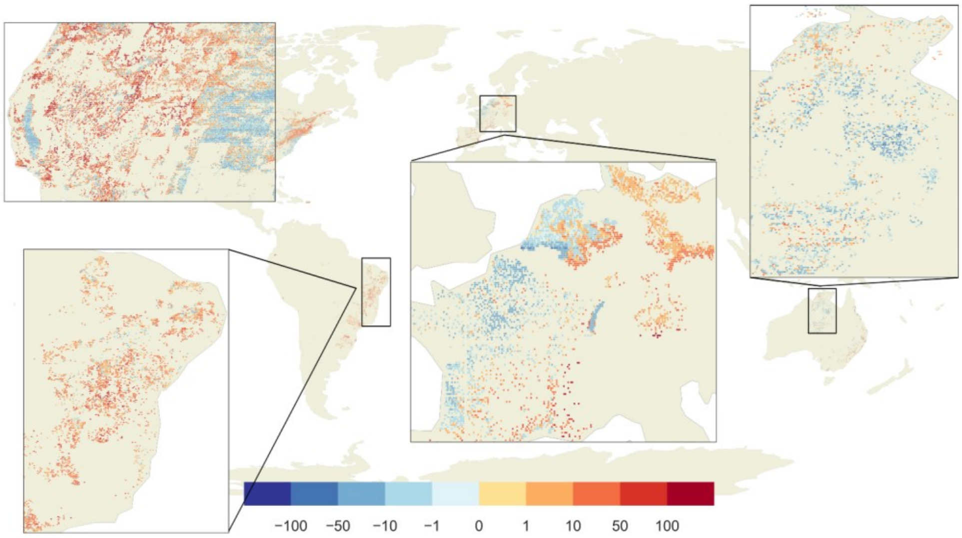

Figure 7 Differences between observed and simulated hydraulic head [***m***]. Red dots show areas where the model simulated deeper GW as observed, blue shallower GW. In grey areas, no observations are available.

To compare performance of G^3M to the steady-state results of two high-resolution model of Fan et al. (2013) and ParFlow (Maxwell et al., 2015), heads in 30" (Fan et al., 2013) and 1 km (ParFlow) grid cells were averaged to the G^3M 5' grid cells. The comparison of 5' observations to the 5' average of ParFlow seem to be consistent with the 1 km model comparison in Maxwell et al. (2015) (their Fig. 5), even though over/under-estimates in the original resolution seemed to be smoothed out by averaging to 5' (not shown). The heads of Fan et al. (2013) fit better to observations than G^3M heads, with less underestimation (Fig. 8b) and an RMSE (Root Mean Square Error) of 26.0 m compared to the 32.4 m RMSE of G^3M. The comparison of G^3M heads to Fan et al. (2013) values for all 5' grid cells, which are also the initial heads of G^3M and the basis to compute river conductances, show that heads computed with the G^3M are mostly much lower except in regions with a shallow GW (Fig. 8c) and an RMSE of 46.7 m. This cannot be attributed to the 100 times lower spatial resolution per se but to the selection of the 30th percentile of the 30'' as the SW drainage level. Outliers in the upper half of the scatter plot, with much larger heads than the initial values, are mainly occurring in steep mountain areas like the Himalayas where the 5' model is not representing smaller valleys with a lower head. For the continental US, the computationally expensive 1-km integrated hydrological model ParFlow (Maxwell et al., 2015), fits much better to observations than G^3M (Figs. 8d, e) 14.3 m (ParFlow) RMSE compared to 34.2 m (G^3M). G^3M produces a generally lower water table (Fig. 8f), the main reason being that ParFlow assumes an impermeable bedrock at a depth of more than 100 m below the land surface elevation.

Commented [o28]: #1.3

The global map of head comparison (Fig. 7) suggests that G^3M performs reasonably well in flat areas compared to mountainous regions. This is corroborated by Fig. 9 that shows the difference between observed and simulated hydraulic heads for five land surface elevation categories. It is evident that model performance deteriorates with increasing land surface elevation.

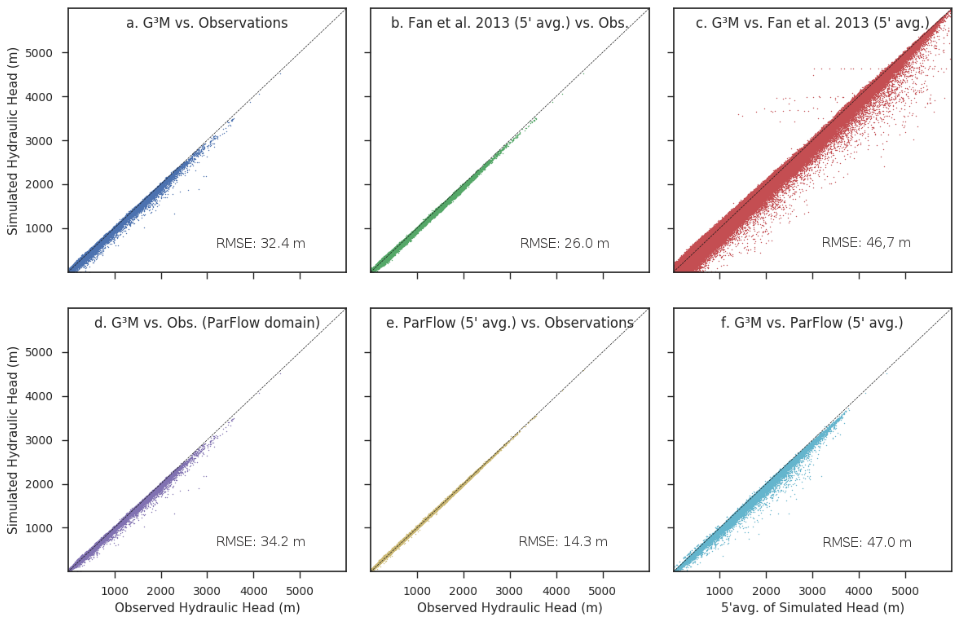

Figure 8 Scatterplots of simulated vs. observed hydraulic head and inter-model comparison of heads. (Upper panel) The steady-state run of G³M vs. observations (a), the 5' average of the equilibrium head of Fan et al. (2013) vs. observations (b) and the avg. equilibrium vs. G³M (c). (Lower panel) The steady-state run of G³M vs. observations only for the ParFlow domain (d), the 5' average of the ParFlow average annual GW table (Maxwell et al., 2015) vs. observations (e) and the steady-state run of G³M vs. 5' average of the ParFlow average annual GW table (f).

Plotting hydraulic head instead of depth to GW has the disadvantage that the goodness of fit is dominated by the topography as the observed heads are calculated based on the surface elevation of the model. Well-observations always measure water table depth and only sometimes contain complementary data specifying the elevation at which the measurements were taken. Even though hydraulic heads are a direct result of the model and are forcing lateral GW flows, depth to GW is more relevant for processes like capillary rise. For G³M, there is almost no correlation between depth to GW observations and simulated values. To our knowledge, no publication on large-scale GW modeling presents correlations of simulated with observed depth to GW.

Commented [o29]: #1.3

Commented [RR30]: #A2.26

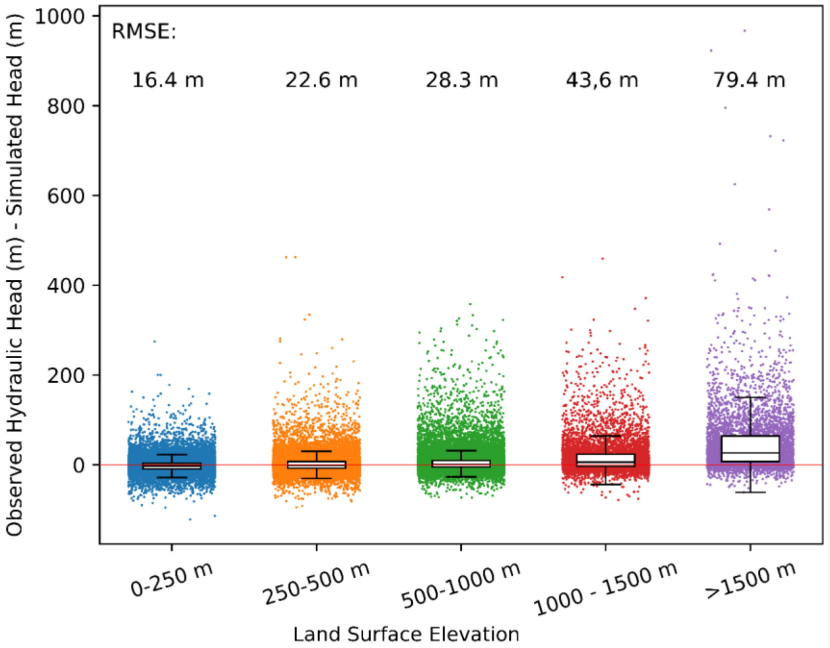

Figure 9 Observed minus simulated hydraulic head for different land surface elevation categories. The whiskers of the boxplots show the interquartile range.

Commented [RR31]: #1.3

3.6 Testing sensitivity of computed steady-state hydraulic heads to parameter values and spatial resolution

Commented [RR32]: #1.4

To limit the computational effort for assessing model sensitivity to both parameters and grid size, we selected New Zealand as a representative "small world" that includes a complex topography and the ocean as a clear boundary condition. All inputs and parameters are the same as in the global 5' model.

3.6.1 Parameter sensitivity

Commented [o33]: #1.5

To determine to which parameters simulated hydraulic heads are most sensitive to, we used the established sensitivity tool UCODE 2005 (Hill and Tiedeman, 2007) to compute composite scaled sensitivity (CSS) values for seven model parameters (S.3). CSS of h_{swb} is orders of magnitude larger than the CSS of the other parameters. This confirms our observations during model development when an appropriate value for h_{swb} had to be found (section 2.2). The second-most important parameter is K_{aq}, with the sensitivity to R_g being only half a large. CSS of c_{Lakes} is one magnitude less than CSS of R_g but as only a few cells contain lakes, the CSS value that averages over all grid cells indicates a large sensitivity to c_{Lakes} for grid cells with lakes. Simulated hydraulic heads were found to be rather insensitive to changes in the other parameters $c_{Wetlands}$, c_{ocean}, and c_{riv}.

3.6.2 Sensitivity to spatial resolution

The extremely high sensitivity of simulated hydraulic heads to the choice of the elevation of the surface water bodies within a grid cell (section 3.6.1) and the better agreement of the continental models with a spatial resolution of (approx.) 30", the Fan et al. (2013) model and ParFlow (section 3.5) indicated motivated us to run G^3M for New Zealand with a spatial resolution of 30", to understand the impact of spatial resolution on simulated hydraulic heads. The 30" model uses the same input as the 5' model except for the land surface elevation, h_{swb} and the location of rivers (thus the 30" model contains more cells without a river). While the total length and width of the rivers is equal in both models, a river is assumed to exist in all 5' grid cells, the river is concentrated, in the 30" model, to a few grid cells with each 5' grid cell. The river cell locations at 30" are determined based on 30" HydroSHEDS (hydrosheds.org) information on flow accumulation. Starting with the 30" cell with the highest number of upstream cells per 5' cell, a river is added to this 30" cell using the width and length information of HydroSHEDS. This is repeated for all 30" per 5' cell until the length and width of the added rivers is equal to the one used in the 5' model. The areal fractions of all other SW bodies from 5' grid data where used for all 30" grid cells within the 5' grid cell. h_{swb} is set to the land surface elevation.

Figure 10 compares the performance of the two model versions. The comparison of simulated hydraulic head to observations (Westerhoff et al., 2018) shows that the overall performance of the 30" model is better, with a smaller RMSE of 26.7 m as compared to an RMSE of 53.8 m in case of the original spatial resolution of 5'. The 30" model results in generally lower simulated hydraulic heads leading to a closer fit to the observed values. This is likely caused by the improved estimation of SW body elevation, which generally leads to lower estimates of h_{swb}. On the other hand, underestimation of observed hydraulic heads prevails in the 30" model, even though of h_{swb} was set to the land surface elevation, indicating that further investigation is necessary.

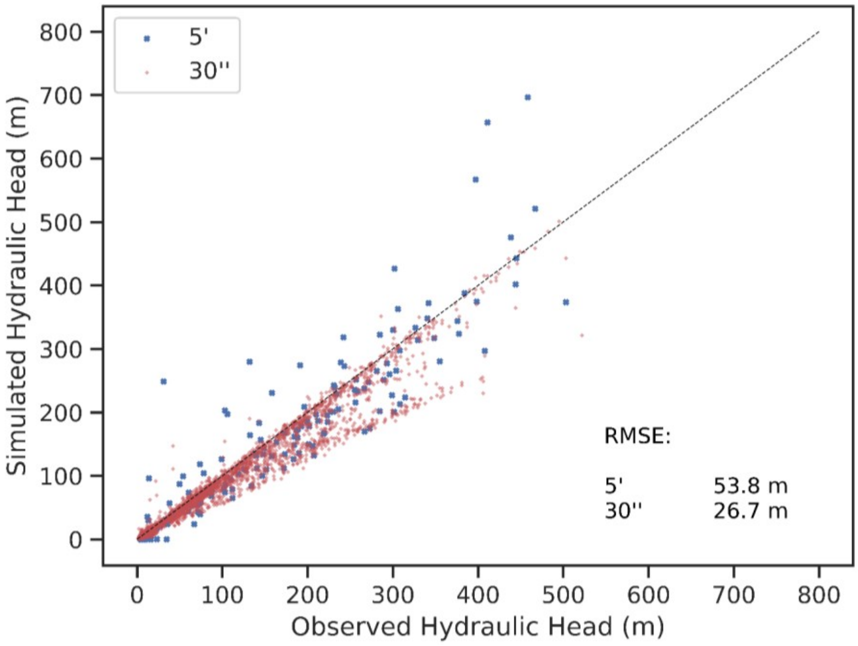

Figure 10 Low (5') vs. high (30'') spatial resolution for New Zealand, comparison of observed vs. simulated hydraulic head for both resolutions. The observed head is the geometric mean per 5' and 30" respectively.

4 Discussion

The objective of global gradient-based groundwater flow modeling with G³M is to better simulate water exchange between SW and GW in the GHM WaterGAP, for example for improved estimation of GW resources in dry regions of the globe that are augmented by focused recharge from SW bodies. The presented steady-state model is the first step in this direction: (1) It aligns with established GW model development practices that helped (2) to understand basic model behavior e.g. the sensitivity to SW body elevation, and the necessary improvement of its parameterization, before moving to the more complex integrated transient model. The reduced runtime of the steady-state model in comparison to a fully integrated transient run (3) supported the investigation of parameter sensitivity and sensitivity to spatial resolution. Additionally, (4) the presented steady-state model can be used in future fully integrated transient runs as initial condition.

A major challenge for simulating GW-SW interactions (but also capillary rise) at the global scale is the large size of grid cells that is required due to computational constraints. Within the 5' grid cells, land surface elevation at the scale of 30" very often varies by more than 20 m, and often by 200 m and more (Fig. S4.1), while the vertical position of the cell and the hydraulic head are approximated in the model by just one value. The question is whether head-dependent flows between grid cells, between GW and SW and from GW to soil (capillary rise) can be simulated successfully at the global scale, i.e. whether an improved quantification of these flows as compared to the simple linear reservoir model currently used in most GHMs can be achieved by this approach. This question cannot be answered yet as we have not yet achieved a dynamic coupling of G³M with a global hydrological model but one may speculate that some innovative approach to take into account the elevation variations within the grid cells may be needed.

Commented [RR34]: #2.1,2.2

It is difficult to assess the performance of the presented steady-state G³M results. Model performance is assessment is hindered by data availability and the coarse model resolution. (1) To our knowledge the data collection of depth to groundwater by Fan et al. (2013) is unique. However, they do not represent steady-state values. Apart from depth to groundwater observations, hardly any relevant data is available at the global scale. Especially exchange between surface water and groundwater is difficult to measure even at the local scale. Therefore, we compared G³M results with the results from other large-scale models. Comparison to the results of catchment-scale groundwater flow models is planned for transient runs that will be possible after integration into WaterGAP. (2) Scale differences make the comparison to point observations of depth to groundwater difficult. Multiple local observations within a 5' cell may strongly vary, possibly just due to land surface elevation variations within the approximately 80 km^2 large cells (compare Fig. S4.1 and S4.8). Often, observations are biased towards alluvial aquifers in valleys. The calculated hydraulic head of the grid cell may represent the average groundwater level per grid cell correctly but can be still far off the local observations of depth to groundwater. As the current model only presents an uncalibrated natural steady-state, a comparison to observations only provides a first indicator where the model and the performance measurements need to be improved as we move to a fully transient model.

The presented comparison to other large-scale models is based on the assumption that same model deficiencies e.g. in available data and scale issues can uncover differences in model decisions e.g. used equations and spatial resolution. A comparison to catchment scale models is challenging as scales can differ by multiple magnitudes. As the model is further developed towards a transient model, comparisons to simulations in data-rich regions need to be extended and temporal changes in interactions with surface water investigated.

Commented [RR35]: #A2.27

The presented development of the uncoupled steady-state global GW flow model enabled us to better understand how the spatial hydraulic head pattern relates to the fundamental drivers topography, climate and geology (Fan et al., 2007) and how the interaction to SW bodies governs the global head distribution. Simulated depth to groundwater is particularly affected by the assumed hydraulic head in SW bodies, the major GW drainage component in the model. As rivers represent natural occurring drainage at the lowest point in a given topography, one would assume that the minimum elevation 30" land surface elevation per 5' grid cell is a reasonable choice. Experiments have shown that this will induce a head distribution well below the average 5' elevation that is much below observations of Fan et al. (2013). We also tested setting h_{swb} to the average elevation of all "blue" cells (with a depth to GW of less than 0.25 m) of the steady-state 30" water table results of Fan et al. (2013) that indicate the locations were GW discharges to the surface. This leads to an overall underestimation of the observed hydraulic heads (Fig. S4.9). Furthermore, it leads to an increase in losing SW bodies (comp. Fig. S4.11 with Fig. 3). However, it is difficult to judge whether this improves the simulation. More stretches of the Nile and its adjacent wetlands and also of the Niger wetlands and rivers in North-eastern Brazil are losing in case of lower h_{swb}, which appears to reasonable. Additionally, choosing the average as SW elevation provides on the on hand a better fit to observations (Fig. S4.9) but leads to a wold wide flooding with largely overestimated heads (Fig. S4.10) and a much longer convergence time due to an increased oscillation between gaining and losing conditions.

The problem is very likely one of scale. This is supported by the fact that both high-resolution models, even the simple one of Fan et al. (2013) fit better to observations than the low-resolution model G^{3}M (Fig. 8). In case of high resolution (e.g. 30''), there are a number of grid cells at an elevation above the average 5' land surface elevation, leading to higher hydraulic heads in parts of the 5' area that drain towards the SW body in a lower 30'' grid cell. In case of the low spatial resolution of 5' in which h_{swb} is set to the elevation of the fine-resolution drainage cell, the 5' hydraulic head is rather close to this (low) elevation (Fig. S4.12), resulting in an underestimation of hydraulic head and thus an overestimation of depth to GW. While it is plausible and necessary to assume that there is SW-GW interaction within each of in the approximately 80 km^2, this is not the case for the two orders of magnitude smaller 30'' grid cells. Thus, with the high resolution, heads are not strongly controlled everywhere by the head in SW bodies. A sensitivity analysis of the parameters and the spatial resolution confirms this hypothesis. Selecting the 30th percentile of the 30'' land surface elevation as h_{swb} was found, by trial-and-error, to lead to a hydraulic head distribution that fits reasonably well to observed head. It avoids that the simulated GW table drops to low while avoiding the excessive flooding that occurs if h_{swb} is set to the average of 30'' land surface elevations, i.e. the 5' land surface elevation (Fig. S4.9).

The constraint that the selected h_{swb} value puts on simulated hydraulic heads is also linked to the conductance of the SW bodies. A higher conductance will lead to aquifer heads closer to h_{swb}. If the hydraulic head drops below the bottom level of the SW body, the hydraulic gradient is assumed to become 1 and the SW body recharges the GW with a rate of K_{aq} per unit SW body area. In case of a K_{aq} value of 10^{-5} m s^{-1}, the SW body would lose approximately 1 m of water each day. It is to be investigated how the sensitivity to choice of SW body elevation and conductance leads to a solution that fits observations best. A lower conductance may lead to a higher groundwater table as SW bodies don't drain as much water; on the other hand, they seem to provide an important recharge mechanism in the steady-state model for some regions preventing an even higher depth to GW. The simple conductance approach applied in G^{3}M could possibly be improved by the approach proposed by Morel-Seytoux et al. (2017).

de Graaf et al. (2015) set their SW head (h_{swb}) to the land surface elevation of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to avoid excessive flooding, and that is not needed in G^{3}M. On the other hand, this adaption allows the drainage of water even if the hydraulic head is below the SW elevation that might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity of SW body elevation. A summary of model differences are shown in Table 2.

As described above, G³M differs from regional groundwater models due to grid cell size which requires G³M to be more conceptual and does not allow for capturing actual variability of topography, aquifer depth (Richey et al., 2015) and (vertical) heterogeneity of subsurface properties. The lack of information about the three-dimensional distribution of hydraulic conductivity is expected to negatively impact the quality of simulated GW flow. For example, the lateral conductivity and connectivity of groundwater along thousands of kms from e.g. the Rocky Mountains in the Central USA to the coast as well as the vertical connectivity is likely to be overestimated by G^{3}M, as vertical faults and interspersed aquitards are not represented; this leads to an underestimation of hydraulic head in those mountainous areas.

5 Conclusions

We have presented the concept and first results of a new global gradient-based 5' GW flow model that is to be integrated into the 0.5° GHM WaterGAP. The uncoupled steady-state model has provided important insights into challenges of global GW flow modeling mainly related to the necessarily large grid cells size (5' by 5') as well as first global maps of SW-GW interactions. Simulated heads were found to be strongly impacted by assumptions regarding the interaction with SW bodies, in particular, the selected elevation of the SW table. We have demonstrated that simulated G^3M hydraulic heads fit better to observed heads than the heads of the comparable steady-state GW model of de Graaf et al. (2015), without requiring additional drainage that would prevent a full coupling to a GHM. Furthermore, we provided insights into how the choice of surface water body elevation affects the model outcome and plan to further investigate how we can use higher resolution topographic data to overcome these challenges by comparing simulation results of a 5', 30'', and 3'' GW model of New Zealand.

The presented results are the first step towards a fully coupled model in which SW heads are computed as a function of surface water hydrology and GW abstractions can be taken into account. Especially the interaction with SW bodies that can run dry will make the model behavior more realistic. The fully coupled model will simulate transient behavior reflecting climate variability and change. Simulated hydraulic head dynamics will be compared to observed head time series as well as to the output of large-scale regional models, while total water storage variations will be compared to GRACE satellite data. However, it will be challenging to judge the quality of simulated GW-SW interactions due to a scarcity of observations.

6 Code and data availability

The model-framework code is available at globalgroundwatermodel.org or at DOI: 10.5281/zenodo.1175540 with a description on how to compile and run a basic GW model. The code is available under the GNU General Public License 3. Model output is available at DOI: 10.5281/zenodo.1315471.

Acknowledgments

Part of this study was funded by a Friedrich-Ebert foundation Ph.D. fellowship. We are very grateful to Ying Fan and Gonzalo Miguez-Macho for fruitful discussions and data provisioning. We thank the reviewers for their thoughtful comments that helped to improve the manuscript. Furthermore, we like to thank Alexander Wachholz for contributing to the New Zealand figures and Rogier Westerhof for providing the observations.

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrological Sciences Journal, 48, 317–337, doi:10.1623/hysj.48.3.317.45290, 2003.

Allen, P. M., Arnold, J. C., and Byars, B. W.: Downstream channel geometry for use in planning-level models, J Am Water Resources Assoc, 30, 663–671, doi:10.1111/j.1752-1688.1994.tb03321.x, 1994.

Belcher, W. R. and Sweetkind, D. S.: Death Valley regional groundwater flow system, Nevada and California-Hydrogeologic framework and transient groundwater flow model, U.S. Geological Survey Professional Paper, 398, 2010.

Belcher, W.R. (E.): Death Valley Regional Ground-water Flow System, Nevada and California-Hydrogeologic Framework and Transient Ground-water Flow Model, U.S. Geological Survey Professional Paper, pp. 408, 2004.

Costa, A. C., Foerster, S., de Araújo, J. C., and Bronstert, A.: Analysis of channel transmission losses in a dryland river reach in north-eastern Brazil using streamflow series, groundwater level series and multi-temporal satellite data, Hydrol. Process., 27, 1046–1060, doi:10.1002/hyp.9243, 2013.

de Graaf, I. E.M. de, Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, Hydrol. Earth Syst. Sci., 19, 823–837, doi:10.5194/hess-19-823-2015, 2015.

de Graaf, I. E.M. de, van Beek, R. L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.P.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, Advances in Water Resources, 102, 53–67, doi:10.1016/j.advwatres.2017.01.011, 2017.

Dogrul, E., Brush, C., and Kadir, T.: Groundwater Modeling in Support of Water Resources Management and Planning under Complex Climate, Regulatory, and Economic Stresses, Water, 8, 592, doi:10.3390/w8120592, 2016.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12, 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B.R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, Journal of Geodynamics, 59-60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, Journal of Hydrology, 270, 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, doi:10.1002/2014WR015595, 2014.

Eisner, S.: Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients, University of Kassel, Kassel, Germany, 2016.

Elsawwaf, M., Feyen, J., Batelaan, O., and Bakr, M.: Groundwater-surface water interaction in Lake Nasser, Southern Egypt, Hydrol. Process., 28, 414–430, doi:10.1002/hyp.9563, 2014.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science (New York, N.Y.), 339, 940–943, doi:10.1126/science.1229881, 2013.

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, J. Geophys. Res., 112, doi:10.1029/2006JD008111, 2007.

FAO: Irrigation potential in Africa: http://www.fao.org/docrep/w4347e/w4347e0i.htm, 1997, last access: 1 May 2018.

Faunt, C. C. (Ed.): Groundwater Availability of the Central Valley Aquifer, California, U.S. Geological Survey Professional Paper, 1776, 225 pp., 2009.

Faunt, C. C., Provost, A. M., Hill, M. C., and Belcher, W. R.: Comment on "An unconfined groundwater model of the Death Valley Regional Flow System and a comparison to its confined predecessor" by R.W.H. Carroll, G.M. Pohll and R.L. Hershey [Journal of Hydrology 373/3–4, pp. 316–328], Journal of Hydrology, 397, 306–309, doi:10.1016/j.jhydrol.2010.11.038, 2011.

Faunt, C. C., Sneed, M., Traum, J., and Brandt, J. T.: Water availability and land subsidence in the Central Valley, California, USA, Hydrogeol J, 24, 675–684, doi:10.1007/s10040-015-1339-x, 2016.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nature Geosci, 9, 161–167, doi:10.1038/ngeo2590, 2016.

Gleeson, T., Moosdorf, N., Hartmann, J., and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898, doi:10.1002/2014GL059856, 2014.

Gleeson, T., Wada, Y., Bierkens, M. F. P., and van Beek, L. P. H.: Water balance of global aquifers revealed by groundwater footprint, Nature, 488, 197–200, doi:10.1038/nature11295, 2012.

Gossel, W., Ebraheem, A. M., and Wycisk, P.: A very large scale GIS-based groundwater flow model for the Nubian sandstone aquifer in Eastern Sahara (Egypt, northern Sudan and eastern Libya), Hydrogeology Journal, 12, 698–713, doi:10.1007/s10040-004-0379-4, 2004.

Harbaugh, A. W.: MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process, US Department of the Interior, US Geological Survey Reston, 2005.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophysics, Geosystems, 13, doi:10.1029/2012GC004370, 2012.

Hill, M. C. and Tiedeman, C. R.: Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty, Wiley-Interscience, Hoboken, N.J., 2007.

Konikow, L. F.: Contribution of global groundwater depletion since 1900 to sea-level rise, Geophys. Res. Lett, 38, doi:10.1029/2011GL048604, 2011.

Krakauer, N. Y., Li, H., and Fan, Y.: Groundwater flow across spatial scales: importance for climate modeling, Environ. Res. Lett., 9, 34003, doi:10.1088/1748-9326/9/3/034003, 2014.

Lange, S.: EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), 2016.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, 296, 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Maxwell, R. M., Condon, L. E., and Kollet, S. J.: 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, 923–937, doi:10.5194/gmd-8-923-2015, 2015.

Mehl, S.: Use of Picard and Newton iteration for solving nonlinear ground water flow equations, Ground Water, 44, 583–594, doi:10.1111/j.1745-6584.2006.00207.x, 2006.

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, J. Geophys. Res., 112, D13108, doi:10.1029/2006JD008112, 2007.

Morel-Seytoux, H. J., Miller, C. D., Miracapillo, C., and Mehl, S.: River Seepage Conductance in Large-Scale Regional Studies, Ground Water, 55, 399–407, doi:10.1111/gwat.12491, 2017.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Naff, Richard L., and Edward R. Banta: The US Geological Survey modular ground-water model-PCGN: a preconditioned conjugate gradient solver with improved nonlinear control, Geological Survey (US), 1331, 2008.

Niswonger, R. G., Panday, S., and Ibaraki, M.: MODFLOW-NWT, a Newton formulation for MODFLOW-2005, US Geological Survey Techniques and Methods, 6, 2011.

Reinecke, R.: G³M-f a global gradient-based groundwater modelling framwork, JOSS, 3, 548, doi:10.21105/joss.00548, 2018.

Richey, A. S., Thomas, B. F., Lo, M.-H., Famiglietti, J. S., Swenson, S., and Rodell, M.: Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework, Water Resour. Res, 51, 5198–5216, doi:10.1002/2015WR017351, 2015.

Saad, Y.: ILUT: A dual threshold incomplete LU factorization, Numer. Linear Algebra Appl., 1, 387–402, doi:10.1002/nla.1680010405, 1994.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, PNAS, 109, 9320–9325, doi:10.1073/pnas.1200311109, 2012.

Sheets, R. A., Hill, M. C., Haitjema, H. M., Provost, A. M., and Masterson, J. P.: Simulation of water-table aquifers using specified saturated thickness, Ground Water, 53, 151–157, doi:10.1111/gwat.12164, 2015.

Stonestrom, D. A., Constantz, J., Ferre, T. P.A., and Leake, S. A.: Ground-water recharge in the arid and semiarid southwestern United States, U.S. Geological Survey Professional Paper 1703, 414 p, 2007.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin, Hydrol. Earth Syst. Sci., 15, 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I., and Treidel, H.: Ground water and climate change, Nature Climate change, 3, 322–329, doi:10.1038/nclimate1744, 2012.

UNESCO-IHP, IGRAC, WWAP: GEF-TWAP Methodology Transboundary Aquifers.: http://isarm.org/sites/default/files/resources/files/TWAP%20Methodology%20Groundwater%20Component%20%28Revised%20Aug%202012%29.pdf, 2012, last access: 9 October 2017.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, doi:10.1029/2010WR009791, 2011.

Vergnes, J.-P., Decharme, B., Alkama, R., Martin, E., Habets, F., and Douville, H.: A Simple Groundwater Scheme for Hydrological and Climate Applications: Description and Offline Evaluation over France, J. Hydrometeor, 13, 1149–1171, doi:10.1175/JHM-D-11-0149.1, 2012.

Vergnes, J.-P., Decharme, B., and Habets, F.: Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model, J. Geophys. Res. Atmos., 119, 11, doi:10.1002/2014JD021573, 2014.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F., and Alcamo, J.: Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe, Journal of Hydrology, 424-425, 238–251, doi:10.1016/j.jhydrol.2012.01.005, 2012.

Wada, Y.: Modeling Groundwater Depletion at Regional and Global Scales: Present State and Future Prospects, Surv Geophys, 37, 419–451, doi:10.1007/s10712-015-9347-x, 2016.

Wada, Y., van Beek, L. P. H., and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: A global assessment, Water Resour. Res., 48, doi:10.1029/2011WR010562, 2012.

Westerhoff, R., White, P., and Miguez-Macho, G.: Application of an improved global-scale groundwater model for water table estimation across New Zealand, Hydrol. Earth Syst. Sci, 22, 6449–6472, doi:10.5194/hess-22-6449-2018, 2018.

Table 1 Model parameter values, input data sources and other information about the steady-state simulation.

Parameter	Symbol	Units	Description	Eq. No.
Landmask	-	-	Location and area of 2161074 cells at 5' resolution based on WaterGAP (Eisner, 2016))	-
GW recharge	R_g	L^3T^{-1}	Mean annual diffuse GW recharge 1901–2013 of WaterGAP 2.2c (Müller Schmied et al., 2014) forced with EWEMBI (Lange, 2016), spatial resolution 0.5° (Fig. S4.4)	2,6,8
Hydraulic conductivity	K_{aq}	LT^{-1}	Derived from Gleeson et al., 2014 (Fig. S4.3)	1,4
Hydraulic head	$h_{(aq)}$	L	Head of the aquifer in a computational cell, initial estimate based on 5' average of 30" head of Fan et al. (2013)	1,3,6
Ocean boundary conductivity	c_{ocean}	L^2T^{-1}	$10\ m^2\ day^{-1} = 0.1\ m\ day^{-1}\ 10\ km\ 10\ km^{-1}\ 100\ m$, with K of $10^{-6}\ m\ s^{-1}$ and a distance of 10 km from the cell center to the boundary with a cell thickness of 100 m	2,3
Ocean boundary head	h_{ocean}	L	Global mean sea-level of 0 m	3
SW head	h_{swb}	L	30 % quantile (P_{30}) of 30" land surface elevation of Fan et al. (2013) per 5' grid cell	4
SW bottom elevation	B_{swb}	L	2 m (wetlands), 10 m (local lakes), 100 m (global lakes) below P_{30}	4
Area of global and local lakes and global and local wetlands	WL	L^2	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016),	5
Length of the river	L	L	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016)	5
Width of the river	W	L	Per 5' grid cell, based d on WaterGAP 3 (Eisner, 2016)	5
River head	h_{riv}	L	h_{swb}	5,6
River bottom elevation	B_{riv}	L	$h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994)	6
Equilibrium hydraulic head	h_{eq}	L	Steady-state hydraulic head of Fan et al. (2013) (averaged to 5' from original spatial resolution of 30")	6
Layers	-	-	2 confined, 100 m thick each	-
Land surface elevation	-	L	5' average of 30" digital elevation map of Fan et al. (2013) (Fig. S4.2)	-
E-folding factor	-	-	Applied only to lower layer for 150 m depth, based on area-weighted average of Fan et al. (2013)	-
Timestep	t	T	Daily timestep	-
Convergence criterion	-	L	$\|\|$hydraulic head residuals$\|\|_{inf} < 10^{-100}$ and max head change < 10 m	-
Inner iterations	-	-	50 inner iterations between Picard iterations (Naff, Richard L., and Edward R. Banta, 2008)	-

Table 2 Comparison of global- and continental-scale groundwater models

Commented [RR36]: #A2.28

Aspect	G^{3}M	de Graaf et al. (2015; 2017)	Fan et al. (2013)	ParFlow
Extent	Global	Global	Global	Continental USA
Resolution	5'	6'	30"	1 km
Software	G^{3}M-f	MODFLOW	Unnamed	ParFlow
Computational expense	Medium	Medium	High	Very high
Flow representation	3D saturated	3D saturated	2D saturated	3D saturated/unsaturated
Time scale	Steady-state/(transient)	Steady-state/transient	Steady-state	Steady-state
Vertical layers	2	2	1	5
Full coupling possible	Yes	No (Conceptual issue)	No	Yes (already coupled)
In-memory coupling	Yes	No	N/A	Yes
Constant saturated thickness	Yes	Yes	No	No
Impermeable bottom	No	No	No	Yes
Surface water body location	In every cell	In almost every cell	No surface water	Created during simulation
Surface water body elevation	P_{30} of 30" DEM	Avg. of 30" DEM	N/A (outflow if depth to GW < 0.25 m)	N/A
Deviation from observations	Large	Very large	Medium	Medium

Supplement

1 Coupling to WGHM

With a spatial resolution of 0.5° by 0.5° (approximately 55 km by 55 km at the equator), the WaterGAP 2 model (Alcamo et al., 2003) computes human water use in five sectors and the resulting net abstractions from GW and SW for all land areas of the globe excluding Antarctica. These net abstractions are then taken from the respective water storages in the WaterGAP Global Hydrology Model (WGHM) (Müller Schmied et al., 2014; Döll et al., 2003; 2012; 2014). With daily time steps, WGHM simulates flows among the water storage compartments canopy, snow, soil, GW, lakes, man-made reservoirs, wetlands, and rivers. As in other GHMs, the dynamic of GW storage (GWS) is represented in WGHM by a linear GW reservoir model, i.e.

$$\frac{dGWS}{dt} = R_g + R_{g\_swb} - NA_g - k_g\, GWS \tag{S1}$$

where R_g $[L^3T^{-1}]$ is diffuse GW recharge from soil, R_{g_swb} $[L^3T^{-1}]$ GW recharge from lakes, reservoirs and wetlands (only in arid and semiarid regions, with a global constant value per SW body area), NA_g $[L^3T^{-1}]$ net GW abstraction. The product $k_g\, GWS$ quantifies GW discharge to SW bodies as a function of GWS and the GW discharge coefficient k_g (Döll et al., 2014). G³M is to replace this linear reservoir model in WGHM. Capillary rise is not included in the presented steady-state simulation, as simulation of capillary rise requires information of soil moisture that is only available when G³M is fully integrated into WGHM.

G³M will be integrated into WGHM by exchanging information on (1) R_{g_swb} and NA_g, (2) soil water content, (3) Q_{cr}, (4) h_{swb}, and (5) Q_{swb}. Figure S1.1 indicates the direction of the information flows. Water flows from the 0.5° cells of WGHM are distributed equally to all 5' G³M grid cells inside a 0.5° cell. Flows transferred from the 5' cells of G³M to WGHM are aggregated. GW recharge and net abstraction from GW together with SW tables are the main drivers of the GW model that will be provided dynamically by WGHM. GW-SW flow volumes computed by G³M will be aggregated and added or subtracted from the SW body volumes in WGHM, and SW body heads will be recalculated. WGHM soil water content together with G³M depth to GW will be used to calculate capillary rise and thus a change of soil water content. WaterGAP includes a one layer soil water storage compartment characterized by land cover specific rooting depth, maximum storage capacity and soil texture (Döll et al., 2014). The water content in the soil storage is increased by incoming precipitation and decreased by evapotranspiration and runoff generation (Döll et al., 2014). Capillary rise is not yet implemented in G³M, and SW heads are currently based on land surface elevation.

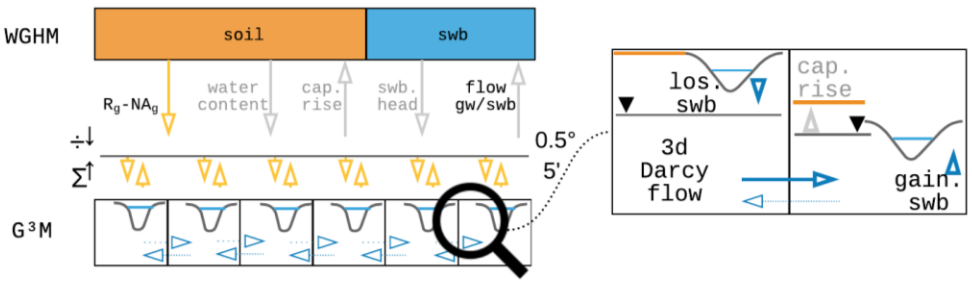

Figure S1.1 Conceptual view of the coupling between WGHM and G³M. WGHM provides calculated GW recharge (R_g) (Döll and Fiedler, 2008) and if the human impact is considered, net abstraction from GW (NA_g) (Döll et al., 2012). G³M spreads this input equally to all 5' grid cells inside a 0.5° cell and calculates hydraulic head and interactions with SW bodies (swb) as well as capillary rise (cap. rise) at the 5' resolution. Grey arrows show information flow that is not yet implemented.

Commented [RR37]: #2.3

Commented [o38]: #A1.10

2 Case study Central Valley

To evaluate G^3M further, its results were analyzed for to a well-studied region, the Central Valley in California, USA. The Central Valley is one of the most productive agricultural regions of the world and heavily relies on GW pumpage to meet irrigation demands (Faunt et al., 2016). GW pumping in the valley increased rapidly in the 1960s (Faunt, 2009). Figure S2.1 shows simulated depth to GW for the Central Valley, the coast and the neighboring Sierra Nevada mountainside as well as parts of the Great Basin. The depth to GW table represents natural conditions without any pumping and is rather small. It roughly resembles the depth to GW assumed in the Central Valley Hydrological Model (CVHM) as initial condition, representing a natural state (Faunt, 2009) (Fig. S2.1b). G^3M correctly computes the shallow conditions with groundwater above the surface in the north, partially in the south of the valley and decreasing towards the Sierra Nevada. The difference in the extent of the flooded area could be due to large wetlands areas still present in the early 60s which are not represented in this extent in the data used by G^3M. Beyond the CVHM domain, depth to GW in mountainous regions is probably overestimated by G^3M. The elevation of neighboring cells may differ up to a 1000 meter resulting in a large gradient (Fig. S4.6b and S4.6e).

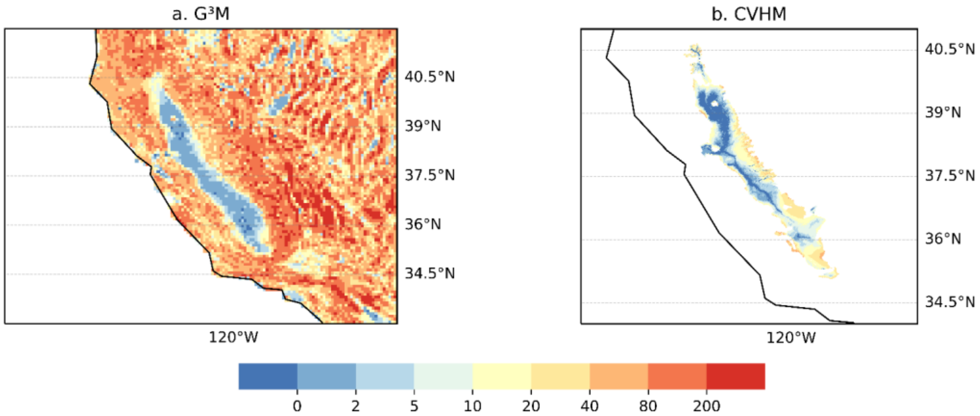

Figure S2.1 Plots of depth to GW [m] as calculated by G^3M for the Central Valley and the Great Basin (a), and as used by CVHM as the natural state and starting condition (Faunt, 2009) (b).

Commented [RR39]: #A1.17

3 Sensitivity Analysis

Sensitivities are calculated using forward differences

$$\frac{\Delta y'_i}{\Delta b_j} = \frac{y'_i(b_j + \Delta b_j) - y'_i(b_j)}{\Delta b_j} \tag{S2}$$

where y'_i is the simulated hydraulic head at position i from ND number of observations and b_j the perturbed parameter (Table S1) in a vector of all parameters of length j. Based on these values the composite scaled sensitivity is computed as

$$CSS_j = \sqrt{\sum_{i=1}^{ND} \left(\frac{\Delta y'_i}{\Delta b_j} \frac{b_j}{\sigma_{y_i}} \right) ND^{-1}} \tag{S3}$$

where σ_{y_i} the standard deviation of an observation at this position. Because the observations are only available in a very small part of New Zealand "artificial observations" for each cell are assumed with 1 m hydraulic head each. Thus, σ_{y_i} is 1 m. The result of CSS is dimensionless and can be compared directly between parameter multipliers (Table S1).

Table S1 Ranges of parameter multipliers used in the local sensitivity analysis and their resulting composite scaled sensitivity values. The multiplier for the wetlands applies to global and local wetlands.

Parameter	Multiplier Range	Composite Scaled Sensitivity
h_{swb}	1.01	39132.1
K_{aq}	1.01	76.8
R_g	1.1	39.8
c_{Lakes}	1.1	3.2
$c_{Wetlands}$	1.1	0.014
c_{riv}	1.1	0.013
c_{ocean}	1.1	0.013

4 Additional Figures

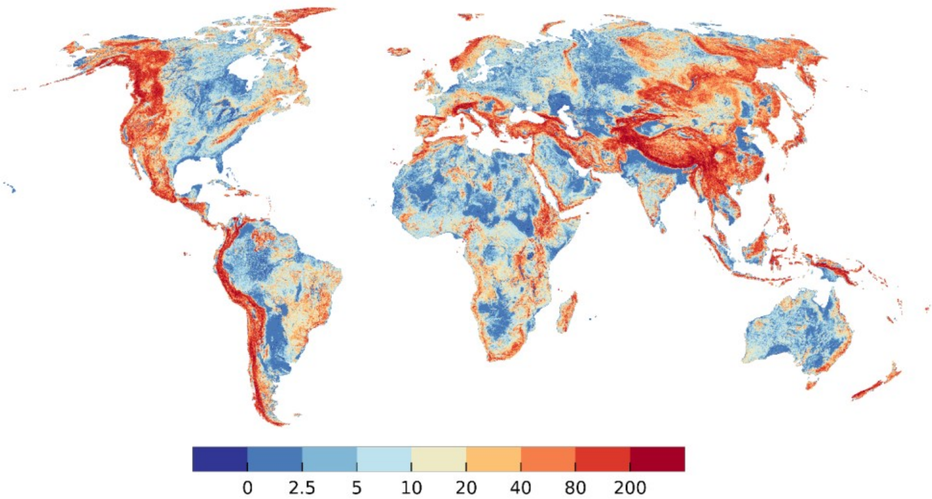

Figure S4.1 Difference [***m***] between mean elevation and P_{30} elevation. Maximum value 1365 m**.**

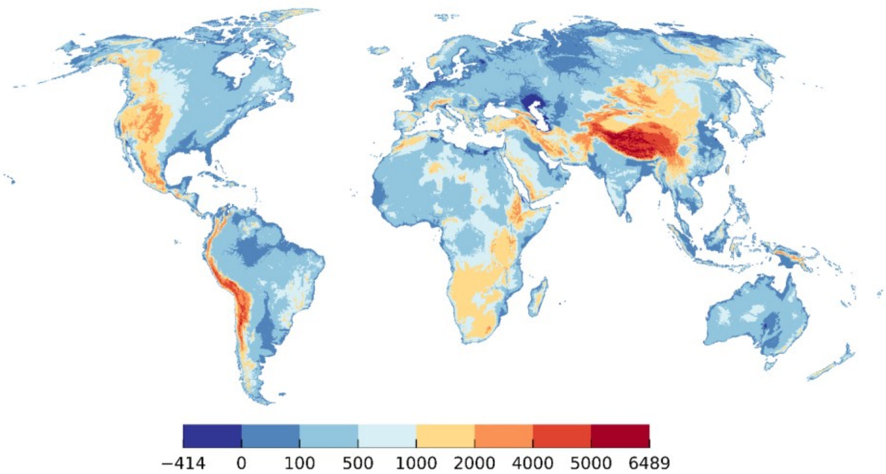

Figure S4.2 Land-surface elevation [***m***] used in G^3M: 5' average of 30'' land surface elevation used in Fan et al. (2013).

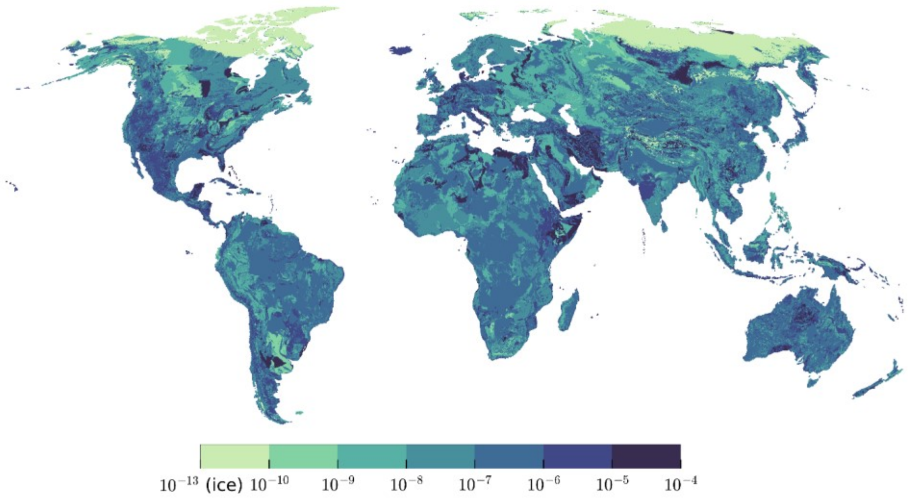

Figure S4.3 Hydraulic conductivity [$\boldsymbol{ms^{-1}}$] derived from Gleeson et al. (2014) by scaling it with the geometric mean to 5'. Very low values in the northern hemisphere are due to permafrost conditions.

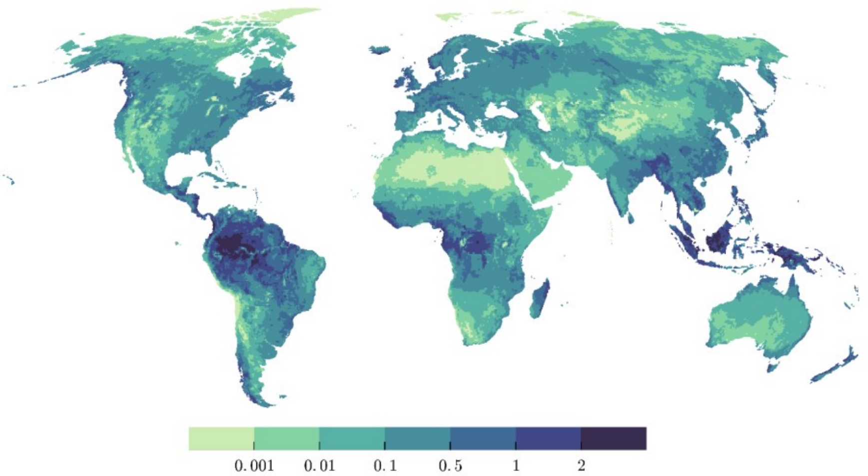

Figure S4.4 Mean annual groundwater recharge [$\boldsymbol{mm\ day^{-1}}$] between 1901-2013, from WaterGAP 2.2c.

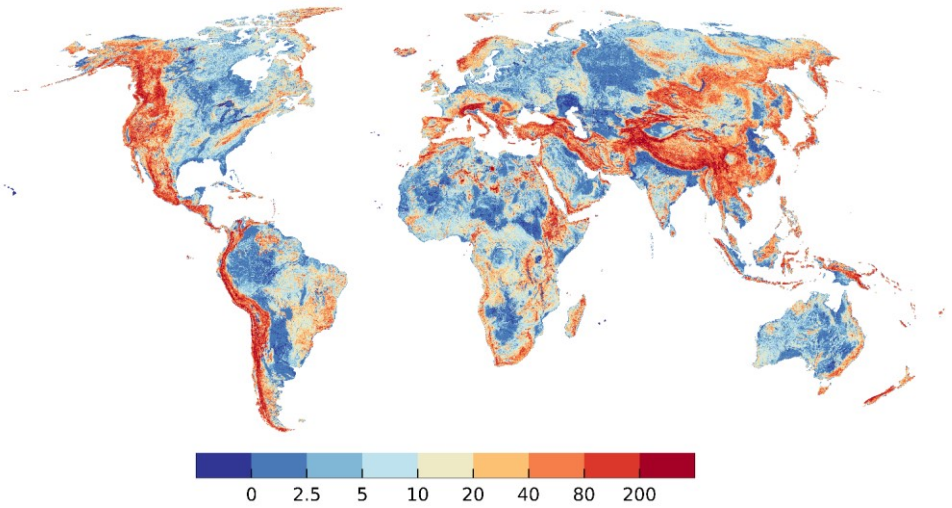

Figure S4.5 Arithmetic mean [***m***] of the 30'' land surface elevation per 5' grid cell minus simulated equilibrium hydraulic head (simulated depth to GW). Maximum value 2070 m, minimum value -414 m (Extremes included in dark blue and dark red).

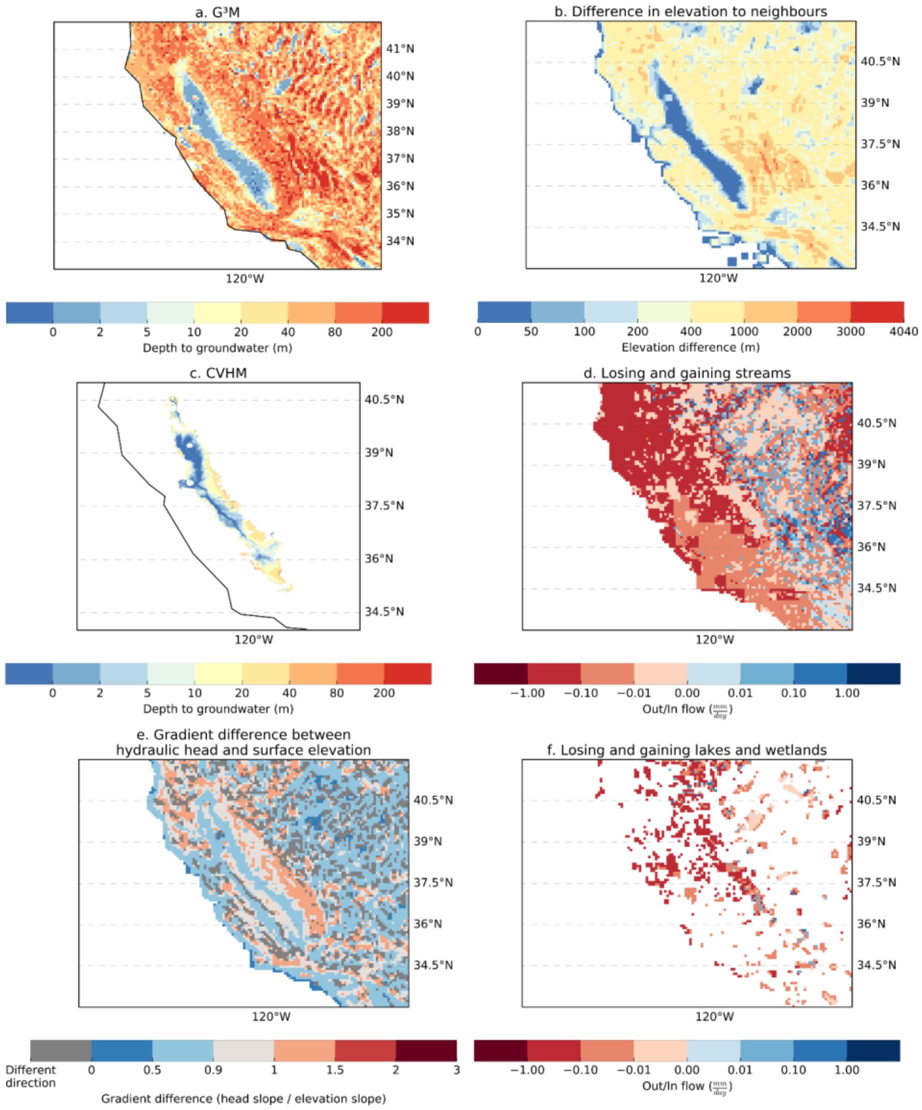

Figure S4.6 Plots of depth to GW as calculated by G³M (a), difference in surface elevation to neighbouring cells (b), depth to GW as used by the CVHM as the natural state and starting condition (Faunt, 2009) (c), losing and gaining streams as calculated by G³M (d), difference in gradient of hydraulic head and surface elevation (e), losing and gaining lakes and wetlands as calculated by G³M for the Central Valley and the Great Basin.

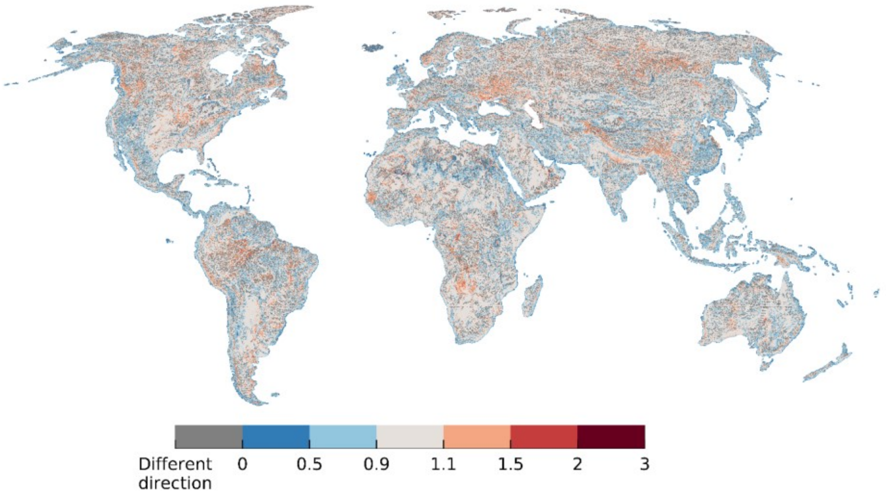

Figure S4.7 Ratio of hydraulic head gradient to 5' mean surface elevation gradient, only computed if the difference in direction of the gradient was smaller than 45°.

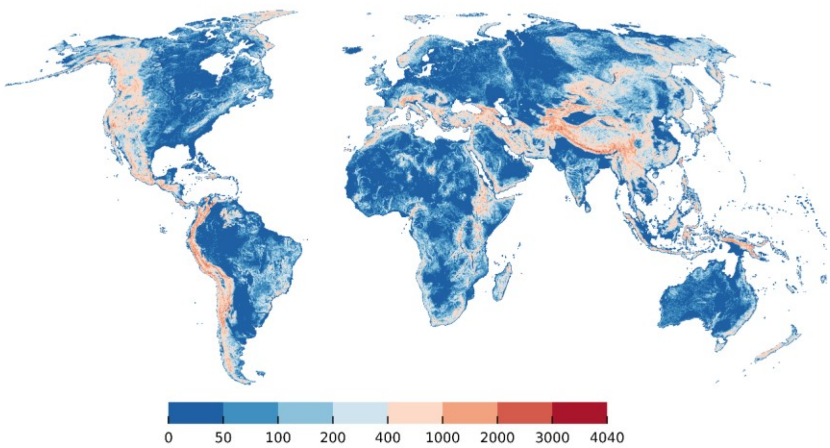

Figure S4.8 Land-surface elevation Difference of 30'' mean land surface elevation in 5' grid cell to mean elevation of neighboring cells [**m**] to mean elevation of neighboring cells on 5' resolution.

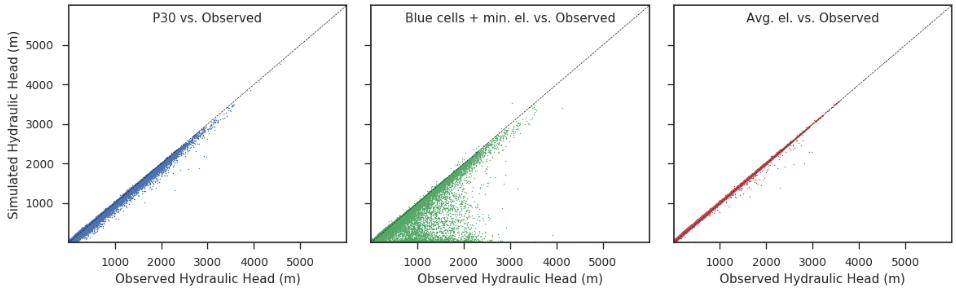

Figure S4.9 Comparison between three alternatives for setting $\boldsymbol{h_{swb}}$. Left to right: Fit of simulated hydraulic heads observations if $\boldsymbol{h_{swb}}$ is set (1) to the 30th percentile of the 30'' land surface elevations (standard model) , (2) alternatively to the average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) or (3) is set to the average of the 30'' land surface elevations. A blue cell has a depth to GW of less than 0.25 m and indicates GW discharge to the surface. If no "blue" cell exists in the 5' cell, the minimum elevation of the 30'' land surface elevation values within the cell was used.

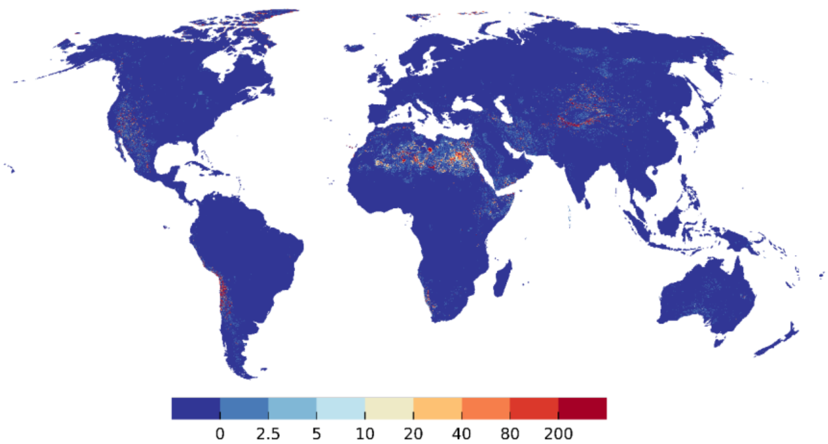

Fig. S4.10 Depth to groundwater [$\boldsymbol{m}$] for SW body elevation at average of 30'' land surface elevations.

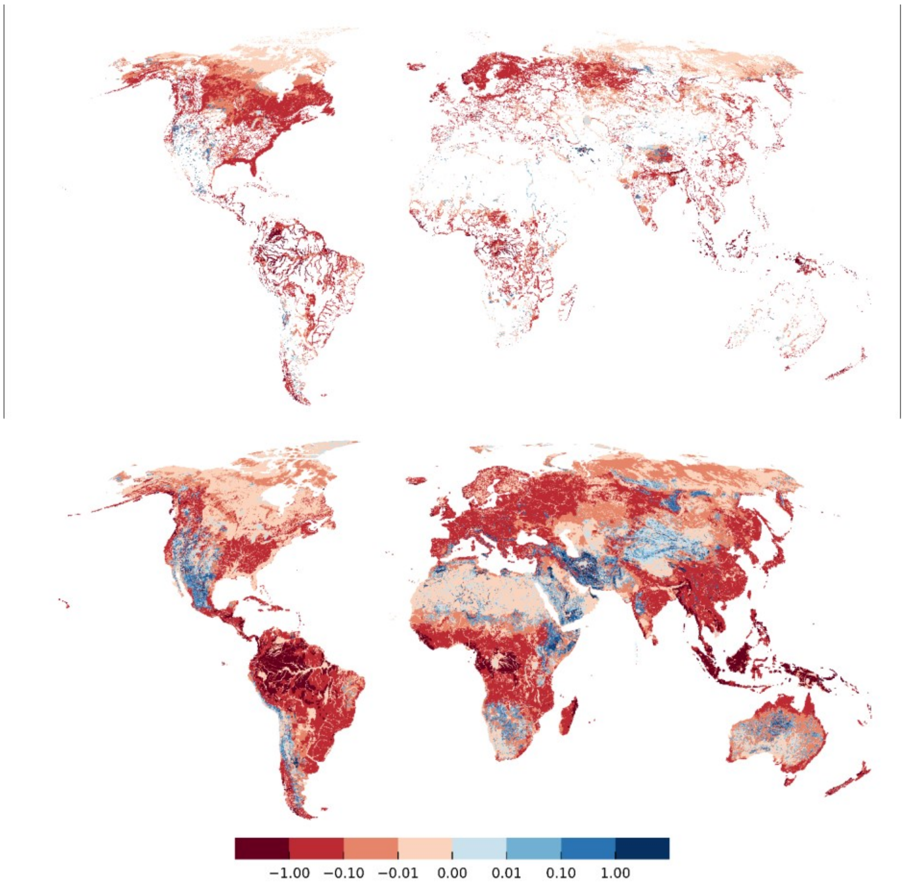

Figure S4.11 Gaining and losing rivers (lower panel) and wetlands and lakes (upper panel) as flow into/out the GW [$\boldsymbol{mm\ day^{-1}}$] if $\boldsymbol{h_{swb}}$ is set to average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) (right). A blue cell is defined as a depth to groundwater of less than 0.25 m. If no "blue" cell exist in the 5' cell, the minimum elevation of the 30'' land surface elevation values is used. Red denotes gaining SW bodies.

Dear Dr. Kurtz,

Thank you for your remarks and the offer to provide additional guidance for revising the manuscript. In the following we are presenting additional changes to the manuscript to address the issues raised by the reviewers, as indicated in your letter.
Our comments to your remarks (numbered) are provided in italics.

(1) Both reviewers question the choice of a confined aquifer for the simulation of river-aquifer exchange. Evidence needs to be provided that the chosen assumption does not bias model results, e.g. by comparison with an unconfined representation of the upper model layer (as used in other available large-scale groundwater models), or by conducting a sensitivity study as suggested by referee 2.

We acknowledge that the approach is counterintuitive and needs to be justified more extensively. Assuming that hydraulic conductivity/transmissivity is independent of the hydraulic head is an established practice in other large-scale GW models e.g. the Rhine Meuse basin model of Sutanudjaja et al. (2011), the Death Valley Regional Flow Model of Belcher (2004) and is applied in the global-scale groundwater flow model of de Graaf et al. (2015; 2017). Furthermore, this assumption, which should more appropriately be called a constant saturated thickness assumption, has been previously investigated by Faunt et al. (2011) and Sheets et al. (2015) showing reasonable performance for a large-scale regional GW model.

We extend section 2.1 (page 5, lines 1-14) with an additional paragraph:

"[..]The simulation of aquifers that contain dry cells and/or cells that oscillate between wet and dry states pose great challenges to the solver (Niswonger et al. 2011). G³M-f implements the traditional wetting approach of Harbaugh (2005) as well as the approach prosed in Niswonger et al. (2011) along with the proposed damping scheme. However, both approaches have proven to be insufficient to simulate hydraulic head-dependent transmisivities (i.e. unconfined conditions) on the global scale. Large mountainous areas would be excluded from the solution as the head is often far below the deepest model layer, resulting in a no-flow condition and causing convergence issues to the matrix solver. Therefore, we chose to simulate both layers with a constant saturated thickness as experiments with large-scale groundwater flow models have shown that the constant saturated thickness assumption is appropriate for large, complex groundwater models (Faunt et al. 2011; Sheets et al. 2015). Given the large uncertainties regarding hydraulic conductivities (possibly an order of magnitude), it is appropriate to choose the computationally more efficient assumption of specified saturated thickness. This approach was also chosen in recent large-scale groundwater flow studies, e.g for the Rhine Meuse basin (Sutanudjaja et al. 2011) (using one confined layer), the Death Valley Regional Flow Model (Belcher 2004), and the global groundwater model of de Graaf et al. (2017) (two layers and partially unconfined conditions are simulated by parameterization of the storage coefficients and not by a head-dependant transmissivity)."

(2) The verification of the model shows considerable deficiencies. There is a significant mismatch between modelled and observed water table depths which is mainly justified by scale effects. Referee 1 rightfully points out that these discrepancies undermine the credibility of the modelling approach, also with respect to future model extensions and asks for a regional-scale validation of the model (which also avoids the problems of comparison of hydraulic heads over large ranges). The provided comparison with the Central Valley model only provides a rough qualitative comparison whereas a more detailed analysis would be needed for validation. If scale is a general obstacle for a proper model validation, then a logical consequence would be to increase model resolution to an appropriate level.

Increasing the model resolution on the global scale is not possible not only due to insufficient data but in particular computational demand hindering the goal of having a model that can be further evaluated,

e.g. with a sensitivity analysis, and calibrated. We are currently investigating these scale impacts (see next remark).
To reflect the challenges that the proposed approach faces we have changed the paper title and abstract:

"Challenges in developing a global gradient-based groundwater model (G^3M v1.0) for integration into a global hydrological model.

To quantify water flows between groundwater (GW) and surface water (SW) as well as the impact of capillary rise on evapotranspiration by global hydrological models (GHMs), it is necessary to replace the bucket-like linear GW reservoir model typical for hydrological models with a fully integrated gradient-based GW flow model. Linear reservoir models can only simulate GW discharge to SW bodies, provide no information on the location of the GW table, and assume that there is no GW flow among grid cells. A gradient-based GW model simulates not only GW storage but also hydraulic head, which together with information on SW table elevation enables the quantification of water flows from GW to SW and vice versa. In addition, hydraulic heads are the basis for calculating lateral GW flow among grid cells and capillary rise.

Global gradient-based modelling of GW is limited to certain resolutions due to available data and computational demands. We identify challenges linked to the coarse resolution, which necessitates the deviation from established processes in regional groundwater modelling as simulation of unsaturated flow and surface water body elevation. We present G^3M, a global gradient-based GW model with a spatial resolution of 5' intended to replace the current linear GW reservoir in the 0.5° WaterGAP Global Hydrology Model (WGHM). The newly developed model framework enables in-memory coupling to WGHM while keeping overall runtime relatively low, allowing sensitivity analyses and data assimilation. This paper presents the G^3M concept and specific model design decisions together with results under steady-state naturalized conditions (neglecting GW abstractions) that can later be used as initial conditions for the fully-coupled WGHM-G^3M runs. Cell-specific conductances of river beds, which govern GW-SW interaction, were determined based on the 30'' steady-state water table computed by Fan et al. (2013). Together with an appropriate choice for the effective elevation of the SW table within each grid cell, this enables a reasonable simulation of drainage from GW to SW such that, in contrast to the GW model of de Graaf et al. (2015; 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G^3M allowing a full coupling to a GHM. Comparison of simulated hydraulic heads to observations around the world shows better agreement than de Graaf et al. (2015). In addition, G^3M output is compared to the output of two established macro-scale models for the Central Valley, California, and the continental United States, respectively. A first analysis of losing and gaining rivers and lakes/wetlands indicates that GW discharge to rivers is by far the dominant flow, draining diffuse GW recharge, such that lateral flows only become a large fraction of total diffuse and focused recharge in case of losing rivers and some areas with very low GW recharge. G^3M does not represent losing rivers in some dry regions. This study clarifies the conceptual approach to gradient-based groundwater modelling that is necessary for global-scale modelling with a coarse spatial resolution. It presents the first steps towards replacing the linear GW reservoir model in a 0.5° GHM with a 5' gradient-based groundwater model, improving on recent efforts (fit to observations, model coupling), while explicating the major challenges related to the model resolution and the need for analysing the applicability of available higher-resolution land surface elevation data to overcome these challenges in the future."

(3) At the current stage, it is not clear what was learned from the modelling exercise and what are the advantages and differences compared to already existing model set-ups and software packages. As advertised in the manuscript, a main purpose of this study is to use a steady state groundwater model to gain first insights into the credibility of the model set-up for future coupled transient simulations. However, this is not streamlined at the moment, i.e. it is not clear what experience you gained from the

steady state simulations to move forward to transient coupled simulations. The problems outlined by the referees regarding model set-up and model verification will prevail for transient coupled simulations and therefore need to be properly addressed at the current stage. It needs to be pointed out more clearly, how this study serves as a basis for the development of the desired coupled WaterGAP model.

In our view the main contributions of that paper to the scientific community apart from the model itself are:

- *The coupling strategy of a global groundwater model to a GHM on a different spatial resolution*
- *The **first time** that it is possible to **fully** couple a GHM to a GW model*
- *Provide insights into the scale challenges (SWB choices, observation comparison, choice in equations)*
- *Largely improve on recent efforts in comparison to observations*
- *We found that a conceptual approach to gradient-based groundwater flow modelling, with virtual layers, is necessary to deal with the large differences in land surface elevations.*
- *We determined that the large horizontal extent of grid cells required at the global scale is a major problem and developed a research plan involving simulation of the steady-state groundwater flow in New Zealand at three different spatial resolutions using globally available land surface elevation data; to identify how to parameterize the 5 min model using higher-resolution land surface elevation data.*

To streamline the innovation of the paper and the gained insights we added the following paragraphs to the conclusion (together with the changed abstract and title in 2):

"[..] Furthermore, we provided insights into how the choice of surface water body elevation affects model outcome and plan to further investigate how we can use higher resolution topographic data to overcome these challenges by comparing simulation results of a 5', 30", and 3" GW model of New Zealand."

(4) You should more clearly explain the advantages and differences of your modelling approach compared to already existing infrastructure as requested by both referees. For example, it is still not clear why G3M would be technically more suitable for in-memory coupling than already existing groundwater models like Modflow, i.e. why is it easier to access and exchange model variables with G3M. As pointed out by referee 1, the global-scale model set-up itself could be implemented with already existing groundwater modelling software. Following referee 2 also a more in-depth analysis on the differences between the proposed model and the model of de Graaf et al. is needed.

We agree that we should point out the advantages of G^3M-f over an existing modelling software. We added the following paragraph to 2.4 (page 8, lines 29-41):

"[..]G^3M-f has the following advantages over using an established GW modelling software such as *MODFLOW. G^3M-f enables an improved coupling capability: (1) as it is intended to be used as a library-like module (unlike MODFLOW it provides a clear development interface to the programmer coupling a model to G^3M-f, can be easily compiled as a library, and provides a clearly separated logic between computation and data read-in/write-out), (2) is written in the same language as the target GHM enabling a straight-forward in-memory access to arrays without the need to write data to disk, required by other global models (a very expensive operation even if that disk is a RAM-disk). Even though, it is possible to call FORTRAN functions from C++ it is complicated to pass file pointers properly, as the I/O implementation of both languages differ substantially and it is widely considered bad practice to handle I/O in two different languages at once. As MODFLOW was never designed to be coupled to other models, it is not possible to separate the I/O logic fully from the computational logic without substantial code changes that are hard to test. To this end, G^3M-f provides a highly modularized framework that is written with extensibility as design goal while implementing all required groundwater mechanisms."*

Concerning the differences to de Graaf et al., we think that we had already addressed the differences to a large extent.

1) We chose to address this difference already in the abstract: "*Together with an appropriate choice for the effective elevation of the SW table within each grid cell, this enables a reasonable simulation of drainage from GW to SW such that, in contrast to the GW model of de Graaf et al. (2015; 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G³M allowing a full coupling to a GHM. Comparison of simulated hydraulic heads to observations around the world shows better agreement than de Graaf et al. (2015).*"

2) In the introduction:
"The first global gradient-based GW model that was run for both steady-state (de Graaf et al. 2015) and transient conditions (de Graaf et al. 2017) was driven by GW recharge and SW data of the GHM PCR-GLOBWB (van Beek et al. 2011). However, there is not yet a two-way coupling of a GW flow model and a GHM. This may be due to the way de Graaf et al. (2015; 2017) modelled river-GW interaction. To achieve plausible hydraulic head results, they found it necessary to add an additional drainage flux to GW drainage driven by the hydraulic head difference between GW and river. This additional drainage, which accounts for about 50% of global GW drainage, is simulated as a function of GW storage above the floodplain, the values of which are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017)) – the model component that the gradient-based model was intended to replace. This prevents a full integration of the global GW flow model of de Graaf et al. (2017) into a GHM, as then, the linear GW reservoir model would be replaced by the GW flow model."

3) Furthermore, we addressed the differences in the discussion:
"de Graaf et al. (2015) set their SW head (h_{swb}) to the land surface elevation of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to avoid excessive flooding, and that is not needed in G³M. On the other hand, this adaption allows the drainage of water even if the hydraulic head is below the SW elevation that might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity of SW body elevation."

For clarity, to describe the differences of G³M to three previous global/continental-scale groundwater modelling approach referred to in our manuscript better we add an additional table in the discussion.

"Table 2 Comparison of global- and continental-scale groundwater models

Aspect	**G³M**	**de Graaf et al. (2015; 2017)**	**Fan et al. (2013)**	**ParFlow**
Extent	Global	Global	Global	Continental USA
Resolution	5'	6'	30"	1 km
Software	G³M-f	MODFLOW	Unnamed	ParFlow
Computational expense	Medium	Medium	High	Very high
Equation	3d Darcy	3d Darcy	2d Darcy	3d Richards
Time scale	Steady-state/(transient)	Steady-state/transient	Steady-state	Steady-state
Vertical layers	2	2	1	5
Full coupling possible	Yes	No (Conceptual issue)	No	Yes (already coupled)
In-memory coupling	Yes	No	N/A	Yes

Constant saturated thickness	Yes	Yes	No	No
Impermeable bottom	No	No	No	Yes
Surface water body location	In every cell	In almost every cell	No surface water	Created during simulation
Surface water body elevation	P_{30} of 30" DEM	Avg. of 30" DEM	N/A (outflow if depth to GW < 0.25 m)	N/A
Deviation from observations	Large	Very large	Medium	Medium

"

(5) The manuscript quite often refers to implementation that are planned in the future (i.e. coupling to WaterGAP, transient simulations). While it is ok to outline these future plans in order to justify the proposed modelling approach for the current global steady-state model, the frequency of references to future work is often misleading (as also acknowledged by referee 2), because it is often not clear which kind of feature is really implemented and which one is intended to be implemented. Therefore, I would ask you to limit references to future work to the necessary minimum.

We acknowledge that remark and think that providing a vision for a newly presented model is necessary to understand the presented decisions. We revised the presented document extensively and reduced references to future development to a minimum.

Kind regards,
Robert Reinecke

References

Belcher, W.R (Ed.). (2004): Death Valley Regional Ground-water Flow System, Nevada and California-Hydrogeologic Framework and Transient Ground-water Flow Model. In: *U.S. Geological Survey Professional Paper* (2004-5205), pp. 408.

de Graaf, Inge E.M. de; Sutanudjaja, E. H.; van Beek, L. P. H.; Bierkens, M. F. P. (2015): A high-resolution global-scale groundwater model. In: *Hydrol. Earth Syst. Sci.* 19 (2), S. 823-837. DOI: 10.5194/hess-19-823-2015.

de Graaf, Inge E.M. de; van Beek, Rens L.P.H.; Gleeson, Tom; Moosdorf, Nils; Schmitz, Oliver; Sutanudjaja, Edwin H.; Bierkens, Marc F.P. (2017): A global-scale two-layer transient groundwater model. Development and application to groundwater depletion. In: *Advances in Water Resources* 102, S. 53–67. DOI: 10.1016/j.advwatres.2017.01.011.

Fan, Y.; Li, H.; Miguez-Macho, G. (2013): Global patterns of groundwater table depth. In: *Science (New York, N.Y.)* 339 (6122), S. 940–943. DOI: 10.1126/science.1229881.

Faunt, Claudia C.; Provost, Alden M.; Hill, Mary C.; Belcher, Wayne R. (2011): Comment on "An unconfined groundwater model of the Death Valley Regional Flow System and a comparison to its confined predecessor" by R.W.H. Carroll, G.M. Pohll and R.L. Hershey [Journal of Hydrology 373/3–4, pp. 316–328]. In: *Journal of Hydrology* 397 (3-4), S. 306-309. DOI: 10.1016/j.jhydrol.2010.11.038.

Harbaugh, Arlen W. (2005): MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process: US Department of the Interior, US Geological Survey Reston.

Niswonger, Richard G.; Panday, Sorab; Ibaraki, Motomu (2011): MODFLOW-NWT, a Newton formulation for MODFLOW-2005. In: *US Geological Survey Techniques and Methods* 6 (A37).

Sheets, Rodney A.; Hill, Mary C.; Haitjema, Henk M.; Provost, Alden M.; Masterson, John P. (2015): Simulation of water-table aquifers using specified saturated thickness. In: *Ground Water* 53 (1), S. 151–157. DOI: 10.1111/gwat.12164.

Sutanudjaja, E. H.; van Beek, Ludovicus P. H.; de Jong, S. M.; van Geer, F. C.; Bierkens, Marc F. P. (2011): Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin. In: *Hydrol. Earth Syst. Sci.* 15 (9), S. 2913–2935. DOI: 10.5194/hess-15-2913-2011.

van Beek, L. P. H.; Wada, Yoshihide; Bierkens, Marc F. P. (2011): Global monthly water stress: 1. Water balance and water availability. In: *Water Resour. Res.* 47 (7). DOI: 10.1029/2010WR009791.

Challenges in~~Beyond the bucket –~~ d~~D~~eveloping a global gradient-based groundwater model (G^3M v1.0) for the integration into a global hydrological model ~~from scratch~~

Kommentiert [RR1]: Editor comment #2 #3

Robert Reinecke[1], Laura Foglia[3], Steffen Mehl[4], Tim Trautmann[1], Denise Cáceres[1], Petra Döll[1,2]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany
[2]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany
[3]Department of Land, Air and Water Resources, University of California, Davis, USA
[4]Department of Civil Engineering, California State University, Chico, USA

Correspondence to: Robert Reinecke (reinecke@em.uni-frankfurt.de)

Abstract. To quantify water flows between groundwater (GW) and surface water (SW) as well as the impact of capillary rise on evapotranspiration by global hydrological models (GHMs), it is necessary to replace the bucket-like linear GW reservoir model typical for hydrological models with a fully integrated gradient-based GW flow model. Linear reservoir models can only simulate GW discharge to SW bodies, provide no information on the location of the GW table, and assume that there is no GW flow among grid cells. A gradient-based GW model simulates not only GW storage but also hydraulic head, which together with information on SW table elevation enables the quantification of water flows from GW to SW and vice versa. In addition, hydraulic heads are the basis for calculating lateral GW flow among grid cells and capillary rise.

Global gradient-based modelling of GW is limited to certain resolutions due to available data and computational demands. We identify challenges linked to the coarse resolution, which necessitates the deviation from established processes in regional -groundwater modelling as simulation of unsaturated flow and SW body elevation.

We present G^3M~~, is~~ a new global gradient-based GW model with a spatial resolution of 5' intended to~~that will~~ replace the current linear GW reservoir in the 0.5° WaterGAP Global Hydrology Model (WGHM). The newly developed model framework enables in-memory coupling to WGHM while keeping overall runtime relatively low, allowing sensitivity analyses and data assimilation. This paper presents the G^3M concept and specific model design decisions together with results under steady-state naturalized conditions, i.e. neglecting GW abstractions, that can later be used as initial conditions for the fully-coupled WGHM-G^3M runs. Cell-specific conductances of river beds, which govern GW-SW interaction, were determined based on the 30" steady-state water table computed by Fan et al. (2013). Together with an appropriate choice for the effective elevation of the SW table within each grid cell, this enables a reasonable simulation of drainage from GW to SW such that, in contrast to the GW model of de Graaf et al. (2015; 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G^3M allowing a full coupling to a GHM. Comparison of simulated hydraulic heads to observations around the world shows better agreement than de Graaf et al. (2015). In addition, G^3M output is compared to the output of two established macro-scale models for the Central Valley, California, and the continental United States, respectively. A first analysis of losing and gaining rivers and lakes/wetlands indicates that GW discharge to rivers is by far the dominant flow, draining diffuse GW recharge, such that lateral flows only become a large fraction of total diffuse and focused recharge in case of losing rivers and some areas with very low GW recharge. G^3M does not represent losing rivers in some dry regions. This study clarifies the conceptual approach to gradient-based groundwater modelling that is necessary for global-scale modelling with a coarse spatial resolution. It presents the first steps towards replacing the linear GW reservoir model in a 0.5° GHM with a 5' gradient-based groundwater model, improving on recent efforts (fit to observations, model coupling), while investigating the challenges related to the resolution of the model and the needed understanding of topographic subscale information to overcome these challenges in the future. ~~It presents the first steps towards replacing the linear GW reservoir model in a GHM while improving on recent efforts, demonstrating the feasibility of the approach and the robustness of the newly developed simulation framework.~~

Kommentiert [RR2]: Editor comment #2 #3

1 Introduction

Groundwater (GW) is the source of about 40% of all human water abstractions (Döll et al., 2014). It is also an essential source of water for freshwater biota in rivers, lakes and wetlands, which are in most cases recharged by GW. GW strongly affects river flow regimes and supplies the majority of river water during ecologically and economically critical periods with little precipitation. GW may receive recharge not only from the soil but also from surface water (SW) bodies. In case of small distances between GW table and land surface, GW enhances evapotranspiration via capillary rise. GW storage and flow dynamics have been altered by human GW abstractions as well as climate change and will continue to change in the future (Taylor et al., 2012). Around the globe, GW abstractions have led to lowered water tables and, in some regions, even GW depletion (Döll et al., 2014; Scanlon et al., 2012; Wada et al., 2012; Konikow, 2011). This has resulted in reduced base flows to rivers and wetlands (with negative impacts on water quality and freshwater ecosystems), land subsidence and increased pumping costs (Wada, 2016; Döll et al., 2014; Gleeson et al., 2012; Gleeson et al., 2016; 2016). The strategic importance of GW for global water and food security will probably intensify under climate change as more frequent and intense climate extremes increase variability of SW flows (Taylor et al., 2012). International efforts have been made to promote sustainable GW management and knowledge exchange among countries, e.g., UNESCO's program on International Shared Aquifer Resources Management (ISARM) (http://isarm.org) and the ongoing GW component of the Transboundary Waters Assessment Programme (TWAP) (http://www.geftwap.org). To support priority setting for investment among transboundary aquifers as well as identification of strategies for sustainable GW management, information on current conditions and possible trends of the GW systems is required (UNESCO-IHP, IGRAC, WWAP, 2012). In a globalized world, an improved understanding of GW systems and their interaction with SW and soil is needed not only at the local and regional but also at the global scale.

To assess GW at the global scale, global hydrological models (GHMs) are used ~~(~~e.g.~~,~~ (Wada et al., 2012; 2016; Döll et al., 2012; 2014)~~)~~. In particular, they serve to quantify GW recharge (Döll and Fiedler, 2008). Like typical hydrological models at any scale, GHMs simulate GW dynamics by a linear reservoir model. In such a model, the temporal change of GW storage in each grid cell is computed from the balance of prescribed inflows and an outflow that is a linear function of GW storage. Linear reservoir models can only simulate GW discharge to SW bodies but not a reversal of this flow, even though losing streams may provide focused GW recharge that allows the aquifer to support ecosystems alongside the GW flow path (Stonestrom et al., 2007) as well as human GW abstractions. This flow direction typically occurs in semi-arid and arid but seasonally also in humid regions. In addition, such linear reservoir models provide no information on the location of the GW table, and assume that GW flow among grid cells is negligible. To simulate the dynamics of water flow between SW bodies and GW in both directions as well as the effect of capillary rise on evapotranspiration, it is necessary to compute lateral GW flows among grid cells as function of hydraulic head gradients and thus the dynamic location of the GW table. To achieve an improved understanding of GW systems at the global scale, and in particular of the interactions of GW with SW and soil, it is therefore necessary to replace the linear GW reservoir model in GHMs by a hydraulic gradient-based GW flow model.

Macro-scale gradient-based GW flow models are still rare and mainly available for data-rich regions, e.g. for the Death Valley (Belcher and Sweetkind, 2010) and the Central Valley (Faunt, 2009; Dogrul et al., 2016) in the USA, but also for large fossil groundwater bodies in arid regions (e.g. the Nubian Aquifer System in North Africa, (Gossel et al., 2004)). However, they are in most cases not integrated within hydrological models that quantify GW recharge based on climate data and provide information on the condition of SW. For North America, Fan et al. (2007) and Miguez-Macho et al. (2007) linked a land surface model with a two-dimensional gradient-based GW model and computed, with a daily time step, gradient-based GW flow, water table elevation, GW–SW interaction and capillary rise, using a spatial resolution of 1.25 km. One challenge was the determination of the river conductance that affects the degree of GW–SW interaction. A computationally very expensive integrated simulation of dynamic SW, soil and GW flow using Richards' equation for variably saturated flow was achieved at a spatial resolution of 1 km for the continental US by applying the ParFlow model (Maxwell et al., 2015). In both studies, GW abstractions were not taken into account.

A first simulation of the steady-state GW table for the whole globe at the very high resolution of 30'' was presented by Fan et al. (2013) and compared to an extensive compilation of observed hydraulic heads. However, there was no head-based interactions with SW; GW above the land surface was simply discarded. Global GW flow modeling is strongly hampered by data availability, including the geometry of aquifers and aquitards as well as parameters like hydraulic conductivity (de Graaf et al., 2017), and by computational restrictions on spatial resolution leading to conceptual problems, e.g., regarding SW-GW interactions (Morel-Seytoux et al., 2017). In the last years, some GW flow models that are in principle applicable for the global scale were developed but were applied only regionally in data-rich regions (Rhine basin: Sutanudjaja et al., 2011; France: Vergnes et al., 2012; 2014). The first global gradient-based GW model that was run for both steady-state (de Graaf et al., 2015) and transient conditions (de Graaf et al., 2017) was driven by GW recharge and SW data of the GHM PCR-GLOBWB (van Beek et al., 2011). However, there is not yet a two-way coupling of a GW flow model and a GHM. This may be due to the way de Graaf et al. (2015; 2017) modelled river-GW interaction. To achieve plausible hydraulic head results, they found it necessary to add an additional drainage flux to GW drainage driven by the hydraulic head difference between GW and river. This additional drainage, which accounts for about 50% of global GW drainage, is simulated as a function of GW storage above the floodplain, the values of which are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017)) – the model component that the gradient-based model was intended to replace. This prevents a full integration of the global GW flow model of de Graaf et al. (2017) into a GHM, as then, the linear GW reservoir model would be replaced by the GW flow model.

In this study, we present the Global Gradient-based Groundwater Model (G^{3}M) that is to be integrated into the GHM WaterGAP 2. With a spatial resolution of 0.5° by 0.5° (approximately 55 km by 55 km at the equator), the WaterGAP 2 model (Alcamo et al., 2003) computes human water use in five sectors and the resulting net abstractions from GW and SW for all land areas of the globe excluding Antarctica. These net abstractions are then taken from the respective water storages in the WaterGAP Global Hydrology Model (WGHM) (Müller Schmied et al., 2014; Döll et al., 2003; 2012; 2014). With daily time steps, WGHM simulates flows among the water storage compartments canopy, snow, soil, GW, lakes, man-made reservoirs, wetlands and rivers. As in other GHMs, the dynamic of GW storage (GWS) is represented in WGHM by a linear GW reservoir model, i.e.

$$\frac{dGWS}{dt} = R_g + R_{g\_swb} - NA_g - k_g\, GWS \quad (1)$$

where R_g $[L^3T^{-1}]$ is diffuse GW recharge from soil, R_{g_swb} $[L^3T^{-1}]$ GW recharge from lakes, reservoirs and wetlands (only in arid and semiarid regions, with a global constant ~~the~~ value per SW body area ~~globally constant~~), NA_g $[L^3T^{-1}]$ net GW abstraction. The product $k_g\, GWS$ quantifies GW discharge to SW bodies as a function of GWS and the GW discharge coefficient k_g (Döll et al., 2014). G^{3}M is to replace this linear reservoir model in WGHM.

The G^{3}M-f framework (Reinecke, 2018) was developed to provide full control over the involved processes and allow an optimal in-memory coupling to WGHM. Our model development approach was to learn from existing large-scale regional models (Faunt, 2009; Vergnes et al., 2014; Maxwell et al., 2015; Dogrul et al., 2016) to gain insights into how the coarse spatial resolution, incomplete data, and conceptual model design affects model results. We want to find out whether we can use gradient-based groundwater modelling at the global scale, when later integrated into a global hydrological model, to improve estimation of flows between SW and GW (affecting both e.g. streamflow and groundwater recharge and thus water availability for humans and ecosystems) and capillary rise (affecting evapotranspiration). In this paper, we present the model concept as well as steady-state model results. Steady-state simulations are a well-established first step in groundwater model development to understand the basic model behaviour limiting model complexity and degrees of freedom, thus providing insights into dominant processes and uncovering possible model-inherent characteristics impossible to observe in a fully coupled transient model. A transient model might obfuscate model inherent trends due to the slow changing nature of groundwater processes. A fully coupled model furthermore adds complexity and uncertainty to the model outcome. In addition, the steady-state solution can be used as initial condition for future fully coupled transient runs. Capillary rise is not included in the presented steady-state simulation, as simulation of capillary rise requires information of soil moisture that is only available when G^{3}M is fully integrated into WGHM. In the next section, the model concept (including the concept for coupling with WGHM) and equations as well as applied data and parameter values are presented. In section 3, we show steady-state results of G^{3}M driven by WGHM data, without any two-way coupling. Simulated hydraulic heads are compared to observations world-wide and to the output of established regional models. We also discuss the influence of scale and modeling decisions and finally draw conclusions.

2 Model description

2.1 G^{3}M model concept

Although G^{3}M is based on principles of the well-known GW flow modelling software MODFLOW (Harbaugh, 2005), G^{3}M differs from traditional local and regional GW models. These models are generally based on rather detailed information on hydrogeology (including aquifer geometry and properties such as hydraulic conductivity derived from pumping tests), topography, pumping wells, location and shape of SW bodies as well as on observations of hydraulic head in GW and SW. Local observations guide the developer in constructing the model such that local conditions and processes can be properly represented. The lateral extent of individual grid cells of such GW flow models is generally smaller or similar to the depth of the aquifer(s) and the size of the SW bodies that interact with the GW. The global GW flow model G^{3}M, however, covers all continents of the Earth except Greenland and Antarctica. At this scale, information listed above is poor or non-existing, and the lateral extent of grid cells needs to be relatively large due to computational (and data) constraints. We selected a grid cell size of 5' by 5' (approx. 9 km by 9 km at the equator), as this size fits well to WaterGAP and is smaller than the suggested 6'' of Krakauer et al. (2014). WaterGAP 3 (Eisner, 2016) has the same cell size, and 36 of such cells fit to into one 0.5° WaterGAP 2 cell. Global climate data are only available for 0.5° grid cells. The landmask of G^{3}M, i.e. location and size of 5' grid cells, is that of WaterGAP 3.

Due to the lack of the distribution of hydrogeological properties, we chose to use, in the current version of G^{3}M, two GW layers with a vertical size of 100 m each (Fig. 1). We performed a sensitivity analysis that confirmed the findings of other studies (de Graaf et al., 2015) that the aquifer thickness has a relatively small impact on the model results. Therefore, selecting a uniform thickness of 100 m (motivated by the assumed depth of validity of the lithology data) (Fig. 1) worldwide for the first layer and also for the second layer is expected to lead to less uncertainties as compared to hydraulic conductivities and the surface water table elevation.

~~Due to the lack of information on the three-dimensional distribution of hydrogeological properties, we chose to use, in the current version of G^{3}M, two GW layers with a vertical size of 100 m each (Fig. 1).~~ G^{3}M focuses on a plausible simulation of water flows between GW and SW ~~bodies and on the simulation of capillary rise~~, and we deemed it suitable to have an upper GW layer that interacts with SW and soil and a lower one in which GW may flow laterally without such interactions. As land surface elevation within each 5' grid cell, with an area of approximately 80 km^2, may vary by more than 200 m (Fig. S1), neighbouring cells in G^{3}M may not be adjacent anymore (Fig. 1), in contrast to (regional) GW models with smaller grid cells. This makes G^{3}M a rather conceptual model in which water exchange between groundwater boxes is driven by hydraulic head gradients but flow can no longer be conceptualized as occuring through continuous pore space. In addition, due to the coarse spatial scale and the possible large variations of land surface elevations within each grid cell, the upper model layers should not be considered to be aligned with an average land surface elevation. The model layers can be rather thought to be vertically aligned with the elevation of the surface water body table, as this prescribed elevation is, together with the sea level, the only elevation included in the groundwater flow equation (Eq. (2)).

The simulation of aquifers that contain dry cells and/or cells that oscillate between wet and dry states pose great challenges to the solver (Niswonger et al., 2011). G³M-f implements the traditional wetting approach from Harbaugh (2005) as well as the approach prosed in Niswonger et al. (2011) allong with the proposed damping scheme. Both approaches have proven to be insufficient to simmulate head-based transmisivities (unconfined contions) on the global scale. Large mountenous areas would be excluded from the beginning of the solution step, as the head is often far below the deepest model layer, resulting in a no-flow condition and imposing convergence issues to the matrix solver. We choose to simulate both layers with a specific saturated thickness even though the upper layer can be expected to decrease in depth and thus in transmissivity (hydraulic conductivity times saturated depth). Model experiments have shown that the specific saturated thickness assumption is accurate for large, complex groundwater models (Faunt et al., 2011; Sheets et al., 2015). Given the large uncertainties regarding hydraulic conductivities (possibly an order of magnitude), it is appropriate to choose the computationally more efficient assumption of specified saturated thickness. This approach is consistent with recent presented large scale studies e.g for the Rhine Meuse basin of Sutanudjaja et al. (2011) (using one confined layer), the Death Valley Regional Flow Model (Belcher, 2004), and the global groundwater model of de Graaf et al. (2017) (two layers and partially unconfined conditions are simulated by parametrization of the model input and not by a head-dependant transmissivity).

Kommentiert [RR3]: Editor comment #1

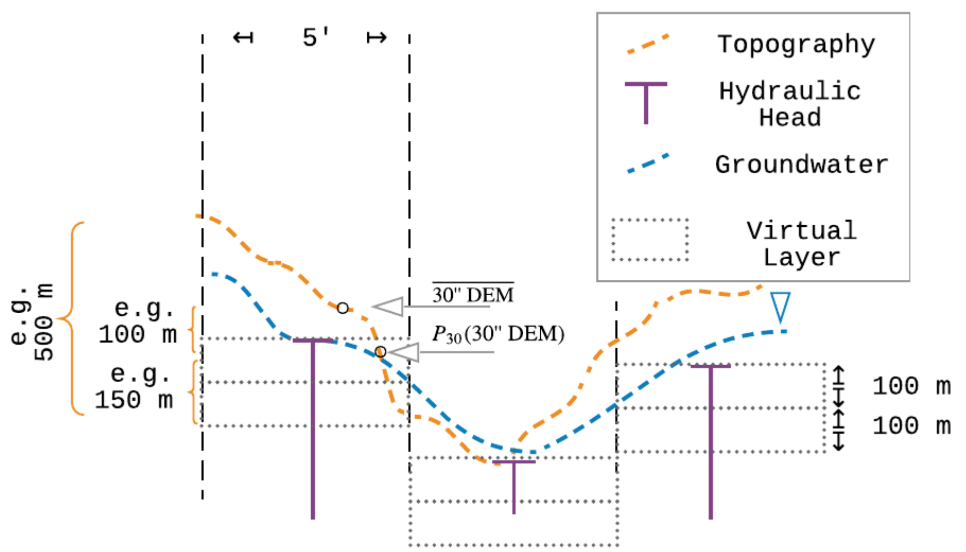

Figure 1 Schematic of G³M's spatial structure, with 5' grid cells, hydraulic head per cell, and the conceptual virtual layers. The underlying variability of the topography changes the perception of simulated depth to groundwater depending on what metrics are used to represent it on a coarser resolution. Layers in G³M are of a conceptual nature and describe the saturated flow between locations of head laterally and vertically.

Three-dimensional groundwater flow is described by a partial differential equation as a function of hydraulic head gradients

$$\frac{dGWS}{dt} = \left(\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) + W\right) \Delta x \Delta y \Delta z = S_s \frac{\partial h}{\partial t} \Delta x \Delta y \Delta z \quad (2)$$

where $K_{x,y,z}$ is the hydraulic conductivity $[LT^{-1}]$ along the x, y, and z axis between the cells (harmonic mean of grid cell conductivity values), S_s the specific storage $[L^{-1}]$, $\Delta x \Delta y \Delta z$ $[L^3]$ the volume of the cell, and h the hydraulic head $[L]$. Inflows in the groundwater are accounted for as

$$W\ \Delta x \Delta y \Delta z = R_g + Q_{swb} - NA_g - Q_{cr} + Q_{ocean} \quad (3)$$

where Q_{swb} is flow between the SW bodies (rivers, lakes, reservoirs and wetlands) and GW $[L^3T^{-1}]$, Q_{cr} is capillary rise, i.e. the flow from GW to the soil, and Q_{ocean} is the flow between ocean and GW[L^3T^{-1}], representing the boundary condition. In case of Q_{swb} and Q_{ocean}, a positive value represents a flow into the groundwater.

The flux across the model domain boundary Q_{ocean} is modeled as a head-dependent flow based on a static head boundary.

$$Q_{ocean} = c_{ocean}(h_{ocean} - h_{aq}) \quad (4)$$

Here h_{ocean} is the elevation of the ocean water table $[L]$, h_{aq} the hydraulic head in the aquifer $[L]$ and c_{ocean} the conductance of the boundary condition $[L^2T^{-1}]$ (Table 1). We assume that density difference to sea-water is negligible at this scale. Q_{cr} is not yet implemented in G^{3}M.

Q_{swb} in Eq. (3) replaces $k_g\ GWS$ and R_{g_swb} in Eq. (1), such that loosing conditions of all types of SW bodies can be simulated dynamically. It is calculated as a function of the difference between the elevation of the water table in the SW bodies h_{swb} $[L]$ and h_{aq} as

$$Q_{swb} = \begin{cases} c_{swb}(h_{swb} - h_{aq}) & h_{aq} > B_{swb} \\ c_{swb}(h_{swb} - B_{swb}) & h_{aq} \leq B_{swb} \end{cases} \quad (5)$$

where c_{swb} is the conductance $[L^2T^{-1}]$ of the SW body bed (river, lake, reservoir or wetland) and B_{swb} the SW body bottom elevation $[L]$.

Conductance of SW bodies is often a calibration parameter in traditional GW models (Morel-Seytoux et al., 2017). Following Harbaugh (2005), it can be estimated by

$$c_{swb} = \frac{K\ L\ W}{h_{swb} - B_{swb}} \quad (6)$$

where K is hydraulic conductivity, L is length and W is width of the SW body per grid cell. For lakes (including reservoirs) and wetlands, c_{swb} is estimated based on hydraulic conductivity of the aquifer K_{aq} and SW body area (Table 1). For gaining rivers, conductance is quantified individually for each grid cell following an approach proposed by Miguez-Macho et al. (2007). The value of river conductance c_{riv}, according to Miguez-Macho et al. (2007), in a GW flow model needs to be set to such a values that, for steady-state conditions, the river is the sink for all the inflow to the grid cell (GW recharge and inflow from neighbouring cells) that is not transported laterally to neighbouring cells such that

$$c_{riv} = \frac{R_g + Q_{eq_{lateral}}}{h_{eq} - h_{riv}} \qquad h_{aq} > h_{riv} \quad (7)$$

For G^{3}M, we computed the equilibrium head h_{eq} as the 5' average of the 30'' steady-state heads calculated by Fan et al. (2013). Using WGHM diffuse GW recharge lateral equilibrium flow $Q_{eq_{lateral}}$ $[L^3T^{-1}]$ is net lateral inflow into the cell computed based on the h_{eq} distribution as well as G^{3}M K_{aq} and cell thickness (Table 1). Elevation of the river water table h_{riv} [L] is to be provided by WGHM. Using a fully dynamic approach, i.e. utilizing the hydraulic head and lateral flows from the current iteration to re-calculate c_{riv} in each iteration towards the steady-state solution, has proven to be too unstable due to its non-linearity affecting convergence. We limit c_{riv} to a maximum of $10^7\ m^2 day^{-1}$; this would be approximately the value for a 10 km long and 1 km wide river with a head difference between GW and river of 1 m and hydraulic conductivity of the river bed of 10^{-5} m/s.

If the river recharges the GW (losing river), Eq. (7) cannot be used as the Fan et al. (2013) high-resolution equilibrium model only models groundwater outflows but not inflows from SW bodies. If h_{aq} drops below h_{riv}, Eq. (5) is used to compute c_{riv}, with K equals to K_{aq}.

2.2 Coupling to WGHM

We intend to couple G³M to WaterGAP 2, i.e. the 0.5° version of WGHM. It will not be coupled to WaterGAP 3 (Eisner, 2016), which has the same spatial resolution as G³M, in order to keep computation time low enough for performing sensitivity analyses and ensemble-based data assimilation and calibration. However, data from WaterGAP 3 were used to set up G³M. Location and area of the 5' grid cells of G³M are the same as in the landmask of WaterGAP 3. In addition, the percentage of the 5' grid cell area that is covered by lakes (including reservoirs) and by wetlands, based on Lehner and Döll (2004), is taken from WaterGAP 3, as well as the length and width of the main river within each 5' grid cell as estimated by WaterGAP 3 (Table 1).

G³M will be integrated into WGHM by exchanging information on (1) R_{g_swb} and NA_g, (2) soil water content, (3) Q_{cr}, (4) h_{swb}, and (5) Q_{swb}. Figure 2 indicates the direction of the information flows. Water flows from the 0.5° cells of WGHM are distributed equally to all 5' G³M grid cells inside a 0.5° cell. Flows transferred from the 5' cells of G³M to WGHM are aggregated. GW recharge and net abstraction from GW together with SW tables are the main drivers of the GW model that will be provided dynamically by WGHM. GW-SW flow volumes computed by G³M will be aggregated and added or subtracted from the SW body volumes in WGHM, and SW body heads will be recalculated. WGHM soil water content together with G³M depth to GW will be used to calculate capillary rise and thus a change of soil water content. Capillary rise is not yet implemented in G³M, and SW heads are currently based on land surface elevation.

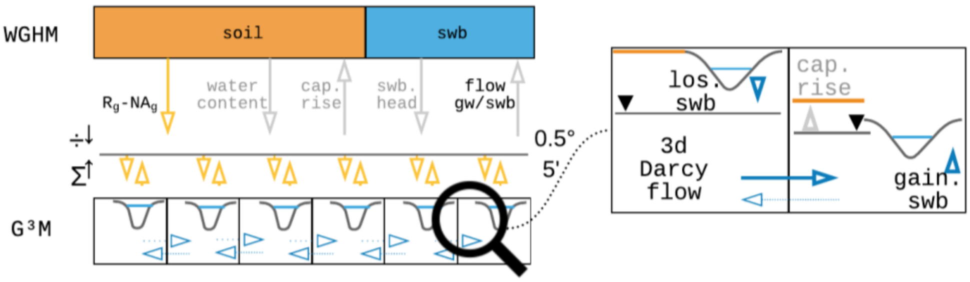

Figure 2 Conceptual view of the coupling between WGHM and G³M. WGHM provides calculated GW recharge (R_g) (Döll and Fiedler, 2008) and if the human impact is considered, net abstraction from GW (NA_g) (Döll et al., 2012). G³M spreads this input equally to all 5' grid cells inside a 0.5° cell and calculates hydraulic head and interactions with SW bodies (swb) as well as capillary rise (cap. rise) at the 5' resolution. Grey arrows show information flow that is not yet implemented.

2.3 The steady-state uncoupled model version

In a first implementation stage, G³M was developed as a steady-state (right-hand side of Eq. (2) is zero) standalone model that represents naturalized conditions (i.e. without taking into account human water use) during 1901-2013. Input data and parameters used are listed in Table 1 and described below.

~~The landmask of G³M, i.e. location and size of 5' grid cells, is that of WaterGAP 3. We performed a sensitivity analysis that confirmed the findings of other studies (de Graaf et al., 2015) that the aquifer thickness has a relatively small impact on the model results. Therefore, selecting a uniform thickness of 100 m (motivated by the assumed depth of validity of the lithology data) worldwide for the first layer and also for the second layer is expected to lead to less uncertainties as compared to hydraulic conductivities and the surface water table elevation. We choose to simulate confined flow conditions in both layers even though the upper layer can be expected to decrease in depth and thus in transmissivity (hydraulic conductivity times saturated depth). Every unconfined aquifer can have an equivalent confined representation assuming a correct saturated thickness (Sheets et al., 2015). However, given the large uncertainties regarding hydraulic conductivities (possibly an order of magnitude) and the lack of knowledge about aquifer thickness, it is appropriate to choose the computationally more efficient assumption of confined conditions.~~

Gleeson et al. (2014) provided a global subsurface permeability data set from which K_{aq} was computed. The data set was derived by relating permeabilities from a large number of local to regional GW models to the type of hydrolithological units (e.g., "unconsolidated" or "crystalline"). The geometric mean permeability values of nine hydrolithological units were mapped to the high-resolution global lithology map GLiM (Hartmann and Moosdorf, 2012). In continuous permafrost areas, a very low permeability value was assumed by Gleeson et al. (2014). The estimated values represent the shallow surface on the scale of 100 m depth. The unique dataset has three inherent problems when used as input for a GW model: (1) At this scale, important heterogeneities such as discrete fractures or connected zones of high hydraulic conductivity controlling the GW flow are not visible. (2) Jurisdictional boundaries due to different data sources in the global lithological map lead to artifacts. (3) The differentiation between coarse and fine-grained unconsolidated deposits is only available in some regions resulting in $10^{-4}\ m\ s^{-1}$ as hydraulic conductivity for coarse-grained unconsolidated deposits. If the distinction is not available, a rather low value of $10^{-6}\ m\ s^{-1}$ is set for unconsolidated porous media (Fig. S3). The original data was gridded to 5' by using an area-weighted average and used as hydraulic conductivity of the upper model layer. For the second layer, hydraulic conductivity of the first layer is reduced assuming that conductivity decreases exponentially with depth. Based on the e-folding factor f used by Fan et al. (2013) (a calibrated parameter based on terrain slope), conductivity of the lower layer is calculated by multiplying the upper layer value by $\exp(-50\ m\ f^{-1})^{-1}$ (Fan et al., 2007).

Mean annual GW recharge computed by WaterGAP 2.2c for the period 1901-2013 is used as input (Fig. S4), while no net abstraction from GW was taken into account. It would not be meaningful to try to derive a steady-state solution under existing net groundwater abstractions that in some regions cause GW depletion with continuously dropping water tables. Regarding the ocean boundary condition, h_{ocean} is set to 0 m and c_{ocean} to $10\ m^2\ day^{-1}$ (Table 1).

It is assumed that there is exchange of water between GW and one river stretch in each 5' grid cell, and in addition where lakes and wetlands exist according to WaterGAP 3, which provides, for each grid cell, the area of "local" and "global" lakes and wetlands. In WaterGAP, "local" SW bodies are only recharged by runoff produced within the grid cell, while "global" SW bodies also obtain inflow from the upstream cell. In an uncoupled model, it is difficult to prescribe the area of lakes and wetlands that affect the flow exchange between SW body and GW. Maps generally show the maximum spatial extent of SW bodies. This maximum extent is seldom reached, in particular in case of wetlands in dry areas. For global wetlands (wetlands greater than one 5' cell), it is therefore assumed in this model version that only eighty percent of their maximum extent is reached. In the transient model SW extends will be changed over time. A further difficulty in an uncoupled model run is that the water table elevation of SW bodies does not react to the GW-SW exchange flows Q_{swb} and that water supply from SW is not limited by availability. A loosing river may in reality dry out due to loss to GW and therefore cease to lose any more water. In case of rivers, B_{swb} is equal to $h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994), where $Q_{bankfull}$ is the bankfull river discharge in the 5' grid cell (Verzano et al., 2012). Globally constant but different values are used for B_{swb} in case of wetlands, local lakes and global lakes (Table 1).

For the steady-state model, river elevation h_{riv} is set in each grid cell to the same elevation as all other SW bodies, h_{swb}. We found that for both gaining and loosing conditions, Q_{swb} and thus computed hydraulic heads are highly sensitive to h_{swb}. The overall best agreement with the hydraulic head observations of Fan et al. (2013) was achieved if h_{swb} (Eq. 5, 6 and 7) was set to the 30th percentile (P_{30}) of the 30" land surface elevation values of Fan et al. (2013) per 5' cell, e.g. the 30" elevation that is exceeded by 70% of the thousand 30" elevation values within one 5' cell. To decrease convergence time we used h_{eq} derived from the high-resolution steady-state hydraulic head distribution of Fan et al. (2013) as initial guess.

2.4 Model implementation

G³M is implemented using a newly developed open-source model framework G³M-f (Reinecke, 2018). The main motivation to develop a new model framework is the efficient in-memory coupling to the GHM and ~~more~~ flexible adaptation to the specific requirements of global-scale modelling. Written in C++14, the framework allows the implementation of global and regional groundwater models alike while providing an extensible purely object-oriented model environment. It is primarily targeted as extension to WGHM but allows an in-memory coupling to any GHM or can be used as a standalone groundwater model. It provides a unit-tested environment offering different modules that can couple results in-memory to a different model or write out data flows to different file formats. G³M-f has the following advantages over using an established GW modelling software such as MODFLOW. G³M-f enables an improved coupling capability: (1) as it is intended to be used as a library-like module (unlike MODFLOW it provides a clear development interface to the programmer coupling a model to G³M-f, can be easily compiled as a library, and provides a clearly separated logic between computation and data read-in/write-out), (2) is written in the same language as the target GHM enabling a straight-forward in-memory access to arrays without the need to write data to disk, required by other global models (an expensive operation even if that disk is a RAM-disk). Even though, it is possible to call FORTRAN functions from C++ it is complicated to pass file pointers properly, as the I/O implementation of both languages differ substantially and it is widely considered bad practice to handle I/O in two different languages at once. As MODFLOW was never designed to be coupled to other models, it is not possible to separate the I/O logic fully from the computational logic without substantial code changes that are hard to test. To this end, G³M-f provides a highly modularized framework that is written with extensibility as design goal while implementing all required groundwater mechanisms.

Kommentiert [RR4]: Editor comment #4

As internal numerical library it uses Eigen3 (eigen.tuxfamily.org). Different from Vergnes et al. (2014), G³M's computations are not based on spherical coordinates directly but on an irregular grid of rectangular cells. Cell sizes are provided by WaterGAP3 and are derived from their spherical coordinates maintaining their correct area and centre location. The model code will be adapted in the future to account for the different length in x and y direction per cell correctly.

Eq. (2) is reformulated as finite-difference equation and solved using a conjugate gradient approach and an Incomplete LUT preconditioner (Saad, 1994). In order to keep the memory footprint low, the conjugate gradient method makes use of the sparse matrix. Furthermore, it solves the equations in parallel (preconditioner currently non-parallel). G³M can compute the presented steady-state solution (with the right-hand side of Eq. (2) being zero and the heads of Fan et al. (2013) as initial guess, Table 1) on a commodity computer with four computational cores and a standard SSD in about 30 minutes while occupying 6 GB of RAM.

Similar to MODFLOW, G³M-f solves Eq. (2) in two nested loops: (1) the outer iteration checks the head and residual convergence criterion and adjusts whether external flows have changed into a different state e.g. from gaining to losing conditions and optimizes the matrix if flows are no longer head dependant. (2) The inner loop primarily consists of the conjugate gradient solver, which runs for a number of iterations defined by the user or until the residual convergence criterion is reached (Table 1), solving the current matrix equation.

Because the switch between Eq. (6) and Eq. (7) that occurs if e.g. h_{aq} drops below h_{riv} from one iteration to the next causes an abrupt change of c_{riv} inducing a nonlinearity that affects convergence we introduced an ϵ = 1 m interval around h_{riv} and interpolate c_{riv} by a cubic hermite spline polynomial over that interval. This allows for a smoother transition between both states, reducing the changes in the solution if a river is in a gaining condition in one iteration and in a losing condition in the next and vice versa.

3 Results

3.1 Global hydraulic head and depth to GW distribution under natural steady-state conditions

As expected, the computed global distribution of steady-state hydraulic head (in the upper model layer) under natural conditions (Fig. 3a) follows largely the land surface elevation (Fig. S2), albeit with a lower range and locally different ratios between the hydraulic head and land surface gradients (Fig. S7). Depth to GW can be computed by subtracting the hydraulic head computed by G³M for the upper layer of each 5' grid cell from the arithmetic mean of the land surface elevations of the 100 30" grid cells within each 5' cells (Fig. S2). The global map of steady-state depth to GW (Fig. S5) clearly resembles the map of differences between surface elevation and P_{30} , the assumed water level of SW bodies h_{swb}, shown in Fig. S1**,** which indicates that simulated depth to GW is strongly governed by the assumed water level in SW bodies.

Deep GW, i.e. a large depth to GW, occurs mainly in mountainous regions (Fig. S5). These high values of depth to GW are mainly a reflection of the steep relief in these areas as quantified either by the differences of mean land surface elevations between neighbouring grid cells (Fig. S8) or the difference between mean land surface elevation and P_{30}, the 30th percentile of the 30'' land surface elevations (Fig. S1). When computed hydraulic head is subtracted not from average land surface elevation but from P_{30}, the assumed water table elevation of SW bodies, the resulting map shows that the groundwater table is mostly above P_{30}, in both flat and steep terrain (Fig. 3b). Thus, high depth to GW values at the 5' resolution do not indicate deep unsaturated zones and loosing rivers but just high land surface elevation variations within a grid cell. In steep terrain, 5'~~'~~ water tables are higher above water level in the surface water bodies than in flat terrain (Fig. 3b). Deep GW tables that are not only far below the mean land surface elevation but also below the water table of surface water bodies are simulated to occur in some (steep or flat) desert area with very low GW recharge.

In 2.1 % of all cells, GW head is simulated to be above the average land surface elevation, by more than 1 m in 0.3 % and by more than 100 m in 0.004 % of the cells. The shallow water table in large parts of the Sahara is caused by losing rivers (and some wetlands) that cannot run dry in the model, causing an overestimation of the GW table (section 2.3). Please note that the computed steady-state depth to GW certainly underestimates the steady depth to GW in GW depletion areas such as the High Plains Aquifer and the Central Valley in the USA (section 3.4), Northwestern India, North China Plain and parts of Saudi Arabia and Iran (Döll et al., 2014) as groundwater withdrawals are not taken into account in the presented steady-state simulation of G^3M.

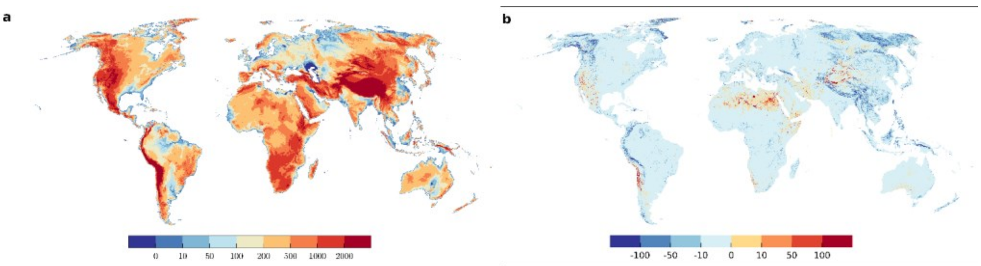

Figure 3 (a) Simulated equilibrium hydraulic head [$\boldsymbol{m}$]. Maximum value 6375 m, minimum value -414 m (Extremes included in dark blue and dark red). (b) Difference between 30th percentile of the 30'' land surface elevation per 5' grid cell (chosen elevation for surface water bodies) and simulated equilibrium hydraulic head. Maximum value 1723 m, minimum value -1340 m (Extremes included in dark blue and dark red).

3.2 Global water budget

Inflows to and outflows from GW of all G^3M grid cells were aggregated according to the compartments ocean, river, lake, wetland, and diffuse GW recharge from soil (Fig. 4). The difference between the global sum of inflows and outflows is~~n~~ less than 10^{-6} %. This small volume balance error indicates the correctness of the numerical solution.

Total diffuse GW recharge from soil is $3.9\ 10^{10}\ m^3\ day^{-1}$ and approximately equal to the drainage of GW to rivers. Rivers are the ubiquitous drainage component of the model~~)~~, followed by wetlands, lakes and the ocean boundary. According to G^3M, the amount of river water that recharges GW is only about a 40th of the drainage to GW, and the relative inflow to GW from lakes and wetlands is even smaller (Fig. 4). Possibly, flow from SW to GW is ~~even~~ overestimated, as outflow from SW is not limited by water availability in the SW, and depending on the hydraulic conductivity, Eqs. (5) and (6) can lead to rather large flows. Inflow from the ocean, which is more than two magnitudes smaller than outflow to ocean, occurs in regions where $h_{swb} = P_{30}$ is below h_{ocean}.

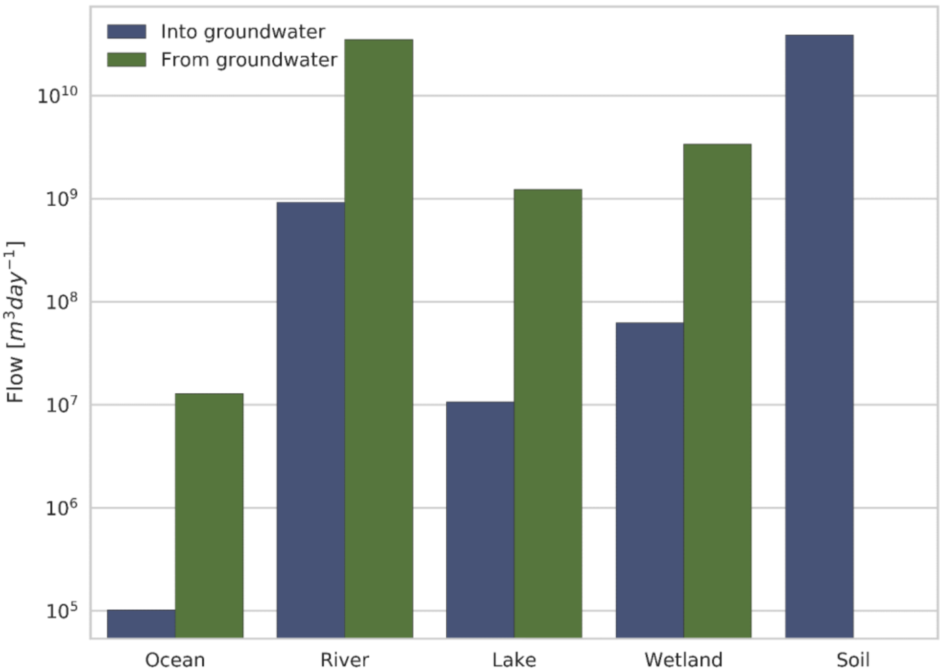

Figure 4 Global sums of flows from different compartments into or from GW at steady state. Flows into the GW are denoted by the color blue, flows out of the GW into the different compartments by green. The compartment soil is the diffuse GW recharge from soil calculated by WaterGAP.

3.3 GW-SW interactions

Figure 5 plots the spatial distribution of simulated flows from and to lakes and wetlands (Fig. 5a) as well as from and to rivers (Fig. 5b). It reveals strong interaction between GW and SW bodies that is dominated by GW discharging to SW bodies. Parallel to the overall budget (Fig. 4), the map reveals the globally large but locally strongly varying influence of lakes and wetlands (Fig. 5a). Rivers with riparian wetlands such as the Amazon River receive comparably small amounts of GW as most of the GW is drained by the wetland (compare Figs. 5a and 5b). Similarly, areas dominated by wetlands and lakes (e.g. parts of Canada and Scandinavia) show less inflow for rivers (Fig. 5b). 93 % of all grid cells contain gaining rivers, and only 7% loosing rivers. Gaining lakes and wetlands are found in 12 % and 11 % of the cells, respectively, whereas only 0.2 % contain a losing lake or wetland. In G^3M, all SW bodies (rivers, lakes and wetlands) in a grid cell either loose or gain water.

Gaining rivers, lakes and wetlands with very high absolute Q_{swb} values over 1 mm day^{-1} (averaged over the grid cell area of approximately 80 km^2) can be found in the Amazon, Congo, Bangladesh, and Indonesia, where GW recharge in very high (Fig. S4). Values below 0.01 mm day^{-1} are present in dry and in permafrost areas where groundwater recharge is small.

Losing SW bodies occur in the model under two conditions, in mountainous regions where depth to GW is high and in arid and semi-arid climates with low diffuse groundwater recharge. Without focused GW recharge, the GW table would drop to even further in the mountains and is necessary to counteract the large hydraulic gradients caused by the large topographic gradients. Rivers lose more than 1 mm day^{-1} in Ethiopia and Somalia, West Asia, Northern Russia, the Rocky Mountains and the Andes whereas lower values can be observed in Australia and in the Sahara. High values of outflow from wetlands and lakes are found in Tibet, the Andes and northern Russia, lower values in the Sahara and Kazakhstan. The river Nile in the Northern Sudan and Egypt is correctly simulated to be a losing river (Fig. 5b), being an allogenic river that is mainly sourced from the upstream humid areas, including the man-made Lake Nasser (Elsawwaf et al., 2014) (Fig. 5a). Furthermore, the following lakes and riparian wetlands are simulated to recharge GW: parts of the Congo River, Lake Victoria, the Ijsselmeer, Lake Ladoga, the Aral Sea, parts of the Mekong Delta, the Great Lakes of North America. On the other hand, no losing stretches are visible at the Niger River and its wetlands and almost none Northeastern Brazil even though that losing conditions are known to occur there (Costa et al., 2013; FAO, 1997).

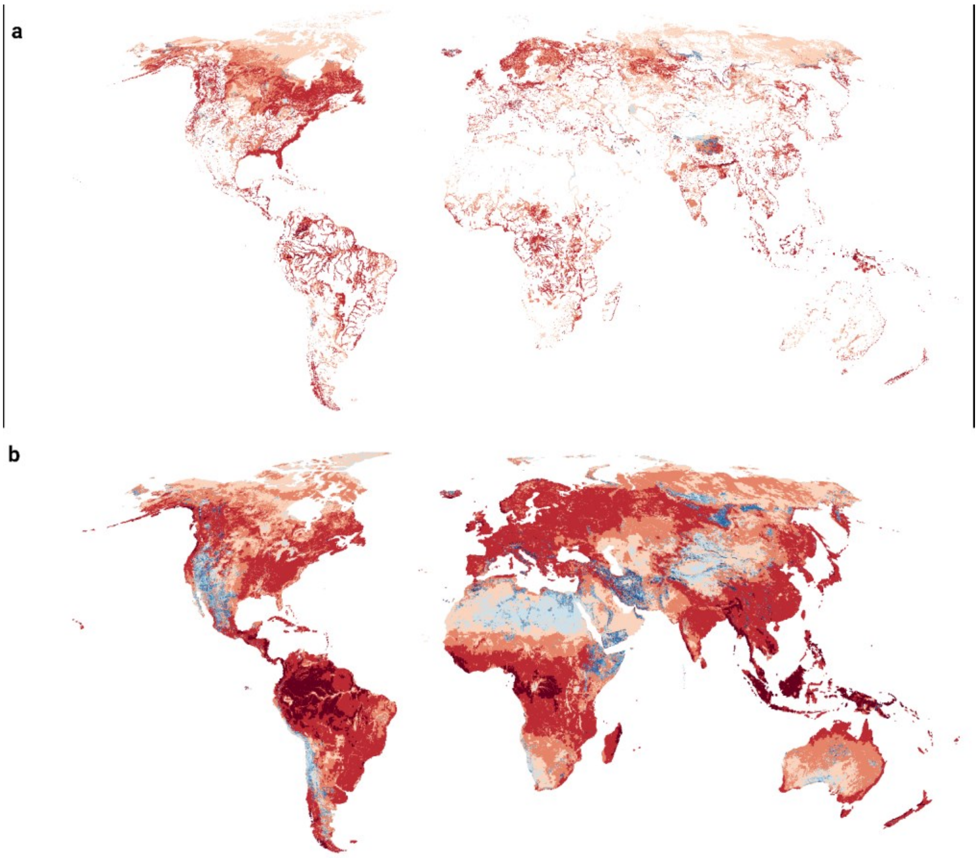

Figure 5 Flow $\boldsymbol{Q_{swb}}$ [$\boldsymbol{mm\ day^{-1}}$] from/to wetlands, lakes (a) and losing/gaining streams (b) with respect to the 5'-grid cell area. Gaining rivers are shown in red, rivers recharging the aquifer in blue. Focused recharge occurs in arid regions, e.g. alongside the river Nile, and in mountainous regions where the average water table is well below the land surface elevation.

Simulated flows between GW and SW depend on assumed conductances for both rivers and lakes/wetlands (Eqs. (5), (6), (7)) shown in Fig. 6. Q_{swb} (Fig. 5) correlates positively with conductance. Conductance for gaining rivers correlates positively with GW recharge (Eq. (7) and Fig. S4). High conductance values are reached in the tropical zone due to a higher GW recharge but are capped at a plausible maximum value of $10^7\, m^2 day^{-1}$s in case of river (section 2.1) (Fig. 6b), while lakes and wetlands, with a larger area, can reach larger values, e.g. in Canada or Florida.

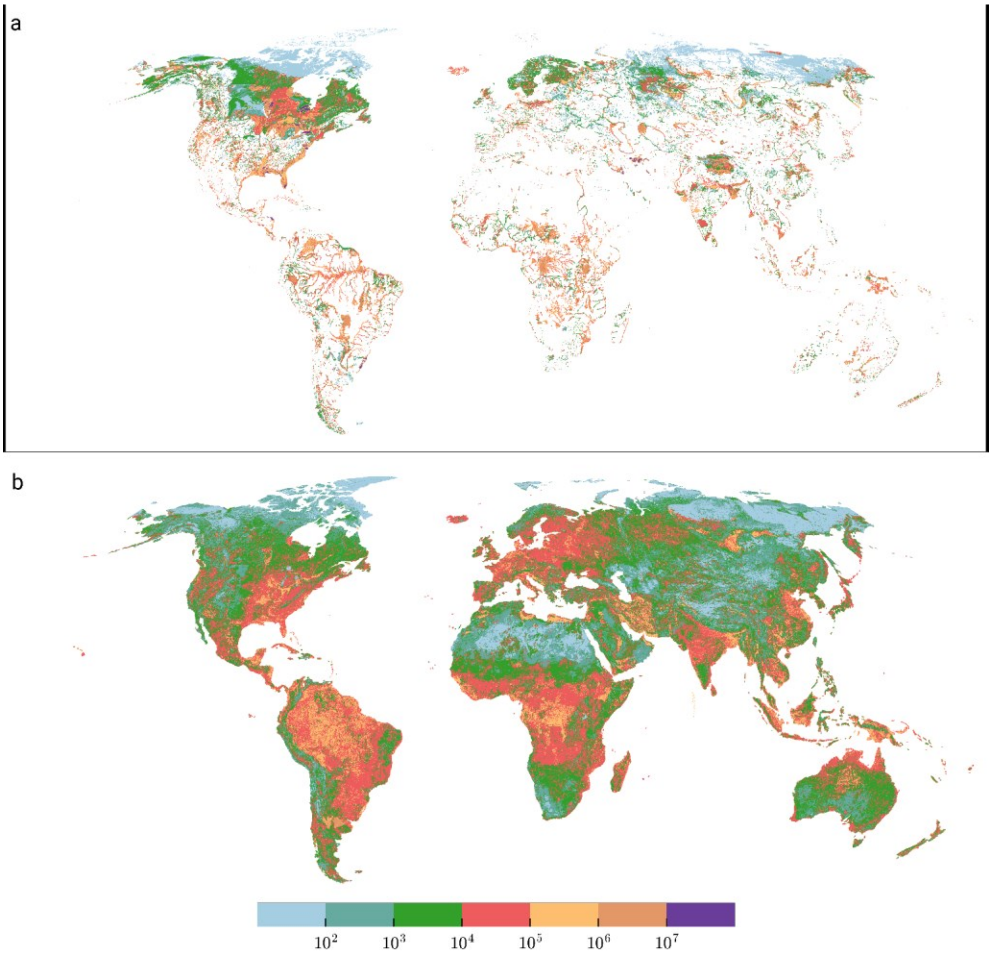

Figure 6 Conductance [$m^2 day^{-1}$] of lakes and wetlands (a) and rivers (b). In regions close to the pole conductance is in general lower due to the influence of the low aquifer conductivity (losing conditions), and relatively small GW recharge due to permafrost conditions (only applies for gaining conditions). Max conductance of wetlands is 10^8 $m^2 day^{-1}$.

3.4 Lateral flows

Figure 7 shows lateral outflow from both model layers in percent of the sum of diffuse GW recharge from soil and GW recharge from SW bodies. The percentage of recharge that is transported through lateral flow to neighbouring cells depends on 5 main factors: (1) hydraulic conductivity (Fig. S3), (2) diffuse GW recharge (Fig. S4), (3) losing or gaining SW bodies (Fig. 5), (4) their conductance (Fig. 6) and (5) the head gradients (Fig. 3).

On large areas of the globe, where GW discharges to SW, the lateral flow percentage is less than 0.5% of the total GW recharge, as GW recharge in a grid cells is simulated to leave the grid cell by discharge to SW. For example, in the permafrost regions, very low hydraulic conductivity limits the outflow to neighbouring cells of the occurring recharge, leading to these very low percent values. Such values also occur in regions with high SW conductances and rather low hydraulic conductivity, e.g. in the Amazon Basin. Values of more than 5% occur where hydraulic conductivity is high even if the terrain in rather flat, such as in Denmark. Higher values may occur for in case of gaining rivers in dry areas like Australia or in mountainous regions where large hydraulic gradients can develop. In mountains with gaining surface water bodies, lateral outflows may even exceed GW recharge of the cell. In grid cells where SW bodies recharge the GW, outflow tends to be a large percentage of total GW recharge as there is no outflow from GW other than in lateral direction, and values often exceed 100% (Fig. 7).

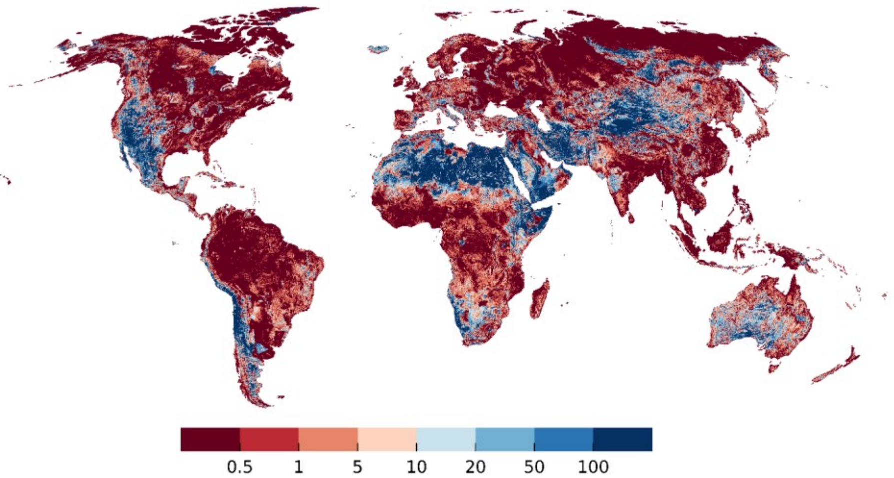

Figure 7 Percentage of GW recharge from soil and surface water inflow that is transferred to neighboring cells through lateral out flow (sum of both layers). Grid cells with zero total GW recharge are shown in white (a few cells in the Sahara and the Andes).

3.5 Comparison to groundwater well observations

Global observations of depth to GW were assembled by Fan et al. (2007; 2013). We selected only observations with known land surface elevation and removed observations where a comparison to local studies suggested a unit conversion error. This left total of 1,070,402 depth to GW observations. An "observed head" per 5' model cell was then calculated by first computing hydraulic head of each observation by subtracting depth to GW from the 5' average of the 30" land surface elevation~~land surface elevation used in G³M~~ and then calculating the arithmetic mean of all observations within the 5' model cell. Multiple obstacles limit the comparability of observations to simulated values. (1) Observations were recorded at a certain moment in time influenced by seasonal effects and abstraction from GW, whereas the simulated heads represent a natural steady-state condition. (2) Observation locations are biased towards river valleys and productive aquifers. (3) Observations may be located in valleys with shallow local water tables too small to be captured by a coarse resolution of 5'.

Simulated steady-state hydraulic heads in the upper model layer are compared to observations in Fig. 8. Shallow GW is generally better represented by the model than deeper GW. Especially the water table in mountainous areas is underestimated, which may be related to observations in perched aquifers caused by low permeability layers (Fan et al., 2013) that are not represented in G^3M due to lacking information. Because the steady-state model cannot take into account the impact of GW abstraction, the computed depth to GW values are considerably smaller than currently observed values in GW depletion areas like the Central Valley in California (where once wetlands existed before excessive GW use depleted the aquifer) and the High Plains Aquifer in the Midwest of the USA. Still, the elevation of the GW table in the non-depleted Rhine valley in

Germany is overestimated, too. Figure 9a shows the hydraulic head comparison as scatter plot. Overall, the simulation results tend to underestimate observed hydraulic head but much less than the steady-state model presented by de Graaf et al. (2015).

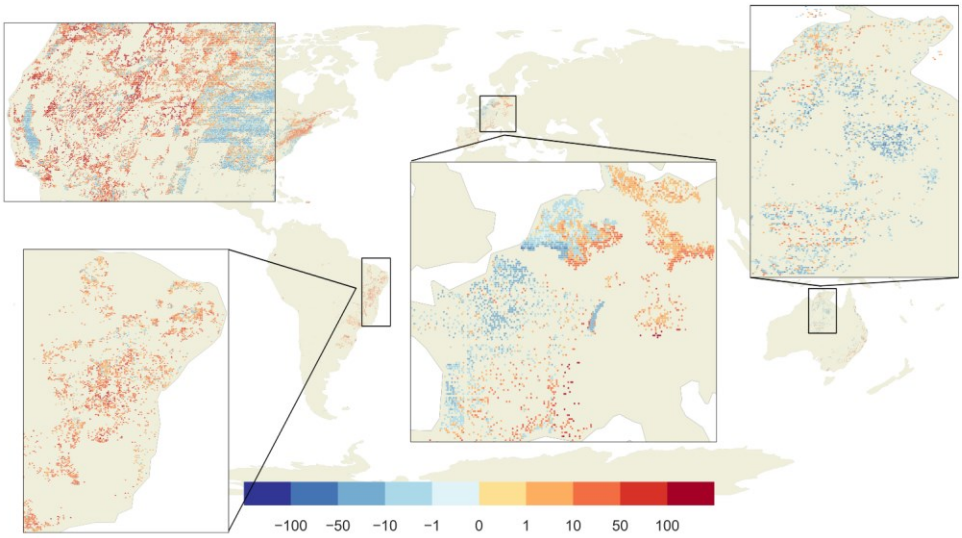

Figure 8 Differences between observed and simulated hydraulic head [***m***]. Red dots show areas where the model simulated deeper GW as observed, blue shallower GW. In the grey areas, no observations are available.

To compare performance of G^3M to the steady-state results of two high-resolution model of Fan et al. (2013) and ParFlow (Maxwell et al., 2015), heads in 30" (Fan et al., 2013) and 1 km (ParFlow) grid cells were averaged to the G^3M 5' grid cells. The comparison of 5' observations to the 5' average of ParFlow seem to be consistent with the 1 km model comparison in Maxwell et al. (2015)~~,~~ (their Fig. 5), even though over/under -estimates in the original resolution seemed to be smoothed out by averaging to 5' (not shown). The heads of Fan et al. (2013) fit better to observations than G^3M heads, with less underestimation (Fig. 9b). The comparison of G^3M heads to Fan et al. (2013) values for all 5' grid cells, which are also the initial heads of G^3M and the basis to compute river conductances, show that heads computed with the G^3M are mostly much lower except in regions with a shallow GW (Fig. 9c). This cannot be attributed to the 100 times lower spatial resolution per se but to the selection of the 30th percentile of the 30" as the SW drainage level. Outliers in the upper half of the scatter plot, with much larger heads than the initial values, are mainly occurring in steep mountain areas like the Himalayas where the 5' model is not representing smaller valleys with a lower head.

For the continental US, the computationally expensive 1-km integrated hydrological model ParFlow (Maxwell et al., 2015), fits much better to observations than G^3M (Figs. 9d, e). G^3M produces a generally lower water table (Fig. 9f), a main reason being that ParFlow assumes an impermeable bedrock at a depth of more than 100 m below the land surface elevation. Plotting hydraulic head instead of depth to GW has the disadvantage that the goodness of fit is dominated by the topography as the observed heads are calculated based on the surface elevation of the model. Even though hydraulic heads are a direct result of the model and are forcing lateral GW flows, depth to GW is more relevant for processes like capillary rise. For G^3M, there is almost no correlation between depth to GW observations and simulated values. To our knowledge, no publication on large-scale GW modeling presents correlations of simulated with observed depth to GW.

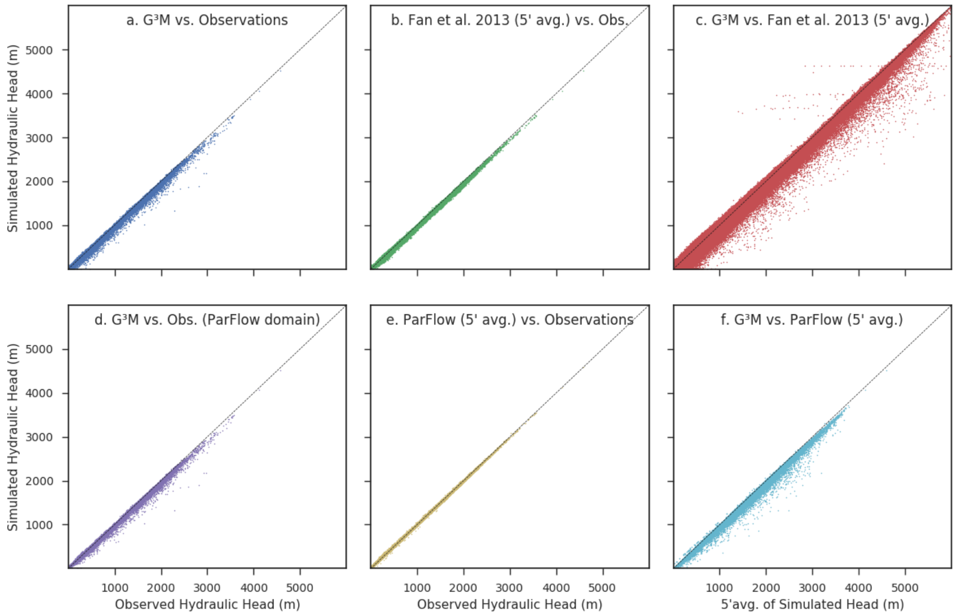

Figure 9 Scatterplots of simulated vs. observed hydraulic head and inter-model comparison of heads. (Upper panel) The steady-state run of G³M vs. observations (a), the 5' average of the equilibrium head of Fan et al. (2013) vs. observations (b) and the avg. equilibrium vs. G³M (c). (Lower panel) The steady-state run of G³M vs. observations only for the ParFlow domain (d), the 5' average of the ParFlow average annual GW table (Maxwell et al., 2015) vs. observations (e) and the steady-state run of G³M vs. 5' average of the ParFlow average annual GW table (f).

3.6 Case study Central Valley

To evaluate G³M further, its results were analysed for to a well-studied region, the Central Valley in California, USA. The Central Valley is one of the most productive agricultural regions of the world and heavily relies on GW pumpage to meet irrigation demands (Faunt et al., 2016). GW pumping in the valley increased rapidly in the 1960s (Faunt, 2009). Figure 10a shows simulated depth to GW for the Central Valley, the coast and the neighboring Sierra Nevada mountainside as well as parts of the Great Basin. The depth to GW table represents natural conditions without any pumping and is rather small. It roughly resembles the depth to GW assumed in the Central Valley Hydrological Model (CVHM) as initial condition, representing a natural state (Faunt, 2009) (Fig. 10b). G³M correctly computes the shallow conditions with groundwater above the surface in the north, partially in the south of the valley and decreasing towards the Sierra Nevada. The difference in the extend of flooded area could be due to large wetlands areas still present in the early 60s which are not represented in this extent in the data used by G³M. Beyond the CVHM domain, depth to GW in mountainous regions is probably overestimated by G³M. The elevation of neighboring cells may differ up to a 1000 meter resulting in a large gradient (Fig. S6b and S6e).

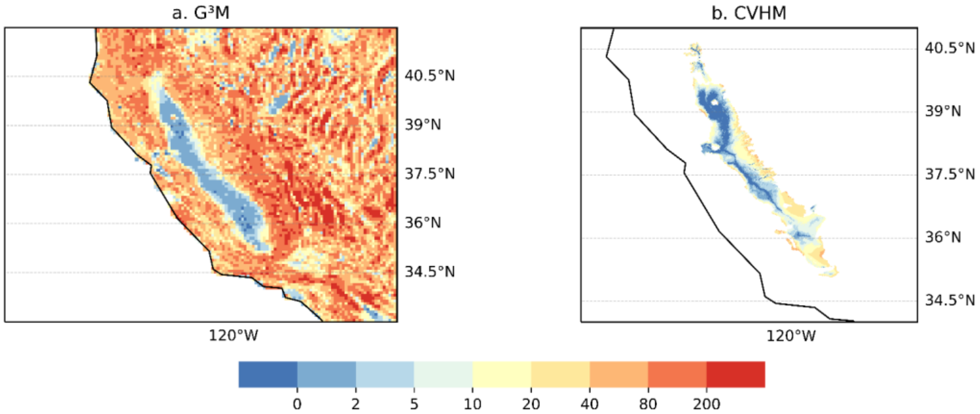

Figure 10 Plots of depth to GW [***m***] as calculated by G³M for the Central Valley and the Great Basin (a), and as used by CVHM as the natural state and starting condition (Faunt, 2009) (b).

4 Discussion

The main aim of global gradient-based groundwater flow modelling with G³M is to better simulate water exchange between SW and GW, for example for an improved estimation of GW resources in dry regions of the globe that are augmented by focused recharge from SW bodies. A major challenge for simulating GW-SW interactions (but also capillary rise) at the global scale is the large size of grid cells that is required due to computational constraints. Within the 5' grid cells, land surface elevation at the scale of 30'' very often varies by more than 20 m, and often by 200 m and more (Fig. S1), while the vertical position of the cell and the hydraulic head are approximated in the model by just one value. The question is whether head-dependent flows between grid cells, between GW and SW and from GW to soil (capillary rise) can be simulated successfully at the global scale, i.e. whether an improved quantification of these flows as compared to the simple linear reservoir model currently used in most GHMs can be achieved by this approach. This question cannot be answered yet as we have not yet achieved a dynamic coupling of G³M with a global hydrological model but one may speculate that some innovative approach to take into account the elevation variations within the grid cells may be needed.

It is difficult to assess performance of the presented steady-state G³M results. Model performance is assessment is hindered by data availability and the coarse model resolution. (1) To our knowledge the data collection of depth to groundwater by Fan et al. (2013) is unique. However, they do not represent steady-state values. Apart from depth to groundwater observations, hardly any relevant data is available at the global scale. Especially exchange between surface water and groundwater is difficult to measure even at the local scale. Therefore, we compared G³M results with the results from other large-scale models. Comparison to the results of catchment-scale groundwater flow models is planned for transient runs that will be possible after integration into WaterGAP. (2) Scale differences make the comparison to point observations of depth to groundwater difficult. Multiple local observations within a 5' cell may strongly vary, possibly~~maybe~~ just due to land surface elevation variations within the approximately 80 km^2 large cells (compare Fig. S1 and S8). Often, observations are biased towards alluvial aquifers in valleys. The calculated hydraulic head of the grid cell may represent the average groundwater level per grid cell correctly but can be still far off the local observations of depth to groundwater. As the current model only presents an uncalibrated natural steady-state, a comparison to observations only provides a first indicator where the model and the performance measurements needs to be improved as w~~m~~e move to a fully transient model.

The presented comparison to other large-scale models is based on the assumption that same model deficiencies e.g. in available data and scale issues can uncover differences in model decision. A comparison to catchment scale models is challenging as scales can differ by multiple magnitudes. As the model is further developed towards a transient model the presented comparison to simulations in data-rich regions need to be extended and temporal changes in interactions with surface water investigated.

The comparison to the initial state (based on historical observations) of the CVHM model presents a first comparison within a data-rich region which provides also the future possibility of comparing transient model results and human impact on a regional scale. G³M is able to reproduce the shallow groundwater table in the early 1960s. Differences are likely due to the steady-state approach and the connected assumptions on surface water bodies.

The presented development of the uncoupled steady-state global GW flow model enabled us to better understand how the spatial hydraulic head pattern relates to the fundamental drivers topography, climate and geology (Fan et al., 2007) and how the interaction to SW bodies govern the global head distribution. Simulated depth to groundwater is particularly affected by the assumed hydraulic head in SW bodies, the major GW drainage component in the model. As rivers represent a natural occurring drainage at the lowest point in a given topography, one would assume that the minimum elevation 30'' land surface elevation per 5' grid cell is a reasonable choice. Experiments have shown that this will induce a head distribution well below the average 5' elevation that is much below observations of Fan et al. (2013). We also tested setting h_{swb} to the average elevation of all "blue" cells (with a depth to GW of less than 0.25 m) of the steady-state 30'' water table results of Fan et al. (2013) that indicate the locations were GW discharges to the surface. This leads to an overall underestimation of the observed hydraulic heads (Fig. S9). Furthermore, it leads to an increase in losing SW bodies (comp. Fig. S11 with Fig. 4). However, it is difficult to judge whether this improves the simulation. More stretches of the Nile and its adjacent wetlands and also of the Niger wetlands and rivers in Northeastern Brazil are losing in case of lower h_{swb}, which appears to reasonable. Additionally, choosing the average as SW elevation provides on the on hand a better fit to observations (Fig. S9) but leads to a wold wide flooding with largely overestimated heads (Fig. S10) and a much longer convergence time due to an increased oscillation between gaining and losing conditions.

The problem is very likely one of scale. This is supported by the fact that both high-resolution models, even the simple one of Fan et al. (2013) fit better to observations than the low-resolution model G³M (Fig. 9). In case of high resolution (e.g. 30''), there are a number of grid cells at an elevation above the average 5' land surface elevation, leading to higher hydraulic heads in parts of the 5' area that drain towards the SW body in a lower 30'' grid cell. In case of the low spatial resolution of 5' in which h_{swb} is set to the elevation of the fine-resolution drainage cell, the 5' hydraulic head is rather close to this (low) elevation (Fig. S12), resulting in an underestimation of hydraulic head and thus an overestimation of depth to GW. While it is plausible and necessary to assume that there is SW-GW interaction within each of in the approximately 80 km^2, this is not the case for the two orders of magnitude smaller 30'' grid cells. Thus, with the high resolution, heads are not strongly controlled everywhere by the head in SW bodies. Selecting the 30th percentile of the 30'' land surface elevation as h_{swb} was found, by trial-and-error, to lead to a hydraulic head distribution that fits reasonably well to observed head. It avoids that the simulated GW table drops to low while avoiding the excessive flooding that occurs if h_{swb} is set to the average of 30'' land surface elevations, i.e. the 5' land surface elevation (Fig. S9).

The constraint that the selected h_{swb} value puts on simulated hydraulic heads is also linked to the conductance of the SW bodies. A higher conductance will lead to aquifer heads closer to h_{swb}. If the hydraulic head drops below the bottom level of the SW body, the hydraulic gradient is assumed to become 1 and the SW body recharges the GW with a rate of K_{aq} per unit SW body area. In case of a K_{aq} value of 10^{-5} m s^{-1}, the SW body would lose approximately 1 m of water each day. It is to be investigated how the sensitivity to choice of SW body elevation and conductance leads to a solution that fits observations best. A lower conductance may lead to a higher groundwater table as SW bodies don't drain as much water; on the other hand, they seem to provide an important recharge mechanism in the steady-state model for some regions preventing an even higher depth to GW. The simple conductance approach applied in G³M could possibly be im~~ap~~proved by the approach proposed by Morel-Seytoux et al. (2017).

de Graaf et al. (2015) set their SW head (h_{swb}) to the land surface elevation of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to avoid excessive flooding, and that is not needed in G^{3}M. On the other hand, this adaption allows the drainage of water even if the hydraulic head is below the SW elevation that might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity of SW body elevation. A summary of model differences are shown in Table 2.

Kommentiert [RR5]: Editor comment #4

As described above, G³M differs from regional groundwater models due to grid cell size in that it is more conceptual and cannot capture actual variability of topography, aquifer depth (Richey et al., 2015) and (vertical) heterogeneity of subsurface properties. The lack of information about the three-dimensional distribution of hydraulic conductivity is expected to negatively impact the quality of simulated GW flow. For example, the lateral conductivity and connectivity of groundwater along thousands of kms from e.g. the Rocky Mountains in the Central USA to the coast as well as the vertical connectivity is likely to be overestimated by G^{3}M, as vertical faults and interspersed aquitards are not represented; this leads to an underestimation of hydraulic head in those mountainous areas.

5 Conclusions

We have presented the concept and first results of a new global gradient-based 5' GW flow model that is to be coupled to the 0.5° GHM WaterGAP. The uncoupled steady-state model has provided important insights into challenges of global GW flow modelling mainly related to the necessarily large grid cells size (5' by 5') as well as first global maps of SW-GW interactions. Simulated heads were found to be strongly impacted by assumption regarding the interaction with SW bodies, in particular the selected elevation of the SW table and the prescribed conductance. We have demonstrated that simulated G^{3}M hydraulic heads fit better to observed heads than the heads of the comparable steady-state GW model of de Graaf et al. (2015), without requiring additional drainage that would prevent a full coupling to a GHM. Furthermore, we provided insights into how the choice of surface water body elevation affects model outcome and plan to further investigate how we can use higher resolution topographic data to overcome these challenges by comparing simulation results of a 5', 30'', and 3'' GW model of New Zealand.

Formatiert: Einzug: Erste Zeile: 0 cm

Kommentiert [RR6]: Editor comment #3

~~We have demonstrated that simulated G^{3}M hydraulic heads fit better to observed heads than the heads of the comparable steady-state GW model of de Graaf et al. (2015),without requiring additional drainage that would prevent a full coupling to a GHM.~~

The presented results are the first step towards a fully coupled model in which SW heads are computed as a function of surface water hydrology and GW abstractions can be taken into account. Especially the interaction with SW bodies that can run dry will make the model behavior more realistic. The fully coupled model will simulate transient behaviour reflecting climate variability and change. Simulated hydraulic head dynamics will be compared to observed head time series as well as to the output of large-scale regional models, while total water storage variations will be compared to GRACE satellite data. However, it will be challenging to judge the quality of simulated GW-SW interactions due to a scarcity of observations.

6 Code and data availability

The model-framework code is available at globalgroundwatermodel.org or at DOI: 10.5281/zenodo.1175540 with a description on how to compile and run a basic GW model. The code is available under the GNU General Public License 3. Model output is available at DOI: 10.5281/zenodo.1315471.

Acknowledgments

Part of this study was funded by a Friedrich-Ebert foundation Ph.D. fellowship. We are very grateful to Ying Fan and Gonzalo Miguez-Macho for fruitful discussions and data provisioning. We thank the two reviewers for their thoughtful comments that helped to improve the manuscript.

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrological Sciences Journal, 48, 317–337, doi:10.1623/hysj.48.3.317.45290, 2003.

Allen, P. M., Arnold, J. C., and Byars, B. W.: Downstream channel geometry for use in planning-level models, J Am Water Resources Assoc, 30, 663–671, doi:10.1111/j.1752-1688.1994.tb03321.x, 1994.

Belcher, W. R. and Sweetkind, D. S.: Death Valley regional groundwater flow system, Nevada and California-Hydrogeologic framework and transient groundwater flow model, U.S. Geological Survey Professional Paper, 398, 2010.

Belcher, W.R. (E.): Death Valley Regional Ground-water Flow System, Nevada and California-Hydrogeologic Framework and Transient Ground-water Flow Model, U.S. Geological Survey Professional Paper, pp. 408, 2004.

Costa, A. C., Foerster, S., de Araújo, J. C., and Bronstert, A.: Analysis of channel transmission losses in a dryland river reach in north-eastern Brazil using streamflow series, groundwater level series and multi-temporal satellite data, Hydrol. Process., 27, 1046–1060, doi:10.1002/hyp.9243, 2013.

de Graaf, I. E.M. de, Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, Hydrol. Earth Syst. Sci., 19, 823–837, doi:10.5194/hess-19-823-2015, 2015.

de Graaf, I. E.M. de, van Beek, R. L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.P.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, Advances in Water Resources, 102, 53–67, doi:10.1016/j.advwatres.2017.01.011, 2017.

Dogrul, E., Brush, C., and Kadir, T.: Groundwater Modeling in Support of Water Resources Management and Planning under Complex Climate, Regulatory, and Economic Stresses, Water, 8, 592, doi:10.3390/w8120592, 2016.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12, 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B.R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, Journal of Geodynamics, 59-60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, Journal of Hydrology, 270, 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, doi:10.1002/2014WR015595, 2014.

Eisner, S.: Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients, University of Kassel, Kassel, Germany, 2016.

Elsawwaf, M., Feyen, J., Batelaan, O., and Bakr, M.: Groundwater-surface water interaction in Lake Nasser, Southern Egypt, Hydrol. Process., 28, 414–430, doi:10.1002/hyp.9563, 2014.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science (New York, N.Y.), 339, 940–943, doi:10.1126/science.1229881, 2013.

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, J. Geophys. Res., 112, doi:10.1029/2006JD008111, 2007.

FAO: Irrigation potential in Africa: http://www.fao.org/docrep/w4347e/w4347e0i.htm, 1997, last access: 1 May 2018.

Faunt, C. C. (Ed.): Groundwater Availability of the Central Valley Aquifer, California, U.S. Geological Survey Professional Paper, 1776, 225 pp., 2009.

Faunt, C. C., Provost, A. M., Hill, M. C., and Belcher, W. R.: Comment on “An unconfined groundwater model of the Death Valley Regional Flow System and a comparison to its confined predecessor” by R.W.H. Carroll, G.M. Pohll and R.L. Hershey [Journal of Hydrology 373/3–4, pp. 316–328], Journal of Hydrology, 397, 306–309, doi:10.1016/j.jhydrol.2010.11.038, 2011.

Faunt, C. C., Sneed, M., Traum, J., and Brandt, J. T.: Water availability and land subsidence in the Central Valley, California, USA, Hydrogeol J, 24, 675–684, doi:10.1007/s10040-015-1339-x, 2016.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nature Geosci, 9, 161–167, doi:10.1038/ngeo2590, 2016.

Gleeson, T., Moosdorf, N., Hartmann, J., and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898, doi:10.1002/2014GL059856, 2014.

Gleeson, T., Wada, Y., Bierkens, M. F. P., and van Beek, L. P. H.: Water balance of global aquifers revealed by groundwater footprint, Nature, 488, 197–200, doi:10.1038/nature11295, 2012.

Gossel, W., Ebraheem, A. M., and Wycisk, P.: A very large scale GIS-based groundwater flow model for the Nubian sandstone aquifer in Eastern Sahara (Egypt, northern Sudan and eastern Libya), Hydrogeology Journal, 12, 698–713, doi:10.1007/s10040-004-0379-4, 2004.

Harbaugh, A. W.: MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process, US Department of the Interior, US Geological Survey Reston, 2005.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophysics, Geosystems, 13, doi:10.1029/2012GC004370, 2012.

Konikow, L. F.: Contribution of global groundwater depletion since 1900 to sea-level rise, Geophys. Res. Lett, 38, doi:10.1029/2011GL048604, 2011.

Krakauer, N. Y., Li, H., and Fan, Y.: Groundwater flow across spatial scales: importance for climate modeling, Environ. Res. Lett., 9, 34003, doi:10.1088/1748-9326/9/3/034003, 2014.

Lange, S.: EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), 2016.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, 296, 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Maxwell, R. M., Condon, L. E., and Kollet, S. J.: 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, 923–937, doi:10.5194/gmd-8-923-2015, 2015.

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, J. Geophys. Res., 112, D13108, doi:10.1029/2006JD008112, 2007.

Morel-Seytoux, H. J., Miller, C. D., Miracapillo, C., and Mehl, S.: River Seepage Conductance in Large-Scale Regional Studies, Ground Water, 55, 399–407, doi:10.1111/gwat.12491, 2017.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Naff, Richard L., and Edward R. Banta: The US Geological Survey modular ground-water model-PCGN: a preconditioned conjugate gradient solver with improved nonlinear control, Geological Survey (US), 1331, 2008.

Niswonger, R. G., Panday, S., and Ibaraki, M.: MODFLOW-NWT, a Newton formulation for MODFLOW-2005, US Geological Survey Techniques and Methods, 6, 2011.

Reinecke, R.: G³M-f a global gradient-based groundwater modelling framwork, JOSS, 3, 548, doi:10.21105/joss.00548, 2018.

Richey, A. S., Thomas, B. F., Lo, M.-H., Famiglietti, J. S., Swenson, S., and Rodell, M.: Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework, Water Resour. Res, 51, 5198–5216, doi:10.1002/2015WR017351, 2015.

Saad, Y.: ILUT: A dual threshold incomplete LU factorization, Numer. Linear Algebra Appl., 1, 387–402, doi:10.1002/nla.1680010405, 1994.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, PNAS, 109, 9320–9325, doi:10.1073/pnas.1200311109, 2012.

Sheets, R. A., Hill, M. C., Haitjema, H. M., Provost, A. M., and Masterson, J. P.: Simulation of water-table aquifers using specified saturated thickness, Ground Water, 53, 151–157, doi:10.1111/gwat.12164, 2015.

Stonestrom, D. A., Constantz, J., Ferre, T. P.A., and Leake, S. A.: Ground-water recharge in the arid and semiarid southwestern United States, U.S. Geological Survey Professional Paper 1703, 414 p, 2007.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin, Hydrol. Earth Syst. Sci., 15, 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I., and Treidel, H.: Ground water and climate change, Nature Climate change, 3, 322–329, doi:10.1038/nclimate1744, 2012.

UNESCO-IHP, IGRAC, WWAP: GEF-TWAP Methodology Transboundary Aquifers.: http://isarm.org/sites/default/files/resources/files/TWAP%20Methodology%20Groundwater%20Component%20%28Revised%20Aug%202012%29.pdf, 2012, last access: 9 October 2017.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, doi:10.1029/2010WR009791, 2011.

Vergnes, J.-P., Decharme, B., Alkama, R., Martin, E., Habets, F., and Douville, H.: A Simple Groundwater Scheme for Hydrological and Climate Applications: Description and Offline Evaluation over France, J. Hydrometeor, 13, 1149–1171, doi:10.1175/JHM-D-11-0149.1, 2012.

Vergnes, J.-P., Decharme, B., and Habets, F.: Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model, J. Geophys. Res. Atmos., 119, 11, doi:10.1002/2014JD021573, 2014.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F., and Alcamo, J.: Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe, Journal of Hydrology, 424-425, 238–251, doi:10.1016/j.jhydrol.2012.01.005, 2012.

Wada, Y.: Modeling Groundwater Depletion at Regional and Global Scales: Present State and Future Prospects, Surv Geophys, 37, 419–451, doi:10.1007/s10712-015-9347-x, 2016.

Wada, Y., van Beek, L. P. H., and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: A global assessment, Water Resour. Res., 48, doi:10.1029/2011WR010562, 2012.

Table 1 Model parameter values, input data sources and other information about the steady-state simulation.

Parameter	Symbol	Units	Description	Eq. No.
Landmask	-	-	Location and area of 2161074 cells at 5' resolution based on WaterGAP (Eisner, 2016))	-
GW recharge	R_g	L^3T^{-1}	Mean annual diffuse GW recharge 1901–2013 of WaterGAP 2.2c (Müller Schmied et al., 2014) forced with EWEMBI (Lange, 2016), spatial resolution 0.5° (Fig. S4)	1,3,7
Hydraulic conductivity	K_{aq}	LT^{-1}	Derived from Gleeson et al., 2014 (Fig. S3)	2,5
Hydraulic head	$h_{(aq)}$	L	Head of the aquifer in a computational cell, initial estimate based on 5' average of 30" head of Fan et al. (2013)	2,4,7,8
Ocean boundary conductivity	c_{ocean}	L^2T^{-1}	$10\ m^2\ day^{-1} = 0.1\ m\ day^{-1}\ 10\ km\ 10\ km^{-1}\ 100\ m$, with K of $10^{-6}\ m\ s^{-1}$ and a distance of 10 km from the cell center to the boundary with a cell thickness of 100 m	3,4
Ocean boundary head	h_{ocean}	L	Global mean sea-level of 0 m	4
SW head	h_{swb}	L	30 % quantile (P_{30}) of 30" land surface elevation of Fan et al. (2013) per 5' grid cell	5
SW bottom elevation	B_{swb}	L	2 m (wetlands), 10 m (local lakes), 100 m (global lakes) below P_{30}	5
Area of global and lcal lakes and global and local wetlands	WL	L^2	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016),	6
Length of the river	L	L	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016)	6
Width of the river	W	L	Per 5' grid cell, based d on WaterGAP 3 (Eisner, 2016)	6
River head	h_{riv}	L	h_{swb}	6,7,8
River bottom elevation	B_{riv}	L	$h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994)	6
Equilibrium hydraulic head	h_{eq}	L	Steady-state hydraulic head of Fan et al. (2013) (averaged to 5' from original spatial resolution of 30")	7
Layers	-	-	2 confined, 100 m thick each	-
Land surface elevation	-	L	5' average of 30" digital elevation map of Fan et al. (2013) (Fig. S2)	-
E-folding factor	-	-	Applied only to lower layer for 150 m depth, based on area-weighted average of Fan et al. (2013)	-
Timestep	t	T	Daily timestep	-
Convergence criterion	-	L	$\|\|$hydraulic head residuals$\|\|_{inf} < 10^{-100}$ and max head change < 10 m	-
Inner iterations	-	-	50 inner iterations between Picard iterations (Naff, Richard L., and Edward R. Banta, 2008)	-

Table 2 Comparison of global- and continental-scale groundwater models

Kommentiert [RR7]: Editor comment #4

Aspect	**G³M**	**de Graaf et al. (2015; 2017)**	**Fan et al. (2013)**	**ParFlow**
Extent	Global	Global	Global	Continental USA
Resolution	5'	6'	30"	1 km
Software	G³M-f	MODFLOW	Unnamed	ParFlow
Computational expense	Medium	Medium	High	Very high
Equation	3d Darcy	3d Darcy	2d Darcy	3d Richards
Time scale	Steady-state/(transient)	Steady-state/transient	Steady-state	Steady-state
Vertical layers	2	2	1	5
Full coupling possible	Yes	No (Conceptual issue)	No	Yes (already coupled)
In-memory coupling	Yes	No	N/A	Yes
Constant saturated thickness	Yes	Yes	No	No
Impermeable bottom	No	No	No	Yes
Surface water body location	In every cell	In almost every cell	No surface water	Created during simulation
Surface water body elevation	P_{30} of 30" DEM	Avg. of 30" DEM	N/A (outflow if depth to GW < 0.25 m)	N/A
Deviation from observations	Large	Very large	Medium	Medium

Supplement

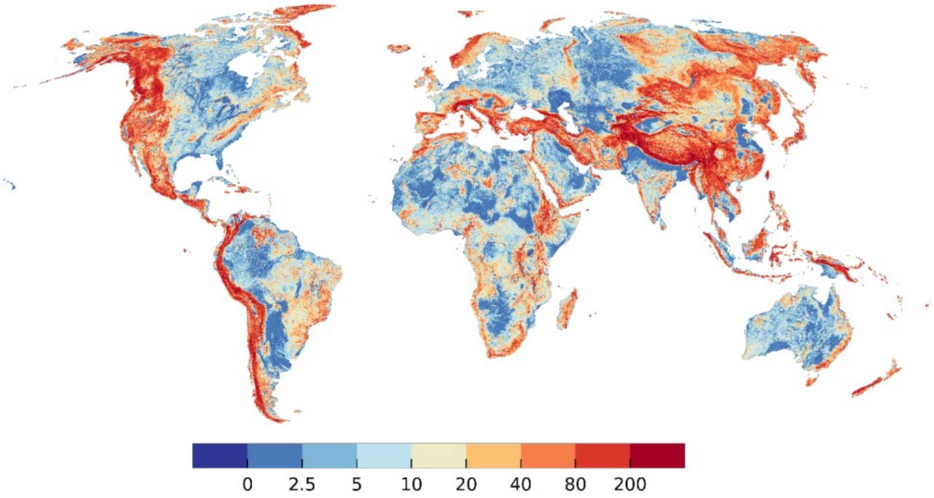

Figure S1 Difference [$\boldsymbol{m}$] between mean elevation and P_{30} elevation. Maximum value 1365 m**.**

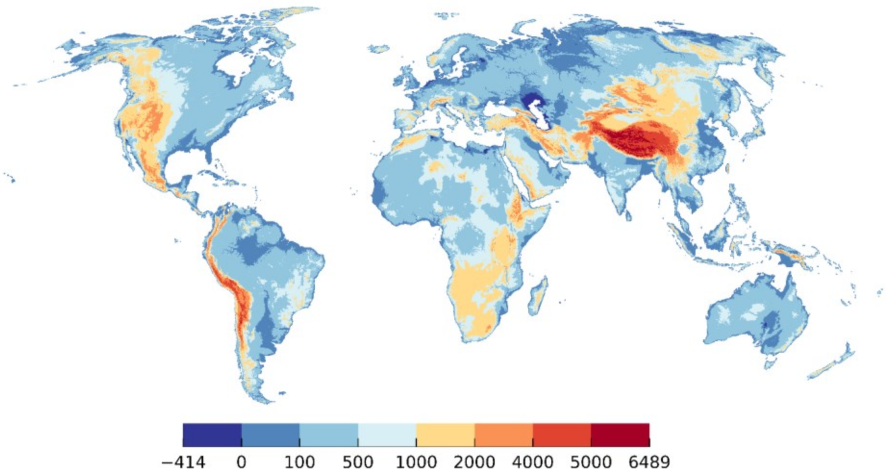

Figure S2 Land surface elevation [$\boldsymbol{m}$] used in G^3M: 5' average of 30" land surface elevation used in Fan et al. (2013).

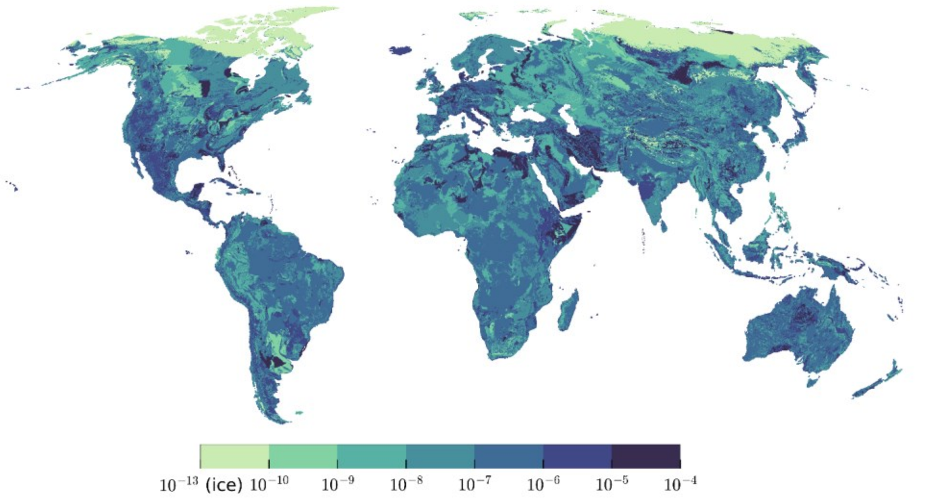

Figure S3 Hydraulic conductivity [$\boldsymbol{ms^{-1}}$] derived from Gleeson et al. (2014) by scaling it with the geometric mean to 5'. Very low values in the northern hemisphere are due to permafrost conditions.

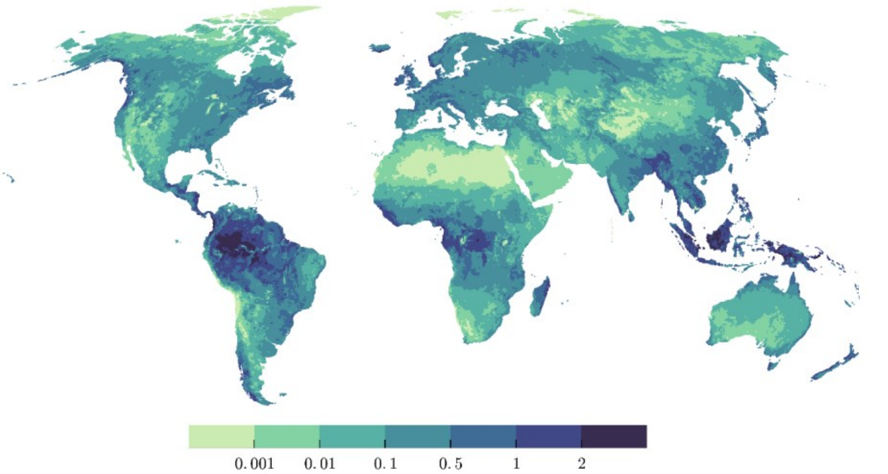

Figure S4 Mean annual groundwater recharge [$\boldsymbol{mm\ day^{-1}}$] between 1901-2013, from WaterGAP 2.2c.

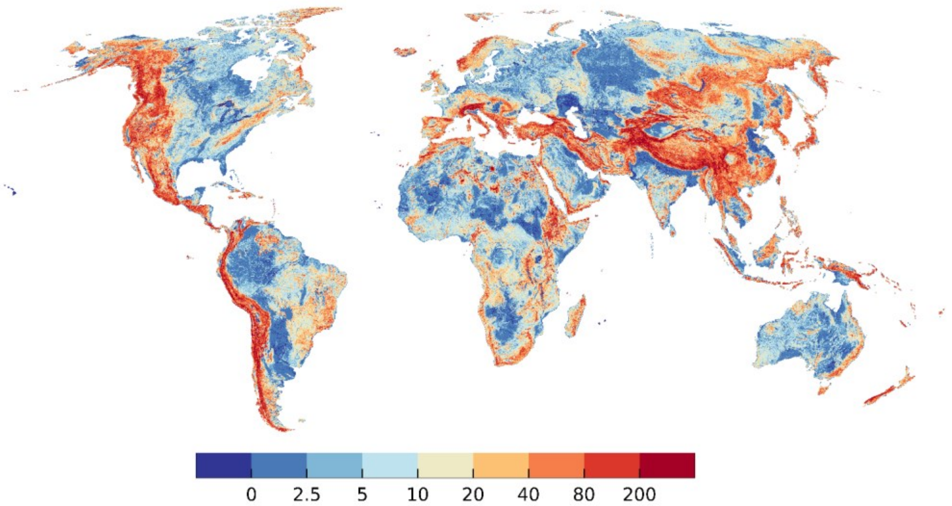

Figure S5 Arithmetic mean [***m***] of the 30'' land surface elevation per 5' grid cell and simulated equilibrium hydraulic head (simulated depth to GW). Maximum value 2070 m, minimum value -414 m (Extremes included in dark blue and dark red).

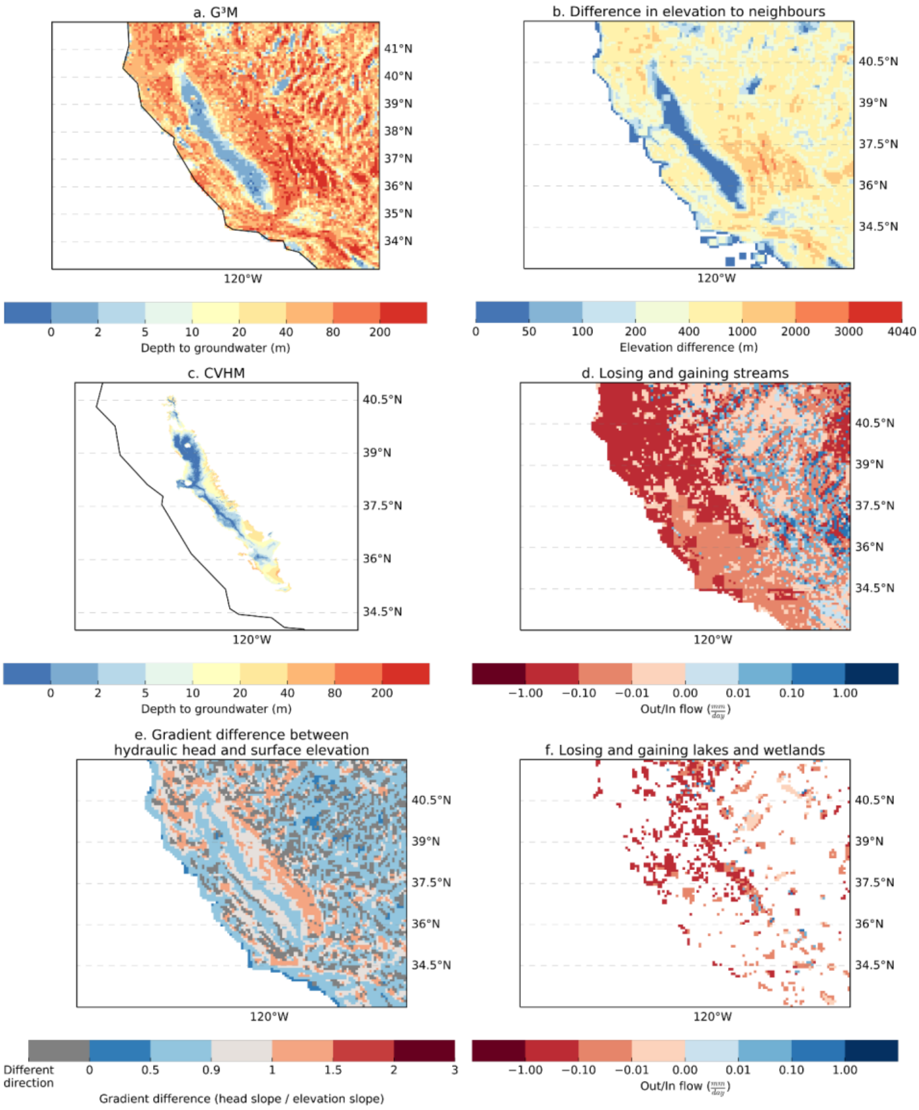

Figure S6 Plots of depth to GW as calculated by G³M (a), difference in surface elevation to neighbouring cells (b), depth to GW as used by the CVHM as the natural state and starting condition (Faunt, 2009) (c), losing and gaining streams as calculated by G³M (d), difference in gradient of hydraulic head and surface elevation (e), losing and gaining lakes and wetlands as calculated by G³M for the Central Valley and the Great Basin.

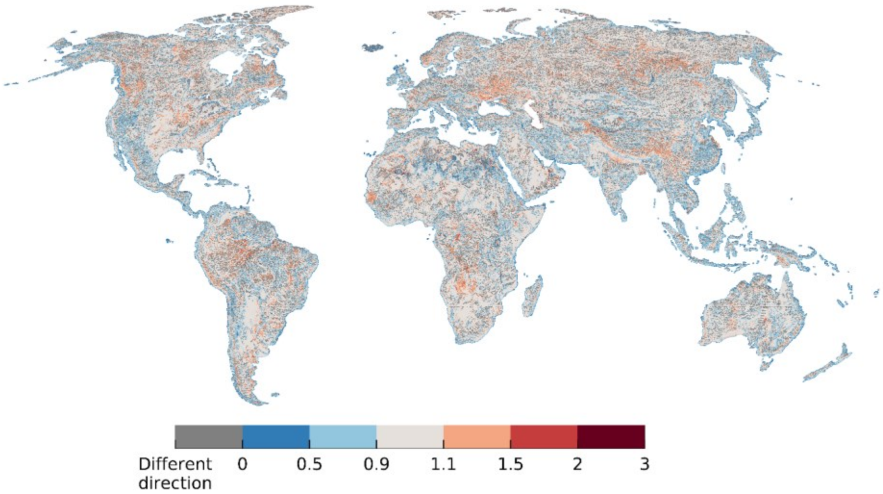

Figure S7 Ratio of hydraulic head gradient to 5' mean surface elevation gradient, only computed if the difference in direction of the gradient was smaller than 45°.

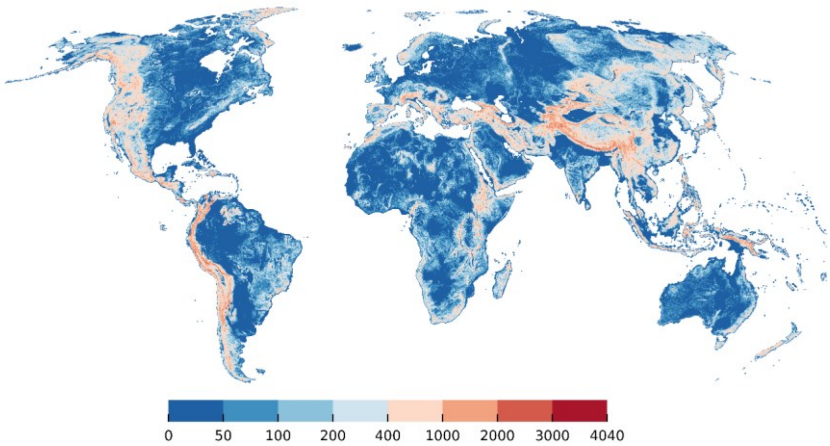

Figure S8 Land surface elevation Difference of 30'' mean land surface elevation in 5' grid cell to mean elevation of neighbouring cells [**m**] to mean elevation of neighboring cells on 5' resolution.

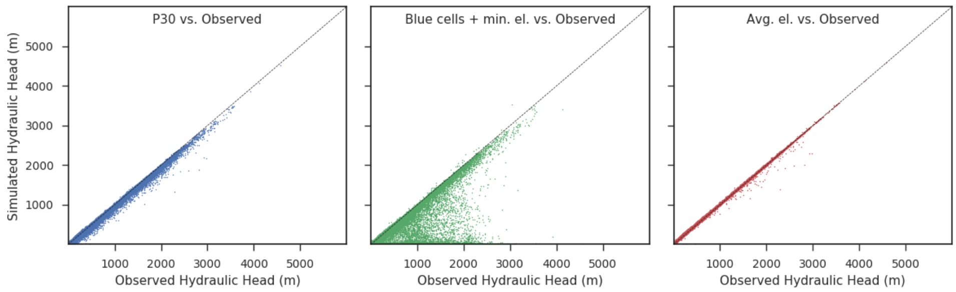

Figure S9 Comparison between three alternatives for setting h_{swb}. Left to right: Fit of simulated hydraulic heads observations if h_{swb} is set (1) to the 30[th] percentile of the 30'' land surface elevations (standard model) , (2) alternatively to the average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) or (3) is set to the average of the 30'' land surface elevations. A blue cell has a depth to GW of less than 0.25 m and indicates GW discharge to the surface. If no "blue" cell exists in the 5' cell, the minimum elevation of the 30'' land surface elevation values within the cell was used.

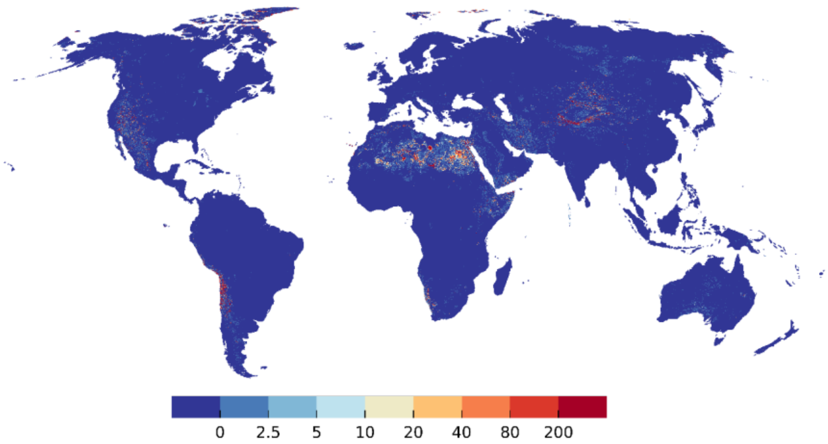

Fig. S10 Depth to groundwater [m] for SW body elevation at average of 30'' land surface elevations.

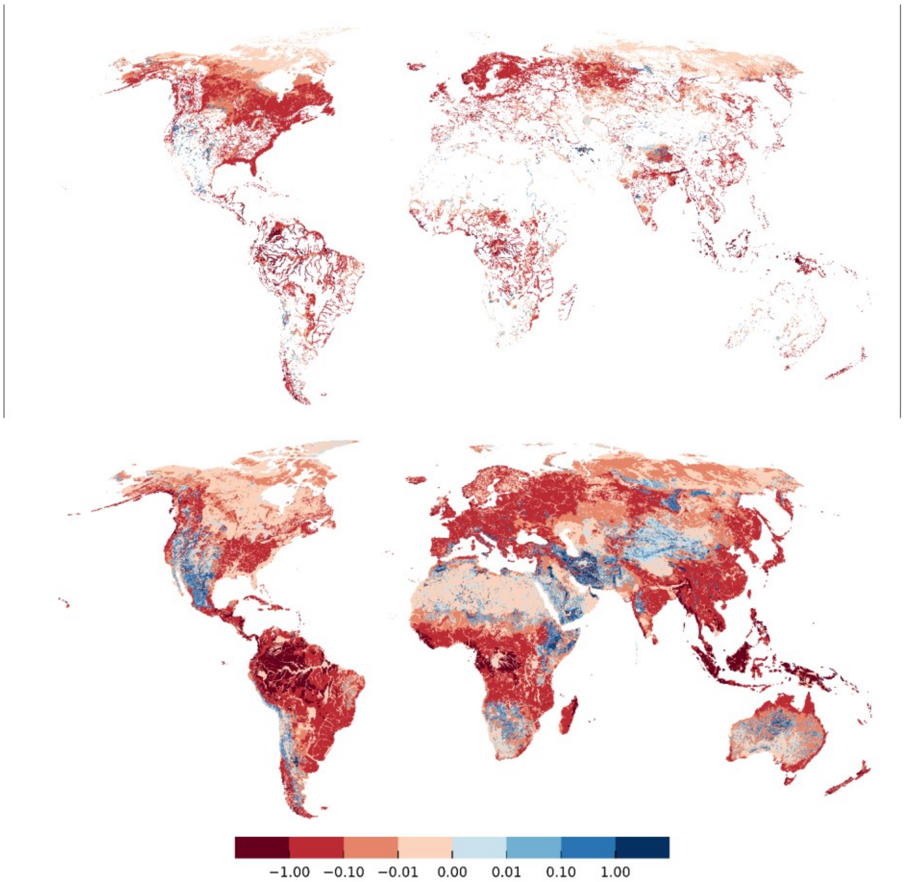

Figure S11 Gaining and losing rivers (lower panel) and wetlands and lakes (upper panel) as flow into/out the GW [$\boldsymbol{mm\ day^{-1}}$] if $\boldsymbol{h_{swb}}$ is set to average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) (right). A blue cell is defined as a depth to groundwater of less than 0.25 m. If no "blue" cell exist in the 5' cell, the minimum elevation of the 30'' land surface elevation values is used. Red denotes gaining SW bodies.

Author comment

We thank both reviewers for the thoughtful comments and questions. They helped us in particular to improve our explanation of the conceptual approach to gradient-based groundwater modeling that is necessary for global-scale groundwater modeling with a coarse spatial resolution, including the choice of simulating unconfined conditions and the conceptual difficulties in defining depth to groundwater table.

Our answers to the referees' comments are written in italics.

Referee #1

#1.1

A fundamental problem with these models is that they are difficult to verify, but this is not at all reflected in the discussion of the results.

We have added a new paragraph to section 4 Discussion.

Changes to manuscript

We added the following paragraph to section 4 Discussion (second paragraph)

"It is difficult to assess performance of the presented steady-state G^3M results. Model performance assessment is hindered by data availability and the coarse model resolution. (1) To our knowledge the data collection of depth to groundwater by Fan. et al (2013) is unique. However, they do not represent steady-state values. Apart from depth to groundwater observations, hardly any relevant data is available at the global scale. Especially exchange between surface water and groundwater is difficult to measure even at the local scale. Therefore, we compared G^3M results with the results from other large-scale models. Comparison to the results of catchment-scale groundwater flow models is planned for transient runs that will be possible after the integration into WaterGAP. (2) Scale differences make the comparison to point observations of depth to groundwater difficult. Multiple local observations within a 5' cell may strongly vary, maybe just due to land surface elevation variations within the approximately 80 km^2 large cells (compare Fig. S1 and S8). Often, observations are biased towards alluvial aquifers in valleys. The calculated hydraulic head of the grid cell may represent the average groundwater level per grid cell correctly but can be still far off the local observations of depth to groundwater. As the current model only presents an uncalibrated natural steady-state, a comparison to observations only provides a first indicator where the model and the performance measurements needs to be improved as we move to a fully transient model."

#1.2

The authors state on line 24 page 1 that these models are useful in areas with little or no data, as they allow to generate robust information. How can anything robust be generated (and how do we know its true?) in the absence of data. The hydrogeological literature is full of examples where even in the data -rich regions different models produce different outcomes.

In the revised version, we have deleted the statement about robust information, and explain now in more detail (on page 2 lines 24 ff) the purpose of our research effort, i.e. global-scale gradient-based groundwater flow modelling.

Changes to manuscript

Page 3, Line 24 ff now reads:

"Our model development approach was to learn from existing large-scale regional models (Faunt, 2009; Vergnes et al., 2014, Maxwell et al., 2015; Dogrul et al., 2016) to gain insights into how the coarse spatial resolution, incomplete data, and conceptual model design effects model outcome. We want to find out whether we can use gradient-based groundwater modelling at the global scale, when later integrated into a global hydrological model, to improve estimation of flows between SW and GW (affecting both e.g. streamflow and groundwater recharge and thus water availability for humans and ecosystems) and capillary rise (affecting evapotranspiration)."

#1.3

We are presented with plots, numbers and graphs and some interpretation, but there is no credible discussion on the reliability of the result obtained. The only indication of model performance is that there is essentially no correlation between simulated and observed depth to groundwater. To me this means simply that the model cannot be used to make these types of predictions.

Comparison between depth to GW derived from simulated steady-state hydraulic head and point-scale observations of (non-steady state) depth to GW is not straightforward at all, such that clear conclusions about the model performance are difficult. The model performance assessment is hindered by two factors: data availability and scale.

(1) To our knowledge the data collection of depth to groundwater by Fan. 2013 is unique. We try to extend that picture with large scale regional models as base for our comparison. We do acknowledge that comparison to a model is not the same as a comparison to observations.

Apart from depth to groundwater observations hardly any relevant data is available. Especially exchange between surface water and groundwater is inherently hart to measure even at the local scale and thus often a calibration parameter in small scale models.

(2) Scale differences make the comparison to depth to groundwater observations difficult. (1) Multiple local observations within a 5 arcmin cell can vary by a large range (2) and they may have been observed at location very different from the average groundwater characteristic within a grid cell - often biased towards alluvial aquifers in valleys. The calculated hydraulic head of the grid cell may represent the average groundwater level per grid cell but can be still far of the local observation.

Furthermore, we observe that the depth to groundwater is highly influenced by the location of the surface water bodies (swb) and perception of depth to groundwater changes if calculated heads are compared to swb elevation and not to the average cell elevation.

Changes to manuscript

To respond to this comment, we added the following paragraph to section 4 Discussion (second paragraph) (refer to comment #1.1)

To show the conceptual difficulty of calculating “simulated” depth to GW from the simulated 5’ grid cell hydraulic head and an effective or mean land surface elevation at this scale, we revised section 3.1 and added, as Fig. 3b, a map showing the difference between P30 (the 30th percentile of the 30'' land surface elevations, the assumed elevation of the surface water body water table) and the computed hydraulic head.

#1.4

It is not useful to plot observed and simulated hydraulic heads over such large scales, even if its just for the sake of model comparison. It is true that other authors have also presented simulated vs observed hydraulic heads over such large scales, but this is simply misleading. Depth to groundwater is the variable that counts for calculating exchanges with surface, amongst many other processes. In this sense none of the available models on a global scale is ready yet. This must not necessarily be a problem, as long as the results are not oversold, as is unfortunately rather often the case.

Hydraulic head is the main model output which is a good reason for showing it as such. And while the simulated heads might not match the observed very well in terms of absolute quantities, there are insights to be gained by looking at trends. Even local-scale models often do not match heads very well, but can be useful to understand the system response (i.e., how/where do aquifer heads change with other changes in the system stresses (pumping increases, recharge decreases, stream flows change, etc.).

Depth to groundwater table is only derived from the model output using some estimate of a representative land surface elevation (see response to the above comment and the ensuing revisions of the manuscript). Calculated depth to groundwater highly depends on how a DEM is used to account for inter cell variability. On the other hand, this is also true for the derived head observations. Plots of simulated head vs. observed head are heavily influenced by the DEM signal and deviations due to difference in depth to water table are obfuscated due to the plot scales.

Furthermore, the interaction of surface water bodies and the groundwater is driven by the gradients between heads. We do agree, however, that simulation of capillary rise requires a good estimate of local-scale depth to GW. Currently the model outcomes are not suitable to perform such a calculation.

We already stated in the original manuscript (p.14 line 19-20) that there is almost no correlation between depth to GW observations and simulated values. So we are transparent about this and think that we do not “oversell” our results.

Changes to manuscript

none

#1.5

The formulation of the equation 2 is for a confined aquifer. The authors justify this conceptually wrong choice on line 20, page 6: “Flow equations are for confined aquifer because it reduces convergence time. “This is a very poor argument, purely based on convenience. To what extent the model should capture the relevant physics should cannot be a question on how difficult it is to solve equations. The goal of this modelling approach is to advance the interaction between the surface and the subsurface across very large scales. Given that the direct interaction with the surface always happens with unconfined aquifers the fundamental basis of the approach is flawed on the most

basic level. While for steady state simulations the term falls out of the equation it is still very concerning that that a model is developed with inadequate flow equations.

The paper presents a conceptual model that differs in many aspects to traditional regional GW models due to the required coarse spatial resolution. Using the flow equation for unconfined conditions, which is typically done for the upper layer of groundwater models (unless confined by aquitards) is done to represent that in case of the same hydraulic gradient, less water can be transported if the hydraulic head and thus the saturated thickness drops. When looking at depth of GW in Fig. (old)3, one may think that in particular in mountainous terrain, the 100 m thick upper layer of the aquifer has fallen (almost) dry and does therefore in reality transfer no more groundwater. However, as shown in section 3.1, the high depth to GW is mainly related to the large land surface elevation differences within the 5′ grid cell, in in almost all cells, the groundwater table is above the elevation of the water table in the surface water bodies (while land surface elevation per se is not part of the flow equation). Thus, given the high uncertainty of assumed hydraulic conductivity values and unknown actual aquifer depth, the assumption of fixed transmissivities seems to be appropriate for our global 5′ model. Using the equation of unconfined conditions cannot be expected to improve the simulations significantly.

Conceptually, at the applied coarse spatial resolution of the GW model, model layers should not be considered to be fixed to a land surface elevation. The model layers can be rather thought to be vertically (somewhat) aligned with the elevation of the surface water body table, and the flow equation rather governs the lateral and vertical fluxes over a thickness of 200 m.

Changes to manuscript

To clarify the difficult but important aspect of the relation between model layers and surface elevation in steep terrain at the spatial resolution of 5′, we revised Figure 1 and added to section 2.1:

"In addition, due to the coarse spatial scale and the possible large variations of land surface elevations within each grid cell, the upper model layers should not be considered to aligned with an average land surface elevation. The model layers can be rather thought to be vertically aligned with the elevation of the surface water body table, as this prescribed elevation is, together with the sea level, the only elevation included in the groundwater flow equation (Eq. 2)."

We added to the second paragraph of section 2.3:

"We choose to simulate confined flow conditions in both layers even though the upper layer can be expected to decrease in depth and thus in transmissivity (hydraulic conductivity times saturated depth). Every unconfined aquifer can have an equivalent confined representation assuming a correct saturated thickness (Sheets et al., 2015). However, given the large uncertainties regarding hydraulic conductivities (possibly an order of magnitude) and the lack of knowledge about aquifer thickness, it is appropriate to choose the computationally more efficient assumption of confined conditions."

#1.6

I did not understand why the authors develop a new model in the first place. They rightfully acknowledge that models such as MODFLOW exist, and these model could potentially do the job. Their argument is that MODFLOW models typically integrate geological data that is not available on a global scale. Therefore, a simplified model is developed. But this is a strange way of reasoning, as with MODFLOW one is not obliged to integrate all the geological complexity. It would have been perfectly possible to use MODFLOW for this project , with several significant advantages: for example, an unconfined aquifer (see below) could have been simulated. In this sense the novelty of the aspects concerning model development is questionable.

The main reason for not using MODFLOW directly but just implementing the MODFLOW approach is an efficient coupling to the existing global hydrological model. The structure of MODFLOW does not allow an efficient in memory coupling that also account for the two different scales without too much computational overhead.

Furthermore, the new model allows a more flexible extension of new components and adaptions to the conceptual nature of the model like an alternative capillary rise or dynamic recalculation of surface waterbody conductance.

Changes to manuscript

Page 8, Line 25-27

"The main motivation to develop a new model framework is the efficient in-memory coupling to the GHM and more flexible adaptation to the specific requirements of global-scale modelling."

#1.7

There are many other problems working on a global scale which are not even mentioned here but will even further undermine the credibility of the model. The three most important ones are: (1) Elevation is the wrong parameter for such a model. The data that should be used is not an ellipsoid-DEM but rather a geoid as the geoundulation is significant. (2) The density of sea-water is different, therefore there should be a density correction. (3) Steady-state conditions are inappropriate assumption that is not justified sufficiently well.

(1) As far as I understand the SRTM-based DEM it is based on a reference ellipsoid (WGS84) and a reference geoid that should already account for geoundulation. We assume that on a 5 arcmin resolution differences in the gravitational field are negligible. Furthermore, other inputs present a much higher uncertainty.

(2) As the model is not intended to be used for studying specifically groundwater-ocean interactions, and given the cell size of 9 km, we assume that the difference in density can be neglected at this scale as other parameterization introduce a higher level of uncertainty.

(3) Presenting a steady-state model is one of the first steps to understand model behaviour before moving towards a fully transient and fully coupled model. This represents a well-established method in developing groundwater models - regional as well as large-scale models.

A steady-state model (1) limits the degrees of freedom and thus model complexity as no time-variation needs to be taken into account and no storage changes need to be tracked. (2) A steady-state uncovers dominant processes and trends clearly that otherwise might have been obfuscated in a transient model due to the slow changing nature of groundwater. It is evident that not all processes can be observed that way and model behaviour will change as we move towards a fully transient model. It represents a first step in the model development process. Furthermore, (3) generated steady-state hydraulic heads can be used as initial state for a transient model spin-up phase in a fully coupled model.

It is true however that surface water bodies do not have a steady-state and that aquifers are ever changing. This is why the presented steady-state represents a first step into the model development as we move towards fully transient and coupled model. We think that it is not meaningful to move to a transient model directly with a completely new model without looking at the steady-state behaviour first.

Changes to manuscript

We added to the last paragraph of the introduction:

"Steady-state simulations are a well-established first step in groundwater model development to understand the basic model behavior limiting model complexity and degrees of freedom, thus providing insights into dominant processes and uncovering possible model-inherent characteristics impossible to observe in a fully coupled transient model. A transient model might obfuscate model inherent trends due to the slow changing nature of groundwater processes. In addition, the steady-state solution can be used as initial condition for future fully coupled transient runs."

(2) Page 5, Line 16

#1.8

Validation is done with other macro-scale models. This is a not an ideal strategy, as these large-scale models suffer from similar deficiencies (even though on less fundamental level). For a solid assessment of model performance a detailed, catchment scale hydrogeological model should be used for a benchmark comparison.

Validation has been achieved by a comparison to global groundwater observations, assumed naturalized conditions in a well-studied area (Central Valley) and by an additional comparison to other large-scale models. Goal of the model development was not the replication of regional groundwater characteristics - at this scale this is not a reasonable goal. Comparison is furthermore likely to be very challenging or impossible as a catchment might span only a couple of cells of the global model. The comparison to other large-scale models however enable a comparison based on similar input data (and input data deficiencies) uncovering how model decisions at this scale affect model outcome.

Changes to manuscript

See changes in response to comment #1.3

Page 17, Line 29 ff.

"The presented comparison to other large-scale models is based on the assumption that same model deficiencies e.g. in available data and scale issues can uncover differences in model decision. A comparison to catchment scale models is challenging as scales can differ by multiple magnitudes. As the model is further developed towards a transient model the presented comparison to simulations in data-rich regions need to be extended and temporal changes in interactions with surface water investigated."

#1.9

On line 28,page 7 the authors highlight that this is ok –" . . .without losing important model behavior. " Transient and steady state is significantly different in both spatial and temporal dynamics.

Reviewer refers to line 28 on page 3. We agree with the reviewer

Changes to manuscript

We revised the sentence. See changes in response to comment #1.2 and #1.7.

#1.10

The description of the conductance is confusing. In MODFLOW L is not the length of the river itself, but the length of the river within a grid cell. But this might just be an imprecise formulation.

This is correct (See table 1). Manuscript has been changed accordingly.

Changes to manuscript

Page 6, Line 5

#1.11

Other aspects also require more justification and discussion. Why only 8 % of wetland surfaces? Where does this number come from? What are the numerical convergence criteria, as well as a wide range of additional model parameters?

Manuscript is describing 80% of wetland area. Available maps of wetland areas show the maximum spatial extent of surface water bodies. As the maximum extent is seldom reached we reduce the extent for the steady-state model to 80% of the area shown in maps. In the fully transient model the wetland area will be adjusted in each time step as a function of wetland water storage.

It is not clear to us what the referee meant by "as well as a wide range of additional parameters". Parameters including convergence criteria are shown in Table 1.

Changes to manuscript

Page 8, Line 12

Referee #2

#2.1

Is 5’ an appropriate resolution at which to simulate groundwater flow? The analysis by Krakauer et al may be useful in determining the appropriate resolution.

Kraukauer et al. (2014) suggests that a grid spacing smaller than 0.1° (6’) for lateral groundwater processes is favourable for models running at a finer resolution than 1°. Thus a 5' seems to be reasonable even though our results suggest that the scale properties of surface water elevation need to be investigated further and that information from subgrid scales might need to be accounted for to improve overall results.

Changes to manuscript

Page 4, Line 12,13

#2.2

The work is coupled to WaterGap at 0.5deg, this is a really large scale discrepancy. How do you think this might alter the model results?

As groundwater recharge is mainly driven by climate inputs that are only available at coarse scales the presented steady-state model is not affected by the scale differences. Moving towards a fully coupled model scale differences between the two models play an important role especially for surface water body coupling. For example, it is not reasonable to calculate a river head change in the 0.5° model and apply that change equally to all 5' grid cells to recalculate the interaction between the surface water and the groundwater. The (future) presentation of a fully transient coupled model needs to discuss this more extensively.

Changes to manuscript

none

#2.3

The comparisons between this study and Fan et al and Maxwell et al are interesting. While pressure head is important, I think the bias from these scatterplots, basically water table depth, is more meaningful (as plotted in Fan et al / Maxwell et al too). The statistics will really be driven by topography which can occlude model performance and differences.

Please refer to our responses and changes to manuscript in response to comments #1.3 and #1.4.

#2.4

The diagram for how the model handles topographic breaks (Fig 1) is super confusing. Basically is water moved between cells even if there is a disconnect?

Yes, this is due to the coarse lateral discretization where in a 5' grid cell with approx. 80 km² area, the elevation differences can be larger than 200 m (as described in the text). Lateral interaction between neighbouring cells is always calculated in the model even if large topographic breaks are present. In order to avoid confusion, we modified Fig. 1 and text in section 2.1 to clarify that the top layer in the model should not be thought of as being located right at the land surface elevation.

Changes to manuscript

To clarify the difficult but important aspect of the relation between model layers and surface elevation in steep terrain at the spatial resolution of 5', we revised Figure 1 and added to section 2.1:

"In addition, due to the coarse spatial scale and the possible large variations of land surface elevations within each grid cell, the upper model layers should not be considered to aligned with an average land surface elevation. The model layers can be rather thought to be vertically aligned with the elevation of the surface water body table, as this prescribed elevation is, together with the sea level, the only elevation included in the groundwater flow equation (Eq. 2)."

#2.5

The assumption of confined conditions really seems hard to justify. This is effectively what de Graaf et al (2015, 2017) do with their two layer MODFLOW model with a stream package connection to PCRGLOB. There are so many assumptions present I think more careful discussion of how sensitivities in these assumptions (e.g. parameters in what amounts to the stream package used here) and feedback back to the WaterGap (which I think is just one-way at this point) would be really important.

Regarding the assumption of confined conditions, we now explain the rationale for it (see changes to manuscript). A sensitivity analysis is beyond the scope of this paper. We are currently preparing a paper that presents an extensive sensitivity analysis of the steady-state G³M presented here.

Changes to manuscript

Regarding the assumption of confined conditions, we added to the second paragraph of section 2.3:

"We choose to simulate confined flow conditions in both layers even though the upper layer can be expected to decrease in depth and thus in transmissivity (hydraulic conductivity times saturated depth). Every unconfined aquifer can have an equivalent confined representation assuming a correct saturated thickness (Sheets et al., 2015). However, given the large uncertainties regarding hydraulic conductivities (possibly an order of magnitude) and the lack of knowledge about aquifer thickness, it is appropriate to choose the computationally more efficient assumption of confined conditions."

#2.6

From Figure 2 it appears that not all the features are implemented in this model, or perhaps not all the features are activated except for recharge. Since the abstract discusses capillary subsides for plant water use but this feature is not described (nor is it entirely clear how that would be implemented as a simple flux), I think a thorough re-working of this discussion and assumptions are needed. Unfortunately, this figure begs the question why is a methods paper in GMD incomplete and not presenting all the model features?

The intention of Fig. 2 was to show how the gradient-based groundwater model G^{3}M is planned to be coupled with/integrated into the global hydrological model WaterGAP. This information is necessary to understand the modelling choices made for the steady-state G^{3}M presented in the manuscript, as a first step towards a fully coupled transient model. We think that a steady-state model is an important first step to justify a newly developed groundwater model and needs to be presented to the scientific community before moving further along to a fully coupled transient model. The steady-state model alone shows the difficulties of simulating groundwater flows at the coarse spatial resolution required for global-scale modelling. The model feature capillary rise is not presented as it cannot work without coupling to the soil compartment of WaterGAP.

Changes to manuscript

We added the following sentence to the last paragraph of the 1 Introduction:

"Capillary rise is not included in the presented steady-state simulation as simulation of capillary rise requires information of soil moisture that is only available when G^{3}M is fully integrated into WGHM."

#2.7

The maps of water table depth seem to have a tremendous shallow bias. It is hard to say because of low figure resolution, but perhaps most of Eastern N America, most of Australia, half of Europe and all of Tropical Africa are under water. I think additional discussion is needed here at least. Could this be due to the steady state assumptions? Confined conditions? The stream aquifer package? Resolution and slope? ET feedbacks?

So your visual impression is wrong, only the darkest blue means "under water", and this happens only in 2.1% of all cells. As we write (already in the first manuscript version) in section 3.1, "In 2.1 % of all cells, GW head is simulated to be above the land surface elevation, by more than 1 m in 0.3 % and by more than 100 m in 0.004 % of the cells.". Still, areas in Eastern N-America, Australia, Europe and tropical Africa present very shallow groundwater tables. This is mainly due to large wetland extends in these areas in connection with the steady-state approach. The extent of all wetlands (global already reduced by 20%) likely is overestimated as the data represents a maximum extend that is rarely reached in reality. Additionally, wetlands don't have a steady-state (or rather no surface water body) thus the interaction with the groundwater is likely overestimated and leads to the observed flooding.

Changes to manuscript

none

#2.8

It's hard to tell what the difference is here between the PRCGlob-MODFLOW model and this current model. More discussion is needed to clarify this distinction. I actually feel it's okay if there are many similar models out there (and both can be good models or bad models, it's not a competition), I would like more dissection of the differences in approach.

Already in the first version, we wrote in the abstract

"Together with an appropriate choice for the effective elevation of the SW table within each grid cell, this enables a reasonable simulation of drainage from GW to SW such that, in contrast to the GW model of de Graaf et al. (2015, 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G^3M. Comparison of simulated hydraulic heads to observations around the world shows better agreement than de Graaf et al. (2015)."

More explanation about this additional drainage required by PCR-GLOBWB but not G^3M is given in the introduction:

"The first global gradient-based GW model that was run for both steady-state (de Graaf et al., 2015) and transient conditions (de Graaf et al., 2017) was driven by GW recharge and SW data of the GHM PCR-GLOBWB (van Beek et al., 2011). However, there is not yet a two-way coupling of a GW flow model and a GHM. This may be due to the way de Graaf et al. (2015, 2017) *modelled river-GW interaction. To achieve plausible hydraulic head results, they found it necessary to add an additional drainage flux to GW drainage driven by the hydraulic head difference between GW and river. This additional drainage, which accounts for about 50% of global GW drainage, is simulated as a function of GW storage above the floodplain, the values of which are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017) – the model component that the gradient-based model was intended to replace. This prevents a full integration of the global GW flow model of de Graaf et al. (2017) into a GHM, as then, the linear GW reservoir model would be replaced by the GW flow model."*

The section in the discussion read

"De Graaf et al. (2015) set their SW head (h_{swb}) to the land surface elevation of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to avoid excessive flooding. On the other hand, this adaption allows the drainage of water even if the hydraulic head is below the SW elevation that might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity of SW body elevation."

Changes to manuscript

We modified the section in the discussion on the comparison to the gw model for PCR-GLOBWB by adding (see bold words): "De Graaf et al. (2015) set their SW head (h_{swb}) to the land surface elevation of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to avoid excessive flooding**, and that is not needed in G^3M**. On the other hand, this adaption allows the drainage of water even if the hydraulic head is below the SW elevation that might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity of SW body elevation.

#2.9

The current model is also completely different from the Central Valley model. This strikes me as odd too. Is it water use? Boundary conditions?

The presented Central Valley model plot show the initial state of the CVHM model and not computed model results. The initial condition represents the close to natural conditions in the early 1960s in the Central Valley with a very shallow groundwater table and large wetlands. Scale is most likely the main driver for the different results. Except for the scale differences G³M correctly computes shallow conditions close to the values assumed by CVHM with groundwater above the surface in the north and partially in the south of the valley. Furthermore, the depth to groundwater decrease towards the Sierra Nevada. Other differences are likely due to the steady-state and the connected assumptions on surface water bodies.

Changes to manuscript

Page 16, Line 14-17

"G³M correctly computes the shallow conditions with groundwater above the surface in the north, partially in the south of the valley and decreasing towards the Sierra Nevada. The difference in the extend of flooded area could be due to large wetlands areas still present in the early 60s which are not represented in this extend in the data used by G³M."

Page 18, Line 3-6

"The comparison to the initial state (based on historical observations) of the CVHM model presents a first comparison within a data-rich region which provides also the future possibility of comparing transient model results and human impact on a regional scale. G³M is able to reproduce the shallow groundwater table in the early 1960s. Differences are likely due to the steady-state approach and the connected assumptions on surface water bodies."

Short comment L. Gross

#3.1

As explained in https://www.geoscientific-model-development.net/about/manuscript_types.html the preferred reference to code release is through the use of a DOI which is then cited in the paper. As the model version is already published on GitHub a DOI can easily be created using for instance Zenodo, see https://guides.github.com/ activities/citable-code/ for details.

Citation of the code in the Open Source Journal was be replaced with a DOI pointing directly to the code.

Changes to manuscript

Page 19, Line 25 ff.

#3.2

As also stated in the guide lines E-mail contact to obtain access is not preferred and simulations and data should be made available as supplement, as DOI or as part of the release.

Model output will be added to the supplementary material.

Changes to manuscript

Page 19, Line 25 ff.

Beyond the bucket – Developing a global gradient-based groundwater model (G^3M v1.0) for a global hydrological model from scratch

Robert Reinecke[1], Laura Foglia[3], Steffen Mehl[4], Tim Trautmann[1], Denise Cáceres[1], Petra Döll[1,2]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany
[2]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany
[3]Department of Land, Air and Water Resources, University of California, Davis, USA
[4]Department of Civil Engineering, California State University, Chico, USA

Correspondence to: Robert Reinecke (reinecke@em.uni-frankfurt.de)

Abstract. To quantify water flows between groundwater (GW) and surface water (SW) as well as the impact of capillary rise on evapotranspiration by global hydrological models (GHMs), it is necessary to replace the bucket-like linear GW reservoir model typical for hydrological models with a fully integrated gradient-based GW flow model. Linear reservoir models can only simulate GW discharge to SW bodies, provide no information on the location of the GW table, and assume that there is no GW flow among grid cells. A gradient-based GW model simulates not only GW storage but also hydraulic head, which together with information on SW table elevation enables the quantification of water flows from GW to SW and vice versa. In addition, hydraulic heads are the basis for calculating lateral GW flow among grid cells and capillary rise.

G³M is a new global gradient-based GW model with a spatial resolution of 5' that will replace the current linear GW reservoir in the 0.5° WaterGAP Global Hydrology Model (WGHM). The newly developed model framework enables in-memory coupling to WGHM while keeping overall runtime relatively low, allowing sensitivity analyses and data assimilation. This paper presents the G³M concept and specific model design decisions together with results under steady-state naturalized conditions, i.e. neglecting GW abstractions, that can later be used as initial conditions for the fully-coupled WGHM-G^3M runs. Cell-specific conductances of river beds, which govern GW-SW interaction, were determined based on the 30" steady-state water table computed by Fan et al. (2013). Together with an appropriate choice for the effective elevation of the SW table within each grid cell, this enables a reasonable simulation of drainage from GW to SW such that, in contrast to the GW model of de Graaf et al. (2015; 2017), no additional drainage based on externally provided values for GW storage above the floodplain is required in G³M. Comparison of simulated hydraulic heads to observations around the world shows better agreement than de Graaf et al. (2015). In addition, G³M output is compared to the output of two established macro-scale models for the Central Valley, California, and the continental United States, respectively. A first analysis of losing and gaining rivers and lakes/wetlands indicates that GW discharge to rivers is by far the dominant flow, draining diffuse GW recharge, such that lateral flows only become a large fraction of total diffuse and focused recharge in case of losing rivers and some areas with very low GW recharge. G³M does not represent losing rivers in some dry regions. This study clarifies the conceptual approach to gradient-based groundwater modelling that is necessary for global-scale modelling with a coarse spatial resolution. It presents the first steps towards replacing the linear GW reservoir model in a GHM while improving on recent efforts, demonstrating the feasibility of the approach and the robustness of the newly developed simulation framework.

1 Introduction

Groundwater (GW) is the source of about 40% of all human water abstractions (Döll et al., 2014). It is also an essential source of water for freshwater biota in rivers, lakes and wetlands, which are in most cases recharged by GW. GW strongly affects river flow regimes and supplies the majority of river water during ecologically and economically critical periods with little precipitation. GW may receive recharge not only from the soil but also from surface water (SW) bodies. In case of small distances between GW table and land surface, GW enhances evapotranspiration via capillary rise. GW storage and flow dynamics have been altered by human GW abstractions as well as climate change and will continue to change in the future (Taylor et al., 2012). Around the globe, GW abstractions have led to lowered water tables and, in some regions, even GW depletion (Döll et al., 2014; Scanlon et al., 2012; Wada et al., 2012; Konikow, 2011). This has resulted in reduced base flows to rivers and wetlands (with negative impacts on water quality and freshwater ecosystems), land subsidence and increased pumping costs (Wada, 2016; Döll et al., 2014; Gleeson et al., 2012; Gleeson et al., 2016). The strategic importance of GW for global water and food security will probably intensify under climate change as more frequent and intense climate extremes increase variability of SW flows (Taylor et al., 2012). International efforts have been made to promote sustainable GW management and knowledge exchange among countries, e.g., UNESCO's program on International Shared Aquifer Resources Management (ISARM) (http://isarm.org) and the ongoing GW component of the Transboundary Waters Assessment Programme (TWAP) (http://www.geftwap.org). To support priority setting for investment among transboundary aquifers as well as identification of strategies for sustainable GW management, information on current conditions and possible trends of the GW systems is required (UNESCO-IHP, IGRAC, WWAP, 2012). In a globalized world, an improved understanding of GW systems and their interaction with SW and soil is needed not only at the local and regional but also at the global scale.

To assess GW at the global scale, global hydrological models (GHMs) are used (e.g., (Wada et al., 2012; 2016; Döll et al., 2012; 2014)). In particular, they serve to quantify GW recharge (Döll and Fiedler, 2008). Like typical hydrological models at any scale, GHMs simulate GW dynamics by a linear reservoir model. In such a model, the temporal change of GW storage in each grid cell is computed from the balance of prescribed inflows and an outflow that is a linear function of GW storage. Linear reservoir models can only simulate GW discharge to SW bodies but not a reversal of this flow, even though losing streams may provide focused GW recharge that allows the aquifer to support ecosystems alongside the GW flow path (Stonestrom et al., 2007) as well as human GW abstractions. This flow direction typically occurs in semi-arid and arid but seasonally also in humid regions. In addition, such linear reservoir models provide no information on the location of the GW table, and assume that GW flow among grid cells is negligible. To simulate the dynamics of water flow between SW bodies and GW in both directions as well as the effect of capillary rise on evapotranspiration, it is necessary to compute lateral GW flows among grid cells as function of hydraulic head gradients and thus the dynamic location of the GW table. To achieve an improved understanding of GW systems at the global scale, and in particular of the interactions of GW with SW and soil, it is therefore necessary to replace the linear GW reservoir model in GHMs by a hydraulic gradient-based GW flow model.

Macro-scale gradient-based GW flow models are still rare and mainly available for data-rich regions, e.g. for the Death Valley (Belcher and Sweetkind, 2010) and the Central Valley (Faunt, 2009; Dogrul et al., 2016) in the USA, but also for large fossil groundwater bodies in arid regions (e.g. the Nubian Aquifer System in North Africa, (Gossel et al., 2004). However, they are in most cases not integrated within hydrological models that quantify GW recharge based on climate data and provide information on the condition of SW. For North America, Fan et al. (2007) and Miguez-Macho et al. (2007) linked a land surface model with a two-dimensional gradient-based GW model and computed, with a daily time step, gradient-based GW flow, water table elevation, GW–SW interaction and capillary rise, using a spatial resolution of 1.25 km. One challenge was the determination of the river conductance that affects the degree of GW–SW interaction. A computationally very expensive integrated simulation of dynamic SW, soil and GW flow using Richards' equation for variably saturated flow was achieved at a spatial resolution of 1 km for the continental US by applying the ParFlow model (Maxwell et al., 2015). In both studies, GW abstractions were not taken into account.

A first simulation of the steady-state GW table for the whole globe at the very high resolution of 30'' was presented by Fan et al. (2013) and compared to an extensive compilation of observed hydraulic heads. However, there was no head-based interactions with SW; GW above the land surface was simply discarded. Global GW flow modeling is strongly hampered by data availability, including the geometry of aquifers and aquitards as well as parameters like hydraulic conductivity (de Graaf et al., 2017), and by computational restrictions on spatial resolution leading to conceptual problems, e.g., regarding SW-GW interactions (Morel-Seytoux et al., 2017). In the last years, some GW flow models that are in principle applicable for the global scale were developed but were applied only regionally in data-rich regions (Rhine basin: Sutanudjaja et al., 2011; France: Vergnes et al., 2012; Vergnes et al., 2014). The first global gradient-based GW model that was run for both steady-state (de Graaf et al., 2015) and transient conditions (de Graaf et al., 2017) was driven by GW recharge and SW data of the GHM PCR-GLOBWB (van Beek et al., 2011). However, there is not yet a two-way coupling of a GW flow model and a GHM. This may be due to the way de Graaf et al. (2015; 2017) modelled river-GW interaction. To achieve plausible hydraulic head results, they found it necessary to add an additional drainage flux to GW drainage driven by the hydraulic head difference between GW and river. This additional drainage, which accounts for about 50% of global GW drainage, is simulated as a function of GW storage above the floodplain, the values of which are computed externally by the linear GW reservoir model of PCR-GLOBWB (Equation 3 of de Graaf et al. (2017)) – the model component that the gradient-based model was intended to replace. This prevents a full integration of the global GW flow model of de Graaf et al. (2017) into a GHM, as then, the linear GW reservoir model would be replaced by the GW flow model.

In this study, we present the Global Gradient-based Groundwater Model (G^3M) that is to be integrated into the GHM WaterGAP 2. With a spatial resolution of 0.5° by 0.5° (approximately 55 km by 55 km at the equator), the WaterGAP 2 model (Alcamo et al., 2003) computes human water use in five sectors and the resulting net abstractions from GW and SW for all land areas of the globe excluding Antarctica. These net abstractions are then taken from the respective water storages in the WaterGAP Global Hydrology Model (WGHM) (Müller Schmied et al., 2014; Döll et al., 2003; 2012; 2014). With daily time steps, WGHM simulates flows among the water storage compartments canopy, snow, soil, GW, lakes, man-made reservoirs, wetlands and rivers. As in other GHMs, the dynamic of GW storage (GWS) is represented in WGHM by a linear GW reservoir model, i.e.

$$\frac{dGWS}{dt} = R_g + R_{g\_swb} - NA_g - k_g\, GWS \quad (1)$$

where R_g $[L^3T^{-1}]$ is diffuse GW recharge from soil, R_{g_swb} $[L^3T^{-1}]$ GW recharge from lakes, reservoirs and wetlands (only in arid and semiarid regions, with the value per SW body area globally constant), NA_g $[L^3T^{-1}]$ net GW abstraction. The product $k_g\, GWS$ quantifies GW discharge to SW bodies as a function of GWS and the GW discharge coefficient k_g (Döll et al., 2014). G^3M is to replace this linear reservoir model in WGHM.

The G^3M framework (Reinecke, 2018) was developed to provide full control over the involved processes and allow an optimal in-memory coupling to WGHM. Our model development approach was to learn from existing large-scale regional models (Faunt, 2009; Vergnes et al., 2014; Maxwell et al., 2015; Dogrul et al., 2016) to gain insights into how the coarse spatial resolution, incomplete data, and conceptual model design affects model results. We want to find out whether we can use gradient-based groundwater modelling at the global scale, when later integrated into a global hydrological model, to improve estimation of flows between SW and GW (affecting both e.g. streamflow and groundwater recharge and thus water availability for humans and ecosystems) and capillary rise (affecting evapotranspiration). In this paper, we present the model concept as well as steady-state model results. Steady-state simulations are a well-established first step in groundwater model development to understand the basic model behaviour limiting model complexity and degrees of freedom, thus providing insights into dominant processes and uncovering possible model-inherent characteristics impossible to observe in a fully coupled transient model. A transient model might obfuscate model inherent trends due to the slow changing nature of groundwater processes. A fully coupled model furthermore adds complexity and uncertainty to the model outcome. In addition, the steady-state solution can be used as initial condition for future fully coupled transient runs. Capillary rise is not included in the presented steady-state simulation as simulation of capillary rise requires information of soil moisture that is only available when G^3M is fully integrated into WGHM. In the next section, the model concept (including the concept for coupling with WGHM) and equations as well as applied data and parameter values are presented. In section 3, we show steady-state results of G^3M driven by WGHM data, without any two-way coupling. Simulated hydraulic heads are compared to observations world-wide and to the output of established regional models. We also discuss the influence of scale and modeling decisions and finally draw conclusions.

Kommentiert [RR1]: #1.2

Kommentiert [RR2]: #1.7

Kommentiert [RR3]: #2.6

2 Model description

2.1 G³M model concept

Although G³M is based on principles of the well-known GW flow modelling software MODFLOW (Harbaugh, 2005), G³M differs from traditional local and regional GW models. These models are generally based on rather detailed information on hydrogeology (including aquifer geometry and properties such as hydraulic conductivity derived from pumping tests), topography, pumping wells, location and shape of SW bodies as well as on observations of hydraulic head in GW and SW. Local observations guide the developer in constructing the model such that local conditions and processes can be properly represented. The lateral extent of individual grid cells of such GW flow models is generally smaller or similar to the depth of the aquifer(s) and the size of the SW bodies that interact with the GW. The global GW flow model G³M, however, covers all continents of the Earth except Greenland and Antarctica. At this scale, information listed above is poor or non-existing, and the lateral extent of grid cells needs to be relatively large due to computational (and data) constraints. We selected a grid cell size of 5' by 5' (approx. 9 km by 9 km at the equator), as this size fits well to WaterGAP and is smaller than the suggested 6' of Krakauer et al. (2014). WaterGAP 3 (Eisner, 2016) has the same cell size, and 36 of such cells fit to into one 0.5° WaterGAP 2 cell. Global climate data are only available for 0.5° grid cells.

Kommentiert [RR4]: #2.1

Due to the lack of information on the three-dimensional distribution of hydrogeological properties, we chose to use, in the current version of G³M, two GW layers with a vertical size of 100 m each (Fig. 1). G³M focuses on a plausible simulation of water flows between GW and SW bodies and on the simulation of capillary rise, and we deemed it suitable to have an upper GW layer that interacts with SW and soil and a lower one in which GW may flow laterally without such interactions. As land surface elevation within each 5' grid cell, with an area of approximately 80 km^2, may vary by more than 200 m (Fig. S1), neighbouring cells in G³M may not be adjacent anymore (Fig. 1), in contrast to (regional) GW models with smaller grid cells. This makes G³M a rather conceptual model in which water exchange between groundwater boxes is driven by hydraulic head gradients but flow can no longer be conceptualized as occuring through continuous pore space. In addition, due to the coarse spatial scale and the possible large variations of land surface elevations within each grid cell, the upper model layers should not be considered to be aligned with an average land surface elevation. The model layers can be rather thought to be vertically aligned with the elevation of the surface water body table, as this prescribed elevation is, together with the sea level, the only elevation included in the groundwater flow equation (Eq. 2).

Kommentiert [RR5]: #1.5, #2.4

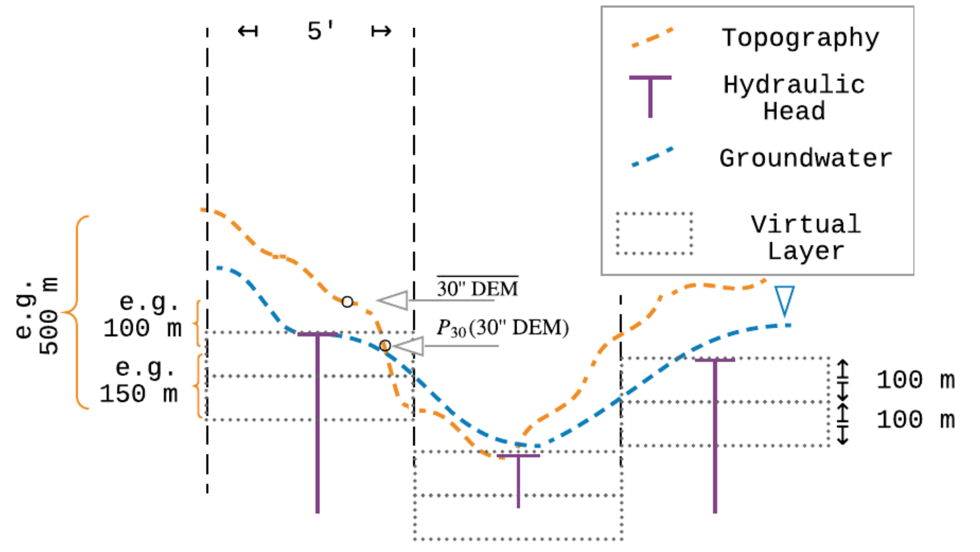

Figure 1 Schematic of G³M spatial structure, with 5' grid cells, hydraulic head per cell and the conceptual virtual layers. The underlying variability of the topography changes the perception of simulated depth to groundwater depending on what metrics are used to represent it on a coarser resolution. Layers in G³M are of a conceptual nature and describe the saturated flow between locations of head laterally and vertically.

Three-dimensional groundwater flow is described by a partial differential equation as a function of hydraulic head gradients

$$\frac{dGWS}{dt} = \left(\frac{\partial}{\partial x}\left(K_x\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z\frac{\partial h}{\partial z}\right) + W\right)\Delta x\Delta y\Delta z = S_s\frac{\partial h}{\partial t}\Delta x\Delta y\Delta z \tag{2}$$

where $K_{x,y,z}$ is the hydraulic conductivity $[LT^{-1}]$ along the x, y, and z axis between the cells (harmonic mean of grid cell conductivity values), S_s the specific storage $[L^{-1}]$, $\Delta x\Delta y\Delta z$ $[L^3]$ the volume of the cell, and h the hydraulic head $[L]$. Inflows in the groundwater are accounted for as

$$W\ \Delta x\Delta y\Delta z = R_g + Q_{swb} - NA_g - Q_{cr} + Q_{ocean} \tag{3}$$

where Q_{swb} is flow between the SW bodies (rivers, lakes, reservoirs and wetlands) and GW $[L^3T^{-1}]$, Q_{cr} is capillary rise, i.e. the flow from GW to the soil, and Q_{ocean} is the flow between ocean and GW$[L^3T^{-1}]$, representing the boundary condition. In case of Q_{swb} and Q_{ocean}, a positive value represents a flow into the groundwater.

The flux across the model domain boundary Q_{ocean} is modeled as a head-dependent flow based on a static head boundary.

$$Q_{ocean} = c_{ocean}\left(h_{ocean} - h_{aq}\right) \tag{4}$$

Here h_{ocean} is the elevation of the ocean water table $[L]$, h_{aq} the hydraulic head in the aquifer $[L]$ and c_{ocean} the conductance of the boundary condition $[L^2T^{-1}]$ (Table 1). We assume that density difference to sea-water is negligible at this scale. Q_{cr} is not yet implemented in G³M.

Q_{swb} in Eq. (3) replaces $k_g\ GWS$ and R_{g_swb} in Eq. (1), such that loosing conditions of all types of SW bodies can be simulated dynamically. It is calculated as a function of the difference between the elevation of the water table in the SW bodies h_{swb} $[L]$ and h_{aq} as

Kommentiert [RR6]: #1.5, #2.4

Kommentiert [RR7]: #1.7

$$Q_{swb} = \begin{cases} c_{swb}(h_{swb} - h_{aq}) & h_{aq} > B_{swb} \\ c_{swb}(h_{swb} - B_{swb}) & h_{aq} \leq B_{swb} \end{cases} \quad (5)$$

where c_{swb} is the conductance $[L^2T^{-1}]$ of the SW body bed (river, lake, reservoir or wetland) and B_{swb} the SW body bottom elevation $[L]$.

Conductance of SW bodies is often a calibration parameter in traditional GW models (Morel-Seytoux et al., 2017). Following Harbaugh (2005), it can be estimated by

$$c_{swb} = \frac{K\ L\ W}{h_{swb} - B_{swb}} \quad (6)$$

where K is hydraulic conductivity, L is length and W is width of the SW body per grid cell. For lakes (including reservoirs) and wetlands, c_{swb} is estimated based on hydraulic conductivity of the aquifer K_{aq} and SW body area (Table 1). For gaining rivers, conductance is quantified individually for each grid cell following an approach proposed by Miguez-Macho et al. (2007). The value of river conductance c_{riv}, according to Miguez-Macho et al. (2007), in a GW flow model needs to be set to such a values that, for steady-state conditions, the river is the sink for all the inflow to the grid cell (GW recharge and inflow from neighbouring cells) that is not transported laterally to neighbouring cells such that

Kommentiert [RR8]: #1.10

$$c_{riv} = \frac{R_g + Q_{eq_{lateral}}}{h_{eq} - h_{riv}} \qquad h_{aq} > h_{riv} \quad (7)$$

For G^{3}M, we computed the equilibrium head h_{eq} as the 5' average of the 30'' steady-state heads calculated by Fan et al. (2013). Using WGHM diffuse GW recharge lateral equilibrium flow $Q_{eq_{lateral}}$ $[L^3T^{-1}]$ is net lateral inflow into the cell computed based on the h_{eq} distribution as well as G^{3}M K_{aq} and cell thickness (Table 1). Elevation of the river water table h_{riv} [L] is to be provided by WGHM. Using a fully dynamic approach, i.e. utilizing the hydraulic head and lateral flows from the current iteration to re-calculate c_{riv} in each iteration towards the steady-state solution, has proven to be too unstable due to its non-linearity affecting convergence. We limit c_{riv} to a maximum of $10^7\ m^2day^{-1}$; this would be approximately the value for a 10 km long and 1 km wide river with a head difference between GW and river of 1 m and hydraulic conductivity of the river bed of 10^{-5} m/s.

If the river recharges the GW (losing river), Eq. (7) cannot be used as the Fan et al. (2013) high-resolution equilibrium model only models groundwater outflows but not inflows from SW bodies. If h_{aq} drops below h_{riv}, Eq. (5) is used to compute c_{riv}, with K equals to K_{aq}.

2.2 Coupling to WGHM

We intend to couple G^{3}M to WaterGAP 2, i.e. the 0.5° version of WGHM. It will not be coupled to WaterGAP 3 (Eisner, 2016), which has the same spatial resolution as G^{3}M, in order to keep computation time low enough for performing sensitivity analyses and ensemble-based data assimilation and calibration. However, data from WaterGAP 3 were used to set up G^{3}M. Location and area of the 5' grid cells of G^{3}M are the same as in the landmask of WaterGAP 3. In addition, the percentage of the 5' grid cell area that is covered by lakes (including reservoirs) and by wetlands, based on Lehner and Döll (2004), is taken from WaterGAP 3, as well as the length and width of the main river within each 5' grid cell as estimated by WaterGAP 3 (Table 1).

G^{3}M will be integrated into WGHM by exchanging information on (1) R_{g_swb} and NA_g, (2) soil water content, (3) Q_{cr}, (4) h_{swb}, and (5) Q_{swb}. Figure 2 indicates the direction of the information flows. Water flows from the 0.5° cells of WGHM are distributed equally to all 5' G^{3}M grid cells inside a 0.5° cell. Flows transferred from the 5' cells of G^{3}M to WGHM are aggregated. GW recharge and net abstraction from GW together with SW tables are the main drivers of the GW model that will be provided dynamically by WGHM. GW-SW flow volumes computed by G^{3}M will be aggregated and added or subtracted from the SW body volumes in WGHM, and SW body heads will be recalculated. WGHM soil water content together with

G³M depth to GW will be used to calculate capillary rise and thus a change of soil water content. Capillary rise is not yet implemented in G³M, and SW heads are currently based on land surface elevation.

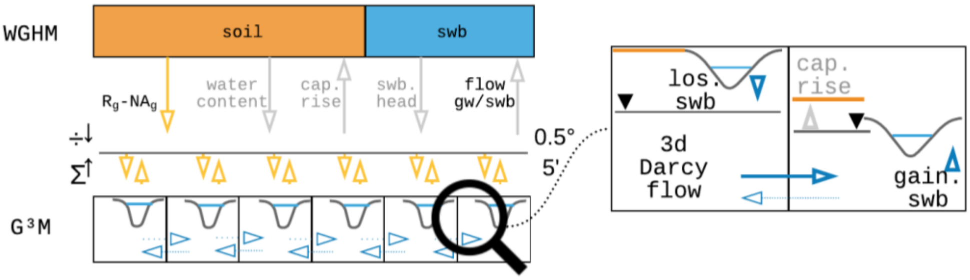

Figure 2 Conceptual view of the coupling between WGHM and G³M. WGHM provides calculated GW recharge (R_g) (Döll and Fiedler, 2008) and if the human impact is considered, net abstraction from GW (NA_g) (Döll et al., 2012). G³M spreads this input equally to all 5' grid cells inside a 0.5° cell and calculates hydraulic head and interactions with SW bodies (swb) as well as capillary rise (cap. rise) at the 5' resolution. Grey arrows show information flow that is not yet implemented.

2.3 The steady-state uncoupled model version

In a first implementation stage, G³M was developed as a steady-state (right-hand side of Eq. (2) is zero) standalone model that represents naturalized conditions (i.e. without taking into account human water use) during 1901-2013. Input data and parameters used are listed in Table 1 and described below.

The landmask of G³M, i.e. location and size of 5' grid cells, is that of WaterGAP 3. We performed a sensitivity analysis that confirmed the findings of other studies (de Graaf et al., 2015) that the aquifer thickness has a relatively small impact on the model results. Therefore, selecting a uniform thickness of 100 m (motivated by the assumed depth of validity of the lithology data) worldwide for the first layer and also for the second layer is expected to lead to less uncertainties as compared to hydraulic conductivities and the surface water table elevation. We choose to simulate confined flow conditions in both layers even though the upper layer can be expected to decrease in depth and thus in transmissivity (hydraulic conductivity times saturated depth). Every unconfined aquifer can have an equivalent confined representation assuming a correct saturated thickness (Sheets et al., 2015). However, given the large uncertainties regarding hydraulic conductivities (possibly an order of magnitude) and the lack of knowledge about aquifer thickness, it is appropriate to choose the computationally more efficient assumption of confined conditions.

Kommentiert [RR9]: #1.5, #2.5

Gleeson et al. (2014) provided a global subsurface permeability data set from which K_{aq} was computed. The data set was derived by relating permeabilities from a large number of local to regional GW models to the type of hydrolithological units (e.g., “unconsolidated” or “crystalline”). The geometric mean permeability values of nine hydrolithological units were mapped to the high-resolution global lithology map GLiM (Hartmann and Moosdorf, 2012). In continuous permafrost areas, a very low permeability value was assumed by Gleeson et al. (2014). The estimated values represent the shallow surface on the scale of 100 m depth. The unique dataset has three inherent problems when used as input for a GW model: (1) At this scale, important heterogeneities such as discrete fractures or connected zones of high hydraulic conductivity controlling the GW flow are not visible. (2) Jurisdictional boundaries due to different data sources in the global lithological map lead to artifacts. (3) The differentiation between coarse and fine-grained unconsolidated deposits is only available in some regions resulting in $10^{-4}\ m\ s^{-1}$ as hydraulic conductivity for coarse-grained unconsolidated deposits. If the distinction is not available, a rather low value of $10^{-6}\ m\ s^{-1}$ is set for unconsolidated porous media (Fig. S3). The original data was gridded to 5' by using an area-weighted average and used as hydraulic conductivity of the upper model layer. For the second layer, hydraulic conductivity of the first layer is reduced assuming that conductivity decreases exponentially with depth. Based on the e-folding factor f used by Fan et al. (2013) (a calibrated parameter based on terrain slope), conductivity of the lower layer is calculated by multiplying the upper layer value by $\exp(-50\ m\ f^{-1})^{-1}$ (Fan et al., 2007).

Mean annual GW recharge computed by WaterGAP 2.2c for the period 1901-2013 is used as input (Fig. S4), while no net abstraction from GW was taken into account. It would not be meaningful to try to derive a steady-state solution under existing net groundwater abstractions that in some regions cause GW depletion with continuously dropping water tables. Regarding the ocean boundary condition, h_{ocean} is set to 0 m and c_{ocean} to $10\ m^2\ day^{-1}$ (Table 1).

It is assumed that there is exchange of water between GW and one river stretch in each 5' grid cell, and in addition where lakes and wetlands exist according to WaterGAP 3, which provides, for each grid cell, the area of "local" and "global" lakes and wetlands. In WaterGAP, "local" SW bodies are only recharged by runoff produced within the grid cell, while "global" SW bodies also obtain inflow from the upstream cell. In an uncoupled model, it is difficult to prescribe the area of lakes and wetlands that affect the flow exchange between SW body and GW. Maps generally show the maximum spatial extent of SW bodies. This maximum extent is seldom reached, in particular in case of wetlands in dry areas. For global wetlands (wetlands greater than one 5' cell), it is therefore assumed in this model version that only eighty percent of their maximum extent is reached. In the transient model SW extends will be changed over time. A further difficulty in an uncoupled model run is that the water table elevation of SW bodies does not react to the GW-SW exchange flows Q_{swb} and that water supply from SW is not limited by availability. A loosing river may in reality dry out due to loss to GW and therefore cease to lose any more water. In case of rivers, B_{swb} is equal to $h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994), where $Q_{bankfull}$ is the bankfull river discharge in the 5' grid cell (Verzano et al., 2012). Globally constant but different values are used for B_{swb} in case of wetlands, local lakes and global lakes (Table 1).

For the steady-state model, river elevation h_{riv} is set in each grid cell to the same elevation as all other SW bodies, h_{swb}. We found that for both gaining and loosing conditions, Q_{swb} and thus computed hydraulic heads are highly sensitive to h_{swb}. The overall best agreement with the hydraulic head observations of Fan et al. (2013) was achieved if h_{swb} (Eq. 5, 6 and 7) was set to the 30th percentile (P_{30}) of the 30" land surface elevation values of Fan et al. (2013) per 5' cell, e.g. the 30" elevation that is exceeded by 70% of the thousand 30" elevation values within one 5' cell. To decrease convergence time we used h_{eq} derived from the high-resolution steady-state hydraulic head distribution of Fan et al. (2013) as initial guess.

2.4 Model implementation

G³M is implemented using a newly developed open-source model framework G³M-f (Reinecke, 2018). The main motivation to develop a new model framework is the efficient in-memory coupling to the GHM and more flexible adaptation to the specific requirements of global-scale modelling. Written in C++, the framework allows the implementation of global and regional groundwater models alike while providing an extensible purely object-oriented model environment. It is primarily targeted as extension to WGHM but allows an in-memory coupling to any GHM or a standalone groundwater model. It provides a unit-tested environment offering different modules that can couple results in-memory to a different model or write out data flows to different file formats. As internal numerical library it uses Eigen3 (eigen.tuxfamily.org). Different from Vergnes et al. (2014), G³M's computations are not based on spherical coordinates directly but on an irregular grid of rectangular cells. Cell sizes are provided by WaterGAP3 and are derived from their spherical coordinates maintaining their correct area and centre location. The model code will be adapted in the future to account for the different length in x and y direction per cell correctly.

Eq. (2) is reformulated as finite-difference equation and solved using a conjugate gradient approach and an Incomplete LUT preconditioner (Saad, 1994). In order to keep the memory footprint low, the conjugate gradient method makes use of the sparse matrix. Furthermore, it solves the equations in parallel (preconditioner currently non-parallel). G³M can compute the presented steady-state solution (with the right-hand side of Eq. (2) being zero and the heads of Fan et al. (2013) as initial guess, Table 1) on a commodity computer with four computational cores and a standard SSD in about 30 minutes while occupying 6 GB of RAM.

Kommentiert [RR10]: #1.11

Kommentiert [RR11]: #1.6

Similar to MODFLOW, G³M-f solves Eq. (2) in two nested loops: (1) the outer iteration checks the head and residual convergence criterion and adjusts whether external flows have changed into a different state e.g. from gaining to losing conditions and optimizes the matrix if flows are no longer head dependant. (2) The inner loop primarily consists of the conjugate gradient solver, which runs for a number of iterations defined by the user or until the residual convergence criterion is reached (Table 1), solving the current matrix equation.

Because the switch between Eq. (6) and Eq. (7) that occurs if e.g. h_{aq}drops below h_{riv} from one iteration to the next causes an abrupt change of c_{riv} inducing a nonlinearity that affects convergence we introduced an ϵ = 1 m interval around h_{riv} and interpolate c_{riv} by a cubic hermite spline polynomial over that interval. This allows for a smoother transition between both states, reducing the changes in the solution if a river is in a gaining condition in one iteration and in a losing condition in the next and vice versa.

3 Results

3.1 Global hydraulic head and depth to GW distribution under natural steady-state conditions

As expected, the computed global distribution of steady-state hydraulic head (in the upper model layer) under natural conditions (Fig. 3a) follows largely the land surface elevation (Fig. S2), albeit with a lower range and locally different ratios between the hydraulic head and land surface gradients (Fig. S7). Depth to GW can be computed by subtracting the hydraulic head computed by G^3M for the upper layer of each 5' grid cell from the arithmetic mean of the land surface elevations of the 100 30" grid cells within each 5' cells (Fig. S2). The global map of steady-state depth to GW (Fig. S5) clearly resembles the map of differences between surface elevation and P_{30} , the assumed water level of SW bodies h_{swb}, shown in Fig. S1**,** which indicates that simulated depth to GW is strongly governed by the assumed water level in SW bodies.

Deep GW, i.e. a large depth to GW, occurs mainly in mountainous regions (Fig. S5). These high values of depth to GW are mainly a reflection of the steep relief in these areas as quantified either by the differences of mean land surface elevations between neighbouring grid cells (Fig. S8) or the difference between mean land surface elevation and P_{30}, the 30th percentile of the 30" land surface elevations (Fig. S1). When computed hydraulic head is subtracted not from average land surface elevation but from P_{30}, the assumed water table elevation of SW bodies, the resulting map shows that the groundwater table is mostly above P_{30}, in both flat and steep terrain (Fig. 3b). Thus, high depth to GW values at the 5' resolution do not indicate deep unsaturated zones and loosing rivers but just high land surface elevation variations within a grid cell. In steep terrain, 5’ water tables are higher above water level in the surface water bodies than in flat terrain (Fig. 3b). Deep GW tables that are not only far below the mean land surface elevation but also below the water table of surface water bodies are simulated to occur in some (steep or flat) desert area with very low GW recharge.

In 2.1 % of all cells, GW head is simulated to be above the average land surface elevation, by more than 1 m in 0.3 % and by more than 100 m in 0.004 % of the cells. The shallow water table in large parts of the Sahara is caused by losing rivers (and some wetlands) that cannot run dry in the model, causing an overestimation of the GW table (section 2.3). Please note that the computed steady-state depth to GW certainly underestimates the steady depth to GW in GW depletion areas such as the High Plains Aquifer and the Central Valley in the USA (section 3.4), Northwestern India, North China Plain and parts of Saudi Arabia and Iran (Döll et al., 2014) as groundwater withdrawals are not taken into account in the presented steady-state simulation of G^3M.

Kommentiert [RR12]: Adapted to new fig03

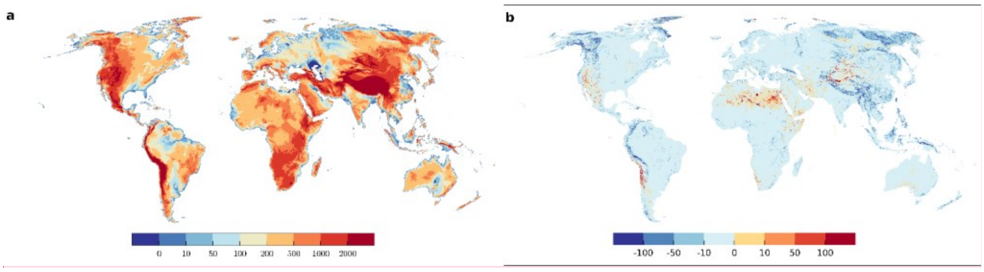

Figure 3 (a) Simulated equilibrium hydraulic head [$\boldsymbol{m}$]. Maximum value 6375 m, minimum value -414 m (Extremes included in dark blue and dark red). (b) Difference between 30th percentile of the 30'' land surface elevation per 5' grid cell (chosen elevation for surface water bodies) and simulated equilibrium hydraulic head. Maximum value 1723 m, minimum value -1340 m (Extremes included in dark blue and dark red).

Kommentiert [RR13]: #1.3, #2.7

3.2 Global water budget

Inflows to and outflows from GW of all G^{3}M grid cells were aggregated according to the compartments ocean, river, lake, wetland, and diffuse GW recharge from soil (Fig. 4). The difference between the global sum of inflows and outflows in less than 10^{-6} %. This small volume balance error indicates the correctness of the numerical solution.

Total diffuse GW recharge from soil is 3.9 $10^{10}\ m^3\ day^{-1}$ and approximately equal to the drainage of GW to rivers. Rivers are the ubiquitous drainage component of the model), followed by wetlands, lakes and the ocean boundary. According to G^{3}M, the amount of river water that recharges GW is only about a 40th of the drainage to GW, and the relative inflow to GW from lakes and wetlands is even smaller (Fig. 4). Possibly, flow from SW to GW is even overestimated, as outflow from SW is not limited by water availability in the SW, and depending on the hydraulic conductivity, Eqs. (5) and (6) can lead to rather large flows. Inflow from the ocean, which is more than two magnitudes smaller than outflow to ocean, occurs in regions where $h_{swb} =$ P_{30} is below h_{ocean}.

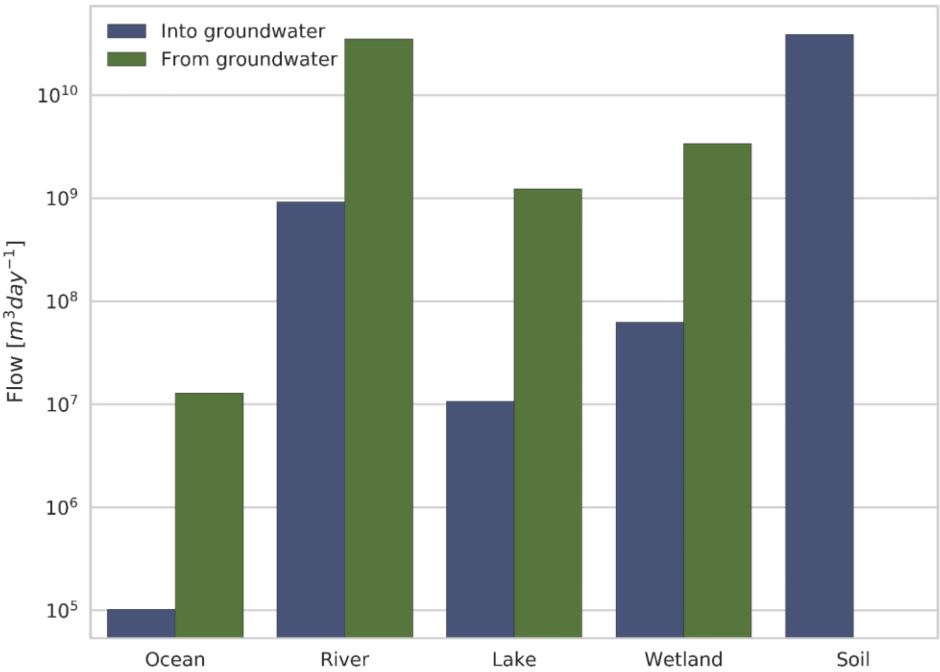

Figure 4 Global sums of flows from different compartments into or from GW at steady state. Flows into the GW are denoted by the color blue, flows out of the GW into the different compartments by green. The compartment soil is the diffuse GW recharge from soil calculated by WaterGAP.

3.3 GW-SW interactions

Figure 5 plots the spatial distribution of simulated flows from and to lakes and wetlands (Fig. 5a) as well as from and to rivers (Fig. 5b). It reveals strong interaction between GW and SW bodies that is dominated by GW discharging to SW bodies. Parallel to the overall budget (Fig. 4), the map reveals the globally large but locally strongly varying influence of lakes and wetlands (Fig. 5a). Rivers with riparian wetlands such as the Amazon River receive comparably small amounts of GW as most of the GW is drained by the wetland (compare Figs. 5a and 5b). Similarly, areas dominated by wetlands and lakes (e.g. parts of Canada and Scandinavia) show less inflow for rivers (Fig. 5b). 93 % of all grid cells contain gaining rivers, and only 7% loosing rivers. Gaining lakes and wetlands are found in 12 % and 11 % of the cells, respectively, whereas only 0.2 % contain a losing lake or wetland. In G3M, all SW bodies (rivers, lakes and wetlands) in a grid cell either loose or gain water.

Gaining rivers, lakes and wetlands with very high absolute Q_{swb} values over 1 mm day^{-1} (averaged over the grid cell area of approximately 80 km^2) can be found in the Amazon, Congo, Bangladesh, and Indonesia, where GW recharge in very high (Fig. S4). Values below 0.01 mm day^{-1} are present in dry and in permafrost areas where groundwater recharge is small.

Losing SW bodies occur in the model under two conditions, in mountainous regions where depth to GW is high and in arid and semi-arid climates with low diffuse groundwater recharge. Without focused GW recharge, the GW table would drop to even further in the mountains and is necessary to counteract the large hydraulic gradients caused by the large topographic gradients. Rivers lose more than 1 mm day^{-1} in Ethiopia and Somalia, West Asia, Northern Russia, the Rocky

Mountains and the Andes whereas lower values can be observed in Australia and in the Sahara. High values of outflow from wetlands and lakes are found in Tibet, the Andes and northern Russia, lower values in the Sahara and Kazakhstan. The river Nile in the Northern Sudan and Egypt is correctly simulated to be a losing river (Fig. 5b), being an allogenic river that is mainly sourced from the upstream humid areas, including the man-made Lake Nasser (Elsawwaf et al., 2014) (Fig. 5a). Furthermore, the following lakes and riparian wetlands are simulated to recharge GW: parts of the Congo River, Lake Victoria, the Ijsselmeer, Lake Ladoga, the Aral Sea, parts of the Mekong Delta, the Great Lakes of North America. On the other hand, no losing stretches are visible at the Niger River and its wetlands and almost none Northeastern Brazil even though that losing conditions are known to occur there (Costa et al., 2013; FAO, 1997).

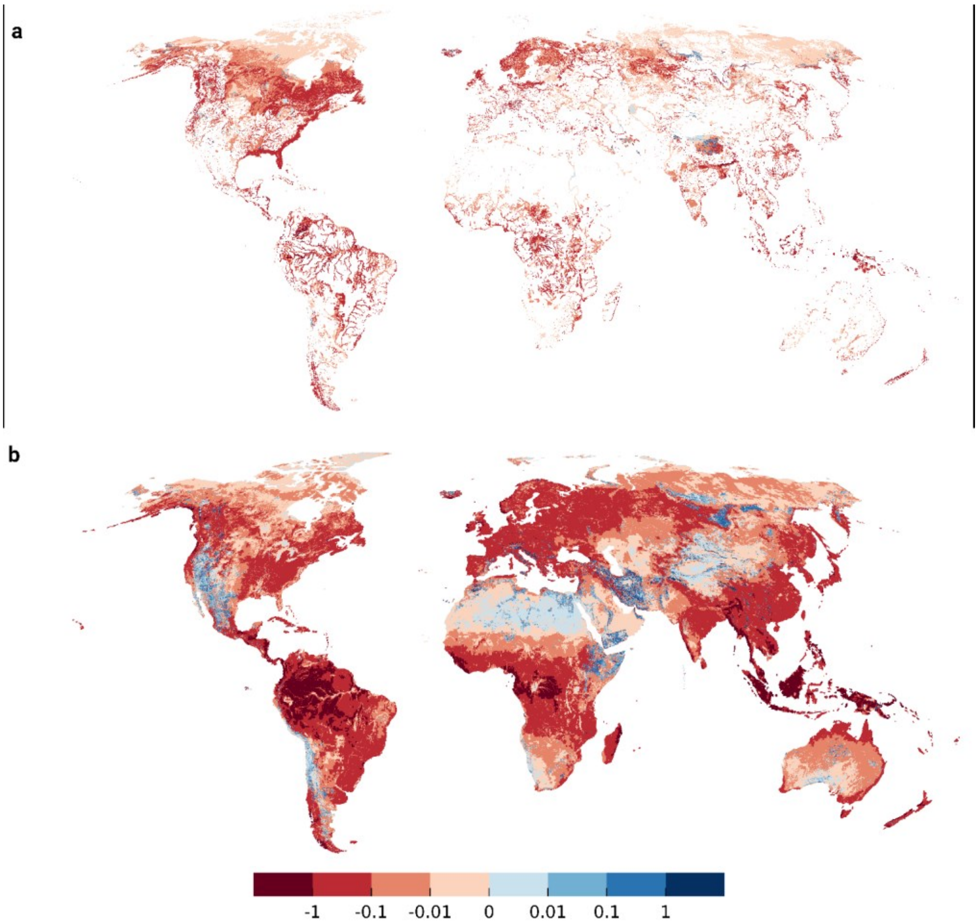

Figure 5 Flow $\boldsymbol{Q_{swb}}$ $[\boldsymbol{mm\ day^{-1}}]$ from/to wetlands, lakes (a) and losing/gaining streams (b) with respect to the 5´grid cell area. Gaining rivers are shown in red, rivers recharging the aquifer in blue. Focused recharge occurs in arid regions, e.g. alongside the river Nile, and in mountainous regions where the average water table is well below the land surface elevation.

Simulated flows between GW and SW depend on assumed conductances for both rivers and lakes/wetlands (Eqs. 5, 6, 7) shown in Fig. 6. Q_{swb} (Fig. 5) correlates positively with conductance. Conductance for gaining rivers correlates positively with GW recharge (Eq. (7) and Fig. S4). High conductance values are reached in the tropical zone due to a higher GW recharge but are capped at a plausible maximum value of $10^7\ m^2 day^{-1}$s in case of river (section 2.1) (Fig. 6b), while lakes and wetlands, with a larger area, can reach larger values, e.g. in Canada or Florida.

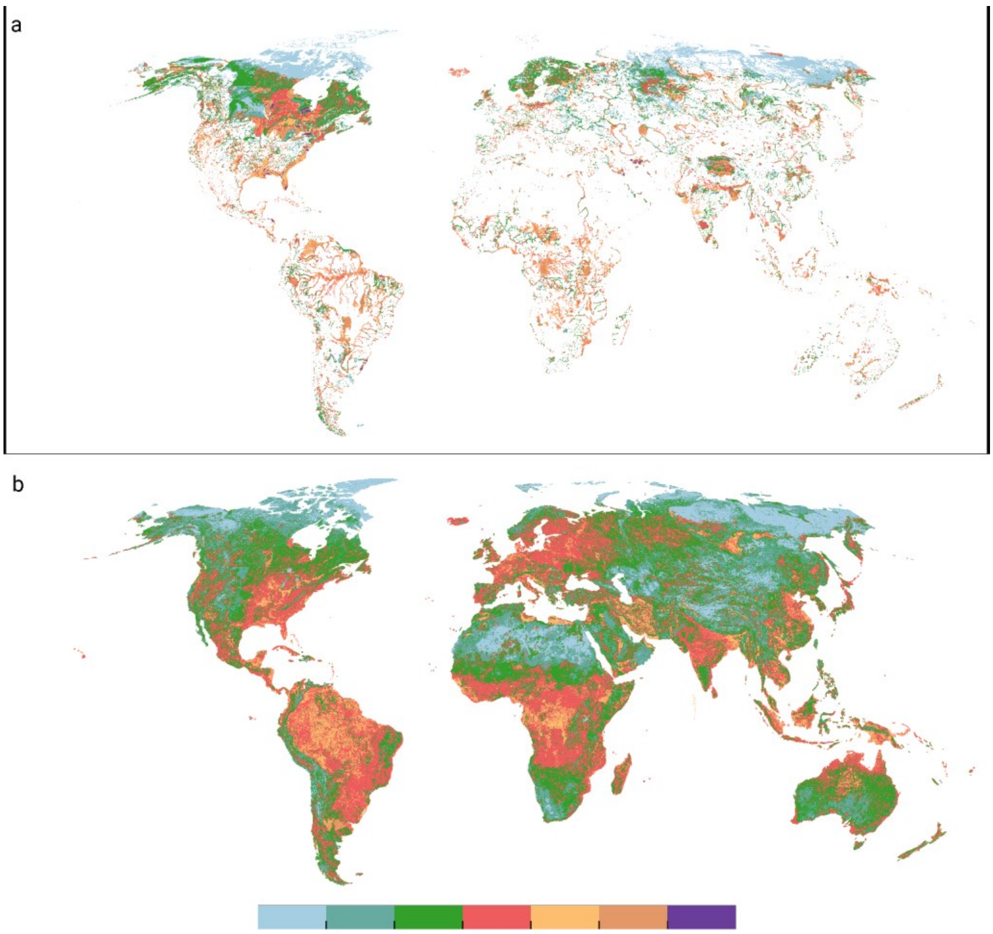

Figure 6 Conductance [$m^2 day^{-1}$] of lakes and wetlands (a) and rivers (b). In regions close to the pole conductance is in general lower due to the influence of the low aquifer conductivity (losing conditions), and relatively small GW recharge due to permafrost conditions (only applies for gaining conditions). Max conductance of wetlands is 10^8.

3.4 Lateral flows

Figure 7 shows lateral outflow from both model layers in percent of the sum of diffuse GW recharge from soil and GW recharge from SW bodies. The percentage of recharge that is transported through lateral flow to neighbouring cells depends on 5 main factors: (1) hydraulic conductivity (Fig. S3), (2) diffuse GW recharge (Fig. S4), (3) losing or gaining SW bodies (Fig. 5), (4) their conductance (Fig. 6) and (5) the head gradients (Fig. 3).

On large areas of the globe, where GW discharges to SW, the lateral flow percentage is less than 0.5% of the total GW recharge as GW recharge in a grid cells is simulated to leave the grid cell by discharge to SW. For example, in the permafrost regions, very low hydraulic conductivity limits the outflow to neighbouring cells of the occurring recharge, leading to these very low percent values. Such values also occur in regions with high SW conductances and rather low hydraulic conductivity, e.g. in the Amazon Basin. Values of more than 5% occur where hydraulic conductivity is high even if the terrain in rather flat, such as in Denmark. Higher values may occur for in case of gaining rivers in dry areas like Australia or in mountainous regions where large hydraulic gradients can develop. In mountains with gaining surface water bodies, lateral outflows may even exceed GW recharge of the cell. In grid cells where SW bodies recharge the GW, outflow tends to be a large percentage of total GW recharge as there is no outflow from GW other than in lateral direction, and values often exceed 100% (Fig. 7).

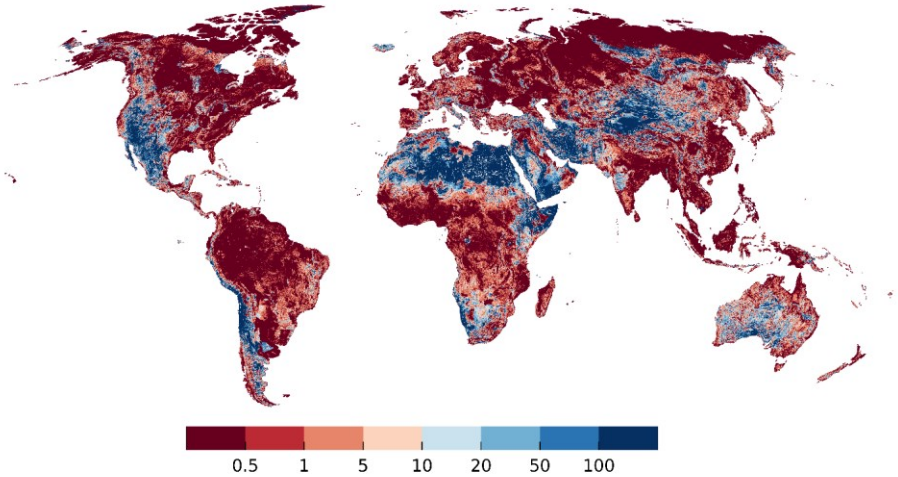

Figure 7 Percentage of GW recharge from soil and surface water inflow that is transferred to neighboring cells through lateral out flow (sum of both layers). Grid cells with zero total GW recharge are shown in white (a few cells in the Sahara and the Andes).

3.5 Comparison to groundwater well observations

Global observations of depth to GW were assembled by Fan et al. (2007; 2013). We selected only observations with known land surface elevation and removed observations where a comparison to local studies suggested a unit conversion error. This left total of 1,070,402 depth to GW observations. An "observed head" per 5' model cell was then calculated by first computing hydraulic head of each observation by subtracting depth to GW from the 5' land surface elevation used in G³M and then calculating the arithmetic mean of all observations within the 5' model cell. Multiple obstacles limit the comparability of observations to simulated values. (1) Observations were recorded at a certain moment in time influenced by seasonal effects and abstraction from GW, whereas the simulated heads represent a natural steady-state condition. (2) Observation locations are biased towards river valleys and productive aquifers. (3) Observations may be located in valleys with shallow local water tables too small to be captured by a coarse resolution of 5'.

Simulated steady-state hydraulic heads in the upper model layer are compared to observations in Fig. 8. Shallow GW is generally better represented by the model than deeper GW. Especially the water table in mountainous areas is underestimated, which may be related to observations in perched aquifers caused by low permeability layers (Fan et al., 2013) that are not represented in G^{3}M due to lacking information. Because the steady-state model cannot take into account the impact of GW abstraction, the computed depth to GW values are considerably smaller than currently observed values in GW depletion areas like the Central Valley in California (where once wetlands existed before excessive GW use depleted the aquifer) and the High Plains Aquifer in the Midwest of the USA. Still, the elevation of the GW table in the non-depleted Rhine valley in Germany is overestimated, too. Figure 9a shows the hydraulic head comparison as scatter plot. Overall, the simulation results tend to underestimate observed hydraulic head but much less than the steady-state model presented by de Graaf et al. (2015).

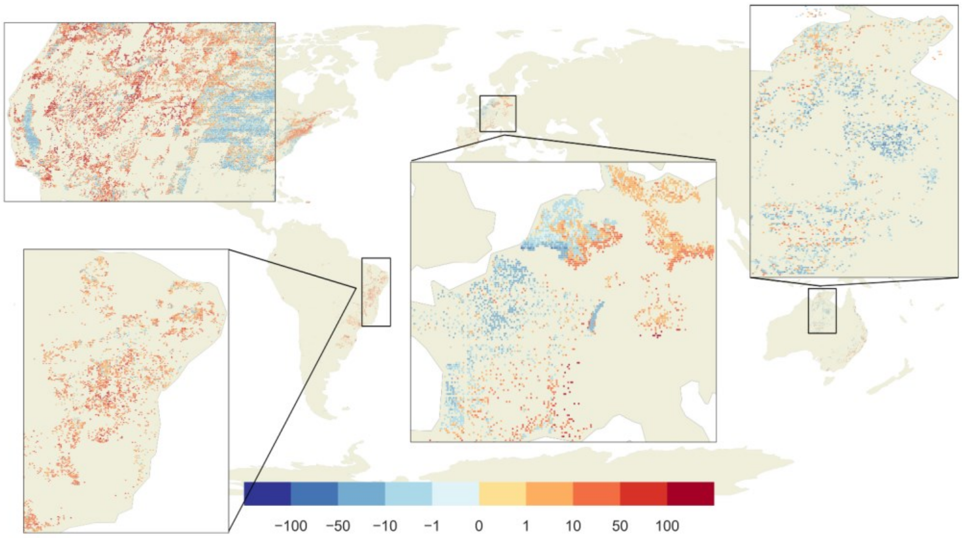

Figure 8 Differences between observed and simulated hydraulic head [***m***]. Red dots show areas where the model simulated deeper GW as observed, blue shallower GW. In the grey areas, no observations are available.

To compare performance of G^{3}M to the steady-state results of two high-resolution model of Fan et al. (2013) and ParFlow (Maxwell et al., 2015), heads in 30'' (Fan et al., 2013) and 1 km (ParFlow) grid cells were averaged to the G^{3}M 5' grid cells. The comparison of 5' observations to the 5' average of ParFlow seem to be consistent with the 1 km model comparison in Maxwell et al., 2015, their Fig. 5, even though over/under -estimates in the original resolution seemed to be smoothed out by averaging to 5' (not shown). The heads of Fan et al. (2013) fit better to observations than G^{3}M heads, with less underestimation (Fig. 9b). The comparison of G^{3}M heads to Fan et al. (2013) values for all 5' grid cells, which are also the initial heads of G^{3}M and the basis to compute river conductances, show that heads computed with the G^{3}M are mostly much lower except in regions with a shallow GW (Fig. 9c). This cannot be attributed to the 100 times lower spatial resolution per se but to the selection of the 30th percentile of the 30'' as the SW drainage level. Outliers in the upper half of the scatter plot, with much larger heads than the initial values, are mainly occurring in steep mountain areas like the Himalayas where the 5’ model is not representing smaller valleys with a lower head.

For the continental US, the computationally expensive 1-km integrated hydrological model ParFlow (Maxwell et al., 2015), fits much better to observations than G^{3}M (Figs. 9d, e). G^{3}M produces a generally lower water table (Fig. 9f), a main reason being that ParFlow assumes an impermeable bedrock at a depth of more than 100 m below the land surface elevation. Plotting hydraulic head instead of depth to GW has the disadvantage that the goodness of fit is dominated by the topography as the observed heads are calculated based on the surface elevation of the model. Even though hydraulic heads are a direct result of the model and are forcing lateral GW flows, depth to GW is more relevant for processes like capillary rise. For G^{3}M, there is almost no correlation between depth to GW observations and simulated values. To our knowledge, no publication on large-scale GW modeling presents correlations of simulated with observed depth to GW.

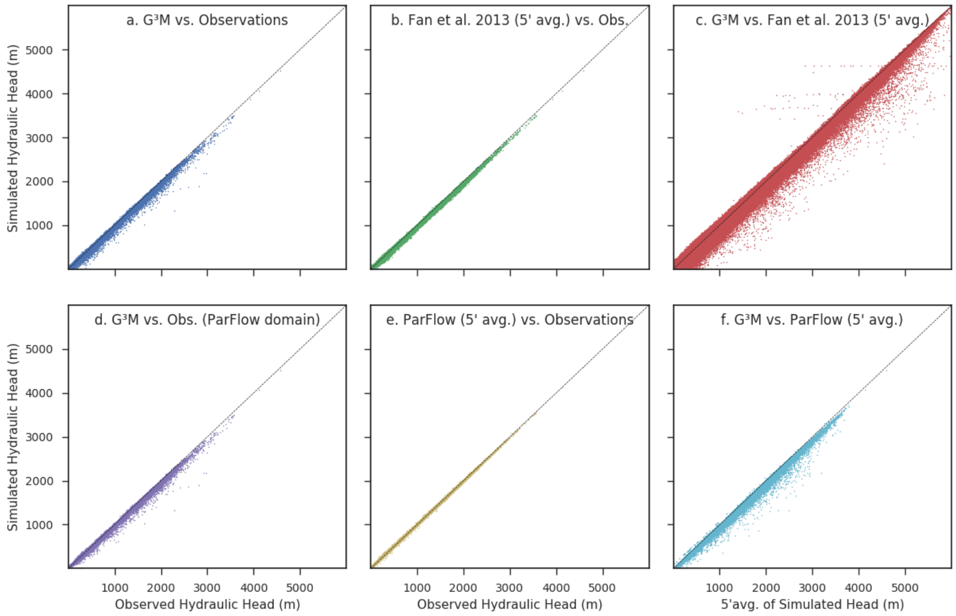

Figure 9 Scatterplots of simulated vs. observed hydraulic head and inter-model comparison of heads. (Upper panel) The steady-state run of G^{3}M vs. observations (a), the 5' average of the equilibrium head of Fan et al. (2013) vs. observations (b) and the avg. equilibrium vs. G^{3}M (c). (Lower panel) The steady-state run of G^{3}M vs. observations only for the ParFlow domain (d), the 5' average of the ParFlow average annual GW table (Maxwell et al., 2015) vs. observations (e) and the steady-state run of G^{3}M vs. 5' average of the ParFlow average annual GW table (f).

3.6 Case study Central Valley

To evaluate G^{3}M further, its results were analysed for to a well-studied region, the Central Valley in California, USA. The Central Valley is one of the most productive agricultural regions of the world and heavily relies on GW pumpage to meet irrigation demands (Faunt et al., 2016). GW pumping in the valley increased rapidly in the 1960s (Faunt, 2009). Figure 10a shows simulated depth to GW for the Central Valley, the coast and the neighboring Sierra Nevada mountainside as well as parts of the Great Basin. The depth to GW table represents natural conditions without any pumping and is rather small. It roughly resembles the depth to GW assumed in the Central Valley Hydrological Model (CVHM) as initial condition, representing a natural state (Faunt, 2009) (Fig. 10b). G^{3}M correctly computes the shallow conditions with groundwater above the surface in the north, partially in the south of the valley and decreasing towards the Sierra Nevada. The difference in the extend of flooded area could be due to large wetlands areas still present in the early 60s which are not represented in this extent in the data used by G^{3}M. Beyond the CVHM domain, depth to GW in mountainous regions is probably overestimated by G^{3}M. The elevation of neighboring cells may differ up to a 1000 meter resulting in a large gradient (Fig. S6b and S6e).

Kommentiert [RR14]: #2.9

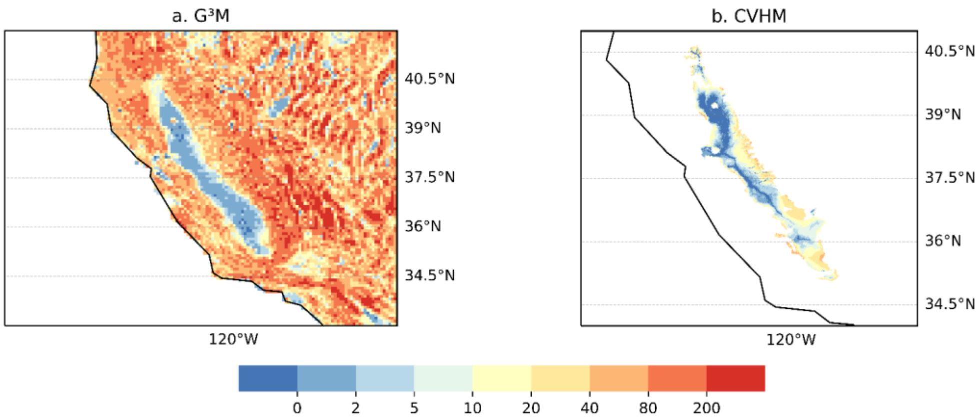

Figure 10 Plots of depth to GW [***m***] as calculated by G³M for the Central Valley and the Great Basin (a), and as used by CVHM as the natural state and starting condition (Faunt, 2009) (b).

4 Discussion

The main aim of global gradient-based groundwater flow modelling with G³M is to better simulate water exchange between SW and GW, for example for an improved estimation of GW resources in dry regions of the globe that are augmented by focused recharge from SW bodies. A major challenge for simulating GW-SW interactions (but also capillary rise) at the global scale is the large size of grid cells that is required due to computational constraints. Within the 5' grid cells, land surface elevation at the scale of 30" very often varies by more than 20 m, and often by 200 m and more (Fig. S1), while the vertical position of the cell and the hydraulic head are approximated in the model by just one value. The question is whether head-dependent flows between grid cells, between GW and SW and from GW to soil (capillary rise) can be simulated successfully at the global scale, i.e. whether an improved quantification of these flows as compared to the simple linear reservoir model currently used in most GHMs can be achieved by this approach. This question cannot be answered yet as we have not yet achieved a dynamic coupling of G³M with a global hydrological model but one may speculate that some innovative approach to take into account the elevation variations within the grid cells may be needed.

It is difficult to assess performance of the presented steady-state G³M results. Model performance is assessment is hindered by data availability and the coarse model resolution. (1) To our knowledge the data collection of depth to groundwater by Fan et al. (2013) is unique. However, they do not represent steady-state values. Apart from depth to groundwater observations, hardly any relevant data is available at the global scale. Especially exchange between surface water and groundwater is difficult to measure even at the local scale. Therefore, we compared G³M results with the results from other large-scale models. Comparison to the results of catchment-scale groundwater flow models is planned for transient runs that will be possible after integration into WaterGAP. (2) Scale differences make the comparison to point observations of depth to groundwater difficult. Multiple local observations within a 5' cell may strongly vary, maybe just due to land surface elevation variations within the approximately 80 km² large cells (compare Fig. S1 and S8). Often, observations are biased towards alluvial aquifers in valleys. The calculated hydraulic head of the grid cell may represent the average groundwater level per grid cell correctly but can be still far off the local observations of depth to groundwater. As the current model only presents an uncalibrated natural steady-state, a comparison to observations only provides a first indicator where the model and the performance measurements needs to be improved as me move to a fully transient model.

Kommentiert [RR15]: #1.1, #1.3

The presented comparison to other large-scale models is based on the assumption that same model deficiencies e.g. in available data and scale issues can uncover differences in model decision. A comparison to catchment scale models is challenging as scales can differ by multiple magnitudes. As the model is further developed towards a transient model the presented comparison to simulations in data-rich regions need to be extended and temporal changes in interactions with surface water investigated.

Kommentiert [RR16]: #1.8

The comparison to the initial state (based on historical observations) of the CVHM model presents a first comparison within a data-rich region which provides also the future possibility of comparing transient model results and human impact on a regional scale. G³M is able to reproduce the shallow groundwater table in the early 1960s. Differences are likely due to the steady-state approach and the connected assumptions on surface water bodies.

Kommentiert [RR17]: #2.9

The presented development of the uncoupled steady-state global GW flow model enabled us to better understand how the spatial hydraulic head pattern relates to the fundamental drivers topography, climate and geology (Fan et al., 2007) and how the interaction to SW bodies govern the global head distribution. Simulated depth to groundwater is particularly affected by the assumed hydraulic head in SW bodies, the major GW drainage component in the model. As rivers represent a natural occurring drainage at the lowest point in a given topography, one would assume that the minimum elevation 30'' land surface elevation per 5' grid cell is a reasonable choice. Experiments have shown that this will induce a head distribution well below the average 5' elevation that is much below observations of Fan et al. (2013). We also tested setting h_{swb} to the average elevation of all "blue" cells (with a depth to GW of less than 0.25 m) of the steady-state 30'' water table results of Fan et al. (2013) that indicate the locations were GW discharges to the surface. This leads to an overall underestimation of the observed hydraulic heads (Fig. S9). Furthermore, it leads to an increase in losing SW bodies (comp. Fig. S11 with Fig. 4). However, it is difficult to judge whether this improves the simulation. More stretches of the Nile and its adjacent wetlands and also of the Niger wetlands and rivers in Northeastern Brazil are losing in case of lower h_{swb}, which appears to reasonable. Additionally, choosing the average as SW elevation provides on the on hand a better fit to observations (Fig. S9) but leads to a wold wide flooding with largely overestimated heads (Fig. S10) and a much longer convergence time due to an increased oscillation between gaining and losing conditions.

The problem is very likely one of scale. This is supported by the fact that both high-resolution models, even the simple one of Fan et al. (2013) fit better to observations than the low-resolution model G³M (Fig. 9). In case of high resolution (e.g. 30''), there are a number of grid cells at an elevation above the average 5' land surface elevation, leading to higher hydraulic heads in parts of the 5' area that drain towards the SW body in a lower 30'' grid cell. In case of the low spatial resolution of 5' in which h_{swb} is set to the elevation of the fine-resolution drainage cell, the 5' hydraulic head is rather close to this (low) elevation (Fig. S12), resulting in an underestimation of hydraulic head and thus an overestimation of depth to GW. While it is plausible and necessary to assume that there is SW-GW interaction within each of in the approximately 80 km², this is not the case for the two orders of magnitude smaller 30'' grid cells. Thus, with the high resolution, heads are not strongly controlled everywhere by the head in SW bodies. Selecting the 30th percentile of the 30'' land surface elevation as h_{swb} was found, by trial-and-error, to lead to a hydraulic head distribution that fits reasonably well to observed head. It avoids that the simulated GW table drops to low while avoiding the excessive flooding that occurs if h_{swb} is set to the average of 30'' land surface elevations, i.e. the 5' land surface elevation (Fig. S9).

The constraint that the selected h_{swb} value puts on simulated hydraulic heads is also linked to the conductance of the SW bodies. A higher conductance will lead to aquifer heads closer to h_{swb}. If the hydraulic head drops below the bottom level of the SW body, the hydraulic gradient is assumed to become 1 and the SW body recharges the GW with a rate of K_{aq} per unit SW body area. In case of a K_{aq} value of 10^{-5} m s^{-1}, the SW body would lose approximately 1 m of water each day. It is to be investigated how the sensitivity to choice of SW body elevation and conductance leads to a solution that fits observations best. A lower conductance may lead to a higher groundwater table as SW bodies don't drain as much water; on the other hand, they seem to provide an important recharge mechanism in the steady-state model for some regions preventing an even higher depth to GW. The simple conductance approach applied in G³M could possibly be approved by the approach proposed by Morel-Seytoux et al. (2017).

de Graaf et al. (2015) set their SW head (h_{swb}) to the land surface elevation of the 6' grid cells minus river depth at bankfull conditions plus water depth at average river discharge. Together with the missing interaction between lakes and wetlands and a different approach to river conductance, this might be a reason for the additional drainage above the floodplain that was necessary to avoid excessive flooding, and that is not needed in G^{3}M. On the other hand, this adaption allows the drainage of water even if the hydraulic head is below the SW elevation that might have led to the global underestimation of hydraulic heads. Thus, the difference in model heads seems to be closely related to the sensitivity of SW body elevation.

Kommentiert [RR18]: #2.8

As described above, G³M differs from regional groundwater models due to grid cell size in that it is more conceptual and cannot capture actual variability of topography, aquifer depth (Richey et al., 2015) and (vertical) heterogeneity of subsurface properties. The lack of information about the three-dimensional distribution of hydraulic conductivity is expected to negatively impact the quality of simulated GW flow. For example, the lateral conductivity and connectivity of groundwater along thousands of kms from e.g. the Rocky Mountains in the Central USA to the coast as well as the vertical connectivity is likely to be overestimated by G^{3}M, as vertical faults and interspersed aquitards are not represented; this leads to an underestimation of hydraulic head in those mountainous areas.

5 Conclusions

We have presented the concept and first results of a new global gradient-based GW flow model that is to be coupled to the GHM WaterGAP. The uncoupled steady-state model has provided important insights into challenges of global GW flow modelling mainly related to the necessarily large grid cells size (5' by 5') as well as first global maps of SW-GW interactions. Simulated heads were found to be strongly impacted by assumption regarding the interaction with SW bodies, in particular the selected elevation of the SW table and the prescribed conductance. We have demonstrated that simulated G^{3}M hydraulic heads fit better to observed heads than the heads of the comparable steady-state GW model of de Graaf et al. (2015),without requiring additional drainage that would prevent a full coupling to a GHM.

The presented results are the first step towards a fully coupled model in which SW heads are computed as a function of surface water hydrology and GW abstractions can be taken into account. Especially the interaction with SW bodies that can run dry will make the model behavior more realistic. The fully coupled model will simulate transient behaviour reflecting climate variability and change. Simulated hydraulic head dynamics will be compared to observed head time series as well as to the output of large-scale regional models, while total water storage variations will be compared to GRACE satellite data. However, it will be challenging to judge the quality of simulated GW-SW interactions due to a scarcity of observations.

6 Code and data availability

The model-framework code is available at globalgroundwatermodel.org or at DOI: 10.5281/zenodo.1175540 with a description on how to compile and run a basic GW model. The code is available under the GNU General Public License 3. Model output is available at DOI: 10.5281/zenodo.1315471.

Kommentiert [RR19]: #3.1, #3.2

Acknowledgments

Part of this study was funded by a Friedrich-Ebert foundation Ph.D. fellowship. We are very grateful to Ying Fan and Gonzalo Miguez-Macho for fruitful discussions and data provisioning. We thank the two reviewers for their thoughtful comments that helped to improve the manuscript.

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrological Sciences Journal, 48, 317–337, doi:10.1623/hysj.48.3.317.45290, 2003.

Allen, P. M., Arnold, J. C., and Byars, B. W.: Downstream channel geometry for use in planning-level models, J Am Water Resources Assoc, 30, 663–671, doi:10.1111/j.1752-1688.1994.tb03321.x, 1994.

Belcher, W. R. and Sweetkind, D. S.: Death Valley regional groundwater flow system, Nevada and California-Hydrogeologic framework and transient groundwater flow model, U.S. Geological Survey Professional Paper, 398, 2010.

Costa, A. C., Foerster, S., de Araújo, J. C., and Bronstert, A.: Analysis of channel transmission losses in a dryland river reach in north-eastern Brazil using streamflow series, groundwater level series and multi-temporal satellite data, Hydrol. Process., 27, 1046–1060, doi:10.1002/hyp.9243, 2013.

de Graaf, I. E.M. de, Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model, Hydrol. Earth Syst. Sci., 19, 823–837, doi:10.5194/hess-19-823-2015, 2015.

de Graaf, I. E.M. de, van Beek, R. L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.P.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, Advances in Water Resources, 102, 53–67, doi:10.1016/j.advwatres.2017.01.011, 2017.

Dogrul, E., Brush, C., and Kadir, T.: Groundwater Modeling in Support of Water Resources Management and Planning under Complex Climate, Regulatory, and Economic Stresses, Water, 8, 592, doi:10.3390/w8120592, 2016.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12, 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B.R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, Journal of Geodynamics, 59-60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, Journal of Hydrology, 270, 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, doi:10.1002/2014WR015595, 2014.

Eisner, S.: Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients, University of Kassel, Kassel, Germany, 2016.

Elsawwaf, M., Feyen, J., Batelaan, O., and Bakr, M.: Groundwater-surface water interaction in Lake Nasser, Southern Egypt, Hydrol. Process., 28, 414–430, doi:10.1002/hyp.9563, 2014.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science (New York, N.Y.), 339, 940–943, doi:10.1126/science.1229881, 2013.

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, J. Geophys. Res., 112, doi:10.1029/2006JD008111, 2007.

FAO: Irrigation potential in Africa: http://www.fao.org/docrep/w4347e/w4347e0i.htm, 1997, last access: 1 May 2018.

Faunt, C. C. (Ed.): Groundwater Availability of the Central Valley Aquifer, California, U.S. Geological Survey Professional Paper, 1776, 225 pp., 2009.

Faunt, C. C., Sneed, M., Traum, J., and Brandt, J. T.: Water availability and land subsidence in the Central Valley, California, USA, Hydrogeol J, 24, 675–684, doi:10.1007/s10040-015-1339-x, 2016.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nature Geosci, 9, 161–167, doi:10.1038/ngeo2590, 2016.

Gleeson, T., Moosdorf, N., Hartmann, J., and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898, doi:10.1002/2014GL059856, 2014.

Gleeson, T., Wada, Y., Bierkens, M. F. P., and van Beek, L. P. H.: Water balance of global aquifers revealed by groundwater footprint, Nature, 488, 197–200, doi:10.1038/nature11295, 2012.

Gossel, W., Ebraheem, A. M., and Wycisk, P.: A very large scale GIS-based groundwater flow model for the Nubian sandstone aquifer in Eastern Sahara (Egypt, northern Sudan and eastern Libya), Hydrogeology Journal, 12, 698–713, doi:10.1007/s10040-004-0379-4, 2004.

Harbaugh, A. W.: MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process, US Department of the Interior, US Geological Survey Reston, 2005.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophysics, Geosystems, 13, doi:10.1029/2012GC004370, 2012.

Konikow, L. F.: Contribution of global groundwater depletion since 1900 to sea-level rise, Geophys. Res. Lett, 38, doi:10.1029/2011GL048604, 2011.

Krakauer, N. Y., Li, H., and Fan, Y.: Groundwater flow across spatial scales: importance for climate modeling, Environ. Res. Lett., 9, 34003, doi:10.1088/1748-9326/9/3/034003, 2014.

Lange, S.: EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), 2016.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, 296, 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Maxwell, R. M., Condon, L. E., and Kollet, S. J.: 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, 923–937, doi:10.5194/gmd-8-923-2015, 2015.

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., and Robock, A.: Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, J. Geophys. Res., 112, D13108, doi:10.1029/2006JD008112, 2007.

Morel-Seytoux, H. J., Miller, C. D., Miracapillo, C., and Mehl, S.: River Seepage Conductance in Large-Scale Regional Studies, Ground Water, 55, 399–407, doi:10.1111/gwat.12491, 2017.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Naff, Richard L., and Edward R. Banta: The US Geological Survey modular ground-water model-PCGN: a preconditioned conjugate gradient solver with improved nonlinear control, Geological Survey (US), 1331, 2008.

Reinecke, R.: G³M-f a global gradient-based groundwater modelling framwork, JOSS, 3, 548, doi:10.21105/joss.00548, 2018.

Richey, A. S., Thomas, B. F., Lo, M.-H., Famiglietti, J. S., Swenson, S., and Rodell, M.: Uncertainty in global groundwater storage estimates in a Total Groundwater Stress framework, Water Resour. Res, 51, 5198–5216, doi:10.1002/2015WR017351, 2015.

Saad, Y.: ILUT: A dual threshold incomplete LU factorization, Numer. Linear Algebra Appl., 1, 387–402, doi:10.1002/nla.1680010405, 1994.

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, PNAS, 109, 9320–9325, doi:10.1073/pnas.1200311109, 2012.

Sheets, R. A., Hill, M. C., Haitjema, H. M., Provost, A. M., and Masterson, J. P.: Simulation of water-table aquifers using specified saturated thickness, Ground Water, 53, 151–157, doi:10.1111/gwat.12164, 2015.

Stonestrom, D. A., Constantz, J., Ferre, T. P.A., and Leake, S. A.: Ground-water recharge in the arid and semiarid southwestern United States, U.S. Geological Survey Professional Paper 1703, 414 p, 2007.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin, Hydrol. Earth Syst. Sci., 15, 2913–2935, doi:10.5194/hess-15-2913-2011, 2011.

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I., and Treidel, H.: Ground water and climate change, Nature Climate change, 3, 322–329, doi:10.1038/nclimate1744, 2012.

UNESCO-IHP, IGRAC, WWAP: GEF-TWAP Methodology Transboundary Aquifers.: http://isarm.org/sites/default/files/resources/files/TWAP%20Methodology%20Groundwater%20Component%20%28Revised%20Aug%202012%29.pdf, 2012, last access: 9 October 2017.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, doi:10.1029/2010WR009791, 2011.

Vergnes, J.-P., Decharme, B., Alkama, R., Martin, E., Habets, F., and Douville, H.: A Simple Groundwater Scheme for Hydrological and Climate Applications: Description and Offline Evaluation over France, J. Hydrometeor, 13, 1149–1171, doi:10.1175/JHM-D-11-0149.1, 2012.

Vergnes, J.-P., Decharme, B., and Habets, F.: Introduction of groundwater capillary rises using subgrid spatial variability of topography into the ISBA land surface model, J. Geophys. Res. Atmos., 119, 11, doi:10.1002/2014JD021573, 2014.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F., and Alcamo, J.: Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe, Journal of Hydrology, 424-425, 238–251, doi:10.1016/j.jhydrol.2012.01.005, 2012.

Wada, Y.: Modeling Groundwater Depletion at Regional and Global Scales: Present State and Future Prospects, Surv Geophys, 37, 419–451, doi:10.1007/s10712-015-9347-x, 2016.

Wada, Y., van Beek, L. P. H., and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: A global assessment, Water Resour. Res., 48, doi:10.1029/2011WR010562, 2012.

Table 1 Model parameter values, input data sources and other information about the steady-state simulation.

Parameter	Symbol	Units	Description	Eq. No.
Landmask	-	-	Location and area of 2161074 cells at 5' resolution based on WaterGAP (Eisner, 2016))	-
GW recharge	R_g	L^3T^{-1}	Mean annual diffuse GW recharge 1901–2013 of WaterGAP 2.2c (Müller Schmied et al., 2014) forced with EWEMBI (Lange, 2016), spatial resolution 0.5° (Fig. S4)	1,3,7
Hydraulic conductivity	K_{aq}	LT^{-1}	Derived from Gleeson et al., 2014 (Fig. S3)	2,5
Hydraulic head	$h_{(aq)}$	L	Head of the aquifer in a computational cell, initial estimate based on 5' average of 30" head of Fan et al. (2013)	2,4,7,8
Ocean boundary conductivity	c_{ocean}	L^2T^{-1}	$10\ m^2\ day^{-1} = 0.1\ m\ day^{-1}\ 10\ km\ 10\ km^{-1}\ 100\ m$, with K of $10^{-6}\ m\ s^{-1}$ and a distance of 10 km from the cell center to the boundary with a cell thickness of 100 m	3,4
Ocean boundary head	h_{ocean}	L	Global mean sea-level of 0 m	4
SW head	h_{swb}	L	30 % quantile (P_{30}) of 30" land surface elevation of Fan et al. (2013) per 5' grid cell	5
SW bottom elevation	B_{swb}	L	2 m (wetlands), 10 m (local lakes), 100 m (global lakes) below P_{30}	5
Area of global and lcal lakes and global and local wetlands	WL	L^2	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016),	6
Length of the river	L	L	Per 5' grid cell, based on WaterGAP 3 (Eisner, 2016)	6
Width of the river	W	L	Per 5' grid cell, based d on WaterGAP 3 (Eisner, 2016)	6
River head	h_{riv}	L	h_{swb}	6,7,8
River bottom elevation	B_{riv}	L	$h_{riv} - 0.349 \times Q_{bankfull}{}^{0.341}$ (Allen et al., 1994)	6
Equilibrium hydraulic head	h_{eq}	L	Steady-state hydraulic head of Fan et al. (2013) (averaged to 5' from original spatial resolution of 30")	7
Layers	-	-	2 confined, 100 m thick each	-
Land surface elevation	-	L	5' average of 30" digital elevation map of Fan et al. (2013) (Fig. S2)	-
E-folding factor	-	-	Applied only to lower layer for 150 m depth, based on area-weighted average of Fan et al. (2013)	-
Timestep	t	T	Daily timestep	-
Convergence criterion	-	L	$\|\|$hydraulic head residuals$\|\|_{inf} < 10^{-100}$ and max head change < 10 m	-
Inner iterations	-	-	50 inner iterations between Picard iterations (Naff, Richard L., and Edward R. Banta, 2008)	-

Supplement

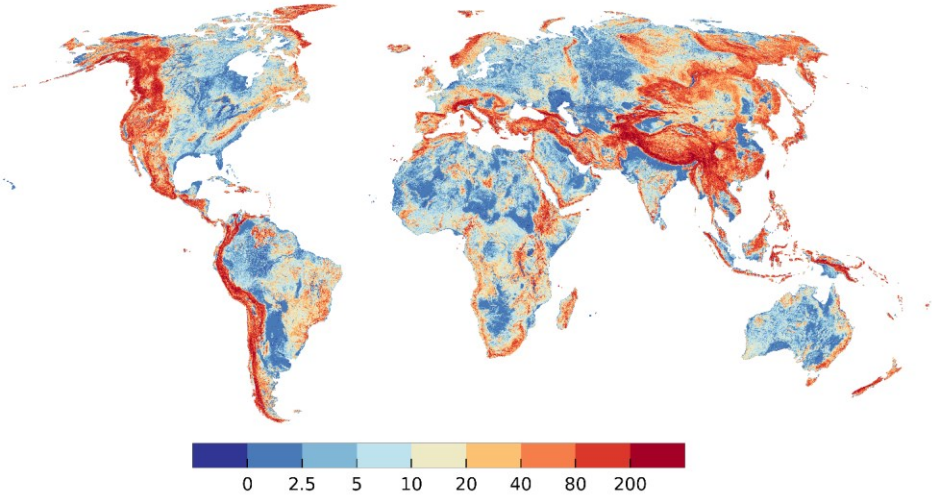

Figure S1 Difference [$\boldsymbol{m}$] between mean elevation and P_{30} elevation. Maximum value 1365 m**.**

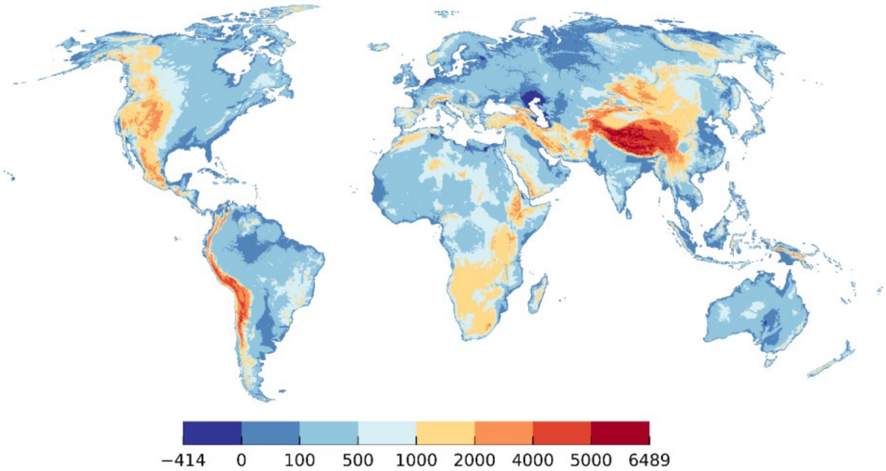

Figure S2 Land surface elevation [$\boldsymbol{m}$] used in G³M: 5' average of 30"land surface elevation used in Fan et al. (2013).

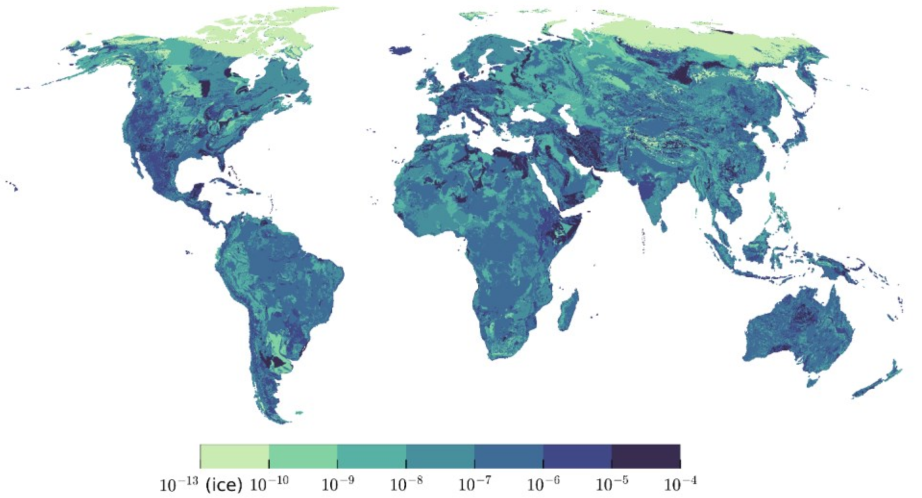

Figure S3 Hydraulic conductivity [ms^{-1}] derived from Gleeson et al. (2014) by scaling it with the geometric mean to 5'. Very low values in the northern hemisphere are due to permafrost conditions.

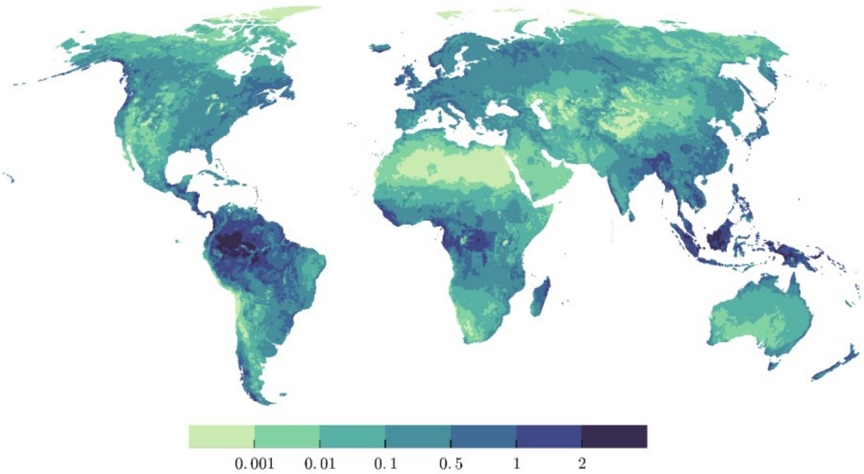

Figure S4 Mean annual groundwater recharge [$mm\ day^{-1}$] between 1901-2013, from WaterGAP 2.2c.

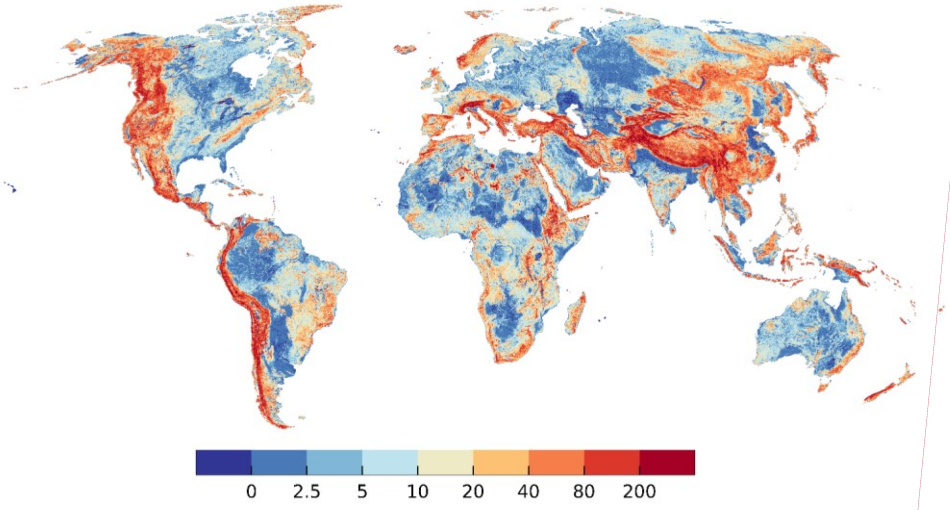

Figure S5 Arithmetic mean [***m***] of the 30'' land surface elevation per 5' grid cell and simulated equilibrium hydraulic head (simulated depth to GW). Maximum value 2070 m, minimum value -414 m (Extremes included in dark blue and dark red).

Kommentiert [RR20]: Replaced by old fig03, S5 was moved to new fig03

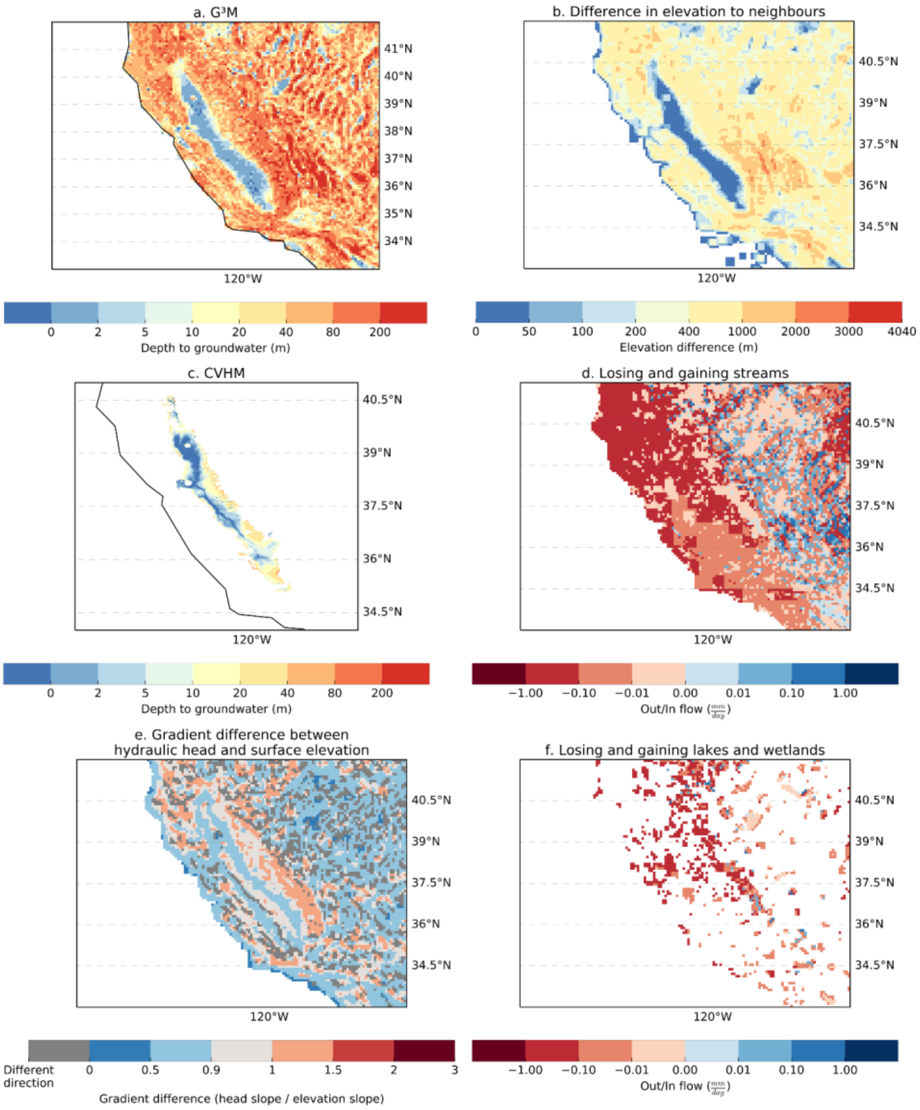

Figure S6 Plots of depth to GW as calculated by G³M (a), difference in surface elevation to neighbouring cells (b), depth to GW as used by the CVHM as the natural state and starting condition (Faunt, 2009) (c), losing and gaining streams as calculated by G³M (d), difference in gradient of hydraulic head and surface elevation (e), losing and gaining lakes and wetlands as calculated by G³M for the Central Valley and the Great Basin.

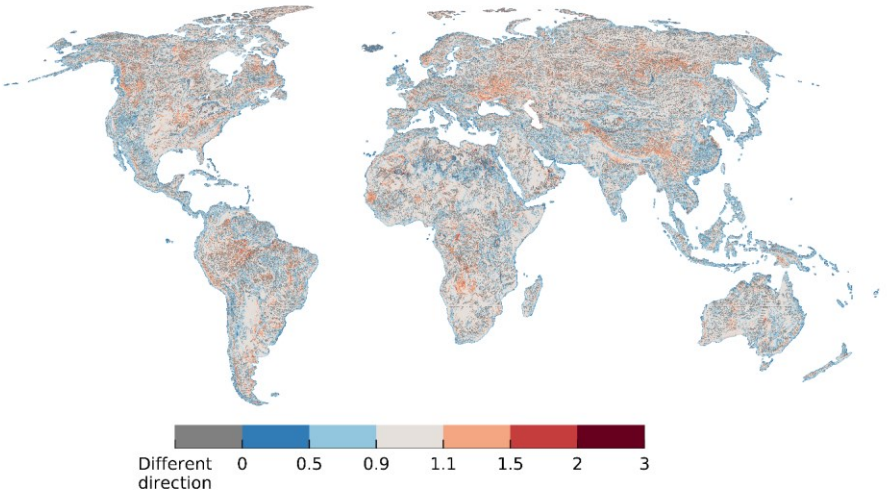

Figure S7 Ratio of hydraulic head gradient to 5' mean surface elevation gradient, only computed if the difference in direction of the gradient was smaller than 45°.

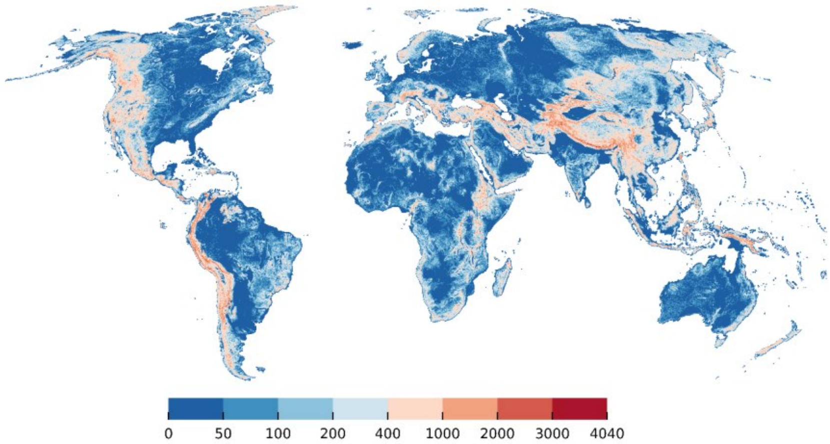

Figure S8 Land surface elevation Difference of 30'' mean land surface elevation in 5' grid cell to mean elevation of neighbouring cells [**m**] to mean elevation of neighboring cells on 5' resolution.

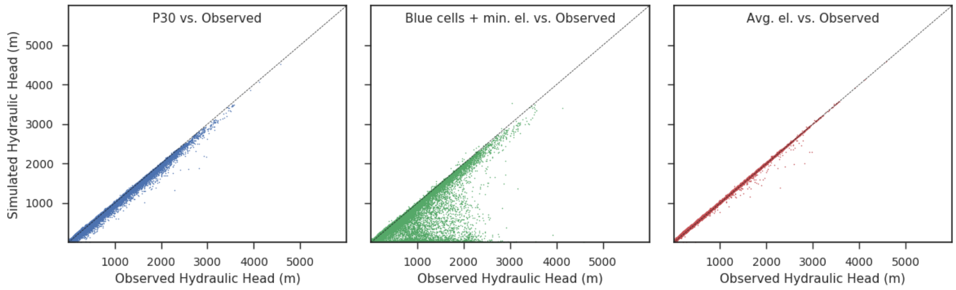

Figure S9 Comparison between three alternatives for setting $\boldsymbol{h_{swb}}$. Left to right: Fit of simulated hydraulic heads observations if $\boldsymbol{h_{swb}}$ is set (1) to the 30th percentile of the 30'' land surface elevations (standard model) , (2) alternatively to the average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) or (3) is set to the average of the 30'' land surface elevations. A blue cell has a depth to GW of less than 0.25 m and indicates GW discharge to the surface. If no "blue" cell exists in the 5' cell, the minimum elevation of the 30'' land surface elevation values within the cell was used.

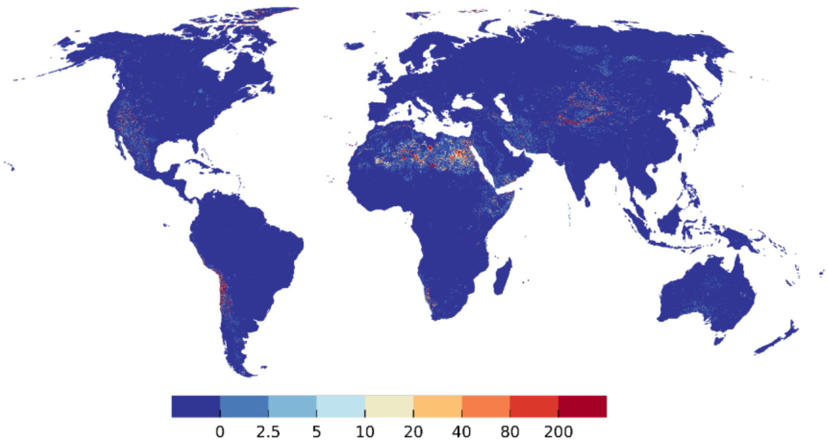

Fig. S10 Depth to groundwater [$\boldsymbol{m}$] for SW body elevation at average of 30'' land surface elevations.

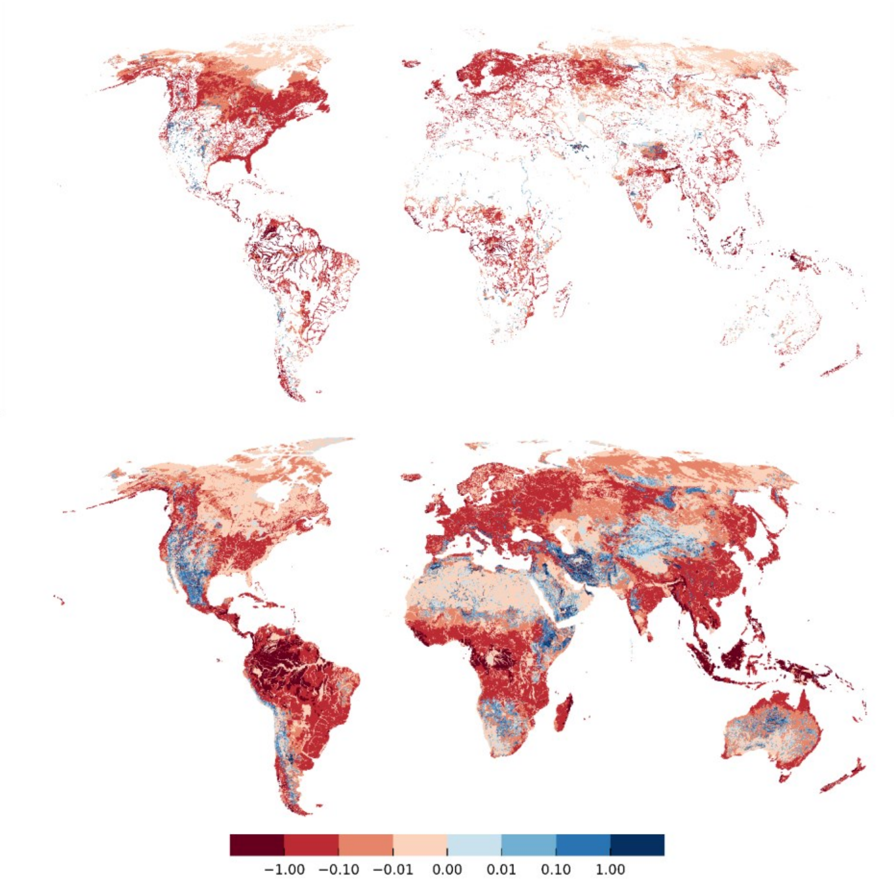

Figure S11 Gaining and losing rivers (lower panel) and wetlands and lakes (upper panel) as flow into/out the GW [$\boldsymbol{mm\ day^{-1}}$] if $\boldsymbol{h_{swb}}$ is set to average elevation of all "blue" cells of the 30'' water table results of Fan et al. (2013) (right). A blue cell is defined as a depth to groundwater of less than 0.25 m. If no "blue" cell exist in the 5' cell, the minimum elevation of the 30'' land surface elevation values is used. Red denotes gaining SW bodies.

Kommentiert [RR21]: S13 moved to new fig03