

Importance of Spatial Resolution in Global Groundwater Modeling

by Robert Reinecke^{1,2}, Alexander Wachholz^{1,3}, Steffen Mehl⁴, Laura Foglia⁵, Christoph Niemann¹, and Petra Döll^{1,6}

Abstract

Global-scale gradient-based groundwater models are a new endeavor for hydrologists who wish to improve global hydrological models (GHMs). In particular, the integration of such groundwater models into GHMs improves the simulation of water flows between surface water and groundwater and of capillary rise and thus evapotranspiration. Currently, these models are not able to simulate water table depth adequately over the entire globe. Unsatisfactory model performance compared to well observations suggests that a higher spatial resolution is required to better represent the high spatial variability of land surface and groundwater elevations. In this study, we use New Zealand as a testbed and analyze the impacts of spatial resolution on the results of global groundwater models. Steady-state hydraulic heads simulated by two versions of the global groundwater model G³M, at spatial resolutions of 5 arc-minutes (9 km) and 30 arc-seconds (900 m), are compared with observations from the Canterbury region. The output of three other groundwater models with different spatial resolutions is analyzed as well. Considering the spatial distribution of residuals, general patterns of unsatisfactory model performance remain at the higher resolutions, suggesting that an increase in model resolution alone does not fix problems such as the systematic overestimation of hydraulic head. We conclude that (1) a new understanding of how low-resolution global groundwater models can be evaluated is required, and (2) merely increasing the spatial resolution of global-scale groundwater models will not improve the simulation of the global freshwater system.

Introduction

Groundwater is the largest source of available freshwater (Gleeson et al. 2016). The assessment of global groundwater resources has been the subject of multiple studies recently, including impacts of

groundwater abstractions (Wada et al. 2010) and climate change (Portmann et al. 2013; Cuthbert et al. 2019). Furthermore, global hydrological models (GHMs) started to replace their bucket-like linear groundwater storage models with (hydraulic) head gradient-based models to better represent the interaction between surface water and groundwater as well as lateral and vertical flows, including capillary rise (de Graaf et al. 2015, 2017, 2019; Reinecke et al. 2019a, 2019b). Compared to established regional models, global groundwater models are a new endeavor for groundwater hydrologists that is mostly unexplored. These models use a rather coarse spatial resolution (1) due to a lack of available global data and (2) to handle the computational demand that arises from fully coupling groundwater models to surface hydrology models. In typical GHMs, each cell has a river, in addition to other possible surface water bodies, that can cause changes in the flows to and from the groundwater. These changes represent an enormous challenge for the numerical solution while sustaining reasonable run times that allow for sensitivity analysis and data assimilation. Currently, the available global groundwater models have grid cell sizes that range from 30'' (arc-seconds, ~900 m by 900 m at the equator) to 5' (arc-minutes, ~9 km by 9 km at the equator) (Reinecke et al. 2019a). In comparison, the spatial resolution of traditional groundwater models ranges from

¹Corresponding author: Institute of Physical Geography, Goethe University Frankfurt, Altenhöferallee 1, 60438 Frankfurt, Germany; reinecke@em.uni-frankfurt.de

²International Centre for Water Resources and Global Change (UNESCO), Federal Institute of Hydrology, 56002, Koblenz, Germany

³Department of Aquatic Ecosystem Analysis, Helmholtz Centre for Environmental Research, Magdeburg, Germany; alexander.wachholz@ufz.de

⁴Department of Civil Engineering, California State University, Chico, CA; smehl@csuchico.edu

⁵Department of Land, Air and Water Resources, University of California, Davis, CA; lfoglia@ucdavis.edu

⁶Senckenberg Leibniz Biodiversity and Climate Research Centre (SBIK-F), Frankfurt, Germany; p.doell@em.uni-frankfurt.de

Article impact statement: Global groundwater models are a new endeavor for hydrologists that face multiple challenges.

Received July 2019, accepted March 2020.

© 2020 The Authors. *Groundwater* published by Wiley Periodicals, Inc. on behalf of National Ground Water Association.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

doi: 10.1111/gwat.12996

centimeters for transport models, for example, García-Gil et al. (2016), over 1 m to 50 m for local models, for example, Limberg et al. (2010), up to 1 mile for extensive regional models like the Central Valley Model CVHM (Faunt et al. 2016).

Unfortunately, global models show unsatisfying performance for simulated head compared with well observations, often by more than 100 m (de Graaf et al. 2017; Reinecke et al. 2019a). The reasons for these significant over- and underestimates are likely the spatial resolution of the model, improper comparisons to observations, lack of high-resolution data, and assumptions of surface water body location and extent. Higher-resolution models (e.g., 30"), like the global steady-state model of Fan et al. (2013), seem to exceed the performance (much smaller difference to global observations) of the coarse-scale (5') models (Reinecke et al. 2019a), leading to the assumption that an increase in spatial resolution will improve model performance. In general, the hydrological community has argued that a higher spatial resolution in all earth system models is necessary to represent essential processes better (Wood et al. 2011). The topic of scales in hydrological systems has also been extensively discussed for fractal scaling of river networks, topography and spatial scaling of conductivity (Mark and Aronson 1984; Rosso et al. 1991; Gupta et al. 2007; Wörman et al. 2007).

One of the primary purposes for integrating gradient-based groundwater models into GHMs is to improve the simulation of the flows between surface water bodies and groundwater. In particular, the dynamic simulation of flows between surface water bodies and groundwater cannot be achieved with the bucket-like linear groundwater reservoir models in current GHMs (Reinecke et al. 2019a). To compute the flow between a surface water body and groundwater, the hydraulic head of the water body is needed. Determination of this head as the elevation of the water surface within a computational grid cell depends on the used digital elevation model (DEM), for example, HydroSHEDS (Lehner et al. 2008) and the simulated temporal storage changes in the surface water bodies. For example, if the GHM only contains one river per computation cell (typical for global models), the river elevation accuracy is limited by the spatial resolution of the GHM and the aggregation method used to determine the elevation. GHMs currently use spatial resolutions varying from 2° (~200 km) up to 5' (~9 km), whereas most GHMs use a resolution of 0.5° (Sood and Smakhtin 2015). Contrarily global DEMs are available up to a 12 m resolution (DLR 2016). Krakauer et al. (2014) suggested a spatial resolution of 6' to simulate groundwater flow at continental scales. The two gradient-based groundwater models, coupled to GHMs, use a spatial resolution of 5' and 6', respectively (de Graaf et al. 2017; Reinecke et al. 2019a).

Due to the coarse spatial resolution used in global-scale groundwater models, one simulated head value represents a large volume of the groundwater system, making a comparison to observations extremely challenging. For

example, it is unlikely that wells are equally distributed inside these large cells. So how can one determine an average observed head for a cell from these wells that is an appropriately aggregated representation of reality? An aggregation to an arithmetic average measurement might misrepresent the reality due to not only a nonuniform distribution inside the cell but even more when the wells also differ in their hydrogeologic characteristics (e.g., different aquifer layer). This may result in the undesirable phenomenon that sparsely available observations falsely indicate a better fit than a dense measurement network (Beven 2000).

Related to this issue is that gradient-based models compute heads, but observations are often available as water table depths (WTDs). Thus, to compare these two values, either (1) the simulated head has to be subtracted from the average land surface elevation (based on an available DEM) for the computation cell or (2) the WTD has to be converted to a hydraulic head. Calculation of the head is achieved by either using a known land surface elevation of the WTD measurement location (which is often not available), a high-resolution DEM, or the land surface elevation assigned to the grid cell of the groundwater model. Either way, both methods include the uncertainty of the DEM combined with the error of averaging observations.

In this article, we reevaluate the pursuit of higher spatial resolution for global models in the context of global groundwater models. New Zealand (NZ) is used in this study as a small-world representation with a well-defined ocean boundary condition and shorter runtimes compared to the global model. We present uncertainties in well observations and surface elevation from DEMs in the context of coarse model resolution. Results show the sensitivity of a global groundwater model (G³M) (Reinecke et al. 2019a) to changes in only spatial resolution from 5' to 30" and model outputs compared to available well observations. We focus on the Canterbury region due to the large density of available measurements. By comparing the simulated steady-state WTD of five different global-scale models in this region, we show that model performance is still not acceptable to reliably compute surface water to groundwater interactions on the global scale, and that our community needs to be clear about the limitations and uncertainties of global groundwater models. Based on previous sensitivity analysis results (Reinecke et al. 2019b), we investigate from where uncertainties in the current model approach originate. Finally, we discuss how the results of macro-scale models should be presented and how the current shortcomings can be addressed.

Methods

The Global Groundwater Model

G³M (Reinecke et al. 2019a) is a global gradient-based groundwater model intended to be coupled with the GHM WaterGAP (Döll et al. 2014; Müller

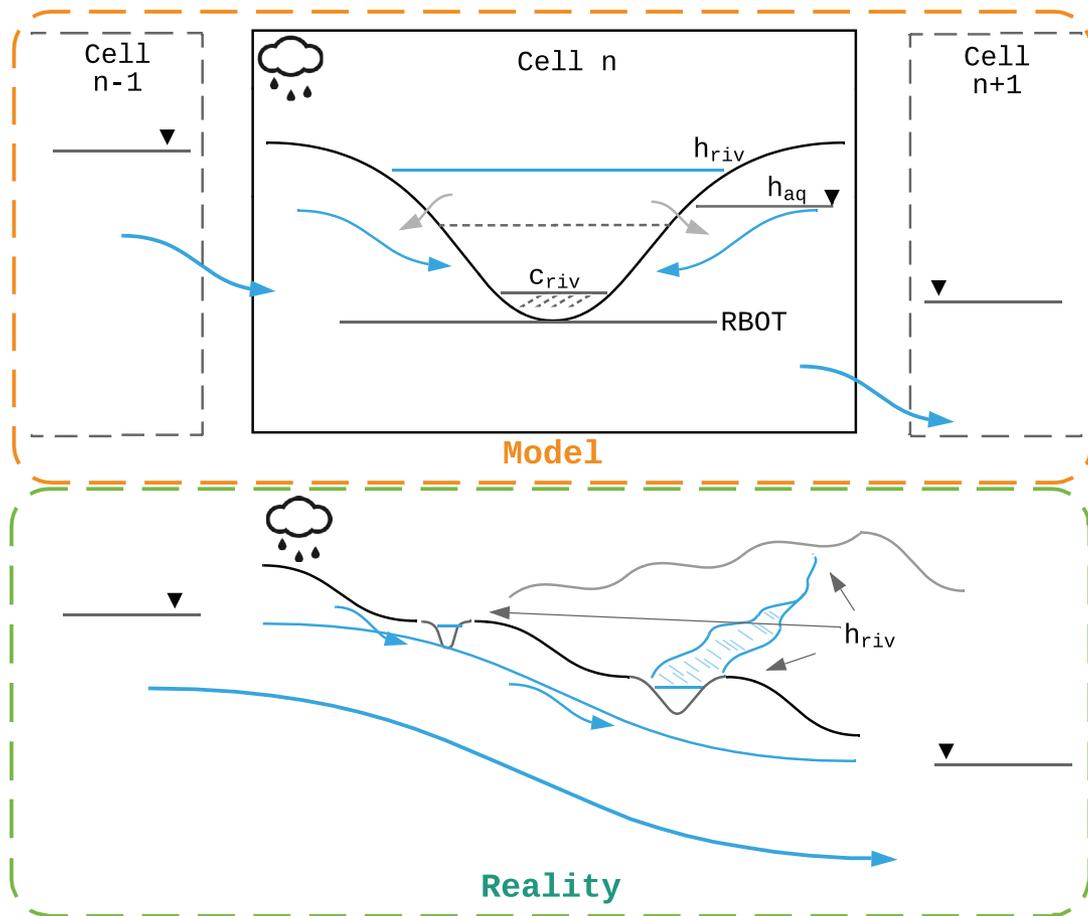


Figure 1. Comparison of the global gradient-based groundwater model approach to equivalent groundwater flow processes. The conceptual groundwater model has only one river in every model grid cell (common in GHMs) and flow between cells is an average over much more complex hydrogeological units and local interactions with surface water. Exchange with the surface water bodies is calculated as a function of the head in the aquifer h_{aq} , the river head h_{riv} , the bottom of the streambed $RBOT$, and a conductance c_{riv} that varies between losing and gaining conditions. Details can be found in Reinecke et al. (2019a).

Schmied et al. 2014) and is based on the Open Source groundwater modeling framework G³M-f (available on globalgroundwatermodel.org) (Reinecke 2018). G³M implements the same saturated groundwater flow equations and processes as many established modeling codes, such as MODFLOW (Harbaugh 2005) (for a detailed discussion see Reinecke et al. 2019a). It computes lateral and vertical groundwater flows as well as exchanges with surface-water bodies for all land areas of the globe (except Antarctica) with a resolution of 5' (ca. 9 km by 9 km at the equator) with two vertical layers with a thickness of each 100 m. The evaluation presented in this study is based on a steady-state variant of the model representing an equilibrium state, without any groundwater pumping (Reinecke et al. 2019a).

Figure 1 shows the conceptual nature of the model. In each model cell, there is one river and possibly multiple other surface water bodies like lakes and wetlands (not shown in Figure 1). Groundwater recharge is based on mean annual groundwater recharge (Döll et al. 2014) computed by WaterGAP 2.2c (Müller Schmied et al. 2014) for the period 1901–2013. This simulated

recharge is uncertain as components leading to this flow are uncertain (e.g., precipitation, computation of evapotranspiration, and total runoff and its partitioning into fast runoff and groundwater recharge) but the calibration constrains total runoff to streamflow observations. WaterGAP also computes consumptive use as net groundwater and surface water abstractions, which can be taken into account as a correction factor on groundwater recharge.

Hydraulic conductivity (K) is derived from GLHYMPS 2.0 (Huscroft et al. 2018). The original data (polygons of different spatial resolution) were gridded to 5' by area-weighted averaging and used as K for the upper model layer. For the second layer, K of the first layer (K_{upper}) is reduced by an e-folding factor f (a calibrated parameter based on terrain slope from Fan et al. [2013]), assuming that K decreases exponentially with depth. K of the lower layer (K_{lower}) is calculated with an equation from Fan et al. (2013), eq. (7) as

$$K_{lower} = \frac{K_{upper}}{\exp\left(-\frac{50}{f}\right)} \quad (1)$$

where 50 is a factor in meters. Application of the e-folding factor reduces the mean K from 10^{-6} m s^{-1} (first layer) to $10^{-11} \text{ m s}^{-1}$ for the second layer.

Testbed: New Zealand

NZ, consisting of a total of 4602 5' cells in two layers, is characterized by a simple continuous boundary condition (the ocean) and diverse topography while providing continuous and dense well observations for some regions. It allows for model exploration with much lower runtimes as compared to simulations for the whole globe. Model parameters and input data are the same as in the global version.

To investigate the sensitivity of model performance to spatial resolution, a 30'' (ca. 900 m by 900 m at the equator) G³M version was set up for NZ, comprising 433,468 model cells. The 30'' model uses the same input as the 5' model except for the land surface elevation and the location of rivers. Total river length and width are equal in both models. Contrary to the 5' model, which contains a river in each model cell, the 30'' model contains the same rivers only in selected cells. These cells are determined using 30'' HydroSHEDS (Lehner et al. 2008) information. From HydroSHEDS we derive the length at 30'' resolution together with information on the drainage relation of the cells expressed as a number of upstream cells. Upstream cells are the number of cells that are assumed to drain into an individual cell. Based on this information Algorithm 1 determines the 30'' cells that contain a river, which are assumed to be those with the largest number of upstream 30'' cells. While it is assumed that all 5' cells contain a river, this is the case only for 14% of the cells in the 30'' model variant.

All other surface water bodies (lakes and wetlands) are incorporated as a fraction of the total grid cell size. For the 5' version, this fraction is based on the area that lakes and wetlands cover in each grid cell, based on the Global Lakes and Wetlands Database GLWD (Lehner and Döll 2004). The same fraction is assigned to each 30'' cell contained in the 5' cell.

In the 5' model, the elevation of the surface water bodies is the 30th percentile (P_{30}) of the 30'' DEM values within the 5' grid cell. This elevation was found to provide the best global fit to observations (Reinecke et al. 2019a). For the 30'' model, the elevation of the surface water bodies was set to the 30'' DEM value.

The Canterbury Region

Our study focuses on the Canterbury region in NZ due to a large number of available groundwater well observations. It is located on the central-eastern South Island (see also Figures 4 and 5). The region covers an area of 44,508 km² and is home to a population of 624,000 (June 2018) (Aotearoa 2018). It consists of the Southern Alps in the hinterland and the Canterbury Plains in the coastal area. While the Southern Alps consist mostly of greywacke and schist, the Canterbury Plains are composed of Quaternary gravel, sand, and silt deposited by rivers and meltwater outwash (Westerhoff and White 2013). In the higher elevations of the Canterbury Plains, aquifers are typically unconfined while in the lower parts, confined conditions occur where marine deposits are interbedded with the terrestrial sediments (Westerhoff and White 2013). The Canterbury Plains is the region with the most groundwater usage in New Zealand (Westerhoff and White 2013). Groundwater recharge occurs through

Algorithm 1: Determining the 30'' inside a 5' cell which contains a river

Result: The 30'' model contains the same total river length and width as the 5' model. Only selected 30'' cells contain a river.

```

foreach 5' cell do
    foreach 30'' cell do
        | order by upstream cells;
    end
    total_l_riv ;                /* The total length of the 5' river */
    l_riv ;                       /* The current length of all 30'' rivers in this cell */
    while l_riv < total_l_riv do
        | select the next 30'' cell with the highest upstream number;
        | l_hydro = river length for this cell from HydroSHEDS;
        | /* l_hydro is the length of the river in the 30'' cell          */
        | /* The width is the same as the 5' river                       */
        | l_riv += l_hydro;
    end
end

```

Table 1
Comparison of Available Global (Steady-State) Models for NZ

	Fan et al. (2013)	Westerhoff et al. (2018)	G ³ M	de Graaf et al. (2015)
Spatial resolution	30" (~900 m)	7.5" (~200 m)	5' (~9 km)	6' (~10 km)
Surface elevation	30" DEM	Avg. of 8 m DEM	Avg. of 30" DEM	Avg. of 30" DEM
River elevation	-	-	P ₃₀ of 30" DEM	Avg. of 30" DEM + calculation based on bankfull flow and naturalized river discharge
Conductivity data	Global lithology (Hartmann and Moosdorf 2012)	QMAP Rattenbury and Isaac (2012)	GLHYMPS 2.0 (Huscroft et al. 2018)	GLHYMPS 1.0 (Gleeson et al. 2014)
Aquifer thickness	Infinite	Infinite	200 m	Calibrated
Layers	1	1	2	2
Groundwater recharge	Mean of multiple GHMs (1961–1990)	Simulated recharge (Westerhoff et al. 2018)	WaterGAP mean (1901–2013)	PCR-GLOBWB (Sutanudjaja et al. 2018) mean (1960–2010)
Calibrated	Manual	Yes	No	Manual

Note: Only Westerhoff et al. (2018) is restricted to NZ—all other models provide worldwide results. Fan et al. (2013) and Westerhoff et al. (2018) have no prescribed river elevation. In both models groundwater is removed as soon as it reaches the land surface elevation. GLHYMPS, GLObal HYdrogeology MaPS; QMAP, Quarter-million MAP.

rainfall, braided rivers, and artificial recharge (Westerhoff and White 2013).

Available Global Groundwater Models

Only three global gradient-based groundwater models are currently available (Fan et al. 2013; de Graaf et al. 2015; Reinecke et al. 2019a). This study analyzes steady-state hydraulic heads simulated by these three models as well as by a regional model for NZ (Table 1). The model by Westerhoff et al. (2018) is a refined version for NZ of the model by Fan et al. (2013), with different hydraulic conductivities, groundwater recharge, and land surface elevations. The main difference between G³M and the models of Fan et al. (2013) and Westerhoff et al. (2018) is that no interaction with surface water bodies is simulated in the latter. Groundwater is removed as soon as it reaches the surface. Fan et al. (2013) and Westerhoff et al. (2018) are only able to simulate steady-state conditions, whereas de Graaf et al. (2015) and G³M can also run transient conditions and simulate interaction with surface water bodies. The simulated heads used in this study are from steady-state simulations.

Observations

The observations used in this study are based on a dataset of piezometric time series from individual groundwater wells by Westerhoff and White (2013) that are temporally aggregated to an average WTD. Time series length is on average 12.4 years, with a median of 7 years. The longest time series is from 1894 to 2013.

Observed head values are calculated using the observed WTD and an 8 m DEM (LINZ 2012) resulting in 4459 observations. This DEM is the DEM with

the highest resolution that is freely available for New Zealand. We assume that it provides the best estimate of the observation elevation we can obtain for the entire model domain. For comparison with simulated values, the observations are aggregated by a geometric mean. The geometric mean was chosen as it is more robust against outliers and because the data are skewed. The number of (aggregated) observations to which the simulated heads can be compared depends on the model resolution (see N^o in Table 2).

It is unknown whether an observation was made from a confined or an unconfined aquifer. Thus, the following analysis does not treat the observations differently, even though it is likely that some observations are obtained from a confined aquifer. We assume that all observations correspond to simulated heads in the upper layer of the models.

Results

Impact of Spatial Resolution and Aggregation on Comparing Simulated Heads Against Observations

Depending on the spatial resolution of the model, there is inherent uncertainty in land surface elevation per model grid cell because each cell represents an aggregated characteristic of the reality (Figure 1), where land surface elevations within the grid cells areas may vary by hundreds of meters. Considering all grid cells in NZ, even in a 30" cell, the average range of 8 m DEM elevations is 266.6 m, and increases to 834.8 m for the 5" cell (Figure 2). These significant variations are important as within each grid cell, there is currently only one value

Table 2
Bias and RMSE Comparison for WTD and Head for Various Steady-State Models (Table 1)

Model	Resolution	N ^o	b(WTD)	b(head)	RMSE(WTD)	RMSE(head)
Westerhoff et al. (2018)	7.5''	3148	-9.7 m	9.1 m	19.3 m	18.8 m
Fan et al. (2013)	30''	2377	-14.4 m	14.7 m	28.7 m	26.9 m
G ³ M	30''	2377	-16.0 m	17.1 m	29.2 m	28.5 m
G ³ M	5'	163	2.5 m	25.9 m	35.5 m	52.0 m
de Graaf et al. (2015)	6'	147	42.3 m	-8.1 m	79.9 m	30.1 m

Note: Simulated WTD is calculated based on the respective model land surface elevation. Observed WTD is the geometric average of all observations in the model grid cells, while observed head was calculated by first computing the hydraulic head at the observations wells by using a high-resolution DEM and then averaging over all observation wells in the model grid cells (see Observations). The bias (b) is calculated as the arithmetic average of the residuals (simulated - observed). N^o is the number of grid cells for which b and RMSE are computed.

for the elevation of the water table of the surface water bodies, and the simulation results strongly depend on the assumed surface water elevations (Reinecke et al. 2019a).

Moreover, determining a value for the land surface elevation also impacts the model performance evaluation as observations are well measurements of WTD that need to be converted to head observations by using the land elevation to compare them to model results. In other words, the simulated head needs to be converted to simulated WTD, assuming an average cell elevation. For some measurements, land surface elevation is available, but for most measurements, including the only available global data set of observed groundwater heads by Fan et al. (2013), land surface elevation must be based on a DEM. To compare simulated groundwater heads to "observed" values, de Graaf et al. (2015, 2017); Reinecke et al. (2019a) used a 30'' DEM to calculate observed head values. All head observations in this study were obtained by first subtracting WTD of individual wells from the 8 m DEM. The range of the observed WTD and head per 30'' cell is relatively low, 2.6 m on average, but gets much higher for 5' cells, with an average variation of 49 m per cell (Figure 2).

Figure 3 shows the range of observed heads in all 5' grid cells that contain at least one observation well, together with the aggregated observed value. For most cells, the range of observed heads in the 5' cell is much larger than the deviation of the simulated head from the aggregated observed value. For lower aggregated observed head values, the range of observed heads decreases likely due to flatter terrain at lower land surface elevations. Aggregated observed heads between 100 m and 400 m have a large variability, with a range of over 200 m. In mountainous areas, the range of observed heads decreases due to the limited number of available head observations existing within the 5' cell.

The Relation Between Land Surface Elevation and Simulated Head at Different Spatial Resolutions along a Profile in the Canterbury Region

To understand the implications of spatial resolution on model performance, Figure 4 shows a land surface elevation profile through the Canterbury region along with simulated heads. Please note that the profile is not

a NS or WE profile; thus, the lengths of the grid cell intersections along the degrees latitude axis vary. The figure shows the cross-section with the most available observations and follows roughly the groundwater flow direction. The majority of observation wells are at lower elevations where groundwater is used for irrigation. This observation bias makes it difficult to assess how the model behaves in challenging mountainous regions.

Averaging the 30'' DEM to 5' reshapes how the region is represented. For example, near the coast (to the right), the 5' surface elevation is almost 50 m above the 30'' DEM. This is because most of the 100 30'' cells within this 5' cell have a higher land surface elevations than the 30'' cells along the profile. Thus, averaging to coarser resolution also impacts elevation gradients.

Because the assumed surface water body elevation (P₃₀ of 30'' DEM) near the coast is close to the observations, simulated heads are relatively close to the observed heads. However, when using the mean 5' land surface elevation to derive simulated WTD, a land surface elevation difference to the 8 m DEM of 50 m translates to directly to a difference of 50 m in WTD. Further upslope, this issue gets more severe. Between 43.24° S and 43.27° S, the 5' average of the surface elevation is almost 100 m away from the 30'' elevation. The simulated head differs by 50 m from the 30'' head at this location. Just due to this scale effect, head and WTD simulated by a 5' resolution model will be off by a large magnitude from any observed value. The 30'' simulated head, on the other hand, follows the 30'' DEM more closely and thus is also closer to the observations in most cases, but by chance, e.g., at 43.22° S, the 5' simulation result fits much better (further discussed in The sensitivity of simulated head to spatial resolution). Comparing the 30'' DEM to the high-resolution 8 m DEM shows that the 30'' DEM is a relatively good approximation of the real topography in most parts, even though it can deviate by ~10 m even at low elevations. More significant deviations can be seen in regions with a large slope, for example, at 43.17° S, where the 8 m DEM is almost 15 m lower.

The Sensitivity of Simulated Head to Spatial Resolution

Figure 5 shows the steady-state heads of the two model variants compared to observations. According to

	Observation uncertainty	Average range per	
		30" cell	5' cell
Land surface elevation (8 m DEM) 	 6 m Based on SRTM + local model	 266.6 m	 834.8 m
Water table depth 	 ~cm Need to assume an average elevation per cell to computed simulated WTD	 2.6 m	 49 m
Hydraulic head 	 ~cm +/- 6 m For most observations no land surface elevation is recorded: Need to use DEM	 2.6 m	 49 m

Figure 2. Uncertainties of land surface elevation, water table, and head observations and the additional uncertainty due to spatial aggregation of observations to the two grid-cell sizes. The DEM measurement uncertainties are based on Rodriguez et al. (2006) for NZ. Head observations are calculated assuming the land surface elevation of the 8 m DEM. Thus the head uncertainty is the sum of the uncertainties of the WTD observation and the DEM. WTD observations are temporally averaged values from Westerhoff et al. 2018 for the Canterbury region (see Observations). Average ranges of values per grid cell over NZ are shown for both 30" and 5' cells.

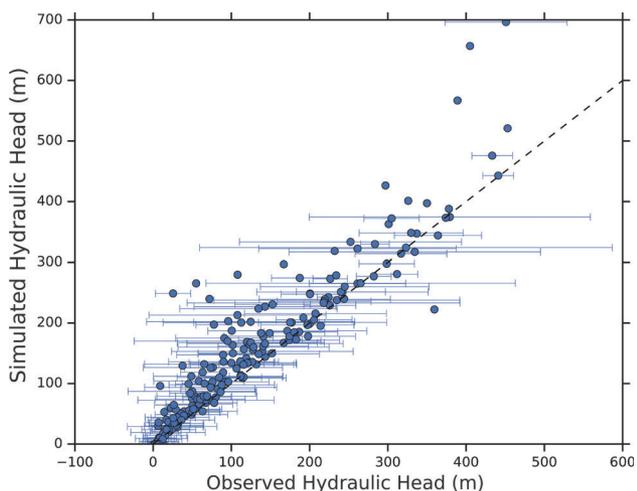


Figure 3. Hydraulic heads simulated by the 5' G³M as compared to the range of observed head values (8 m DEM – observed WTD) values in the 5' model cells in the Canterbury region of NZ, including 4459 observations in 163 cells. Dots show geometric means of observed head values (see also Observations).

the RMSE, the 30" model, with a value of 28.54 m, outperforms the 5' model with a value of 52.00 m. These numbers can be somewhat misleading regarding the model performance. As the well density is higher at lower land surface elevations where fit is generally better. The number of aggregated observation values increases more strongly at lower land elevations than at higher land

surface elevations with less dense observations. Therefore, the lower RMSE of the 30" model does not necessarily indicate a better overall model performance. Both scatter plots show a persistent overestimation of heads with a similar magnitude and distribution. The 5' model produces fewer underestimates but one with a higher magnitude. The underestimates of the 30" model are few compared to the overestimates. Due to the scale, the underestimates are hardly visible on the map. The 60 m underestimate of the 30" model is located close to the Waikamari river and the elevation profile. In general, the underestimates of the 5' model appear at higher elevations, whereas the 30" model shows increased underestimates at lower elevations; this can also be observed on the maps. They also show a better agreement (deviation from the aggregated observed head smaller than 10 m) at lower elevations toward the coast.

The overestimates in the 5' model are mainly located where the contours of observed water table bulge at the location of two major rivers—the Ashburton River and the Rakaia River. In 1974, both rivers were found to lose water to the groundwater in their lower parts close to the coast (Mandel 1974). For the Ashburton River, this agrees with the simulated contours but not with the observed contours that throughout show gaining conditions. For the Rakaia, both the simulations and observations indicate losing conditions close to the coast. Overestimation of observed heads could be due to high levels of groundwater pumping or underestimation of hydraulic conductivities (Westerhoff et al. 2018).

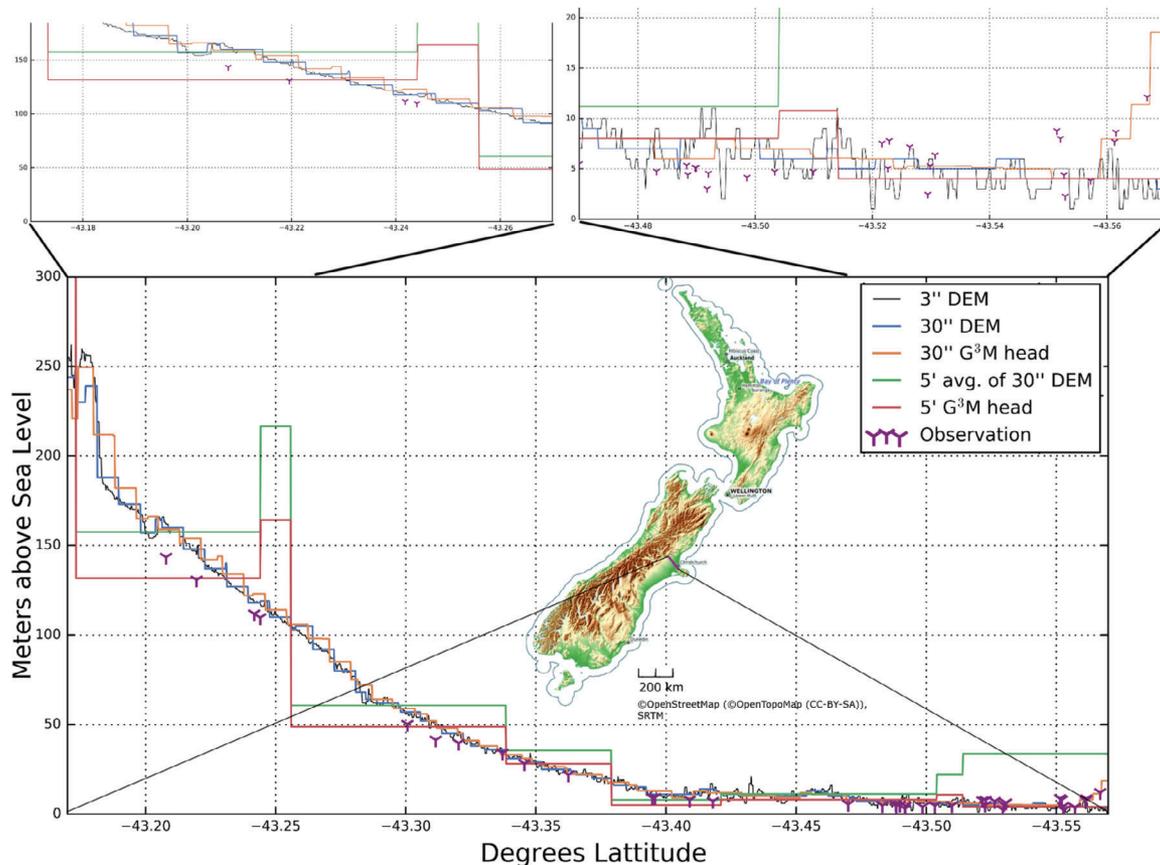


Figure 4. Elevation and simulated head profile at different spatial resolutions through Canterbury. Observed heads are located within a distance of 200 m from the cross section. Observed heads are calculated based on observed WTD and an 8 m DEM.

Comparison to Other Models

The systematic overestimation of heads in the Canterbury region is not unique to the G³M model but persists in the global model of Fan et al. (2013) and the higher resolution and locally refined adapted version of this model by Westerhoff et al. (2018) (Figure 6, see also Table 1). Again, the higher spatial resolution of the Westerhoff et al. (2018) model has a lower RMSE for the observed heads, but like G³M and the Fan et al. (2013) model tends to overestimate.

Westerhoff (2017, 227 to 240) attributed the overestimation not only to the pumping of groundwater but also to an underestimation of hydraulic conductivity in his model compared to the local estimates of Broadbent and Callander (1991). Glacial outwash gravels are known to occur especially adjacent to the rivers (e.g., Ashburton and Rakaia River) (Mandel 1974), and vertical vein-like flow through unconfined gravels was reported (Mandel 1974). Neither are considered by any of the investigated models and could explain the overestimation of the hydraulic head. GLHYMPS 2.0 (Huscroft et al. 2018) (the hydraulic conductivity data set used in G³M) contains homogeneous K for the study area, and in G³M vertical K is assumed to be 10% of the horizontal K ; an assumption which is questionable for unconsolidated sands and gravel deposits.

Simulation of WTD

Gradient-based groundwater models formulate the groundwater flow equations in terms of the hydraulic head relative to a specified datum. However, estimates of WTD are required for comparison with observations (see Observations) as well as for assessing the impact of capillary rise on evapotranspiration in GHMs (Reinecke et al. 2019a). Most GHMs do not implement capillary rise even though the impact of groundwater on the water budget simulated at the land surface–atmosphere interface has been widely studied (Vergnes et al. 2014). WTD is calculated as the surface elevation of the model minus the simulated head. Figure 7 shows that all global models require substantial improvements to represent adequately WTD in the Canterbury region. Only the WTD simulated by the high-resolution model of (Westerhoff et al. 2018) shows a (weak) correlation with observed WTD. All other models tend to deviate by 50 m or more. All high resolution (30'' or higher) models underestimate the WTD, thus simulate a water table that is too shallow, while the coarser-scale models tend to both over- and underestimate WTD. Both 30'' models, even though they differ regarding the interaction with surface water bodies, simulate low WTD everywhere, no matter if observed WTD is shallow or up to 100 m deep. Only the 30'' G³M model also overestimates WTD (obfuscated in Figure 7 due to the

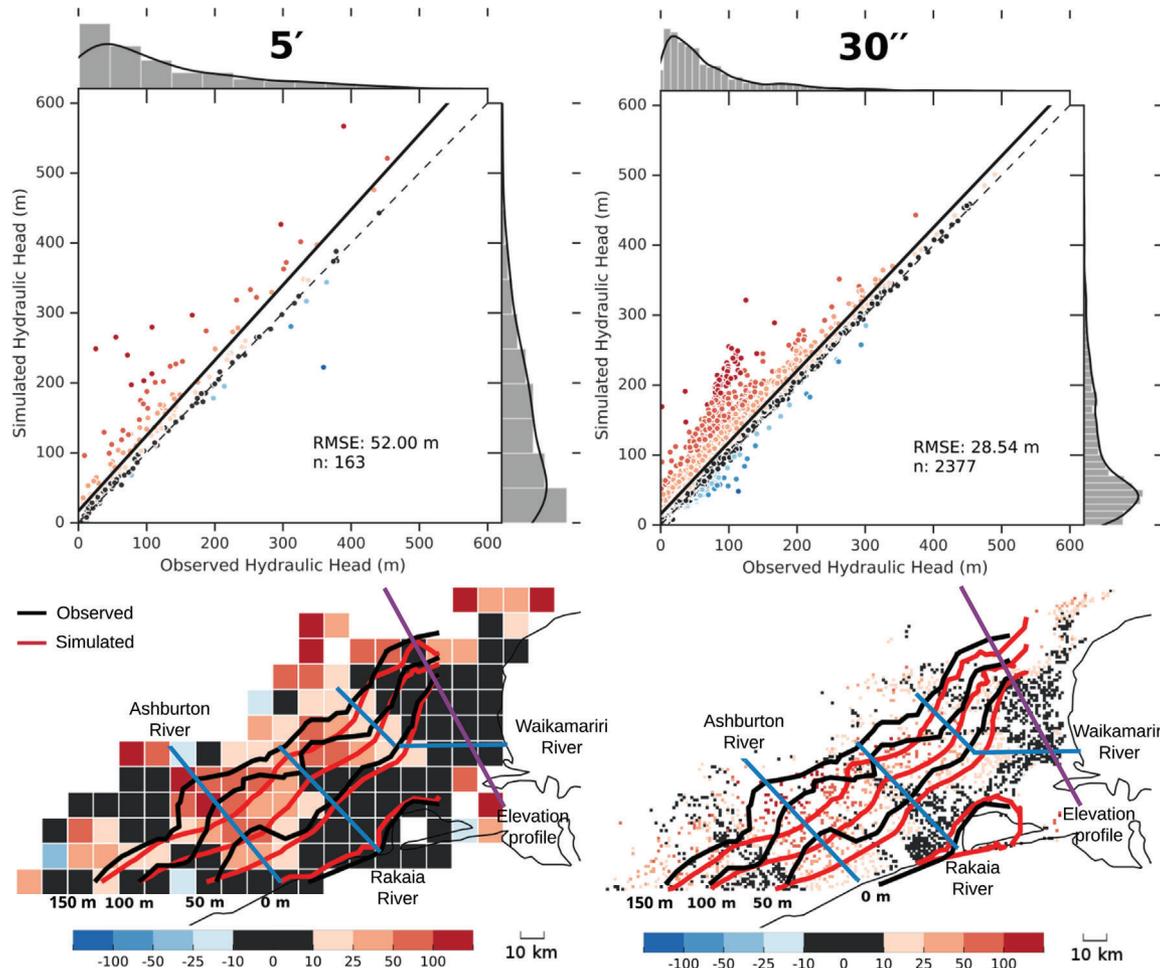


Figure 5. Simulated heads compared with observed heads for two different model resolutions (5' and 30'') for the Canterbury region. The scatterplot shows the 1:1 line as dashed and the regression as solid line. The normalized histogram indicates the number of the respective values of simulated and observed heads. The colors of the scatter points correspond to the colors of the grid cells on the maps. The maps show deviations between simulated and (geometric mean) aggregated observed head (simulated – observed). On the 30'' map the blue cells are hardly visible due to the high resolution and overlaps with the contour lines. Contours for the observed head (black) in both maps are based on the 5' average as the 30'' resolution led to artifacts in the interpolation method. Contours of simulated head (red) are shown for the corresponding spatial resolution of 5' and 30''. The purple line shows the cross section of the elevation profile in Figure 4.

number of scatter points, also compare Figure 5). Some of the overestimates in the 30'' G³M are above the land surface. These overestimates are likely caused by the large gradient between the upland and the lowlands. The model of de Graaf et al. (2015) seems to strongly overestimate (by more than 300 m) the WTD in most cases, while the 5' G³M shows wrong estimates in both directions with slightly less severe overestimates.

Impact of Model Parameter Values on Simulated Heads

Groundwater abstractions heavily influence the study region. However, groundwater abstractions are not represented in any of the steady-state groundwater models. The impact of groundwater abstractions can be represented by decreasing natural groundwater recharge by net abstractions (abstractions minus return flow to groundwater). A decreased effective groundwater recharge is expected to lead to lower heads, and thus a better representation of heads observed in the Canterbury region. However, when

subtracting abstractions simulated by WaterGAP (Müller Schmied et al. 2014) from natural groundwater recharge applied to G³M, we found that they are too small to reduce the overestimation significantly. This is even the case if total water abstractions from surface water and groundwater, as computed by WaterGAP, are assumed to be abstracted from groundwater.

Evaluating 50 experiments with varying hand-tuned parameter sets, we determined by trial and error, with the 5' G³M variant, which variations of the three most influential parameters can significantly reduce the overestimation of heads in the Canterbury region. Using the Method of Morris (Reinecke et al. 2019b), groundwater recharge, hydraulic conductivity, and elevation of surface water bodies were found to be most influential for the simulated 5' head in New Zealand (Figure S1). Therefore, the 5' baseline experiment (same as Figure 5) with an RMSE of 52 m was compared to experiments where (1) a factor of 10³ homogeneously reduces groundwater

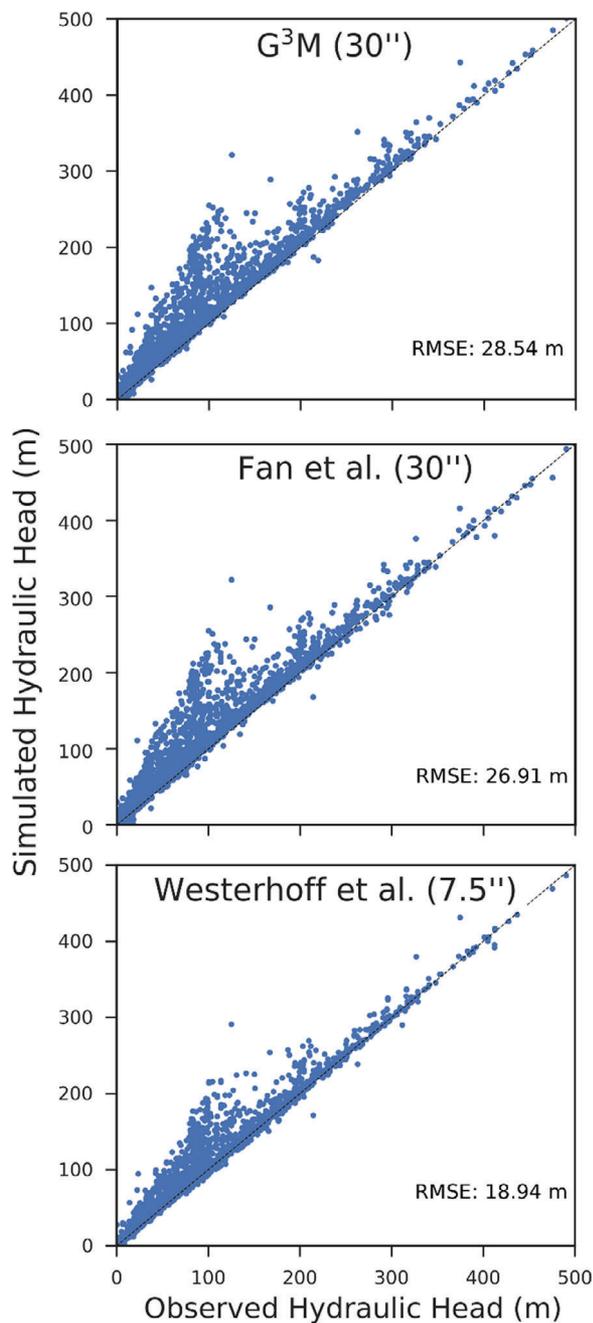


Figure 6. Simulated head (m) of two 30'' (G³M and Fan et al. [2013]) and one 7.5'' (Westerhoff et al. [2018]) groundwater models compared to observations in Canterbury. The observed heads are derived from observed WTD (see Observations) and a 8 m DEM. Multiple observations inside one cell are aggregated with a geometric mean to 30'' and 7.5'' (depending on the model they are compared to).

recharge, (2) K is increased by a factor of $10^{4.5}$ in the mountain cells (average elevation <400 m and average slope $>5^\circ$) and 100 in the other cells or (3) an alternative subgrid parameterization (SP) for determining the surface water body elevation is used. With SP, surface water body elevation is not determined as the 30th percentile of the 30'' land surface elevations within the 5' cell but as the mean of the elevations of all existing 15'' rivers within the 5' cell. These rivers are defined by the

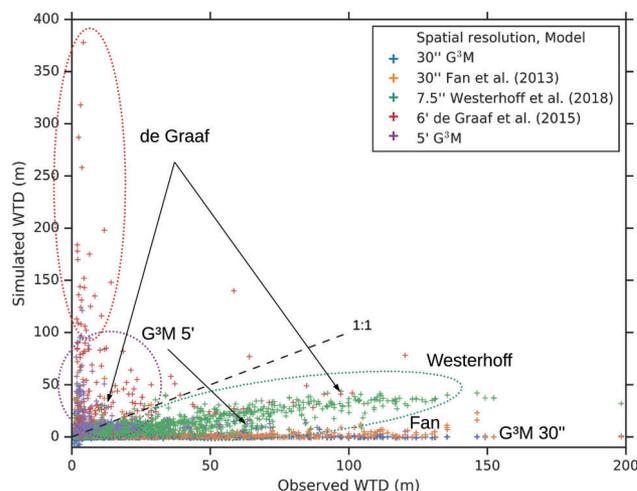


Figure 7. Simulated WTD in Canterbury compared to observed WTD for various steady-state groundwater models: G³M (5' and 30''), Fan et al. (2013) (30''), de Graaf et al. (2015) (6'), and Westerhoff and White (2013) (7.5''). The 7.5'' values are aggregated to 30'' for better visibility. The aggregation only improves the readability of the plot and does not change the general trend of the results.

15'' HydroSHEDS river network map (Lehner et al. 2008). Also, combinations of two and all three alternatives are evaluated (Figure S1).

With the extreme reduction of groundwater recharge to almost zero, the simulated heads show better agreement compared to the aggregated observed heads (RMSE: 39.57 m), but no reduction of overestimation is visible. The extreme increase in K also leads to a smaller RMSE (43.48 m) but adds more underestimates of the observed head. Estimation of the surface water body elevation by using higher-resolution DEM and derived river network data in a physically meaningful way (SP), however, reduces the fit to observations (RMSE: 60.31 m). A combination of SP with an increased K (RMSE: 43.48 m) shows no differences to only increasing the K (RMSE: 43.48 m). Combining SP with the groundwater recharge reduction, the RMSE (40.90 m) is similar to the value achieved by recharge reduction alone. A similar pattern results if a reduced recharge is combined with an increased K , leading to a higher RMSE of 41.59 m but less visible overestimates than recharge reduction alone, but overall much more severe (300 m) underestimates. The combination of all three changes yields similar results with even more underestimates, an RMSE of 41.45 m, and no overestimates larger than 200 m. Extreme reduction of groundwater recharge produces the lowest RMSE, and the experiments show that there are nonlinear relationships between the parameters. A systematic model calibration may lead to a better agreement (e.g., smaller RMSE) with observations.

Discussion

Inclusion of gradient-based instead of bucket-like groundwater models in GHMs can improve the simulation

of exchanges with surface water bodies and evapotranspiration as both require simulation of sufficiently accurate groundwater heads. While GHMs compute water flows and storages independent of any absolute elevations (Müller Schmied et al. 2014), gradient-based groundwater models, as well as GHMs coupled to gradient-based groundwater models, require determination of groundwater heads that depend on absolute elevations of the location on which they occur, elevation of the water level in surface water bodies, and the elevation of the land surface. This poses significant challenges that have not been encountered yet in global hydrological modeling.

Representation of spatially strongly variable land surface elevations in coarsely discretized global groundwater models is inherently incomplete, uncertain, and a function of grid-cell size (Figure 4). The model input land surface elevation as well as simulated head and WTD for a grid cell are an average representation of, in reality, a distribution of land surface elevation, heads and WTDs that may strongly vary within the cell.

We need to contemplate how global-scale models should be designed so that they can estimate water flows between the groundwater and surface water bodies and between groundwater and the unsaturated zone adequately, at least in a way that fits hydrologists' perceptual model. According to Beven (2012), a perceptual model is a representation of a watershed (in this case, not only a watershed but large-scale groundwater processes spanning multiple watersheds) that is based on our understanding of real-world processes and is translated into a numerical model. Importantly, our understanding, especially at these large scales and subsurface systems, might be weak and highly uncertain (Neuman 2002).

Furthermore, to what extent is a comparison of hydraulic heads computed by macro-scale groundwater models with large grid cells to well observations meaningful for evaluating model performance? Publications on macro- and global-scale groundwater models often show scatterplots comparing simulated head values to aggregated observations of the hydraulic head, for example, de Graaf et al. (2015, 2017, 2019), Fan et al. (2013), Maxwell et al. (2015), and Reinecke et al. (2019a). Such scatterplots can provide a rough visualization of model biases (Figure 6) but data points close to the 1:1 line and high values of the coefficient of determination (R^2) or the Nash-Sutcliffe coefficient (modeling efficiency) are not indicative of a high model quality as the variation of the heads is a strong function of the topography in nonarid climates (Toth 1963; Haitjema and Mitchell-Bruker 2005). The ability to simulate both flows between surface water and groundwater and capillary rise depends on the ability to simulate WTD rather than hydraulic head. If land surface elevation is the same within the whole model domain, then scatterplots of heads and WTD would look the same. The larger the variation of land surface elevation within the spatial domain, the more the fit of simulated and observed heads shown by scatterplots becomes visually better just because land surface elevations dominate the visualization. Due to an

increased variation of the observed head, both the coefficient of determination and the Nash-Sutcliffe coefficient increase, too. This is an issue only in the case of simulating larger spatial domains such as NZ or the globe as these spatial domains contain a much larger range of land surface elevations than the smaller spatial domains for which groundwater modeling is typically done.

Before this article, no publication on global groundwater models showed a direct comparison of observed and simulated WTD. The comparison of WTD in this study reveals the abysmal performance of all global/macroscale models (Figure 7). This corresponds to large RMSE values for WTD between 18.9 m and 52.0 m and biases between a mean WTD underestimation of 16.0 m and a WTD overestimation of 42.3 m (Table 2). The model of Westerhoff et al. (2018), in which a global groundwater model algorithm is applied for NZ only, shows the best WTD performance. This might be due to a combination of factors: the calibration of the model, its spatial resolution, and the use of high-resolution conductivity data. For the three high-resolution models, the two performance criteria are almost the same regarding WTD and head, but they are different for the two lower-resolution models, the 5' G³M and the 6' de Graaf et al. (2015) model. For the 5' G³M, performance of head is much better than performance of WTD, while the opposite is true for the de Graaf et al. (2015) model (Table 2). Performance values likely differ a lot between head and WTD for the lower-resolution models due to the much larger number of well observations that are aggregated to one "observed" value within the cell as compared to the high-resolution models (Table 2). Both performance criteria would be equal for WTD and head if the same land surface elevations were used to compute "simulated" WTD or "observed" head but we preferred to calculate best estimates of "observed" head by using the best estimate of land surface elevation to compute "observed" head from actual observed WTD. However, it is not possible to use these spatially varying land surface elevations at the well locations to compute simulated WTD from actual simulated head; instead, the land surface elevation of the grid cell is used. An alternative for determining "simulated" WTD, just for determining model performance, is to calculate "simulated" WTD as the difference between the mean land surface elevations of the observation wells within the model grid cell and the simulated head, but this was not explored.

Suspecting that the opposite performance of the 5' G³M and de Graaf et al. (2015) model may be caused by differences in the grid cell extents that encompass different observation wells, we determined RMSE values that would result if random samples of only fractions of all available 4459 observations were used for performance testing. However, the opposite behavior persists for all fractions and the 100 random samples for each fraction class (Figure 8). RMSE values of the two models for WTD and head positively correlate with the absolute values of the biases, but it is unclear why for example, the absolute WTD bias of 5' G³M is so much smaller than the corresponding value of the de Graaf et al. (2015) model

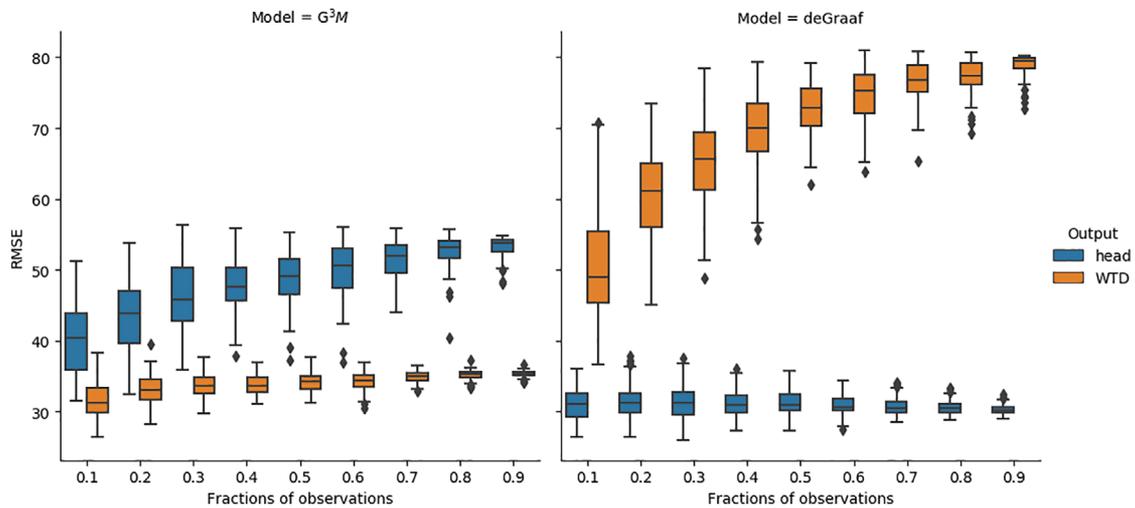


Figure 8. Box plot of RMSE as a function of fractions of observations for head and WTD [m]. Each box is based on 100 repeated random samples (for the according fraction) from all available observations that are then aggregated to the according spatial resolution. Whiskers represent the 1.5 interquartile range.

(with the opposite being true for the absolute head bias). This behavior may be related to the different assumptions about the elevation of the surface water bodies (Reinecke et al. 2019a, 2019b).

Figure 8 reveals that values of model performance indicators strongly depend on the available observations. RMSE could vary by more than 10 m or even 30 m for different sets of observation wells. Interestingly, RMSE tends to increase with an increasing number of observations (Figure 8).

Differences between simulated and observed heads and WTDs can depend strongly on which point observation the simulated result is compared to. Unless there is (1) an extensive number of rather homogeneously distributed observation wells within each large model grid cell, (2) with observations that characterize the average groundwater distribution in such a large area, observations cannot be aggregated in a way that is meaningful for a comparison to simulated heads or WTD. In most cases, it is unlikely that both conditions are fulfilled.

In conclusion, while we found that increasing spatial resolution of the groundwater model improves the fit to observations (compare the RMSE of five models in Table 2), we suspect that this is partially due to decreased variability of point observation values in the smaller grid cells, and thus a better representation of average conditions within the grid cells by the observed point values. Regardless, the bias, a general overestimation of observed heads, remained in the testbed region.

Based on the results presented, one cannot conclude that an increased spatial resolution is by itself sufficient for improving the model's ability to simulate heads and flows accurately. Somewhat surprisingly, the 30'' variant of G³M shows similar performance to the contrasting 30'' model of Fan et al. It is unclear why this is the case.

Additional research is necessary to investigate which deviations from observations occur in transient simulations and if the fit to temporal variations is better than the fit to WTD. Testing the simulation of temporal variations might be more meaningful for coarse-scale groundwater models than testing steady-state WTD.

A limitation of this study is the steady-state nature of the model output and the comparison to observations in a region that is profoundly impacted by groundwater abstractions. Abstractions are another challenge for the evaluation of global groundwater models as we can expect a high density of observations mostly in regions with intensive use of groundwater. At the same time, we do not have any global data of groundwater pumping and therefore rely on simulation results of GHMs like WaterGAP (Müller Schmied et al. 2014). A possible approach in future studies would be to use known elevations of springs and wetlands to estimate the WTD in regions where no observations are available.

To find an explanation for the overestimated heads, we analyzed an ensemble of model variants with a different combination of the most critical model parameters but the model fit was not substantially improved. While automated calibration could be a solution, Westerhoff (2017) showed that this might lead to implausibly high *K* values. Additionally, other calibration efforts of global models led to nonphysical parameters (de Graaf et al. 2017) and automated calibration on the global scale might not be feasible due to computational limitations. Furthermore, it is questionable whether calibration against spatially aggregated heads is meaningful.

Conclusions

Global-scale simulation of groundwater heads and flows is challenging. This is predominantly due to high spatial variability of the elevations of the land

surface and the water table as their variations (or spatial gradients) govern resulting water flows, but cannot be easily represented with the rather large grid cells required in global-scale groundwater modeling. The large grid cells also make a meaningful comparison to point observations of groundwater level difficult.

Due to the coarse resolution, global groundwater models struggle in correctly representing the flow of water through the subsurface because the range of land surface elevation within one computational cell may be, for example, 1000 m more than the thickness of the aquifers. In addition, due to the lack of high-quality global data, model input is highly uncertain. Both issues strongly differentiate global groundwater modeling from the well-established local or regional basin-scale groundwater modeling.

We need new strategies for evaluating macro-scale groundwater models as we demonstrated in this study that comparison of simulated grid cell values to well observations, that need to be aggregated to the model cell size, is problematic. The number and range of observed values per cell can vary greatly and affects both how observation data relate to a particular model and comparisons of models with different spatial resolutions. A suitable approach to represent the subgrid variability of elevations and hydraulic heads within large grid cells is needed to improve evaluation of macro-scale groundwater models.

Gradient-based groundwater models are required to improve the established GHMs, but the community needs to be open about their shortcomings when advancing them. Increasing the spatial resolution of global groundwater models will not necessarily result in better models—only more copious amounts of output. To accurately assess these complex models, we require not only more high-quality hydraulic head observations but also estimates of (preferably large-scale) exchange flows between groundwater and surface water. International efforts for combining already available local groundwater-related data into global data sets should be intensified.

Acknowledgments

The model framework code can be obtained from globalgroundwatermodel.org. We like to thank Rogier Westerhof for providing the steady-state observations for Canterbury New Zealand and additional insights into his model and we like to thank Inge de Graaf, Ying Fan, and Gonzalo Miguez-Macho for providing their model output. Moreover, we like to thank the three anonymous reviewers and Scott James for their insightful comments that improved this manuscript.

Author Contributions

RR led the conceptualization, formal analysis, methodology, software development, visualization, and writing of the original draft. AW built the 30'' data

sets, ran the models, led the initial analysis, and did initial visualizations. AW, LF, SM, CN supported review and editing. PD supervised the work of RR and made suggestions regarding the analysis, structure, and wording of the text and design of tables and figures.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally *not* peer reviewed.

Appendix S1: Supporting Information

References

- Aotearoa, S.N.T. 2018. Population numbers. data retrieved from Stats NZ Tauranga Aotearoa. <https://www.stats.govt.nz/information-releases/subnational-population-estimates-at-30-june-2018-provisional> (accessed June 2019).
- Beven, K.J. 2000. Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences* 4, no. 2: 203–213.
- Beven, K. 2012. *Rainfall-Runoff Modelling: The Primer*, Second ed. 15. Wiley, West Sussex, UK.
- Broadbent, M., and P.F. Callander. 1991. A resistivity survey near waimakariri river, Canterbury plains, to improve understanding of local groundwater flow and of the capabilities of the survey method. *New Zealand Journal of Geology and Geophysics* 34, no. 4: 441–453.
- Cuthbert, M., T. Gleeson, N. Moosdorf, K. Befus, A. Schneider, J. Hartmann, and B. Lehner. 2019. Global patterns and dynamics of climate–groundwater interactions. *Nature Climate Change* 9, no. 2: 137–141.
- DLR. 2016. Tandem-x. <https://tandemx-science.dlr.de> (accessed June 2019).
- Döll, P., H. Müller Schmied, C. Schuh, F.T. Portmann, and A. Eicker. 2014. Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and grace satellites. *Water Resources Research* 50, no. 7: 5698–5720.
- Fan, Y., H. Li, and G. Miguez-Macho. 2013. Global patterns of groundwater table depth. *Science* 339, no. 6122: 940–943.
- Faunt, C.C., M. Sneed, J. Traum, and J.T. Brandt. 2016. Water availability and land subsidence in the central valley, California, USA. *Hydrogeology Journal* 24, no. 3: 675–684.
- García-Gil, A., J. Epting, C. Ayora, E. Garrido, E. Vázquez-Suñé, P. Huggenberger, and A.C. Gimenez. 2016. A reactive transport model for the quantification of risks induced by groundwater heat pump systems in urban aquifers. *Journal of Hydrology* 542: 719–730.
- Gleeson, T., N. Moosdorf, J. Hartmann, and L. Van Beek. 2014. A glimpse beneath earth's surface: Global hydrogeology maps (glhymps) of permeability and porosity. *Geophysical Research Letters* 41, no. 11: 3891–3898.
- Gleeson, T., K.M. Befus, S. Jasechko, E. Luijendijk, and M.B. Cardenas. 2016. The global volume and distribution of modern groundwater. *Nature Geoscience* 9, no. 2: 161–167.
- de Graaf, I.E., E. Sutanudjaja, L. Van Beek, and M. Bierkens. 2015. A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences* 19, no. 2: 823–837.
- de Graaf, I.E., R.L. van Beek, T. Gleeson, N. Moosdorf, O. Schmitz, E.H. Sutanudjaja, and M.F. Bierkens. 2017.

- A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. *Advances in Water Resources* 102: 53–67.
- de Graaf, I.E., T. Gleeson, L.R. van Beek, E.H. Sutanudjaja, and M.F. Bierkens. 2019. Environmental flow limits to global groundwater pumping. *Nature* 574, no. 7776: 90–94.
- Gupta, V.K., B.M. Troutman, and D.R. Dawdy. 2007. Towards a nonlinear geophysical theory of floods in river networks: An overview of 20 years of progress. In *Nonlinear Dynamics in Geosciences*, 121–151. New York, NY: Springer New York.
- Haitjema, H.M., and S. Mitchell-Bruker. 2005. Are water tables a subdued replica of the topography? *Groundwater* 43, no. 6: 781–786.
- Harbaugh, A.W. 2005. *MODFLOW-2005, the US Geological Survey Modular Ground-Water Model: The Ground-Water Flow Process*. VA: US Department of the Interior, US Geological Survey Reston.
- Hartmann, J. and Moosdorf, N. 2012. Global Lithological Map Database v1.0 (gridded to 0.5° spatial resolution).
- Huscroft, J., T. Gleeson, J. Hartmann, and J. Börker. 2018. Compiling and mapping global permeability of the unconsolidated and consolidated earth: Global hydrogeology maps 2.0 (glhyps 2.0). *Geophysical Research Letters* 45, no. 4: 1897–1904.
- Krakauer, N.Y., H. Li, and Y. Fan. 2014. Groundwater flow across spatial scales: Importance for climate modeling. *Environmental Research Letters* 9, no. 3: 034003.
- Lehner, B., and P. Döll. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* 296, no. 1: 1–22.
- Lehner, B., K. Verdin, and A. Jarvis. 2008. New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union* 89, no. 10: 93–94.
- Limberg, A., U. Hörmann, and H. Verleger. 2010. Modellenwicklung zur berechnung des höchsten grundwasserstandes im land berlin. *Brandenburg. Geowiss. Beitr* 17, no. 1–2: 23–37.
- LINZ, d. s. 2012. Land information new zealand 8 m dem. data retrieved from LINZ data service, <https://data.linz.govt.nz/layer/51768-nz-8m-digital-elevation-model-2012> (accessed June 2019).
- Mandel, S. 1974. The groundwater resources of the Canterbury plains. Technical Report, Lincoln College. New Zealand Agricultural Engineering Institute.
- Mark, D.M., and P.B. Aronson. 1984. Scale-dependent fractal dimensions of topographic surfaces: An empirical investigation, with applications in geomorphology and computer mapping. *Journal of the International Association for Mathematical Geology* 16, no. 7: 671–683.
- Maxwell, R.M., L.E. Condon, and S.J. Kollet. 2015. A high-resolution simulation of groundwater and surface water over most of the continental us with the integrated hydrologic model parflow v3. *Geoscientific Model Development* 8, no. 3: 923–937.
- Müller Schmied, H., S. Eisner, D. Franz, M. Wattenbach, F.T. Portmann, M. Flörke, and P. Döll. 2014. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences* 18, no. 9: 3511–3538.
- Neuman, S. 2002. Accounting for conceptual model uncertainty via maximum likelihood bayesian model averaging. *Acta Universitatis Carolinae - Geologica* 46, no. 2–3: 529–534.
- Portmann, F.T., P. Döll, S. Eisner, and M. Flörke. 2013. Impact of climate change on renewable groundwater resources: Assessing the benefits of avoided greenhouse gas emissions using selected CMIP5 climate projections. *Environmental Research Letters* 8, no. 2: 024023.
- Rattenbury, M., and M. Isaac. 2012. The qmap 1: 250 000 geological map of New Zealand project. *New Zealand Journal of Geology and Geophysics* 55, no. 4: 393–405.
- Reinecke, R. 2018. G³m-f a global gradient-based groundwater modelling framework. *JOSS* 3: 548.
- Reinecke, R., L. Foglia, S. Mehl, T. Trautmann, D. Cáceres, and P. Döll. 2019a. Challenges in developing a global gradient-based groundwater model (g³m v1.0) for the integration into a global hydrological model. *Geoscientific Model Development* 12, no. 6: 2401–2418.
- Reinecke, R., L. Foglia, S. Mehl, J.D. Herman, A. Wachholz, T. Trautmann, and P. Döll. 2019b. Spatially distributed sensitivity of simulated global groundwater heads and flows to hydraulic conductivity, groundwater recharge, and surface water body parameterization. *Hydrology and Earth System Sciences* 23, no. 11: 4561–4582.
- Rodriguez, E., C.S. Morris, and J.E. Belz. 2006. A global assessment of srtnm performance. *Photogrammetric Engineering & Remote Sensing* 72: 249–260.
- Rosso, R., B. Bacchi, and P. La Barbera. 1991. Fractal relation of mainstream length to catchment area in river networks. *Water Resources Research* 27, no. 3: 381–387.
- Sood, A., and V. Smakhtin. 2015. Global hydrological models: A review. *Hydrological Sciences Journal* 60, no. 4: 549–565.
- Sutanudjaja, E.H., R. van Beek, N. Wanders, Y. Wada, J.H.C. Bosmans, N. Drost, R.J. van der Ent, I.E.M. de Graaf, J.M. Hoch, K. de Jong, D. Karssenber, P. López López, S. Peßenteiner, O. Schmitz, M.W. Straatsma, E. Vannamtee, D. Wisser, and M.F.P. Bierkens. 2018. Pcr-globwb 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development* 11, no. 6: 2429–2453.
- Toth, J. 1963. A theoretical analysis of groundwater flow in small drainage basins. *Journal of Geophysical Research* 68, no. 16: 4795–4812.
- Vergnes, J.-P., B. Decharme, and F. Habets. 2014. Introduction of groundwater capillary rises using subgrid spatial variability of topography into the isba land surface model. *Journal of Geophysical Research: Atmospheres* 119, no. 19: 11,065–11,086.
- Wada, Y., L.P.H. van Beek, C.M. van Kempen, J.W.T.M. Reckman, S. Vasak, and M.F.P. Bierkens. 2010. Global depletion of groundwater resources. *Geophysical Research Letters* 37, no. 20: L20402. <https://doi.org/10.1029/2010GL044571>.
- Westerhoff, R.S. 2017. Satellite remote sensing for improvement of groundwater characterisation. PhD thesis, University of Waikato.
- Westerhoff, R.S., and P.A. White. 2013. Application of equilibrium water table estimates using satellite measurements to the Canterbury region, New Zealand. *GNS Science Report* 43: 25.
- Westerhoff, R., P. White, and G. Miguez-Macho. 2018. Application of an improved global-scale groundwater model for water table estimation across New Zealand. *Hydrology and Earth System Sciences* 22, no. 12: 6449–6472.
- Wood, E.F., J.K. Roundy, T.J. Troy, L.P.H. van Beek, M.F.P. Bierkens, E. Blyth, A. de Roo, P. Döll, M. Ek, J. Famiglietti, D. Gochis, N. van de Giesen, P. Houser, P.R. Jaffé, S. Kollet, B. Lehner, D.P. Lettenmaier, C. Peters-Lidard, M. Sivapalan, J. Sheffield, A. Wade, and P. Whitehead. 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water. *Water Resources Research* 47: W05301. <https://doi.org/10.1029/2010WR010090>
- Wörman, A., A.I. Packman, L. Marklund, J.W. Harvey, and S.H. Stone. 2007. Fractal topography and subsurface water flows from fluvial bedforms to the continental shield. *Geophysical Research Letters* 34: L07402. <https://doi.org/10.1029/2007GL029426>