

1 **Response Reviewer Comment 2:**

2 The manuscript describes further development of a global groundwater model  
3 (GLOBGM v1.1). This updated model incorporates key advances on its predecessor  
4 (GLOBGM v1.0, Verkaik et al., 2024) that include model calibration and downscaling of  
5 groundwater recharge. The manuscript is well structured, written and presented.

6 **Major comments:**

7 As outlined by the authors, the updated model marks an advance over its predecessor.  
8 However, the more pertinent and central question, especially given the authors'  
9 arguments in the Introduction, Discussion and Conclusions, is whether this advance  
10 marks a progression such that the updated, calibrated model provides meaningful  
11 simulations of observed changes in groundwater storage over the historical period (1960–  
12 2014) and thus projected changes in groundwater storage under socio-economic scenarios  
13 (i.e. 2015–2100: SSP1-RCP2.6, SSP3-RCP7.0, SSP5-RCP8.5). What is meant by  
14 “meaningful” may well depend upon one’s perspective but it would broadly be expected  
15 to demonstrate skill in representing historical observations. Objectively, the authors rely  
16 on the application of the non-parametric form of the Kling-Gupta Efficiency (KGE-  
17 NP). Beyond reference to the threshold value of -0.41 cited from Knoben et al. (2019), the  
18 authors need to make a stronger case as to why this threshold is appropriate, together  
19 with an evaluation of the implications of the uncertainty in defining this threshold.  
20 Indeed, greater explicit quantitative discussion is needed in section 4 of the consequences  
21 of the observed uncertainty in model simulations indicated by the CDFs and their  
22 (limited) spatial distribution in correlations in Figure 5.

23 Given the hyper-resolution (~1 km) scale of the model, I can understand the authors'  
24 necessary focus on reconciling the GLOBGM v1.1 to measurements of groundwater head  
25 from piezometric observations. I am surprised, however, not to see an effort to reconcile  
26 the latest model to evidence of changes in terrestrial water storage, albeit at lower  
27 resolution (aggregating outcomes to larger scales), from satellite gravimetry  
28 (GRACE/GRACE-FO).

29 We thank Referee #2 for the review and providing constructive comments and suggestions,  
30 which will help to improve the quality of the manuscript. We agree with the reviewer that  
31 "meaningful" performance is based on the readers perspective. It should ideally depend on the  
32 intended use of the model, which for an Earth System type of model as ours pertains to the  
33 ability to map patterns and space-time trends that are realistic and consistent with observations.  
34 As can be seen from the additional evaluation in this response, this is mostly the case.

35 The reviewer’s concerns regarding the use of -0.41 as threshold for KGE-NP is valid, since it  
36 was originally formulated for the standard KGE (Knoben et al. 2019). This is a valid point  
37 which we will address in the revised manuscript. We aimed to replicate the rationale presented  
38 by Knoben et al. (2019) for the KGE-NP score. However, given that the KGE-NP of a mean

39 benchmark simulation depends on coefficient of mean absolute deviation of the observed  
40 record, a universal threshold cannot be defined. Instead of applying a global threshold, we  
41 instead rely on the  $KGE-NP_{skill}$  score to evaluate whether GLOBGM is better or worse than  
42 using a mean benchmark. This is calculated individually for each observation time series and  
43 rely on a metric suitable for global-scale evaluation and multi-model comparison.

44 When it comes to judging whether the model development is deemed an advance compared to  
45 previous effort, we argue that the skill and accuracy of GLOBGM, or any model for that matter,  
46 should be judged relative to the current state-of-the-art. We acknowledge that this comparison  
47 was absent from the original manuscript. We will address this directly and add a model  
48 intercomparison to the revised manuscript, in which we compare the performance of the  
49 GLOBGM v1.1 against previous global groundwater models that have been successfully  
50 applied to predicting steady-state groundwater patterns and trends (Figure RC2 1).

51 For the steady-state comparison, Fan et al. (2017) achieves the lowest mean absolute bias  
52 overall (18.11 m), the CDFs intersect at approximately 20 m, indicating a lack of overall  
53 dominance. GLOBGMv1.0, GLOBGMv1.1 and ML-GLOBGMv1.1 exhibit a higher  
54 proportion of sites with absolute bias below 20 m, yet beyond this threshold Fan et al. (2017)  
55 provides more accurate estimations. This crossover suggests that Fan et al. (2017) achieves  
56 its lower mean absolute bias primarily through superior performance at locations with  
57 moderate-to-large errors, while GLOBGMv1.1 more effectively captures water table depth at  
58 the majority of locations (approximately 75 %) but at the cost of a heavier error tail (Figure  
59 RC2 1a,b).

60 Compared to the 5 arc-minutes counterparts (de Graaf et al. 2017) the 30 arc-seconds  
61 simulations show marked improvements (Figure RC2 1c,d). Collectively, different versions of  
62 GLOBGM display increases of approximately 19 and 15 percentage points in the number of  
63 locations improve upon the mean benchmark ( $KGE-NP_{skill} > 0$ ), at the monthly and annual  
64 resolution, respectively (Figure RC2 1c,d). Between the different version of GLOBGM,  
65 GLOBGMv1.1 represents an improvement over v1.0, particularly at the monthly resolution, 8  
66 and 1 percentage points increase at the monthly and annual resolution, respectively (Figure  
67 RC2 1c,d). This comparison demonstrates that the updated, calibrated model shows improved  
68 skill relative to the current state-of-the-art, providing a stronger basis for confidence in our  
69 simulations. We thank the reviewer for highlighting this gap.

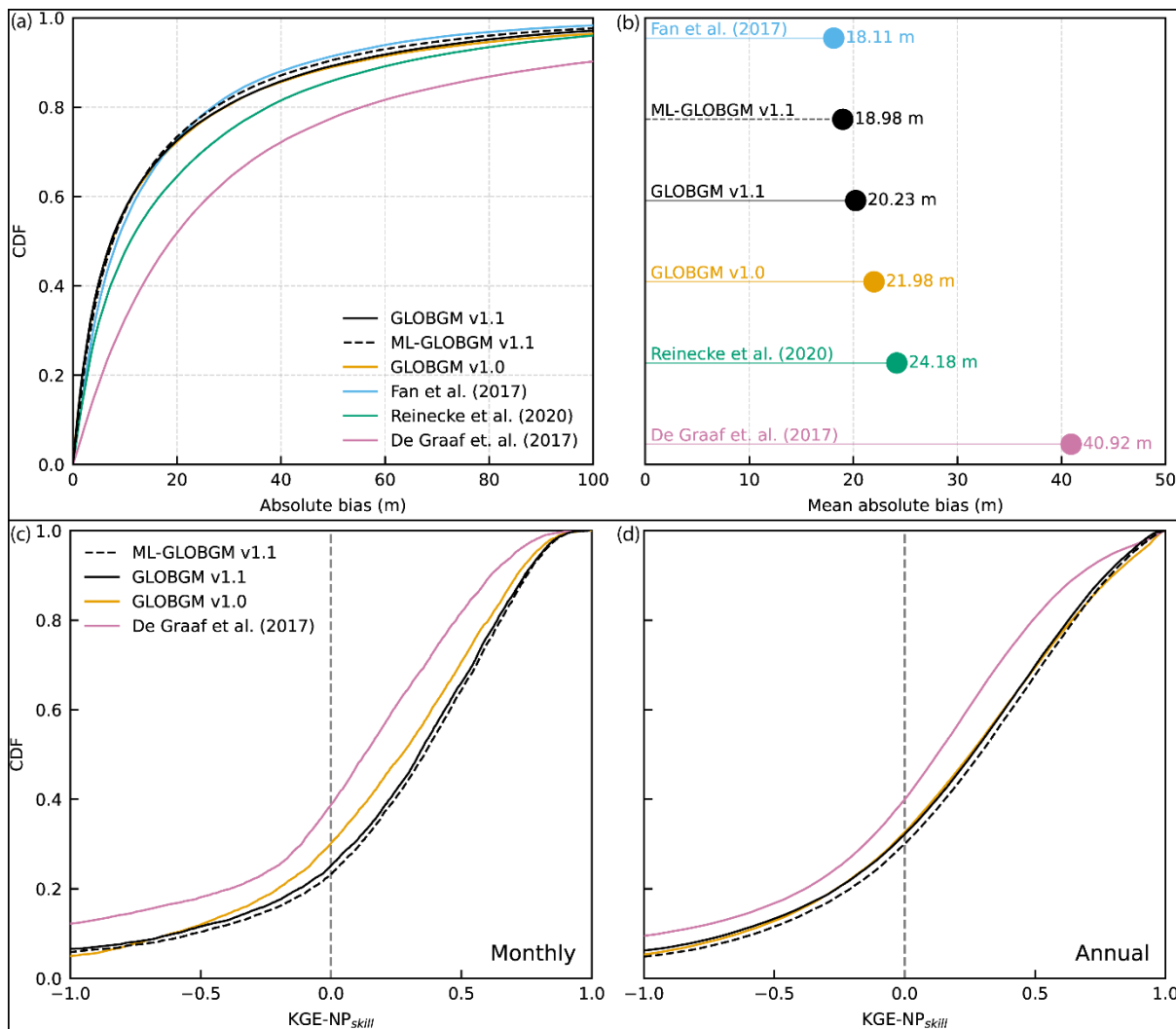


Figure RC2 1: (a) CDF and (b) mean absolute error absolute bias in simulated steady-state water table depth validated against observations. (c) Monthly and (d) annual KGE- $NP_{skill}$  scores for transient simulated hydraulic heads validated against observations.

70 We agree with the reviewer that the discussion requires more focus on the uncertainty in light  
 71 of the limited spatial distribution of observed data and the suggestion to include  
 72 GRACE/GRACE-FO for evaluation, which will allow us to assess accuracy where point-based  
 73 observations are lacking. For fair comparison we performed a spatial moving averaging  
 74 approach of the yearly groundwater levels with a Gaussian spatial filter with a standard  
 75 deviation of 75 km resembling approximately the GRACE/GRACE-FO footprint  
 76 (Chandanpurkar et al. 2025). We will compare the 5 arc-minute PCR-GLOBWB TWS trends  
 77 by which the groundwater component is replaced with GLOBGM estimates and the  
 78 groundwater storage trends from GLOBGM only separately. We have included draft figure of  
 79 this comparison (Figure RC2 2).

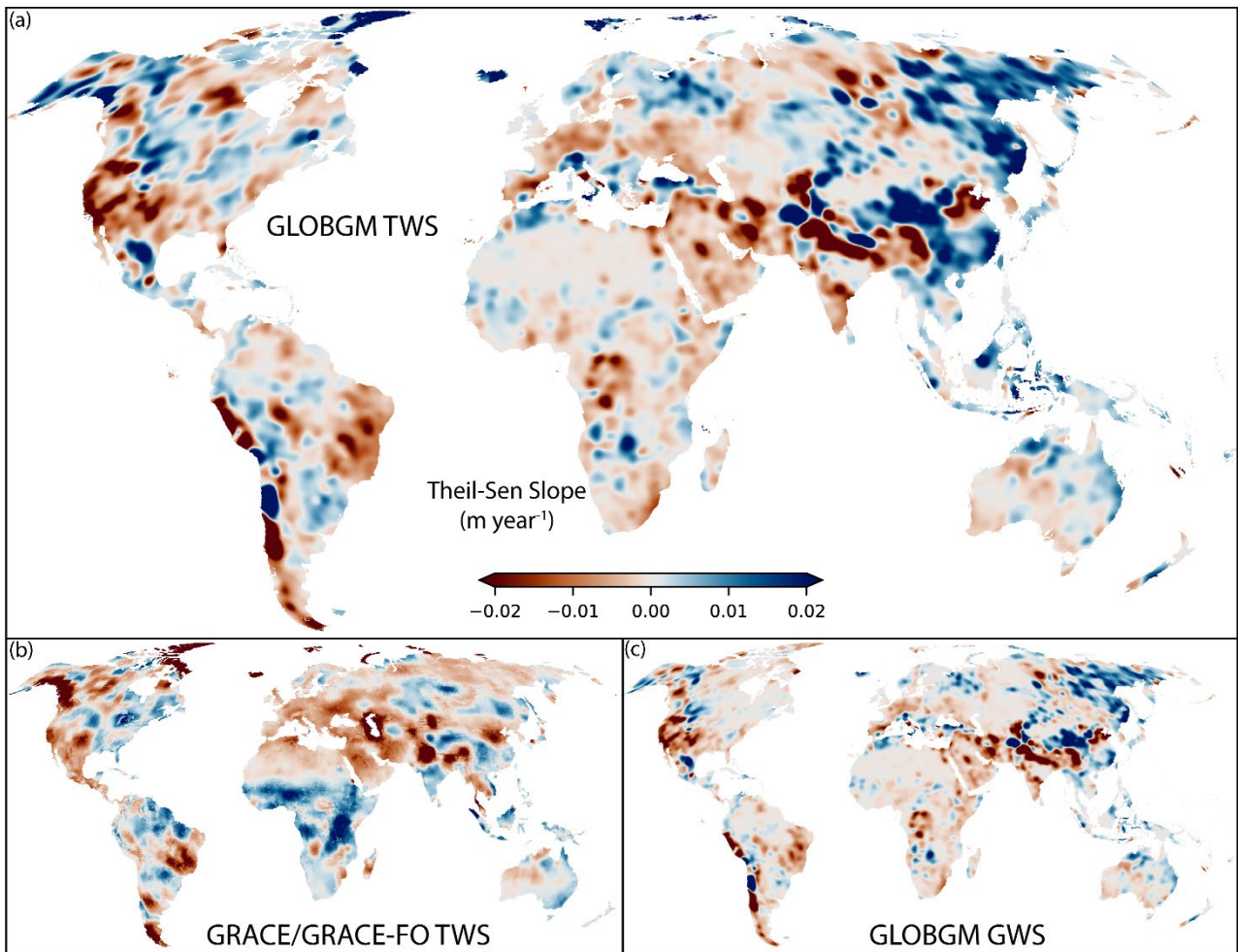


Figure RC2 2: Annual Theil-Sen slope estimates (2002-2019) for the (a) Terrestrial Water Storage trends from the GLOBGM model. (b) observed TWS trends derived from GRACE/GRACE-FO satellite data. (c) simulated Groundwater Storage (GWS) trends from GLOBGM.

81 It is well accepted that what the authors seek to achieve (i.e. a global hyper-resolution  
 82 groundwater model) is hugely challenging and of very considerable importance. It is also  
 83 heartening to see groundwater specialists leading this effort. The question is whether  
 84 GLOBGMv1.1 marks an important technical advance on previous global groundwater  
 85 models or whether the model has now reached a stage where it can be applied to simulate  
 86 meaningfully the impact of human interventions and climate change; the current text  
 87 attests the latter but current outcomes suggest the former.

88 We thank the reviewer for the kind words and for drawing the distinction between technical  
 89 advances that GLOBGM represents and the suitability of GLOBGM for assessing  
 90 socioeconomic impacts. We agree that this is an important distinction to make.

91 We reiterate that "important technical advance" must be assessed relative to the available state-  
 92 of-the-art alternatives. The model comparison that will be added demonstrates that  
 93 GLOBGMv1.1 performance is comparable to or better than existing global groundwater  
 94 models, which are themselves widely applied to socioeconomic and climate impact assessment  
 95 in the peer-reviewed literature.

96 **Minor comments:**

97 **1. There appears to be imprecision in mapped features in Figures 1 and 4 that**  
98 **presumably derive from the databases used to generate these. In Figure 1 for example,**  
99 **aquifers throughout Bangladesh are depicted as confined whereas, in reality, confining**  
100 **conditions are restricted to areas overlain by Pleistocene terrace deposits**  
101 **(Shamsudduha et al., 2022). In the same figure in Canada, mapping of confined aquifers**  
102 **(overlain thick glacial tills, especially in the prairies) are absent. In Figure 4, permafrost**  
103 **regions – certainly for Canada - are of much greater extent than depicted.**

104 We agree that this could be improved. We are working on a new global hydrogeological  
105 schematization. This is a significant task worthy of a separate paper and not ready for some  
106 time. In the meantime, we relied on the same schematization as developed by De Graaf et al.  
107 (2017).

108 **2. To better understand challenges to model performance, might the authors present**  
109 **(e.g. Supplementary Material) the employed spatio-temporal distribution in**  
110 **observations of groundwater head over from 1960 to 2019**

111 This will be added, to the revised manuscript. Below we have included a draft figure (Figure  
112 RC2 3).

113 **3. In Figure 2b, it is curious to see the exclusion of the United Kingdom from the**  
114 **analysis of water table depth bias. As these data are publicly available, do these not**  
115 **feature in the dataset of 34 800 observations wells from IGRAC?**

116 The UK was excluded from the calibration step since it is an important hotspot for  
117 groundwater use. We calibrated the steady state model for conductivity, anisotropy and  
118 riverbed conductance. The steady state model can be well informed by these resistances while  
119 steady state allows much more model evaluations that take significantly less computation  
120 time than the transient runs. We did not further change the storage coefficients with transient  
121 calibration, as initial experiments showed limited sensitivity to changing these. Consequently,  
122 during the calibration we excluded the observation wells in the areas with significant  
123 groundwater withdrawal where persistent trends can be expected. Thereafter, we performed  
124 the spin-up and simulations with the transient model which was evaluated against the  
125 observation time series, including those from the UK (Figure 5 and 6).

126

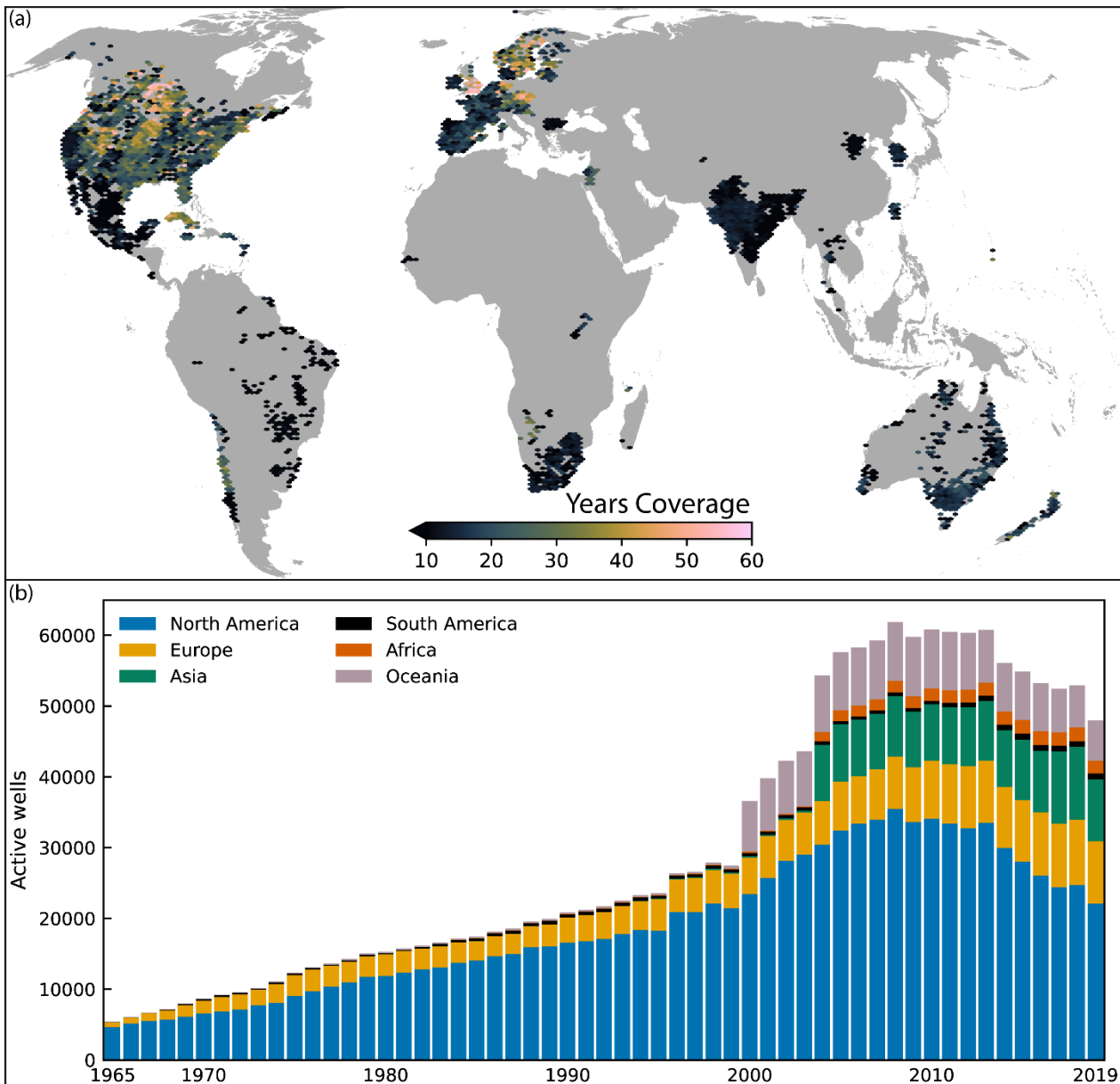


Figure RC2 3: Global distribution and temporal coverage of groundwater observations. (a) Spatial distribution of active groundwater monitoring wells worldwide, colored by the number of years of available data. (b) Number of active groundwater monitoring wells by continent from 1965 to 2019.

127

128 **References**

129 Chandanpurkar, H. A., et al.: Unprecedented continental drying, shrinking freshwater  
 130 availability, and increasing land contributions to sea level rise, *Sci. Adv.*, 11, eadx0298,  
 131 <https://doi.org/10.1126/sciadv.adx0298>, 2025.

132 de Graaf, I. E. M., van Beek, R. L. P. H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja,  
 133 E. H., and Bierkens, M. F. P.: A global-scale two-layer transient groundwater model:  
 134 Development and application to groundwater depletion, *Adv. Water Resour.*, 102, 53–67,  
 135 <https://doi.org/10.1016/j.advwatres.2017.01.011>, 2017.

- 136 Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic  
137 regulation of plant rooting depth, *Proc. Natl. Acad. Sci. U.S.A.*, 114, 10572–10577,  
138 <https://doi.org/10.1073/pnas.1712381114>, 2017.
- 139 Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not?  
140 Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*  
141 *Discuss.*, <https://doi.org/10.5194/hess-2019-327>, 2019.
- 142 Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D., and Döll, P.: Challenges in  
143 developing a global gradient-based groundwater model (G3M v1.0) for the integration into a  
144 global hydrological model, *Geosci. Model Dev.*, 12, 2401–2418, [https://doi.org/10.5194/gmd-](https://doi.org/10.5194/gmd-12-2401-2019)  
145 [12-2401-2019](https://doi.org/10.5194/gmd-12-2401-2019), 2019.
- 146 Shamsudduha, M., Taylor, R.G., Haq, M.I., Nowreen, S., Zahid, A. and Ahmed, K.M. The  
147 Bengal Water Machine: quantification of freshwater capture in Bangladesh. *Science*  
148 377(6612), 1315-1319, <https://doi.org/10.1126/science.abm4730>, 2022.
- 149 Verkaik, J., Sutanudjaja, E. H., Oude Essink, G. H. P., Lin, H. X., and Bierkens, M. F. P.:  
150 GLOBGM v1.0: a parallel implementation of a 30 arcsec PCR-GLOBWB-MODFLOW  
151 global-scale groundwater model, *Geosci. Model Dev.*, 17, 275–300,  
152 <https://doi.org/10.5194/gmd-17-275-2024>, 2024.