Response to reviewer comments for:

Long-term Hydro-economic Analysis Tool for Evaluating Global Groundwater Cost and Supply: Superwell v1.0

Hassan Niazi*, Stephen B. Ferencz, Neal T. Graham, Jim Yoon, Thomas B. Wild, Mohamad Hejazi, David J. Watson, and Chris R. Vernon

** correspondance at: hassan.niazi@pnnl.gov*

General response

We sincerely thank the reviewers for their thoughtful comments and constructive suggestions. We have thoroughly addressed each point raised by the reviewers in the form of detailed explanations in this response document. Due to the discussion forum format requested by the GMD editorial team, we have prepared this document summarizing the changes we *plan* to make in order to address the reviewer comments. These proposed changes will be fully implemented once we receive feedback from the Editor to proceed with revisions. As our responses below describe, the revised paper will clarify several areas where reviewer comments indicated the need for more specificity on model design or additional information to support modeling assumptions. Additionally, a major modeling improvement that will be incorporated in the revised paper is the addition of recharge. This capability has already been added to the Superwell code and we are in the process of generating updated figures based on this code update. The revised manuscript will incorporate the substantial reviewer feedback, including but not limited to those described above, and will be significantly improved from the initial draft. Once again, we would like to thank the reviewers for the time they invested in providing valuable feedback on our work.

Key for this document: black = comments; blue = responses

Reviewer # 1

General Comments Reviewer 1

This manuscript outlines the development of a tool that can estimate the cost of groundwater pumping across the globe. The cost and availability of groundwater pumping is a timely and important area of research, particularly in the context of climate change and increased pressures on our water supplies. My expertise is in hydrogeology and water management, and as such I will be primarily commenting on those aspects.

This manuscript is very well written and brings aspects of hydrogeology and economics together in a clear manner. I do have several concerns with respect to the methodology, particularly the explanation and description of the drawdown assessments, and with the exclusion of any other aspect of the hydrologic cycle within the analysis.

R1.1. I don't feel that this tool evaluates groundwater supply, as described by the title and implemented in the research. It seems to provide one static quantification of availability but doesn't include any other aspect of the hydrologic cycle that effects the ever evolving groundwater volumes and availability across the globe. I think you could argue that this tool can evaluate changes in supply due to pumping, but the lack of connection with any other part of the hydrologic cycle makes the claim of evaluating groundwater supply very thin.

Response. Thank you for this comment. Regarding whether Superwell evaluates groundwater supply, we agree that Superwell does not produce location-specific estimates of groundwater supply (and cost) for meeting current or future anticipated water demands within each grid-cell. As designed, Superwell provides information about extractable volumes and their associated costs of extraction and does not attempt to replicate historical groundwater depletion or project rates of groundwater extraction into the future aligned with expected water demands. However, as mentioned in the Discussion, these are both potential areas of future development. The revised text will make sure that this distinction is clear. Our preference is to keep the "supply" in the title due to improved hydrological dynamics (such as incorporating recharge as detailed below) and to maintain the relevance of the tool (and the title) for hydroeconomists and economists.

The second part of the comment to improve the connection to other parts of the hydrological cycle, even if to evaluate extractable volumes and unit costs, is a very sound suggestion that would improve model's reliability and realism. To that end, based on this comment, we have updated Superwell to include natural recharge, which we source from the widely used Döll and Fiedler (2008) dataset of long-term global recharge (the dataset itself was sourced from Gleeson et al., (2016) who provide the Döll and Fiedler (2008) in the paper data repository). The global recharge dataset has been gridded to the 0.5° Superwell input grid using spatial averaging. The revised paper will include updated results with this new recharge capability added into the model. The recharge capability is flexible, and the user will be able to replace the Döll and Fiedler (2008) dataset with a different dataset of their preference. The addition of recharge improves the surface-groundwater connection and enables the model to produce results that account for regional differences in natural groundwater recharge that can influence the long-term sustainability and cost of groundwater.

Döll, P., & Fiedler, K. (2008). Global-scale modeling of groundwater recharge. *Hydrol. Earth Syst. Sci.*, *12*(3), 863-885. https://doi.org/10.5194/hess-12-863-2008

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., & Cardenas, M. B. (2016). The global volume and distribution of modern groundwater. *Nature Geoscience*, *9*(2), 161-167. https://doi.org/10.1038/ngeo2590

R1.2. Recharge needs to be mentioned WAY before the very end. There are various sources of estimated recharge rates across the globe. I understand that it would bring in a lot of uncertainty, but this is already wrought with uncertainty, I'm not sure how much it would change it. This is connected back to the exclusion of any other part of the hydrologic cycle.

Response. Thank you for this point which ties into the last comment and its response. Overall, incorporating recharge was a common thread across the Reviewer feedback we received. To address this, we have now incorporated recharge in our evaluations. Sections “2.1 Global Hydrogeologic Input Data” and a newly created section “2.2.3 Incorporating Natural Recharge” (before 2.2.4 Hydro-economics) will mention the processing of recharge data and detailed methodology regarding how recharge impacts the pumping and storage dynamics in Superwell. Briefly, we take gridded long-term annual averaged recharge rates from Döll and Fiedler (2008) and Gleeson et al., (2016), regrid them to match the Superwell’s grid using an area-weighted approach, and adjust the ponded depth targets and depth to groundwater based on recharge rates. We implement the effect of recharge in two ways to incorporate the response in shallow subsurface that potentially reduces the pumping requirement and in deep storage that increases groundwater stocks on a longer-timescales to potentially reduce depth to groundwater. The relative contribution to shallow versus deep parts of the aquifer could be controlled by the user, where recharge rates determine the magnitudes of ponded depth target reduction and depth to groundwater reduction. In short, we have incorporated recharge dynamics into Superwell to account for the impacts on the pumping dynamics and cost accounting, and evidently removed it from the “recommendations” section. We have also updated text throughout the manuscript to reflect this change. We will also consider updating the introduction and the abstract to include aspects related to the assessment of varying groundwater costs under climate impacts as enabled by the incorporation of recharge components that alter the pumping requirements and deep aquifer storage level.

R1.3. Lateral inflows are inherently part of Theis – it is assumed that there are infinite sources of water available laterally. Saying you aren’t including them is erroneous unless you have modified Theis to include boundary conditions of some sort.

Response. We would like to clarify that this statement is referring to the absence of lateral flows *between* the 0.5°x 0.5° grid cells. We have not modified the Theis equation or used image wells to represent no-flow boundaries between grid cells. While wells hypothetically located along the boundary of a cell could experience a larger amount of drawdown than wells within the interior of the grid cell, the overall effect given the pumping duration (100 days) and inverse relationship between well spacing and well pumping rate we impose (higher capacity wells are spaced further apart), results in negligible additional drawdown at wells that would be located near cell boundaries. Furthermore, the overall water balance of each grid cell is imposed at annual time steps where the total pumped volume from all wells is summed and converted into an equivalent decrease in saturated thickness across the entire grid cell. The addition of recharge, discussed in R1.2, partially offsets the depletion from pumping and is also accounted for at the annual time scale. We will make sure to clarify this in the main text in Section “2.2.1 Modeling Well Hydraulics”.

R1.4. If the wells are pumped for 100 days (which may be very short for many parts of the globe), are they in recovery for the remaining part of the year? Is that simulated or do you just pause the groundwater levels after 100 days and start from there the next year? Both have obvious assumptions and limitations but it is not clear from the manuscript which approach is taken. I would hope that recovery is enabled through inclusion of modified Theis.

Response. Thank you for flagging this ambiguity, we will clarify this in Section “2.2.1 Modeling Well Hydraulics”. Groundwater head recovery was not simulated using superposition in time by simulating an equivalent injection rate commencing at time = 100 days. We assume that the remaining 265 days allow the groundwater head to mostly re-equilibrate to an initial state. The validity of this assumption is supported by Theis modeling we performed using superposition in time to represent the head response at the well for scenarios of well pumping ceasing after 100 days and also for scenarios of 150 and 200 days of pumping (Figure 1 R1.4). The analysis presented is for a full-factorial sample ($n = 735$) across a wide range of potential hydrogeologic values with aquifer thickness of 50, 75, 100, 125 and 150 m, hydraulic conductivity K values of 0.1, 0.25, 0.5, 1, 2, 5, and 10 m/d, S_y values of 0.1, 0.2, 0.3, and pumping rates of 100, 200, 300, 400, 500, 600, and 700 gpm (Figure 1 R1.4). Before plotting, the results were post-processed to filter out results that the violation check in Superwell would have screened as non-plausible and would have reduced the pumping rate before simulating pumping drawdown.

For the 100-day pumping scenario (Figure 1 R1.4 d), it can be seen that the head at the well location ($r = 0.2\ m$) mostly recovers by day 365. We also plot maximum fraction drawdown versus recovery error (Figure 1 R1.4 a), which is the percent difference between full recovery being achieved by day 365 (i.e., returning to initial reference head of 0 m for this test case) and the final head value resulting from the pumping/injection superposition in time to represent pumping ceasing at day 100. The recovery error shows that the recovery error was less than 1% of saturated thickness, which supports that simulating the recovery is not necessary and that it is reasonable to assume full recovery at the end of each annual time step. At the end of each annual time step the depth to water is updated to reflect the grid cell level depletion (total pumping – recharge) and it is assumed the drawdown at each well has recovered. The results also show how the recovery assumption becomes less safe at longer durations of pumping and suggests that if 200 days were considered that it might be necessary to reduce the fractional drawdown limit to 0.3, which would keep the recovery error below 2%.

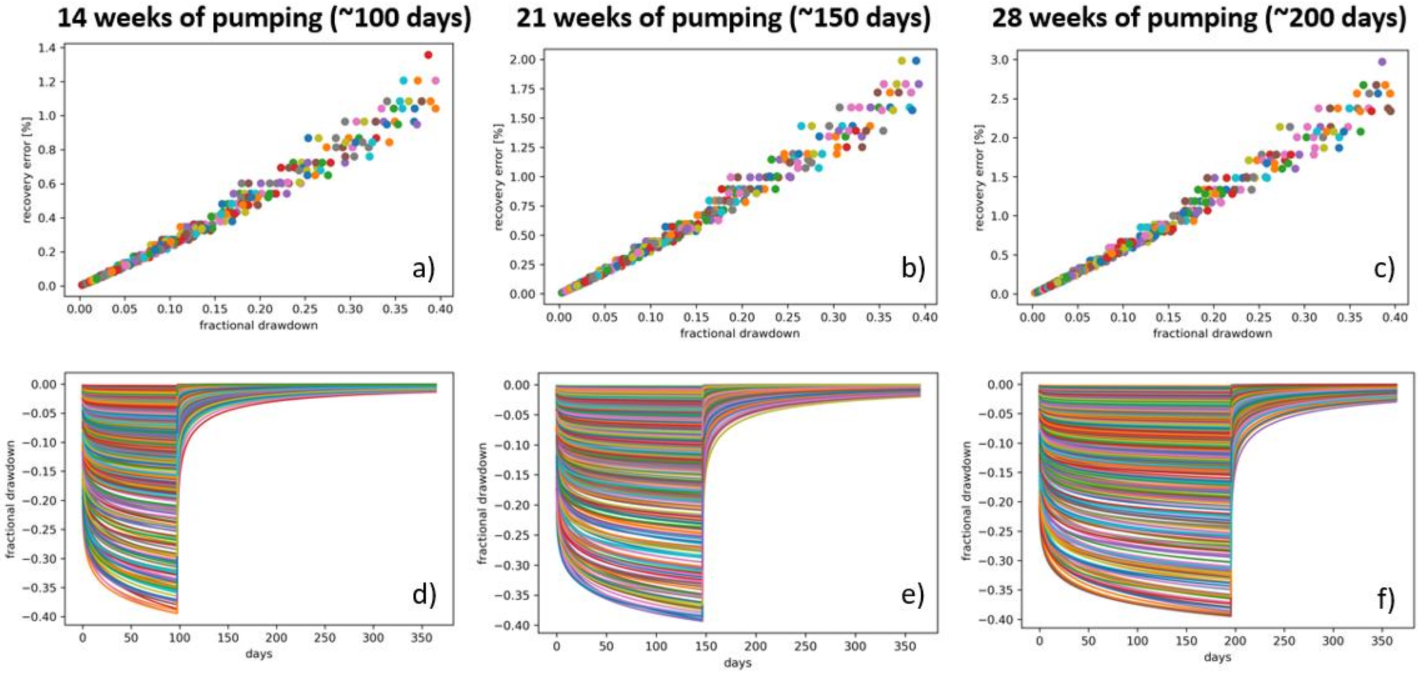


Figure 1 R1.4: Results for end of year head recovery for 100, 150, and 200 days of pumping (a-c), and corresponding fractional drawdown time series at the well (d-f).

The 100-day assumption is based on upper bounds for annual average days of irrigation well pumping from US Department of Agriculture Farm and Ranch Irrigation Survey data (Figure 2 R1.4 2). Bierkens et al. 2022 also use the assumption of 100 days of pumping for their global analysis of groundwater use for irrigation. Additionally, domestic wells or wells used for municipal supply are typically not operated 24 hours a day, 7 days a week so it seemed like a reasonable assumption to represent pumping for ~30% of the hours of the year. However, our analysis in Figure 1 R1.4 supports that bumping up the days of pumping to 150 days would not substantially adversely impact the assumption of recovery. The days of pumping can easily be adjusted in a .csv file outside of the model script and users could decide to evaluate different days of pumping either based on local information or as part of a sensitivity analysis.

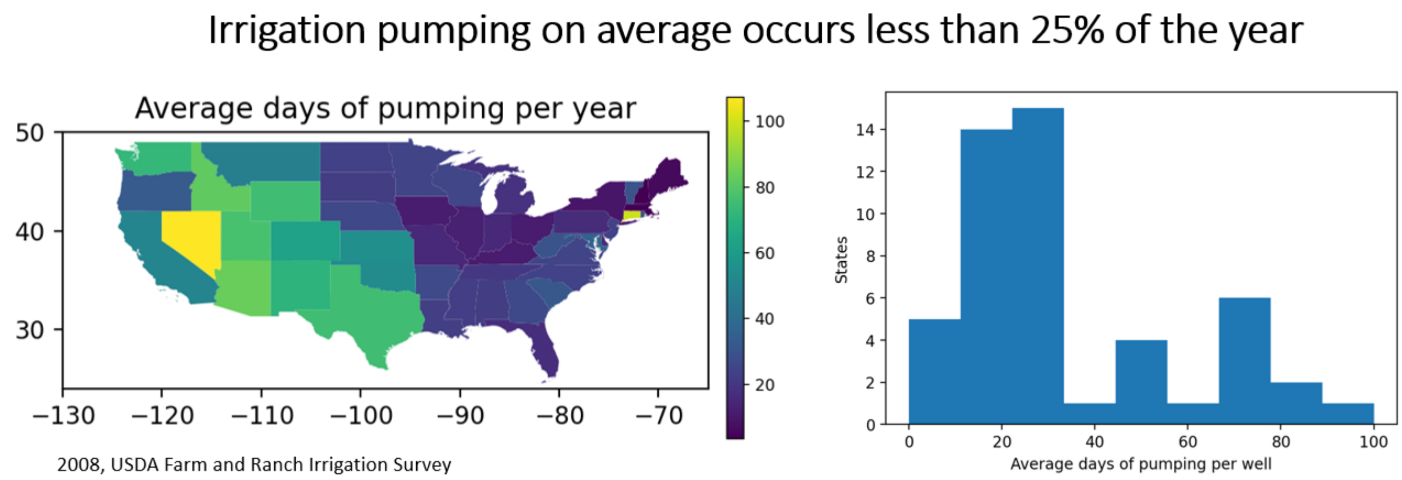


Figure 2 R1.4: Reported annual average days of irrigation well pumping by state in the continental United States (a) and the histogram of the annual average pumping data (b).

We will add the two figures included in this response to the Supplement of the revised paper so the reader can see the basis for these assumptions. We will also add additional description of the basis for our 100-day pumping assumption and also clarify the assumption about aquifer recovery to the Methods section of revised text.

Bierkens, M., De Graaf, I. E., Lips, S., Perrone, D., Reinhard, A. S., Jasechko, S., van der Himst, T., and van Beek, R. (2022). Global Economic Limits of Groundwater When Used as a Last Resort for Irrigation. https://doi.org/10.21203/rs.3.rs-1874539/v1

R1.5. Some discussion of the uncertainty in all of the datasets you use as inputs would be beneficial. Particularly because they are dependent on data that is now over a decade old. For example, how does Fan et al. (2013) capture different aquifer units? There are many instances where irrigators use deeper aquifer units that are overlain by shallow, unconfined units. In addition, how does Gleeson et al. (2014) capture this same issue? What about fractured rock aquifer which are prolific in many parts of the work and are very productive.

Response. The revised text will acknowledge thee limitations and uncertainty in global datasets more clearly and discuss the inherent simplification/abstractions of the real-world system present them. We agree there is substantial uncertainty in how well the global datasets capture aquifer properties. For example, the Fan et al. (2013) data only provides global scale estimates of pre-development unconfined water depths. Similarly, the Gleeson et al. (2014) data also only provides estimates of the dominant surficial lithology (unconsolidated sediment, sedimentary rock, volcanic rock, etc.). Given the limitation of these datasets, it is not possible to resolve layered aquifer units because we lack the hydro-stratigraphy and accompanying water depth/confined aquifer head data to represent such systems. We note that we are able capture fractured rock aquifers as these lithologies are in the Gleeson et al. (2014) dataset. For regions that rely on deeper confined aquifers for the majority of their water supply, local scale studies and information would be more appropriate, which can be noted in the revised paper.

As a result of the concerns above, in addition to some more specific, yet related comments provided below, I suggest this manuscript be returned for major revisions.

Specific Comments Reviewer 1

R1.6. Figure 3: Do you check to see if the well interference and Jacob correction result in a violation?

Response. Part of the model workflow within each simulation year is to check if the combined effects of drawdown at the well and interference from 4 adjacent wells would exceed the conservative drawdown limit of >40% saturated thickness or greater than 80 m for the given year (saturated thickness is updated each annual time step to reflect total pumping withdrawals + the effect of recharge) after 100 days of pumping (or whatever days of pumping have been specified for each annual time step). If the current well pumping rate would create a violation, a new lower pumping rate is determined, following the 2-year continuous pumping that would not have fractional drawdown >0.4 and total drawdown of 80m, before the current annual time step is simulation. This step was not fully properly illustrated in Figure 3, and we will update the wire diagram in the revised paper. We also should clarify that the Jacob correction is not applied to the screening limits, but only on the drawdown + interference timeseries outputs for each annual period to convert to equivalent unconfined drawdown values. The fractional drawdown limit of 0.4 ensures that the Jacob correction will always have a solution. We will make this distinction clear in the revised paper.

R1.7. Line 220: The accepted definition of saturated thickness is the depth from water table to a bottom confining unit, not to the depth of the well. The well can draw water from below as it follows pressure gradients. It is fine to keep this definition, but I would be clear that you are defining it much differently than the convention.

Response. Thank you for pointing this out. We will include a more qualified version of the definition of saturated thickness in the revised manuscript for clarity. We will do this in Section “2.1 Global Hydrogeologic Input Data” where we introduce input datasets and in Section “3.2.1 Diagnostics for Well Hydraulics” where we present diagnostics involving saturated thickness so readers interpret results in line with our definition.

R1.8. Line 241-244: You are assuming that there are always adjacent wells? Or do you do this when the number of wells in the grid meet a certain criteria?

Response. There are always multiple wells within a grid cell. We currently assume square packing for calculating drawdown interference from adjacent wells, which means four adjacent wells of similar sizes are placed around the well in the middle of the same size, contributing to the well's drawdown in the middle. However, upon reconsideration, in light of this comment, we have revised our approach to consider circular packing i.e., six adjacent wells contribute to the drawdown of the well in the middle, which seemed more appropriate given the radial influence of each well over the farm area served. We will include this change in the text and modify Equation 6.

R1.9. Line 252-254: You should provide the main categories of aquifers that you use – I presume they are unconfined and confined? The reader should be provided this information without having to look through supplemental information. I would guess this category determines whether the correction is needed.

Response. Given the global dataset limitations discussed in R1.5, this work assumes that all aquifers are unconfined. Additionally, de Graaf et al., (2017) also suggest that most (80 to 96%) of the global aquifer areas are characterized as unconfined aquifers. Following this assumption, the Jacob correction is always applied to correct the well drawdown resulting from Theis to the equivalent drawdown in an unconfined aquifer. We currently note this assumption of unconfined aquifers in our description of Jacob correction (Page 7), but we will include this earlier in the text as well (e.g., in high-level description of the approach in Section 2).

An additional aspect relating to the types of aquifers that has a large bearing on the cost accounting is the hydrogeological classification of the aquifers in terms of their hydrogeological complexity impacting the well installation costs. We have used the aquifer classes and their HYGE property from WHYMAPs datasets from Richts et al., (2011) to determine the complexity, i.e., WHYMAP Class 10 was major groundwater basins, Class 20 was local and shallow aquifers and Class 30 was complex hydrogeological structures (currently noted in line 159 and shown in Fig S1).

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

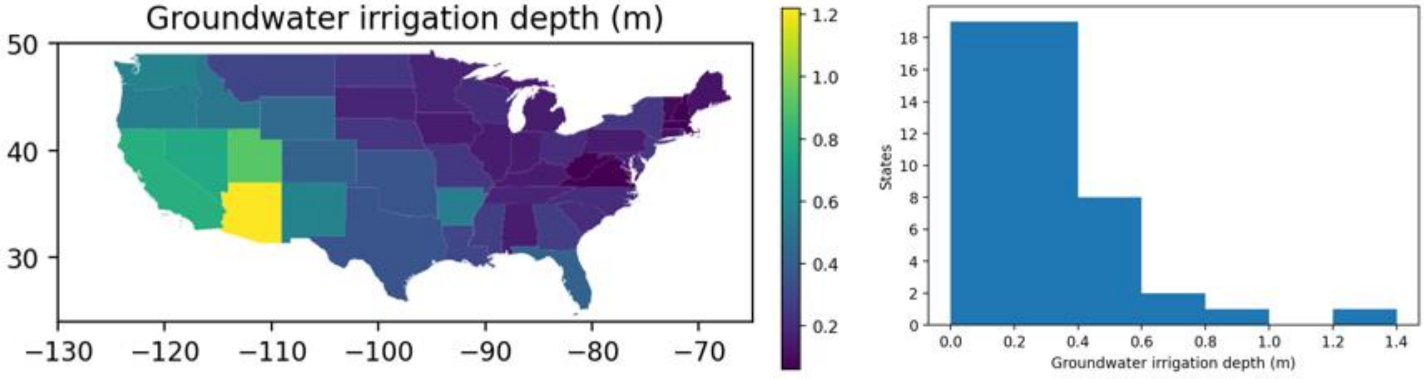
Richts, A., Struckmeier, W. F., & Zaepke, M. (2011). WHYMAP and the Groundwater Resources Map of the World 1:25,000,000. In J. A. Jones (Ed.), *Sustaining Groundwater Resources: A Critical Element in the Global Water Crisis* (pp. 159-173). Springer Netherlands. https://doi.org/10.1007/978-90-481-3426-7_10

R1.10. Line 264: Here you mention an 'off period' – is this simulated as recovery (as per comment above)?

Response. Given the ratio of 100 days of pumping, followed by 265 days of no pumping, it is not necessary to rely on Theis to calculate the recovery. See our comprehensive response to R1.4.

R1.11. Line 362: Some context for these two depths would strengthen this work – do they correlate with particular crops?

Response. Thank you for the suggestion, we will include the motivation behind the choice of 0.3m and 0.6m as exploratory ponded depths in the same section. Ponded depth targets are exploratory variables (i.e., user-defined to explore pumping scenarios) that determine the annual pumping target. We modeled 0.3m and 0.6m keeping in view the crop water needs and effective root zone depth of major crops. These two values are supported by US Department of Agriculture data on annual average GW irrigation depths (Figure R1.11). Based on this data, the 0.3 m value is a reasonable median value, while 0.6 m is a reasonable upper quartile value. However, this parameter could be fine-tuned for each crop in each run separately to model costs associated with extraction for a particular crop. The model is set up flexibly to ingest the change of parameters, including the irrigation depth, in a .csv file outside of the code.



2008, USDA Farm and Ranch Irrigation Survey

Figure R1.11: Inferred irrigation depths (m/year) from US Department of Agriculture Irrigation Survey data mapped (a) and plotted as a histogram (b).

R1.12. Line 373-375: Some regions of the Ogallala are already well beyond these values, some having already depleted most of their resources. This treatment of the aquifer as one large unit is inconsistent with how it actually works. To account for this, just reword to say that on average, it was 30% depleted – if you want upper and lower bounds you can look at recent Kansas Geological Survey reports to see the ranges within Kansas – this would communicate to the reader that you understand that these units do not operate as one big bathtub.

Response. Thanks for catching this error – our statement about depletion levels in the Ogallala lacked sufficient specificity. We will revise the statement to include both local depletion values (for example from McGuire and Strauch, 2024) and also total aquifer-level depletion estimated (for example from Konikow, 2013).

McGuire, V.L., and Strauch, K.R., 2024, Water-level and recoverable water in storage changes, High Plains Aquifer, predevelopment to 2019 and 2017 to 2019: U.S. Geological Survey Scientific Investigations Report 2023–5143, 15 p., https://doi.org/10.3133/sir20235143

Konikow, L. F. (2013), Groundwater depletion in the United States (1900–2008), U.S. Geol. Surv. Sci. Invest. Rep., 2013-5079, 63 p., doi:10.1111/gwat.12306

R1.13. Section 3.2.1: I struggle with this whole section because nothing here is new or novel to the hydrogeology community. I understand that this manuscript is reaching a multidisciplinary audience but given the length of the manuscript I think it could be moved to supplemental information.

Response. Thank you for your comment. We agree that parts of the model evaluation section and the insights therein may not be new for more experienced hydrogeologist readers. To address that, we could move Section "3.2.1 Diagnostics for Well Hydraulics" and corresponding subplots a-c from Figure 4 to the Appendix. However, we think leaving behind only "3.2.2 Diagnostics for Cost Dynamics" would make the model evaluation section incomplete. Therefore, in the interest of completeness, our preference is to keep section 3.2.1 in the main text to familiarize the non-hydrogeologist readers (especially economists) with some of the key relationships, which may be difficult or impossible to infer in the gridded maps, but are important for driving the findings presented in the Results.

R1.14. Line 453: How does the Vavailable term change with time? Again, the problem with this is that you are completely removing GW from the hydrologic cycle. There is data and research that can support bringing it back in (e.g. inclusion of recharge), and I don't feel that doing so is an unreasonable request.

Response. Thank you for this important question. As of the model presented at the time of this comment, $V_{available}$ did not change with time. It is calculated using the global datasets to give the reader a sense of how much groundwater is initially in storage according to those datasets. However, $V_{available}$ is not used in any part of the Superwell model workflow – i.e., it is not a decision variable in the model wire diagram shown in Figure 3. As mentioned in response to R1.2, we have added recharge to Superwell which updates the deep storage volume and reduces the depth to groundwater based on recharge rates contributing to the deep storage (the user can control this relative contribution based on observations). So, at the grid-cell level, $V_{available}$ (if it were a metric tracked during simulation) would now evolve over time as the net result of depletion due to pumping and annual recharge. Thus, in very arid regions groundwater depletion rates are almost equivalent to total pumping, while in wetter regions a non-negligible portion of groundwater depletion is offset by recharge and pumping can be sustained for much longer before depletion limits are reached. We will include this clarification in our description of recharge methodology and in the explanation of Figure 5 (showing $V_{available}$).

R1.15. Figure 5: When is this representative of? Groundwater supply is not stationary and constant. Also interesting that the Great Lakes are not removed from this reporting, as with other inland lake regions - was there a reason for this?

Response. The ponded depth shown in Figure 5 is the total groundwater water storage at time = 0 and is derived from the global datasets presented in Section 2.1 and Figure 2. Ponded depth (in meters) is calculated by dividing the water volume for each grid cell by grid cell area; the same value can be derived by multiplying the total saturated thickness by the porosity (which we assume to be the specific yield; see Equation 17 in the paper). The figure is meant to provide the reader a sense of the geographic distribution of initial groundwater volume as defined by the global input datasets. We agree with the suggestion of excluding land areas inundated with major lakes. We will approach this by expanding our screening criteria, which already filters out grid cells with certain characteristics (e.g., very small grid areas or very high aquifer thicknesses). If we are unable to define screening criteria based on input datasets (and their combinations), we will then explore geospatially intersecting Superwell grid with major lakes and exclude overlapping regions to fully avoid grid cells with lakes. We will use HydroLAKES dataset to do that (https://www.hydrosheds.org/products/hydrolakes). This poses a challenge of separating out an input dataset that calculates the available volume and an input dataset that is used for pumping simulations, hence our preference for improving the screening criteria.

R1.16. Section 6.1: Since you highlight the ability to work at a variety of scales (Figure 11), I would think that the first step could be to calibrate against smaller-scale depletion – for example the well documented depletion in the High Plains Aquifer that you discussed earlier.

Response. This, too, is a helpful suggestion for how to improve Section 6.1. The revised paper will be more specific in describing how one could approach implementing historically informed groundwater extraction and calibrated groundwater depletion. We will incorporate the suggestion of starting in a small number of aquifers where sufficient historical data exists for validation, such as portions of the High Plains Aquifer or the Central Valley, CA.

R1.17. Section 6.3: As described in several previous comments, I think this is a bigger issue than this one paragraph insinuates. I don't have additional comments beyond those given in sections above, but rather point to this as one place that can be extended to better capture the implications of the rest of the hydrologic cycle on this work.

Response. Thank you for this comment about incorporating recharge. We have now incorporated recharge in our evaluations. Please see the details in response to the R1.1 and R1.2 that gives an overview of the changes in the code and manuscript. We will remove this paragraph from the recommendations.

Reviewer # 2

General Comments Reviewer 2

Niazi and colleagues present a distributed model to estimate the cost of groundwater extraction based on the available water volume. Many regions in the world rely on groundwater extraction for their water supply. Meanwhile, climate change and anthropogenic activity have pushed the groundwater balance out of statistical stationarity. Thus, the topic of estimating groundwater extraction cost under changing environmental conditions is timely and of interest to the readership of the journal.

The manuscript is well-written and easy to follow. The introduction gives a concise and clear overview of the model's place in the existing state of research. The methodology is clearly outlined and the results give a good impression of the model's capability.

Response. Thank you for the accurate assessment of our work and for reinforcing statements regarding the need and utility of the model. Please find point-by-point responses to your comments below, including actions we intend to take to incorporate your feedback to improve the model and the manuscript.

Specific Comments Reviewer 2

R2.1. My major concern is that the spatio-temporal dynamics of groundwater recharge and non-anthropogenic losses such as phreatic root water uptake are neglected, which—in my opinion—makes the model operate under rather constraining assumptions that are unsuitable for long term predictions.

Response. Thank you for this comment about including more hydrological processes. We are happy to share that we have now incorporated recharge in our evaluations. We source natural recharge data from the widely used Döll and Fiedler (2008) dataset of long-term global recharge (the dataset itself was sourced from Gleeson et al., (2016) who provide the Döll and Fiedler (2008) in the paper data repository). We take gridded long-term annual averaged recharge rates from Döll and Fiedler (2008) and Gleeson et al., (2016), regrid them to match the Superwell's grid using an area-weighted approach, and adjust the ponded depth targets and depth to groundwater based on recharge rates. We implement the effect of recharge in two ways to incorporate the response in shallow subsurface that potentially reduces the pumping requirement and in deep storage that increases groundwater stocks on a longer-timescales to potentially reduce depth to groundwater. This improves the spatio-temporal dynamics of extractable volumes and associated capital, maintenance, and energy costs. The relative contribution to shallow versus deep parts of the aquifer could be controlled by the user, where recharge rates determine the magnitudes of ponded depth target reduction and depth to groundwater reduction.

The revised paper will include updated results with this new recharge capability added into the model. Specifically, Sections "2.1 Global Hydrogeologic Input Data" and a newly created section "2.2.3 Incorporating Natural Recharge" (before 2.2.4 Hydro-economics) will mention the processing of recharge data and detailed methodology regarding how recharge impacts the pumping and storage dynamics in Superwell. The recharge capability is flexible, and the user will be able to replace the Döll and Fiedler (2008) dataset with a different dataset of their preference. The addition of recharge improves the surface-groundwater connection and enables the model to produce results that account for regional differences in natural groundwater recharge that can influence the long-term sustainability and cost of groundwater. In short, we have incorporated recharge dynamics into Superwell to account for the impacts on the pumping dynamics and cost accounting, and correspondingly will remove it from the "recommendations" section. We will also consider updating the introduction and the abstract to include aspects related to the assessment of varying groundwater costs under climate impacts as enabled by the incorporation of recharge components that alter the pumping requirements and depletion rates of aquifer storage.

Regarding the comment about phreatophyte groundwater use, we interpreted this comment in two ways: 1) first this could mean representing shallow groundwater uptake by vegetation. In the Superwell framework, this would only occur/be possible for very shallow groundwater that would in a decade or two be depleted beyond the vegetation root depth after pumping depletion has resulted in dropping water tables and would also pose a challenge of how to parameterize average annual plant groundwater uptake based on local phreatophyte specifies and climate, this comment could also be referencing 2) the effect of vegetation on net groundwater recharge rates (the amount of recharge that plants intercept and use and therefore does not become deep groundwater recharge). This would also be extremely difficult to globally parameterize. In summary, we think the representation and parameterization of phreatophytes is outside the scope of this effort. The issue of phreatophyte groundwater use is something we can note in the Discussion as something this work is not designed to address, but future work could be focused on refining such representations and assessing the implications for groundwater depletion trends.

Döll, P., & Fiedler, K. (2008). Global-scale modeling of groundwater recharge. *Hydrol. Earth Syst. Sci.*, *12*(3), 863-885. https://doi.org/10.5194/hess-12-863-2008

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., & Cardenas, M. B. (2016). The global volume and distribution of modern groundwater. *Nature Geoscience*, *9*(2), 161-167. https://doi.org/10.1038/ngeo2590

R2.2. I appreciate the parsimony of the presented model and am aware of the additional complexity that adding (eco)hydrological processes would bring. But I think it is crucial to address this issue. Perhaps the most feasible way to address this issue would be a model coupling (one- or two-directional) with a global hydrological model that could account for the seasonal and long term spatio-temporal dynamics of groundwater.

Response. Thank you for this comment, please see our comprehensive response to comment R2.1 that details the description of the changes we have made to improve the connection with hydrological processes. Here, we would like to note that a one-way coupling with a global hydrological model (in which vegetation is prescribed) may not align with one of the primary purposes of our modeling, which is simulating groundwater extraction in a light-weight, flexible framework to provide volume and cost relations for large models to use this information for more rigorous modeling of demand-driven groundwater depletion and cost dynamics to understand human response in cropping patterns. The long-term limits of extraction and associated costs of production are folded into temporally aggregated supply-cost curves relating cumulative volume produced to unit cost of pumping. We are confident that the changes we have made during incorporation of natural recharge on an annual, gridded scale would improve long-term spatio-temporal dynamics of groundwater extractability and its associated capital, maintenance, and energy cost of production.

R2.3. Regarding the cost model, I have some minor questions:

1. How do you account for inflation?
2. For global and long-term predictions, how do you account for variation in the cost of energy?

Response. Thank you for this question. Starting with a second question that has a simpler response, the cost of energy is varied by country and is sourced from the International Energy Agency's country electricity rates. A processed version of this data has been provided as part of the default model package: https://github.com/JGCRI/superwell/blob/main/inputs/GCAM_Electricity_Rates.csv

For the first question, yes we do implicitly include the impact of inflation and we will add the clarification in the revised manuscript, thank you for pointing this out. We incorporate the impact of inflation through our capital cost calculation that uses interest rates to determine amortization factors. Specifically, a transformed version of the annuity equation (Equation 11 in the paper; reproduced below) converts the total cost of a well into a series of uniform annual payments over the well's lifetime, while taking into account the combined impact of interest rate and inflation. Here, we assume an interest rate of 10% that includes the impact of cost of capital financing and inflation. To fully disaggregate the interest rate and inflation, we can update the formulation to calculate nominal interest rates, but this will not have any bearing on the capital cost and unit calculations, and will not alter the results presented in this paper, since the interest rate value is an assumption and includes the combined impact of cost of financing and inflation. A more comprehensive next step could be to collect and process country-specific interest rates and long-term inflation estimates and use the nominal interest rate calculation before calculating amortization factors.

$$Ammortization\ Factor\ = \frac{i \times (1+i)^t}{(1+i)^t - 1}$$

where i = interest rate and t = well lifetime. To incorporate inflation adjustment, we can add the following equation to endogenously determine the nominal interest rate and use the equation above to calculate the amortization factor:

$$i_{nominal} = (1 + i_{real})(1 + \gamma) - 1$$

where $i_{nominal}$ is the inflation adjusted interest rate, i_{real} is the interest rate of that year, and γ is the inflation rate.

We will make this clear in the manuscript and add the possibility of collecting and processing country-specific interest rates and long-term inflation estimates in Section "6.5 Improved, Disaggregated, and Downscaled Datasets".

To address my major comment, I recommend major revision of the manuscript.

Reviewer # 3

General Comments Reviewer 3

The well-written paper by Niazi et al is about a physics-based groundwater extraction and cost accounting model (Superwell) that allows the computation of GW supply-cost curves at a global 50km x 50km resolution. The paper fills the gap between the available purely physical and purely economic datasets related to GW. Physical datasets include remotely sensed products, in-situ sensor networks, proxy measurements, numerical model outputs, and other methods for storage and fluxes that do not consider the costs of pumping and the related human feedback. Many of these datasets do not have global coverage. Economic datasets are usually lumped and do not consider the hydrogeological properties of the underlying aquifers or the long term dynamics of the resource. As a result, most agent-based models that aim to study water management issues in a basin do not incorporate well-hydraulics. Similarly, basin-scale integrated assessment models that can be used for planning or scenario generation have rudimentary GW components due to the missing coupling between the physical and economic components of GW pumping.

As someone interested in coupled human-water systems and socio-hydrological studies at multiple scales, the reviewer is pleased to receive this work as it would allow the incorporation of the GW cost curves as simple look-up tables without the need to deal with the extreme spatiotemporal complexity of groundwater flows that couple with equally complex water usages. Superwell is based on a simplified first-principles model that requires physical parameters such as porosity, permeability, aquifer thickness etc. Some other required parameters such as transmissivity, drawdown depth, hydraulic conductivity, radius of the well's influence, etc. have been derived from the basic parameters. Similarly, economic parameters such as capital and maintenance costs, energy costs etc have been used. Some more complex aspects such as pumping behavior, regulations, transportation costs, treatment costs have not been modeled or simplified. For example, it is assumed that each well pumps for 100 days / year and the remaining 265 days are assumed for recovery. The hydro-economic computations are similarly simple and easy to track.

For each grided cell, 6 scenarios have been generated by crossing 2 ponded depths with 3 intensities of volume depletion. As a non-expert, I am unable to assess whether these choices are appropriate or if more scenarios should have been generated.

The simplicity of the model and its potential for immediate incorporation into socio-hydrological models seems to be the model's biggest strength. For example, the authors have indicated the ease of incorporation in an ABM-based study for irrigation decisions in the US (ref: Yoon et al. 2024) by simple look-up tables.

The various applications mentioned in the paper are very exciting. The application paper by authors in Nature Sustainability (Niazi 2024d) opens exciting directions for future work and further development of the model.

The compilation and availability of both input and output datasets and code are themselves a major contribution, for which the authors must be congratulated.

Overall, this is a well written paper and it is recommended for publication with minor revisions.

Response. Thank you for your appreciation of the model and the gaps it aims to fill. We fully agree with your assessment of its intended use and potential utility for a range of modeling communities. We have also noted down some of the complex processes you accurately mentioned were absent in the model as potential areas of improvement for future model development efforts. Please find specific point-by-point responses to your comments below.

Specific Comments Reviewer 3

Some specific comments/questions are below:

R3.1. The biggest concern is that the GW model does not have a recharge component. Also, the interactions with natural surface flows and irrigation is missing. This is acknowledged by the authors but it makes the non-expert whether the hydrology can be relied upon.

Response. Thank you for the identification of the missing recharge component. We are happy to share that we have now incorporated recharge in our evaluations. We source natural recharge data from the widely used Döll and Fiedler (2008) dataset of long-term global recharge (the dataset itself was sourced from Gleeson et al., (2016) who provide the Döll and Fiedler (2008) in the paper data repository). We take gridded long-term annual averaged recharge rates from Döll and Fiedler (2008) and Gleeson et al., (2016), regrid them to match the Superwell's grid using an area-weighted approach, and adjust the ponded depth targets and depth to groundwater based on recharge rates. We implement the effect of recharge in two ways: 1) to incorporate the response in shallow subsurface that potentially reduces the annual pumping requirement and 2) recharge to deep groundwater storage that increases groundwater stocks on a longer-timescales and can offset groundwater depletion from pumping. This improves the spatio-temporal dynamics of extractable volumes and associated capital, maintenance, and energy costs. The relative contribution to shallow versus deep parts of the aquifer will be able to be controlled by the user.

The revised paper will include updated results with this new recharge capability added into the model. Specifically, Sections "2.1 Global Hydrogeologic Input Data" and a newly created section "2.2.3 Incorporating Natural Recharge" (before 2.2.4 Hydro-economics) will mention the processing of recharge data and detailed methodology regarding how recharge impacts the pumping and storage dynamics in Superwell. The recharge capability is flexible, and the user will be able to replace the Döll and Fiedler (2008) dataset with a different dataset of their preference. The addition of recharge improves the surface-groundwater connection and enables the model to produce results that account for regional differences in natural groundwater recharge that can influence the long-term sustainability and cost of groundwater. In short, we have incorporated recharge dynamics into Superwell to account for the impacts on the pumping dynamics and cost accounting, and evidently removed it from the "recommendations" section. We will also consider updating the introduction and the abstract to include aspects related to the assessment of varying groundwater costs under climate impacts as enabled by the incorporation of recharge components that alter the pumping requirements and deep aquifer storage level.

This model does not have a connection to surface water flow modeling and therefore there is not a direct way to incorporate surface flows into the model. However, our additional of surface recharge as a way to reduce annual groundwater demand and reduce groundwater depletion is a substantial improvement towards representing the role of surface water infiltration on groundwater sustainability and groundwater demand.

R3.2. The method for model calibration / evaluation is a bit complicated to follow. It is totally appreciated that the model only provides a plausible range of future pumping rates (hence the value of the multiple scenarios). But why historical data in limited geographies and limited timespans cannot be used for validation is not clear (If it is due to the unavailability of global data, perhaps a regional downscaled study should be planned). While the significance of expert-based evaluation is fully appreciated, the methodology referred by the authors (Gleeson et al 2021) suggests all three methods simultaneously, i.e. observation based-comparisons, expert evaluations and model-based evaluations and their interdependence. Furthermore, they advocate uncertainty quantification for a robust evaluation.

Response. Thanks for this note. We agree that historical data in limited geographies and limited timespans could be used to validate the model, but it poses significant challenges as we describe below, and we will include a condensed version of these clarification in the revised manuscript as to why this is not attempted in this work.

Structurally, the model is set up in a way to run such experiments by updating parameters in a .csv file outside the code of the model and utilizing the flexibility of the code to run based on the configuration specified by the user. However, doing such an exercise to replicate a region's pumping dynamics and associated costs will be highly constrained by the available data. For instance, validating for a small region will be conditional to data availability of not just hydrogeological properties, which we already use in our model, but also to gridded data of well properties such as well depths and pumping rates, farm areas served per well, etc., as well as the human choices and the cost data, including pumping rates, prevailing interest rates for financing, fuel prices, and labor and material costs for the installation of wells.

A major difficulty with region-specific historical validation would be in validating groundwater costs. In many cases in the U.S., the best available information from the United States Department of Agriculture (USDA) is the “pumping” cost for groundwater, which does not incorporate the infrastructure investments of drilling or maintaining the well. The USDA data is unique compared to the rest of the world in being openly available and easily accessed. Another aspect is if we were simply validating the depletion trends, that would mostly be a result of the water balance (the difference between total historical annual pumping timeseries and annual recharge timeseries) and not reflect well attributes or model representation. In the revised paper, we will acknowledge this in the model evaluation section that explains why “validation” even in one part of the US would be such a challenge.

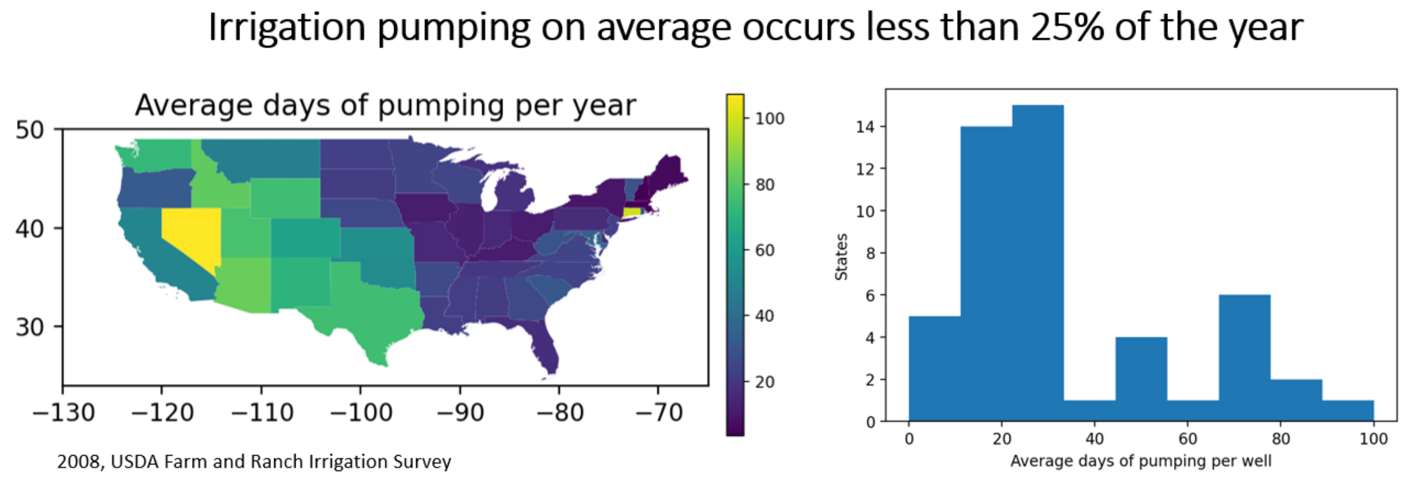
In addition, one of the main purposes of the model is to produce cost curves, relating the unit cost of production and the volume produced. The top end of the cost curve (with the highest cumulative production and highest unit cost) could only be estimated in an exploratory way i.e., no region in our knowledge, with abundant data available for us to run a simulation, has reached the physical limits of groundwater extraction (i.e., run out of groundwater). In the absence of such extreme cases, full cost curves may not be validated even with historical data. In this study, we present Table B1 with unit cost estimates in Appendix B which shows Superwell cost estimates being in the range of previously reported unit costs of groundwater production from active groundwater supply aquifers.

We agree with the importance of the comment regarding the exploration of uncertainty. We do explore the uncertainty around varying depletion limits and pumping targets using a full factorial global sensitivity analysis of pumping targets and depletion limits. More discrete instances could be set up to explore the outcome space, however, we believe the results would remain within the ranges we have already presented in this study. That said, we agree that the sensitivity analysis could be expanded to other variables such as interest rates, installation cost per unit of well depth, maintenance cost factor, well lifetime, and pump efficiency, among others. The model is set up flexibly to ingest these parameters in a .csv file outside the code. We will make modifications to the model to improve the ease of setting up large experiments without manual launching of the simulations. We will create a new section “6.6 Sensitivity for Groundwater Production and Costs“ to highlight the need for a more comprehensive sensitivity analysis for follow-up work to understand how model parameter assumptions and uncertainty in global datasets impact groundwater production and cost estimates and eventually human-Earth system outcomes in water supply, agriculture, municipal, and industry sectors.

R3.3. Some behavioral assumptions, e.g. the well-deepening feature of the model and the choice of 100 days for pumping seem arbitrary. Can the authors provide some strong justifications for these choices?

Response. Thank you for this note. We will provide justification of these choices in the Supplementary information of the paper. Well deepening and well replacements are standard adaptive response of farmers to depleting aquifers and/or increasing water demands. We agree that the choice of deepening 50m per every deepening instance is arbitrary. However, barring a negligible number of edge cases where 50m deepening would be too much, it has no bearing on the unit cost estimates because if the initial 50m weren’t enough the grid cell will deepen again in the following years to reach the pumping target.

The 100-day assumption is based on upper bounds for annual average days of irrigation well pumping from US Department of Agriculture Farm and Ranch Irrigation Survey data (Figure 1 R3.3). Bierkens et al. 2022 also use the assumption of 100 days of pumping for their global analysis of groundwater use for irrigation. Additionally, domestic wells or wells used for municipal supply are typically not operated 24 hours a day, 7 days a week so it seemed like a reasonable assumption to represent pumping for ~30% of the hours of the year. However, our analysis in Figure 2 R3.3 supports that increasing pumping to 150 days would not substantially adversely impact the assumption of recovery between annual time steps. We will note in the revised paper that pumping days/year can easily be adjusted in a .csv file outside of the model script and users could decide to evaluate different days of pumping either based on local information or as part of a sensitivity analysis.



2008, USDA Farm and Ranch Irrigation Survey

Figure 1 R3.3: Reported annual average days of irrigation well pumping by state in the continental United States (a) and the histogram of the annual average pumping data (b).

We assume that the remaining 265 days allow the groundwater head to mostly re-equilibrate to an initial state. The validity of this assumption is supported by Theis modeling we performed using superposition in time to represent the head response at the well for scenarios of well pumping ceasing after 100 days and also for scenarios of 150 and 200 days of pumping (Figure 2 R3.3). The analysis presented is for a full-factorial sample ($n = 735$) across a wide range of potential hydrogeologic values with aquifer thickness of 50, 75, 100, 125 and 150 m, hydraulic conductivity K values of 0.1, 0.25, 0.5, 1, 2, 5, and 10 m/d, S_y values of 0.1, 0.2, 0.3, and pumping rates of 100, 200, 300, 400, 500, 600, and 700 gpm (Figure 2 R3.3). Before plotting, the results were post-processed to filter out results that the violation check in Superwell would have screened as non-plausible and would have reduced the pumping rate before simulating pumping drawdown.

For the 100-day pumping scenario (Figure 2 R3.3 d), it can be seen that the head at the well location ($r = 0.2\ m$) mostly recovers by day 365. We also plot maximum fraction drawdown versus recovery error (Figure 2 R3.3 a), which is the percent difference between full recovery being achieved by day 365 (i.e., returning to initial reference head of 0 m for this test case) and the final head value resulting from the pumping/injection superposition in time to represent pumping ceasing at day 100. The recovery error shows that the recovery error was less than 1% of saturated thickness, which supports that simulating the recovery is not necessary and that it is reasonable to assume full recovery at the end of each annual time step. At the end of each annual time step the depth to water is updated to reflect the grid cell level depletion (total pumping – recharge) and it is assumed the drawdown at each well has recovered. The results also show how the recovery assumption becomes less safe at longer durations of pumping and suggests that if 200 days were considered that it might be necessary to reduce the fractional drawdown limit to 0.3, which would keep the recovery error below 2%.

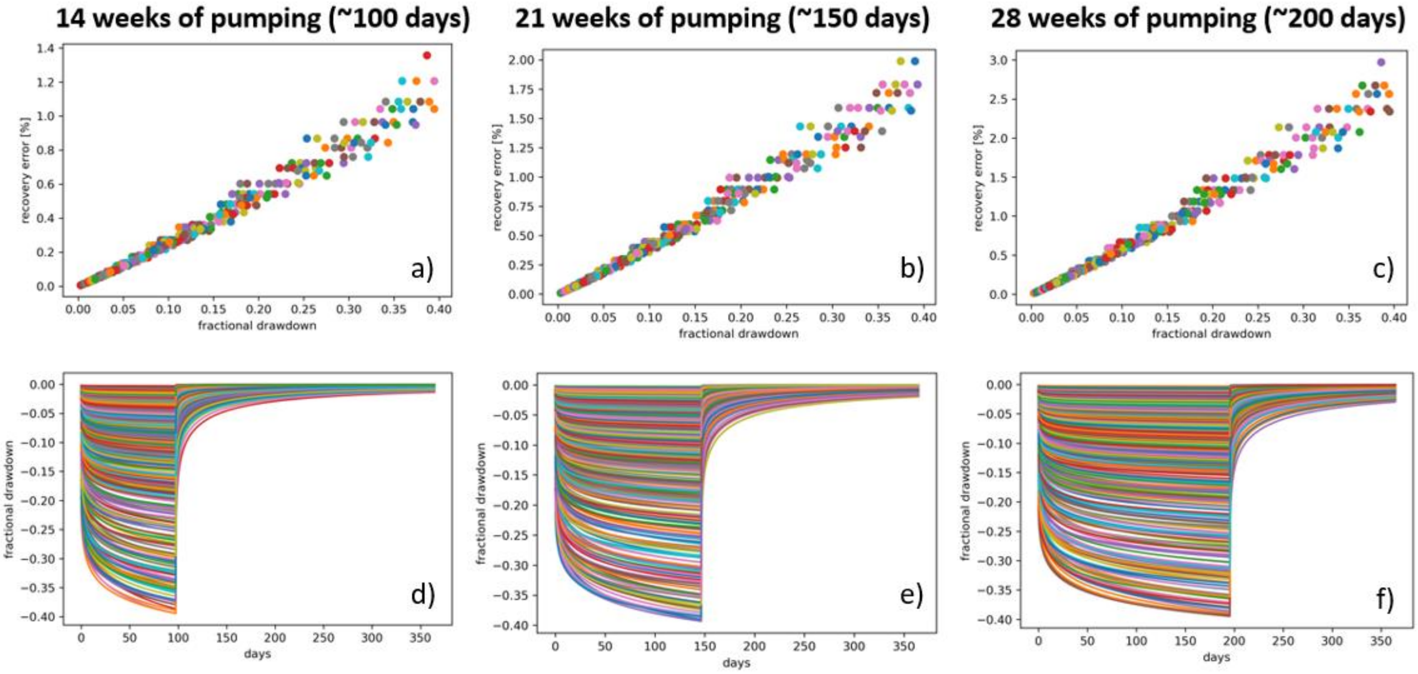


Figure 2 R3.3: Results for end of year head recovery for 100, 150, and 200 days of pumping (a-c), and corresponding fractional drawdown time series at the well (d-f).

We will add the two figures included in this response to the Supplement of the revised paper so the reader can see the basis for these assumptions. We will also add additional description of the basis for our 100-day pumping assumption and also clarify the assumption about aquifer recovery to the Methods section of revised text.

Bierkens, M., De Graaf, I. E., Lips, S., Perrone, D., Reinhard, A. S., Jasechko, S., van der Himst, T., and van Beek, R.: Global Economic Limits of Groundwater When Used as a Last Resort for Irrigation, https://doi.org/10.21203/rs.3.rs-1874539/v1, 2022.