

Peer Review File

Manuscript Title: Widespread potential loss of streamflow into underlying aquifers across USA

Editorial Notes: *none*

Reviewer Comments & Author Rebuttals**Reviewer Reports on the Initial Version:**

Referee #1 (Remarks to the Author):

I appreciate the work in compiling all of the data. This is a very strong point compared to pure modelling studies across large spatial scales. Unfortunately, I do not support publication in this specific outlet. I find the conclusions are already very well known and common sense. For example, line 170-172: Three insights are highlighted, but for none of them this study is required and they are very well known already. Yes, if water is pumped close to a river, the gradient can change from gaining to losing. This is common sense. Also, arid- and semi-arid areas essentially always have lower water tables compared to humid areas, so obviously, losing streams are more common. The same is true for the third point: Topographical gradients in steeper catchments are much more pronounced than the hydraulic gradients of the water table, so you consistently find higher depths to groundwater in mountainous areas. Some other points are:

- For water resources management and all the questions outlined in the introduction (line 54-59) the fluxes are the most essential information. Gaining or losing conditions only show the direction, but tell nothing about fluxes. I thus cannot see how this paper improves management at a local scale, or at a large scale assessment for scientific purposes.
- In the supplementary section S2 it is mentioned that the median of the water tables was used in case multiple readings were available. Rivers can change their state from losing to gaining on a seasonal (and sometimes even daily) basis. I am thus not convinced that the median is an informative indicator.
- As far as I understood the water tables were calculated by subtracting the depth groundwater from topography, and the information of topography comes from rastered DEMs. Even with a high resolution, topographical variations within a pixel might exceed the pointwise measured depth to groundwater, providing a substantial amount of noise to the data.
- I missed data on screening depth. Alluvial settings often are associated with complex nested and layered aquifers. Obviously, only head differences of systems within an aquifer system directly underlying the river are relevant. I understand that the sensitivity analysis was partly done to address this problem. But in fact, it is not the maximum depth of the well that should be analysed, but the depth of the screening which is systematically above the depth of the well itself.
- A wording issue on line 52: groundwater flows according to a gradient of a hydraulic head, which is composed of both pressure and gravity. Purely gravity-driven conditions are mainly found in the unsaturated zone.

Referee #2 (Remarks to the Author):

This paper presents the results of a laborious evaluation of aquifer water levels in wells and their relationship to the water levels in the nearest stream/river - IN *MILLIONS* OF WELLS ALL OVER THE CONTINENTAL US! I emphasize this because the numbers here are an important aspect of the study. The authors find that *most* of the wells have water levels that are at a water level below the river and therefore the river is losing (i.e., the hydrologic gradient goes from the river outward toward banks and beyond). They then assess the geography of the locations of the wells with water levels above the nearest stream/river and those below. Some general patterns emerge in which the aquifer levels higher than river elevations are mainly found in mountainous watersheds and those with heads below river elevations occur in drier climates and in regions with active groundwater extraction. This is really a key finding of this paper and one that is really interesting. It makes sense, but until we had such extensive evidence, we might not otherwise know this.

I appreciate the methods used here and trust the results as presented. Or rather, I should say that I cannot think of a better way to conduct these analyses with the scope of the data included and its availability. My main concern with this paper is the inferred conclusion that a hydraulic gradient away from the river is assumed to be a losing river. I am basing my concern in part on Darcy's Law (as a simple model) – the flux of water through the bed is determined by the product of the hydraulic conductivity and the hydraulic gradient. Whereas I agree that the gradient has to be established in the correct direction for the flux to occur, the presence of the gradient is not evidence of the flux by itself. Thus, I find it hard to call most of the streams/ivers 'losing'. That said, they certainly have the *potential to lose*.

The writing wavers between being 'cautious', i.e., the title, and in lines 104, 105, 126-7, 157. But in several places, reaches are defined as losing or gaining (lines 45-46, 49, 85, 124). It is inappropriate to state that the reaches are gaining or losing without additional confirmation/evidence. Thus, the simplest approach to resolving this is to modify the sentences in question above with an adjective like 'potential'.

This now leads me to evaluate the bottomline – does this paper stand on its own as a significant contribution to science if it is left to focus on 'potential' gaining and losing river reaches? In my opinion, the answer is yes for two important reasons. First, the typical expectation of streams and rivers is that they grow in size and discharge down the river network. The sources of increasing discharge is an expected mix of tributaries and distributed non-point inflows. The hydraulic gradients away from rivers is inconsistent with this expectation. Secondly, there are significant implications of these depressed aquifers around rivers when it comes to evaluating how to manage them as a coupled resource (as the authors indicate). It is not surprising that these depressed aquifers occur in drier climates, flat terrain, and where they are experiencing extractions. But again, the power in this analysis is in the numbers – millions of wells and nearby rivers cannot be wrong... This study is novel as it reveals a widespread condition that has not been previously understood and has huge implications for management of water across the US.

Figures – the figures are clear and on par with the expectations of Nature figures being worth more than 1000 words... Any of these could be used in a graduate hydrology course to illustrate several concepts, a qualification I often consider unique to Nature papers.

References are appropriate, and the writing is clear, with the exceptions noted above.

Review by M. Gooseff

Referee #3 (Remarks to the Author):

General comments to the authors:

The manuscript: "Widespread potential for streamflow losses into underlying aquifers", by Scott Jasechko, Hansjörg Seybold, Debra Perrone, Ying Fan, and James W. Kirchner presents a data-driven analysis of hydraulic head differences between near-stream wells ($< 1\text{ km}$ distance from the stream) and the stream across the contiguous United States (US). A large number of wells was used in conjunction with a hydrographic data base for the US to assess the potential for losing streams based on the direction of the head gradient. The data collection effort and the used data sets are impressive and I see the value of analyses of this type to elucidate general patterns and linkages. The generated national-scale maps reveal interesting and mostly consistent patterns. However, the authors should be cautious not to overstate the results of their analysis. Exchange between groundwater and streams is a complex process, which can be spatially and temporally quite variable depending on geologic heterogeneity, landscape topography, the geometry of aquifers, which defines zones of flow convergence as well as climatic drivers. Assessing such a complex process purely on the basis of head gradients between wells and the stream is challenging and results can be misleading. Hence I would suggest to further refine the analysis to make the identified patterns and general conclusions more meaningful. Some of the general conclusion, like the notion that groundwater and surface water need to be managed as a single resource, seem trivial and are not new. This has been stated repeatedly elsewhere (perhaps most prominently in works by the USGS, e.g. Winter et al. 1998, 1999). Some other general conclusions, e.g. that flat topography seems to generally promote losing conditions, I would at least question. In the following I will outline my points of concern, followed by some line by line comments.

The title should state that this a study on the continental US. Although the used data set is very large and will certainly allow some generalizations for the US, climatic and geologic conditions as well as land use and groundwater use patterns and history in other regions of the world can be too different to allow broad generalizations beyond the US.

The authors argue that the large gaps between detailed local studies have prevented a large scale assessment of patterns of losing versus gaining streams. I generally agree with this statement. However, there has also been work beyond the local studies, which has outlined some more general, landscape-scale controls (e.g. Winter et al. 1999, Relation of streams, lakes, and wetlands to groundwater flow systems, *Hydrogeology Journal*, 7:28–45). This work can provide a good basis for identifying meaningful proxies to assess patterns of groundwater-stream exchange at larger scales and should be discussed in the context of the proposed methodology.

The authors further argue that attempts for larger-scale assessments of groundwater-stream exchange using models have the problem of being difficult to verify as the model boundary conditions and parameters are uncertain and difficult to constrain. Although this is generally true for regional models, physically-based models (if they are set up correctly) typically reflect the general movement of water to drainage points (e.g. in the stream network) well and they should conserve mass. If the model is also thoroughly calibrated against stream flows and other water fluxes this should ensure a reasonable representation of the exchange with the stream network, which has evolved in response to the need of a catchment to remove excess water from the landscape. The data-driven, statistical methodology, which is proposed as an alternative to overcome the problems of models, has its own limitations (see further comments below) and in my opinion should be seen as complimentary to models (e.g. also in terms of providing calibration and validation data for large-scale models as the

authors outline in lines 75 to 77) rather than as a substitute.

As stated earlier I think that it is very difficult to evaluate the state of connection between groundwater and a stream/river purely on the basis of groundwater head data in a set of wells and stream stage, especially in landscapes with higher topographic relief. In mountainous terrain with complex topo-graphy and geology, groundwater flow is typically complex with nested local and regional flow cells (see Winter et al. 1998, 1999). The hydraulic head in a shallow riparian well could indicate a head gradient from the stream to the well, but at the same time the stream can receive inflow from a deeper regional groundwater flow cell and be net gaining. Or a groundwater head in a well outside the alluvial aquifer in the riparian zone (alluvial valley fill aquifers can be much narrower than 1km) may indicate gaining, while the groundwater head represents hydraulic conditions in a groundwater flow cell that does not or only minimally contribute flow to the stream and the stream could be net losing to the alluvial aquifer. Also flow losses and gains along streams in mountainous watersheds can coexist over short stream reaches (e.g. Covino, T. P., and B. L. McGlynn (2007), Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry, *WRR*, 43, W10431, doi:10.1029/2006WR005544, Bergstrom, A., K. Jencso, and B. McGlynn (2016), Spatio-temporal processes that contribute to hydrologic exchange between hillslopes, valley bottoms, and streams, *WRR*, 52, 4628–4645, doi:10.1002/2015WR017972), making a clear classification into losing or gaining difficult. In flatter landscapes, which tend to be characterized by spatially more extensive and thicker aquifers, groundwater flow is usually more uniform in space and head data in wells even further from the stream might give a more reliable assessment of the state of connection. This is not to say that a data driven assessment, as presented in this manuscript, could not provide insights into larger-scale patterns, but the outlined problems should be discussed/addressed to put the robustness of the large scale assessment into perspective.

One conclusion by the authors is that losing conditions are more common in drier climates, flatter landscapes and regions with extensive pumping. While I agree that drier climate and extensive pumping (two things, which also tend to be correlated as the authors correctly state) will generally promote the evolution of losing conditions, I would question that for flatter landscapes. Flat landscapes are often found around the outlet of larger catchments. Here all the water that cannot be stored in the catchment or has been lost to ET has to leave the catchment via the stream outlet. Therefore the main stem of the river network usually functions as a drainage structure unless the aquifers are heavily pumped or the climate is so dry that that large fractions of the water inputs evaporate and transpire before they reach the outlet. In fact I would argue that this drainage function can be seen in your data (Fig. 3), which for the major rivers in flat areas, even in dry climates such as in California (San Joaquin and Sacramento rivers) indicates gaining conditions (similarly for the larger river reaches in southeastern Texas and southern Mississippi in Fig. 3). This tendency, I would argue, would be even more pronounced in flat landscapes with humid climates, which may be underrepresented in your data set as well as in the local studies of your meta-analysis, where there are no studies with very flat terrain and humid climate (left upper quadrant in Fig. 1). The correlation between topography and frequency of losing conditions might be biased for two reasons: i) a general bias for drier climates in flat landscapes across the US (over the range of latitudes in the US), ii) the larger number of lower-order river reaches compared to higher-order reaches in the data base, which in dry climates will be more prone to become or to be losing than the higher order reaches. In summary, I think the conclusion that losing conditions are generally more common in flat landscapes needs to be revised or further specified. Perhaps stream order and the relative frequency of reaches of different order in the data set could be additional valuable variables to further refine the analysis.

Line by line comments:

Line 37 to 38: Some earlier references on the aspect of groundwater contributions to stream flow and

associated environmental flows would be Tabidian, M.A., and D.T. Pederson. 1995. Impact of irrigation wells on baseflow of the Big Blue River, Nebraska. *Water Resources Bulletin* 31, no. 2: 295–306 or Fleckenstein J.H., Anderson M, Fogg G.E., Mount J. (2004) Managing surface water-groundwater to restore fall flows in the Cosumnes River, *Journal of Water Resources Planning and Management*, 130(4): 301-310 and Fleckenstein J.H., Niswonger R.G., Fogg G.E. (2006) River-aquifer interactions, geologic heterogeneity, and low flow management, *Ground Water*, 44(6): 837-852.

Line 43: I would disagree with the general notion that losing conditions are more common in flat landscapes. Do you have references or data to back-up this part of the statement? In flat landscapes with humid climate streams and ditches are important structures to drain the landscape, while in steeper landscapes topographic and geologic structures can frequently cause local losing sections. The data you present in Fig. 1 for humid climates seems to show a prevalence of losing conditions in steeper terrain. The data for the drier climates shows the opposite. The latter might be related to the fact that flat landscapes are typically characterized by more extensive aquifers which in drier climates are typically more intensively pumped for water supply. This can cause spatially extensive hydraulic disconnection between aquifers and rivers. However, this is not due to the flat landscape but the more intensive pumping. Perhaps a better general statement with respect to flat landscapes would be that they are generally more susceptible to groundwater level declines caused by human activity and associated stream flow losses to groundwater.

Line 49: The words “near groundwater wells” are a significant constraint to the more general title of the manuscript, which seems to imply an even broader prevalence. Doesn’t the statement in this line also imply that human activities and their groundwater abstractions are one if not THE main reason for the observed prevalence? I think this would be a justifiable conclusion, which could be stressed more in the manuscript.

Lines 69 to 77: I agree with the statement that it is difficult to bridge between patchy local studies, but I think the assessment based on wells within 1km of the streams has its problems too. SW-GW exchange depends on many aspects in addition to just general hydraulic gradients, including topography and geology. Therefore exchange patterns can be locally quite diverse with nested flow regimes. Such patterns are difficult to generalize easily over larger scales using just local point information on GW head and stream stage elevations.

Lines 86 to 87: As stated above I challenge the hypothesis that flat landscapes promote losing conditions.

Lines 91 to 92: How could you ensure that that wells were not pumped at or shortly before the time of monitoring? Are pumping cycles recorded in the well data base? How were seasonal patterns accounted for?

Lines 89 to 106: How did you deal with the elevation of the well screens? If a groundwater well is only screened over a certain depth interval (which is usually the case) the head in this well and the head difference with respect to the stream may not tell you much about groundwater stream connectivity. In thick alluvial aquifer systems we typically have significant vertical hydraulic gradients, in particular close to the streams. Hence the head in a well may only give you limited information on the overall state of connectivity between groundwater and the stream. I think you need to provide more information on how the head data from the wells was assembled and what assumptions were made (depth of screens, time periods of measurements and seasonal variations etc.). I could not find sufficient information on those aspects, neither in the short methods section nor in the supplements.

Lines 121 to 124: As stated earlier your analysis is biased towards areas with human groundwater

abstractions. Perhaps that should be stated earlier on in the manuscript to avoid false conclusions. For a regional assessment of the effects of human activities (e.g. groundwater abstractions) on groundwater-stream connectivity versus a baseline case with no pumping, even over a range of landscape types (e.g. flat versus steep), a modeling study would be better suited than the data driven approach presented here.

Lines 135 to 137: I would consider this a significant bias in the study, which should be addressed earlier in the manuscript. It also calls for some more information on how the head data from the wells was assembled and what assumptions and simplifications were made (see also one of my earlier comments).

Lines 139 to 142: What do you mean with “not widely publicized”? Do you want to imply that groundwater depletion has not yet been fully recognized in those regions or that the problems that go along with it have not been seen as so severe?

Line 154: As said before, I do not believe that flatter topography alone promotes losing streams.

Lines 162 to 164: I fully agree with this statement and it is important for interpreting your results. Even if the partial regressions imply a statistically significant correlation between topographic slopes and the frequency of losing conditions I would doubt a general mechanistic link between the two variables.

Line 178: I would avoid the term “water table” in this context. If stream and groundwater are hydraulically connected but losing, one would not refer to the groundwater head (or hydraulic potential) below the stream as the water table. Perhaps: “...by lowering the hydraulic head below the stream...”

Lines 181 to 184: I agree with this statement. Groundwater depletion, e.g in the Central Valley of California, has been going on for decades and has disconnected many streams from their underlying aquifer some time ago.

Lines 200 to 202: This is a common feature in rivers that transition from mountainous upper watersheds via alluvial fans in mountain forelands to lowlands towards the basin outlet (see also: Covino, T. P., and B. L. McGlynn (2007), Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry, WRR, 43, W10431, doi:10.1029/2006WR005544). Often these rivers become gaining again past the alluvial fans towards their watershed outlet in the deepest/flattest part of the basin (e.g. the Cosumnes River towards its confluence with the Mokelumne River and later the San Joaquin, Fleckenstein et al. GW 2006). I think this can be seen nicely in the close-up of California in Fig. 3c of your manuscript for the rivers draining the Sierra Nevada into the Central Valley).

Lines 205 to 207: In California the notion that losing rivers are a main source for groundwater recharge in the agricultural areas of the Central Valley has been acknowledged for some time (e.g. Criss, R. E., and Davisson, M. L. (1996) Isotopic imaging of surface water/groundwater interactions, Sacramento Valley, California, J. Hydrol., 178, 205–222) and it is embedded in water management strategies including water banking (e.g. Kocis, T. N., & Dahlke, H. E. (2017). Availability of high-magnitude streamflow for groundwater banking in the Central Valley, California. Environmental Research Letters 12, 084009).

Figure 1: incorrect spelling of “Cosumnes River” in the figure (not “Consumnes”)

Supplements:

Section 1.2. How did you choose the radius of 1km? Is it sufficiently large to capture topographic features that can affect groundwater flow to the stream?

Figure 2: Why are there no studies in very flat humid landscapes (upper left quadrant of the figure)? Is that type of landscape rare in the US or not studied (perhaps because there are no problems of concern with respect to GW-SW interactions)?

Table 1, L49: "Cosumnes River" is incorrectly spelled (not "Consumnes")

Table 4: in the header of the right column it should be: "...wells have water levels that lie above the study stream (consistent with gaining conditions).", and not "lie below the stream"

Figure 15: in the text boxes "fraction of" appears twice and one instance should be removed.

Author Rebuttals to Initial Comments (please note that the authors quote the reviewers in black and respond in blue):

Referee #1 (Remarks to the Author):

I appreciate the work in compiling all of the data. This is a very strong point compared to pure modelling studies across large spatial scales. Unfortunately, I do not support publication in this specific outlet. I find the conclusions are already very well known and common sense. For example, line 170-172: Three insights are highlighted, but for none of them this study is required and they are very well known already. Yes, if water is pumped close to a river, the gradient can change from gaining to losing. This is common sense. Also, arid- and semi-arid areas essentially always have lower water tables compared to humid areas, so obviously, losing streams are more common.

We thank Referee #1 for their comments. We've made a number of revisions to our manuscript after considering their input, and our revised work is better for it. Our analyses represent an observation-based demonstration of what is occurring here and now. Our point is not that hydraulic gradients that characterize losing streams *can* occur, but that they *do* occur, and that they occur much more frequently, and in a wider range of landscapes, than anyone has previously believed (ourselves included). Our goal is not to show that losing streams occur, but to shed light on how extensive and how common losing streams are, based on observations alone with minimal assumptions.

Also, arid- and semi-arid areas essentially always have lower water tables compared to humid areas, so obviously, losing streams are more common.

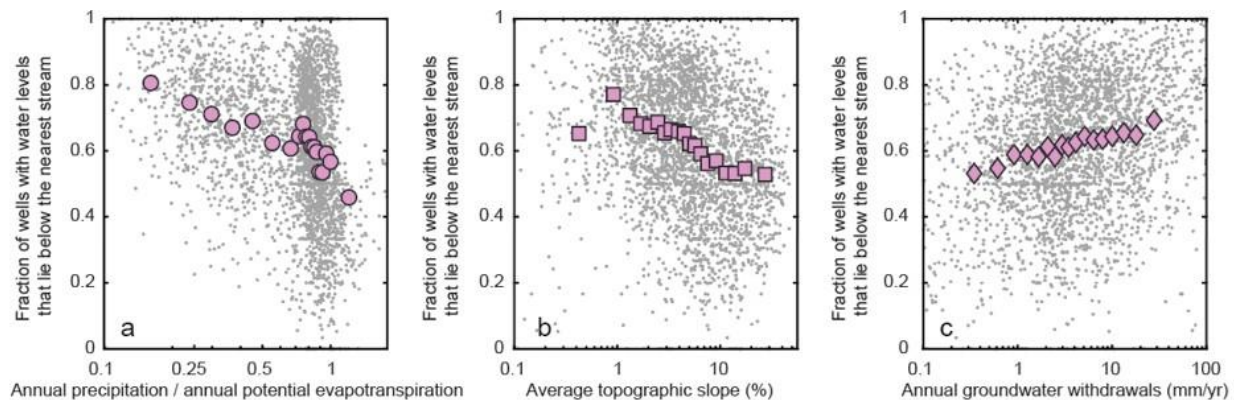
We agree that this is already understood, and this particular finding is only one of the results of our work. Our work demonstrates that there is also widespread potential for losing streams *outside* of arid and semi-arid regions (including, for example, in the Midwest, the Plains States, coastal Texas, and the Mississippi Embayment; Figures 2 and 3). In recognition that this comment highlights the need for better

communication from our side, we include the following sentence in our main text:

Lines 143-146: *“Our analysis also reveals a substantial likelihood of losing rivers in regions that are not widely considered to be groundwater depletion hotspots, including Midwestern states near the Great Lakes, Idaho’s Snake River Plain, central Washington State, and Oregon’s Willamette Valley (Fig. 3).”*

The same is true for the third point: Topographical gradients in steeper catchments are much more pronounced than the hydraulic gradients of the water table, so you consistently find higher depths to groundwater in mountainous areas.

We agree that water tables tend to be deeper in mountainous areas. Our analysis shows that although water tables in steeper areas may generally be deeper below the surface, they nonetheless tend to more frequently lie above, not below, the nearest stream (essentially because the land surface gradient is typically steeper than the water table gradient). Thus, although water tables tend to be deeper in steeper terrain, hydraulic gradients tend to more frequently favor gaining, rather than losing, streams in such landscapes. In an effort to make this latter point clearer we have added a new supplementary figure designed to help readers visualize the significant yet far-from-perfect correlation between topographic slope and the fraction of wells that lie below the nearest stream:



Supplementary Figure 24. Cross plots of each explanatory variable against the fraction of well water levels that lie below the nearest stream. Each grey point in each of the three plots represents one county. The large pink points represent the mean (average) fraction of wells in the county that have a water level that lies below the nearest stream elevation (i.e., average y-axis value) binned at intervals that capture 5% of the data as determined by the x-axis values (i.e., x-axis values have been binned at intervals designed to contain 5% of the grey points displayed on the plot). These plots suggest that the fraction of wells with water levels that lie below the nearest stream tends to be higher in (a) more arid counties (i.e., lower county-wide average values of annual precipitation divided by evapotranspiration), (b) counties that have lower average topographic slopes, and (c) counties where groundwater withdrawal rates are higher (i.e., higher groundwater pumping, expressed in a way that normalizes to the area of a county: units of mm of groundwater pumped per year).

Some other points are:

- For water resources management and all the questions outlined in the introduction (line 54-

59) the fluxes are the most essential information. Gaining or losing conditions only show the direction, but tell nothing about fluxes.

We agree that the hydraulic head data we analyze only provide insights into the *potential* for streamflow losses or gains; how much loss or gain *actually* occurs will depend on the permeability of the subsurface, which, together with the hydraulic gradient, determines the flux. Where permeability is low, steep hydraulic gradients will not translate to high rates of net exchange between the river and the aquifer. Unfortunately, continental scale data are not available to calculate fluxes, especially fluxes in complex aquifer systems. We address this important point via revisions and new text added to our revised paper:

Lines 107-112: *“Although these water level differences can identify potentially gaining and losing rivers, quantifying stream-aquifer fluxes would require high-resolution three- dimensional hydrogeologic data (e.g., permeability) that are presently unavailable at continental scale. The lack of high-resolution geologic data also means that we cannot resolve two-dimensional cross sections of groundwater flow fields, and thus cannot capture flow patterns in some complex aquifer systems.”*

Additionally, one of the conceptual models that we profile with an illustration in our Supplementary Fig. 1e (Brazos River) supports this point. To complement this supplementary figure, we state in our Supplementary Information that:

Supplementary Information Page 4” *“e) how low-permeability sediments can impede groundwater influxes to streams, even where the water table adjacent to the stream overlies the stream surface, creating a hydraulic gradient from the aquifer to the stream (see Rhodes, K.A., Proffitt, T., Rowley, T., Knappett, P.S., Montiel, D., Dimova, N., Tebo, D. & Miller, G.R. (2017). The importance of bank storage in supplying baseflow to rivers flowing through compartmentalized, alluvial aquifers. Water Resources Research, 53(12), 10539-10557) – Supplementary Figure 1e;”*

Regarding water resources management, co-author Dr. Perrone is in the process of a decade-long analysis of the legal evolution of groundwater law and management in the West. Together with Dr. Rebecca Nelson (University of Melbourne), Dr. Perrone has systematically reviewed more than 250 groundwater permitting protocols in the western 17 states. Their preliminary results agree fully with Referee 1 that fluxes between aquifers and rivers are critical to successful water resources management; it would be ideal if jurisdictions sought such information to inform management. Yet just as aquifer physical characteristics vary with space, so do groundwater management approaches and specifically the degree to which interactions between aquifers and rivers are considered. Water quantity is mostly the matter of state law; that means that each state and sub-state area has its own approach to managing groundwater withdrawals. Dr. Perrone’s preliminary (Nelson, R., Perrone, D. The role of permitting regimes in

Western United States groundwater management. *Groundwater* 54(6), 761-764) and ongoing research focused on the western 17 states identifies that:

- (1) Not all states require the consideration of impacts of groundwater pumping on surface water when considering new groundwater withdrawals;
- (2) While controls may be legally available for use, they are not always mandated, and therefore, they are not always used (i.e., implementation varies);
- (3) In places where there is a general acknowledgement of groundwater-surface water interactions reflected within legal powers to manage groundwater withdrawals, the information used to inform exercised or established groundwater withdrawals vary considerably. Two examples include:
 - a. Kansas: KAR 5-3-11, uses a distance threshold plus “calculated recharge” defined as “25% of the average annual rainfall in the area..”
 - b. Colorado: non-tributary groundwater is the withdrawal of groundwater which *would not* deplete the surface water system, within 100 years, at an annual rate greater than 0.1% of the rate of withdrawal

We agree that fluxes are critical, but also acknowledge that sometimes groundwater-surface water interactions are not accounted for or not mandated in the legal controls used to manage groundwater. Additionally, when consideration is given to the impacts of pumping on surface waters, the legal mechanism is not always the same, such that a comprehensive assessment of which fluxes are used would warrant its own publication, likely taking years to amass the management data through legal analysis.

Our main finding—that well water levels are often below nearby stream surfaces, implying the existence of a hydraulic gradient that is consistent with a losing stream—highlights the importance of large scale assessments that use locally relevant data (like ours) to highlight the importance and ubiquity of hydraulic potentials that make exchange between rivers and aquifers the norm, not the exception. Pertaining directly to water resources management, our findings demonstrate the importance of developing new laws and policies *and/or* further implementing existing laws and policies that require consideration of groundwater-surface water connections. To do a better job of linking our research more directly to groundwater management research, we have added another reference to groundwater law research led by Dr. Rebecca Nelson (citation #40 is: Nelson R. & Quevauviller P. Groundwater Law. in: Integrated Groundwater Management. (eds) Jakeman, A. J., Barreteau, O., Hunt, R. J., Rinaudo, J. D., Ross, A. Springer, Cham. pp. 173-196 (2016)) in our revised manuscript’s main text, that now reads:

Line 218-220 *"In some jurisdictions, groundwater pumping permit applications are evaluated for their likely effects on streamflow⁴⁰, providing an opportunity to avoid or limit streamflow depletion."*

Lastly, qualitative knowledge is critical, powerful, and often the first step, toward understanding complex systems. Qualitative knowledge, such as conceptual models before we build numerical models, are more important because they represent a first-order and higher level of understanding of a system. In our case, knowing where losing streams are prevalent and why, and knowing that in a particular region the direction of this exchange, has direct relevance to resource managers. I thus cannot see how this paper improves management at a local scale, or at a large scale assessment for scientific purposes.

It would be wonderful to be able to quantify fluxes between rivers and aquifers at the millions of points where we have water table measurements, but no reliable data on streambed or aquifer permeability exist at continental scale. Our analysis nonetheless informs management by demonstrating that the potential for losing streams is widespread across the US, and by quantifying the influence of several causal factors. Thus, our analysis should help focus managers' attention where it is most clearly needed, even if we cannot quantify exactly which reaches of which streams are losing exactly how much.

To be clearer about these points, we have made the following changes:

- 1) We added a new sentence to the introduction of our manuscript stating:

Lines 86-88 *"Here, we use the term "losing" to refer to cases where water levels in aquifers lie below their adjacent rivers, but we stress that the magnitude of any actual streamflow losses will depend on the permeability of the underlying strata."*

- 2) We modified the abstract of our manuscript to make it more conservative; for example, we revised the text that appeared originally on Lines 45-46 of our initial submission to now read in the revised manuscript as "potentially losing":

Lines 46-50 *"Recent modeling studies¹⁰ have suggested that losing streams could become common in future decades, but our direct observations show that many rivers are already potentially losing, highlighting the importance of managing groundwaters and surface waters as the connected resources that they are¹."*

- 3) We also reinforce that the present-day lack of high-resolution 3D hydrogeologic data limits our ability to confidently calculate fluxes by adding the following new text to a new Methods section titled "Limitations of our analyses":

Lines 259-261 *“Because no locally relevant, high-resolution three-dimensional database of permeability exists at continental scale, only the potential for rivers to lose or gain (rather than the fluxes of actual losses or gains) can be calculated with confidence across the USA.”*

- **In the supplementary section S2 it is mentioned that the median of the water tables was used in case multiple readings were available. Rivers can change their state from losing to gaining on a seasonal (and sometimes even daily) basis. I am thus not convinced that the median is an informative indicator.**

We used the median as an unbiased and robust estimator of the central tendency in water levels, consistent with widespread scientific practice. We agree that rivers can change from gaining to losing and back again, as evidence by studies profiled and detailed in our Supplementary Information:

Supplementary Information Page 3: *“For example, local-scale studies highlight how certain stream reaches can lose water to aquifers during one time interval but gain via groundwater discharges at other times (e.g., see Zimmer, M. A., & McGlynn, B. L. (2017).*

Bidirectional stream–groundwater flow in response to ephemeral and intermittent streamflow and groundwater seasonality. Hydrological Processes 31, 3871-3880). Changes in the directions of flow between aquifers and streams can arise in numerous ways, including “bank flow”, where floodwaters infiltrate streambanks and cause stream reaches that usually leak into aquifers to temporarily gain flow as floodwaters in streambanks seep back into the streams from whence they came (e.g., see Lamontagne, S., Leaney, F. W., & Herczeg, A. L. (2005). Groundwater–surface water interactions in a large semi-arid floodplain: implications for salinity management. Hydrological Processes 19, 3063-3080).”

In a new Methods subsection titled “Limitations to our analyses”, we added the following statement to explicitly mention temporal variations in gaining-versus-losing conditions:

Line 261-265: *“Another limitation of our analysis arises from the lack of water level time series for most wells; because most of our wells have just one water level measurement, we cannot evaluate temporal variations in gaining versus losing conditions, nor can we assess the impact of seasonal fluctuations in climate and groundwater withdrawals on river- aquifer exchanges.”*

- **As far as I understood the water tables were calculated by subtracting the depth groundwater from topography, and the information of topography comes from rastered DEMs. Even with a high resolution, topographical variations within a pixel might exceed the pointwise measured depth to groundwater, providing a substantial amount of noise to the data.**

Yes, we agree. This is why we tested the potential magnitude of this issue in the Supplementary Information. As a thought experiment, if we assume a relatively steep topographic slope of 5%, the topographic variability in a 10 m grid cell could contribute a maximum of 0.25 m of

uncertainty in the water table elevation (because the well or stream location could be a up to 5 m from the grid center). For a 30 m grid cell the corresponding maximum uncertainty is 0.75 m (and again, this assumes a topographic gradient of 5%, which is much steeper than most of the continental US). Our sensitivity analyses presented in Supplementary Figure 19 show that even if all of our measurements were systematically biased by these amounts (0.5 ± 0.25 m), our main findings would not change. But the real effect will be much smaller, because within-pixel topographic variations will primarily introduce random noise rather than persistent bias, and thus the random noise will not systematically alter the fraction of well levels that lie below vs. above the adjacent stream.

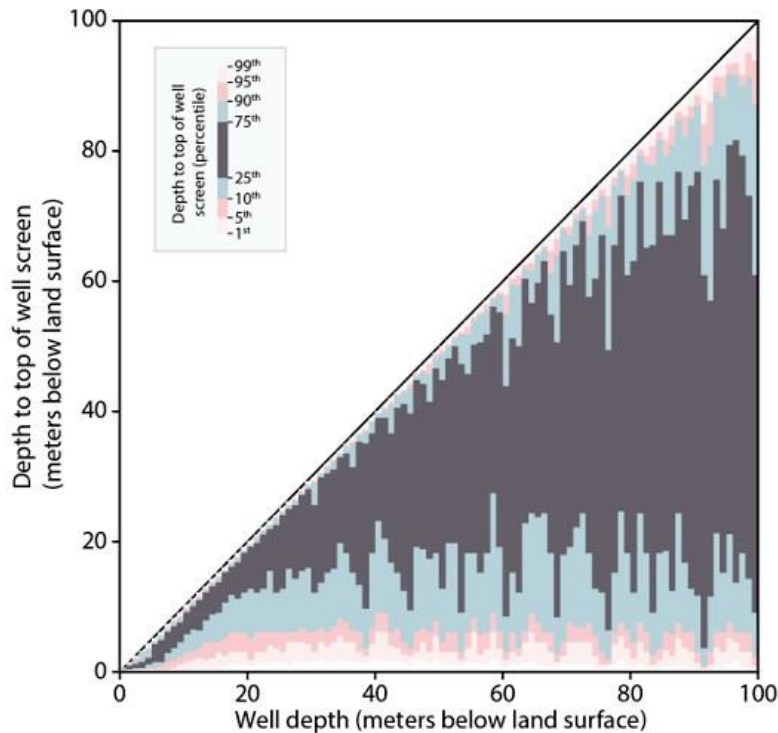
We have added the following statement to the Limitations section in the Methods:

Line 265-269: *“Furthermore, imprecise well location data and the coarseness of the digital elevation data we analyze (30 m by 30 m, or 10 m by 10 m) will lead to uncertainty in land surface elevations and, therefore, in our estimate of hydraulic heads, although these random uncertainties should not substantially bias our assessment of the relative prevalence of losing and gaining conditions.”*

- **I missed data on screening depth. Alluvial settings often are associated with complex nested and layered aquifers. Obviously, only head differences of systems within an aquifer system directly underlying the river are relevant. I understand that the sensitivity analysis was partly done to address this problem. But in fact, it is not the maximum depth of the well that should be analysed, but the depth of the screening which is systematically above the depth of the well itself.**

Our well water level data derive from 64 different state level databases, not all of which record well screen intervals. Even those databases that do report well screen interval information are incomplete, with some of their records reporting only the top of the screened interval, and other records lacking well screen information entirely.

However, we thank the reviewer for bringing up this important point. Using the well screen information that is available, we have added a new figure to the supplementary information comparing total well depths and the top of the shallowest recorded screen interval:



Supplementary Figure 9. Comparison of total recorded well depths (x-axis values; expressed in meters below land surface) and the recorded depth to the top of the perforated interval (y-axis values; expressed in meters below land surface). The diagonal black line marks a 1:1 relationship. Shaded areas mark the percentiles of depths to top of perforations (see legend in upper left corner): dark grey represents 25th-75th percentile range, light blue represents 10th-25th and 75th-90th percentile ranges, dark pink represents 5th-10th and 90th-95th percentile ranges, and light pink represents 1st-5th and 95th-99th percentile ranges (percentiles determined by binning the x-axis well depth data into 1 m intervals). We lack screen interval data for many well completion records, motivating us to use well depth data (which are recorded more widely than well screen interval data) in our sensitivity analyses described in section S4.2. Our comparison of the total depth of wells (x-axis) and the depth to the top of their perforations (y-axis) demonstrates that many wells (for which we have screen interval data) have perforated intervals that begin at considerably shallower depths (e.g., 75% of wells that have total depths of between 99m and 100m have a perforated interval that begins 9.1 m beneath the land surface). Therefore, the depth threshold presented in our sensitivity analyses in Supplementary Section 4.2 should be interpreted as the maximum (i.e., deepest) possible perforated interval.

This figure suggests that deeper wells tend to have longer screened sections, thus the water levels reported for these deep wells provide a vertically integrated measure of the hydraulic head, not just the bottom of the well. This fact makes the deep wells useful for our purpose because their water levels may also reflect the hydraulic head of the shallow aquifers.

We have added a new statement in the main text directing readers to this new figure and its ramifications: that the total well depth represents a maximum value for the potential perforated interval, but that water levels recorded in deep wells can also capture conditions at shallower depths when their screened intervals lie above their bottoms.

Lines 275-276: *“for a comparison of well depths and the depths of the top of the shallowest screen interval see Supplementary Fig. 9”*

We have also added a new statement to our Supplementary Information (section S4.2) describing our rationale for using total well depth as a threshold in our analysis:

Supplementary Information Page 78: *“We analyze total well depth data rather than perforated interval data because the latter is unavailable for many of the wells in our dataset. We emphasize that the total well depth represents the maximum depth at which perforations exist for a given well, but that many production wells in our database have screened intervals far shallower than their bottom (see Supplementary Figure 9).”*

- **A wording issue on line 52: groundwater flows according to a gradient of a hydraulic head, which is composed of both pressure and gravity. Purely gravity-driven conditions are mainly found in the unsaturated zone.**

Agreed. Our revised manuscript (with the statement “gravity-driven” now removed) reads as:

Lines 51-53: *“Hydraulic gradients and subsurface permeability control the exchanges of water between aquifers and rivers¹⁻³, increasing streamflow where groundwaters feed rivers (‘gaining rivers’), and decreasing streamflow where river waters seep into underlying aquifers (‘losing rivers’).”*

Referee #2 (Remarks to the Author):

This paper presents the results of a laborious evaluation of aquifer water levels in wells and their relationship to the water levels in the nearest stream/river - IN *MILLIONS* OF WELLS ALL OVER THE CONTINENTAL US! I emphasize this because the numbers here are an important aspect of the study. The authors find that *most* of the wells have water levels that are at a water level below the river and therefore the river is losing (i.e., the hydrologic gradient goes from the river outward toward banks and beyond). They then assess the geography of the locations of the wells with water levels above the nearest stream/river and those below. Some general patterns emerge in which the aquifer levels higher than river elevations are mainly found in mountainous watersheds and those with heads below river elevations occur in drier climates and in regions with active groundwater extraction. This is really a key finding of this paper and one that is really interesting. It makes sense, but until we had such extensive evidence, we might not otherwise know this.

We thank Dr. Gooseff for these comments. We particularly appreciate the point that there is an important difference between suspecting that something (such as losing relationships between streams and groundwaters) may be the case in some places, and demonstrating, with data, that it is actually a widespread phenomenon that characterizes a wide range of landscapes.

I appreciate the methods used here and trust the results as presented. Or rather, I should say that I cannot think of a better way to conduct these analyses with the scope of the data included and its availability. My main concern with this paper is the inferred conclusion that a hydraulic gradient away from the river is assumed to be a losing river. I am basing my concern in part on Darcy's Law (as a simple model) – the flux of water through the bed is determined by the product of the hydraulic conductivity and the hydraulic gradient. Whereas I agree that the gradient has to be established in the correct direction for the flux to occur, the presence of the gradient is not evidence of the flux by itself. Thus, I find it hard to call most of the streams/rivers 'losing'. That said, they certainly have the *potential to lose*.

The writing wavers between being 'cautious', i.e., the title, and in lines 104, 105, 126-7, 157. But in several places, reaches are defined as losing or gaining (lines 45-46, 49, 85, 124). It is inappropriate to state that the reaches are gaining or losing without additional confirmation/evidence. Thus, the simplest approach to resolving this is to modify the sentences in question above with an adjective like 'potential'.

We agree and appreciate the suggestion to communicate more consistently that our analysis identifies rivers with the potential to gain or lose flow via groundwater discharge or recharge. We have therefore modified each of the sentences Dr. Gooseff specifically identifies:

Lines 45-46 of our original manuscript now reads:

Lines 46-50: *“Recent modeling studies¹⁰ have suggested that losing streams could become common in future decades, but our direct observations show that many rivers are already potentially losing, highlighting the importance of managing groundwaters and surface waters as the connected resources that they are¹.”*

Line 49 of our original manuscript now reads:

Lines 44-46: *“Our results thus imply that climatic factors, geological conditions, and historic groundwater pumping jointly contribute to the widespread risk that streams lose flow into their surrounding aquifers instead of gaining flow from them.”*

Line 85 of our original manuscript now reads:

Lines 83-86: *“Here we analyze millions of well water level observations to evaluate the prevalence and spatial distribution of potentially gaining and losing rivers across the contiguous United States, and to test how the prevalence of potentially losing streams is shaped by climate, physiography, and groundwater pumping.”*

Line 124 of our original manuscript now reads:

Lines 128-130: *“..our analysis inherently focuses on regions where wells pump groundwater, and we cannot quantify the prevalence of potentially losing rivers where wells are absent.”*

This now leads me to evaluate the bottomline – does this paper stand on its own as a significant contribution to science if it is left to focus on ‘potential’ gaining and losing river reaches? In my opinion, the answer is yes for two important reasons. First, the typical expectation of streams and rivers is that they grow in size and discharge down the river network. The sources of increasing discharge is an expected mix of tributaries and distributed non-point inflows. The hydraulic gradients away from rivers is inconsistent with this expectation. Secondly, there are significant implications of these depressed aquifers around rivers when it comes to evaluating how to manage them as a coupled resource (as the authors indicate). It is not surprising that these depressed aquifers occur in drier climates, flat terrain, and where they are experiencing extractions. But again, the power in this analysis is in the numbers – millions of wells and nearby rivers cannot be wrong... This study is novel as it reveals a widespread condition that has not been previously understood and has huge implications for management of water across the US.

Figures – the figures are clear and on par with the expectations of Nature figures being worth more than 1000 words... Any of these could be used in a graduate hydrology course to illustrate several concepts, a qualification I often consider unique to Nature papers.

References are appropriate, and the writing is clear, with the exceptions noted above.

Review by M. Gooseff

We thank Dr. Gooseff for these comments. We particularly appreciate the point that the conventional wisdom would hold that rivers should be generally gaining, and that this conventional wisdom is contradicted by our discovery that water tables often lie below adjacent streams (at least where wells exist, allowing those hydraulic heads to be measured).

Referee #3 (Remarks to the Author):**General comments to the authors:**

The manuscript: “Widespread potential for streamflow losses into underlying aquifers”, by Scott Jasechko, Hansjörg Seybold, Debra Perrone, Ying Fan, and James W. Kirchner presents a data-driven analysis of hydraulic head differences between near-stream wells (< 1km distance from the stream) and the stream across the contiguous United States (US). A large number of wells was used in conjunction with a hydrographic data base for the US to assess the potential for losing streams based on the direction of the head gradient. The data collection effort and the used data sets are impressive and I see the value of analyses of this type to elucidate general patterns and linkages.

We thank Referee #3 for their thorough review. We are pleased they see the value of analyses of this kind.

The generated national-scale maps reveal interesting and mostly consistent patterns. However, the authors should be cautious not to overstate the results of their analysis. Exchange between groundwater and streams is a complex process, which can be spatially and temporally quite variable depending on geologic heterogeneity, landscape topography, the geometry of aquifers, which defines zones of flow convergence as well as climatic drivers.

Assessing such a complex process purely on the basis of head gradients between wells and the stream is challenging and results can be misleading. Hence I would suggest to further refine the analysis to make the identified patterns and general conclusions more meaningful. Some of the general conclusion, like the notion that groundwater and surface water need to be managed as a single resource, seem trivial and are not new. This has been stated repeatedly elsewhere (perhaps most prominently in works by the USGS, e.g. Winter et al. 1998, 1999). Some other general conclusions, e.g. that flat topography seems to generally promote losing conditions, I would at least question. In the following I will outline my points of concern, followed by some line by line comments.

Below we detail how we have revised our manuscript in consideration of each suggestion made by Referee #3. Doing so made our manuscript better; thank you.

The title should state that this a study on the continental US. Although the used data set is very large and will certainly allow some generalizations for the US, climatic and geologic conditions as well as land use and groundwater use patterns and history in other regions of the world can be too different to allow broad generalizations beyond the US.

We agree that this is an important change to make that will help to convey the limitations—in this case spatial—of our analyses. The revised title incorporating Referee #3’s suggestion is (additional text in green):

“Widespread potential for streamflow losses into underlying aquifers across the USA”

We made a similar change elsewhere in our main text, too (as the reviewer makes an important point that we should clarify the spatial extent of our study), adding “across the USA” to the following sentence in our manuscript:

Lines 190-191: *“Our analysis demonstrates the significant ($p < 0.0001$) relationship between groundwater pumping and losing streams across the USA (Fig. 4).”*

The authors argue that the large gaps between detailed local studies have prevented a large scale assessment of patterns of losing versus gaining streams. I generally agree with this statement. However, there has also been work beyond the local studies, which has outlined some more general, landscape-scale controls (e.g. Winter et al. 1999, Relation of streams, lakes, and wetlands to groundwater flow systems, Hydrogeology Journal, 7:28–45). This work can provide a good basis for identifying meaningful proxies to assess patterns of groundwater-stream exchange at larger scales and should be discussed in the context of the proposed methodology.

We thank Referee #3 for the suggestion to incorporate the physiographic frameworks T.C. Winter developed, which continue to guide conceptual models of river-aquifer exchanges to present day. It was important to us that we acknowledge previous local-scale studies in our manuscript and profile them in the main text, which is why we used Fig. 1 to highlight works that have come before us (in addition to profiling several conceptual models based on these previous works in Supplementary Fig. 1). We recognize the importance of local-scale research and the resulting conceptual understanding of likely gaining/losing behavior under different physiographic settings. In an effort to make this point clearer, we have added a new sentence in our discussion of local-scale studies:

Lines 68-70: *“Local-scale studies provide the foundation for our understanding of groundwater interactions with surface waters, providing conceptual frameworks that are relevant beyond just these individual field sites²³.”* (*note: reference 23 in our revised manuscript points to the highly appropriate and recommended citation: Winter, T. C. Relation of streams, lakes, and wetlands to groundwater flow systems. Hydrogeology Journal 7, 28-45 (1999))

We also note that physiographic proxies may be insufficient to diagnose gaining and losing conditions—that these conceptual models provide guidelines that may not capture actual gaining versus losing conditions in areas with similar physiographic features to those profiled in Winter (1999). Specifically, simple physiographic classifications would only be correct a certain fraction of the time, as highlighted by the existence of gaining and losing conditions in settings that have similar topographic and climate conditions (e.g., gaining and losing studies overlap in numerous parts of the plot shown in Fig. 1b). There are signals in this parameter space (e.g., gaining reaches appear more common in more humid settings; Fig. 1b), but there are also

places where actual conditions (i.e., gaining, losing) do not fit with expectations for a given physiographic setting, highlighting the importance of analyzing *in-situ* observations. In addition, physiographic settings do not account for anthropogenic withdrawals, which have likely become the strongest driver in aquifer-stream exchange in the large agricultural regions.

The authors further argue that attempts for larger-scale assessments of groundwater-stream exchange using models have the problem of being difficult to verify as the model boundary conditions and parameters are uncertain and difficult to constrain. Although this is generally true for regional models, physically-based models (if they are set up correctly) typically reflect the general movement of water to drainage points (e.g. in the stream network) well and they should conserve mass. If the model is also thoroughly calibrated against stream flows and other water fluxes this should ensure a reasonable representation of the exchange with the stream network, which has evolved in response to the need of a catchment to remove excess water from the landscape. The data-driven, statistical methodology, which is proposed as an alternative to overcome the problems of models, has its own limitations (see further comments below) and in my opinion should be seen as complimentary to models (e.g. also in terms of providing calibration and validation data for large-scale models as the authors outline in lines 75 to 77) rather than as a substitute.

We fully agree with the reviewer that both models and observations have their pros and cons. We did not take the modeling approach because continental-scale models with sufficient resolution to simulate stream reaches and local groundwater gradients at scales of 10s of meters are not yet feasible, not only because of the high computation demand, but also because such high-resolution models require comparable high-resolution input data, which do not exist at continental scale. The reviewer suggests that this problem can be circumvented through calibration against stream flows and other water fluxes, yet (a) these also are not available (indeed, if we had the subsurface water fluxes then we wouldn't need the model in the first place), and (b) much is known on the pitfalls of model calibration and parameter estimation (often getting good fits to whatever data are available, while failing to constrain the most important parameters). A key input dataset, for example, would be the hydraulic conductivity of the geologic material connecting or isolating local aquifers and streams, at the stream reach scales. Hydraulic conductivities vary by orders of magnitude and no validated high-resolution conductivity data set exists at the scales that would be required for a continental-scale modeling effort. Our well water level observations do not depend on what materials we believe exist or what we believe their permeabilities might be, giving a more objective assessment of the potential and directions for stream-aquifer exchange.

Another critical piece of information that models require to generate the right hydraulic head is adequate information on groundwater withdrawals, which presently are estimated at a county level (i.e., in many jurisdictions across the USA, we do not require metering or reporting of groundwater withdrawals, so continental scale water withdrawals through the USGS for

groundwater are often estimated, then aggregated at the county scale) but not directly measured at a local scale (i.e., the stream reach level). But these withdrawals are reflected in the water level observations. In short, we are relying on the most direct and objective information we have, with minimum assumptions, to address a first-order question at the continent scale.

We agree with Referee #3's point that our analyses may help to test and improve continental-scale models. We will attempt to support such efforts by:

- (1) uploading primary well water level data for states that have provided us with permission to repost these data; our revised data availability statement reads:

Lines 285-290: *"Some states require consent to share groundwater-well data, either through requests to their various agencies, or through public records requests. We have permission to share state-wide groundwater well construction data for California, Colorado, Idaho, Kentucky, Mississippi, Montana, Nevada, Oklahoma, South Carolina, Texas, Utah and Washington and share these data via: [hyperlink to publicly available dataset to be added here, in the event that our submitted article is accepted for publication; here is a temporary link to a set of .csv files for the purpose of peer review [redacted – temporary link]."*

- (2) publishing (as supporting data files) our estimates of potentially gaining and losing rivers (as displayed in main text Fig. 3 and 4) as databases that can be directly incorporated into hydrologic models of the contiguous United States (see revised Data Availability statement)

As stated earlier I think that it is very difficult to evaluate the state of connection between groundwater and a stream/river purely on the basis of groundwater head data in a set of wells and stream stage, especially in landscapes with higher topographic relief. In mountainous terrain with complex topography and geology, groundwater flow is typically complex with nested local and regional flow cells (see Winter et al. 1998, 1999). The hydraulic head in a shallow riparian well could indicate a head gradient from the stream to the well, but at the same time the stream can receive inflow from a deeper regional groundwater flow cell and be net gaining. Or a groundwater head in a well outside the alluvial aquifer in the riparian zone (alluvial valley fill aquifers can be much narrower than 1km) may indicate gaining, while the groundwater head represents hydraulic conditions in a groundwater flow cell that does not or only minimally contribute flow to the stream and the stream could

be net losing to the alluvial aquifer. Also flow losses and gains along streams in mountainous watersheds can coexist over short stream reaches (e.g. Covino, T. P., and B. L. McGlynn (2007), Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry, WRR, 43, W10431, doi:10.1029/2006WR005544, Bergstrom, A., K. Jencso, and B. McGlynn (2016), Spatio-temporal processes that contribute to hydrologic exchange between hillslopes, valley bottoms, and streams, WRR, 52, 4628–

4645,doi:10.1002/2015WR017972), making a clear classification into losing or gaining difficult. In flatter landscapes, which tend to be characterized by spatially more extensive and thicker aquifers, groundwater flow is usually more uniform in space and head data in wells even further from the stream might give a more reliable assessment of the state of connection. This is not to say that a data driven assessment, as presented in this manuscript, could not provide insights into larger-scale patterns, but the outlined problems should be discussed/addressed to put the robustness of the large scale assessment into perspective.

We agree that individual river/aquifer interactions can be complex. The complexities of these individual situations are unlikely to significantly alter our main results, which concern the overall prevalence of potentially losing streams, and how this varies with climate, physiography, and groundwater extraction. We will make clear that we cannot resolve 2D flow patterns (in a cross-section view, with both lateral and vertical flow components) due to the lack of high- resolution geologic (e.g., permeability) data. We have added new text to our manuscript that reads:

Lines 107-112: *“Although these water level differences can identify potentially gaining and losing rivers, quantifying stream-aquifer fluxes would require high-resolution three- dimensional hydrogeologic data (e.g., permeability) that are presently unavailable at continental scale. The lack of high-resolution geologic data also means that we cannot resolve two-dimensional cross sections of groundwater flow fields, and thus cannot capture flow patterns in some complex aquifer systems.”*

We would also like to point out that our well level data reflect a vertically integrated hydraulic head, because the deeper wells tend to have longer screened sections (see the new Supplementary Figure 9 and shown later), sampling both the deeper and shallower aquifers. That is, they give a multi-aquifer average hydraulic potential, collapsing a 2D problem into 1D, allowing direct comparisons between river and well water levels.

One conclusion by the authors is that losing conditions are more common in drier climates, flatter landscapes and regions with extensive pumping. While I agree that drier climate and extensive pumping (two things, which also tend to be correlated as the authors correctly state) will generally promote the evolution of losing conditions, I would question that for flatter landscapes. Flat landscapes are often found around the outlet of larger catchments. Here all the water that cannot be stored in the catchment or has been lost to ET has to leave the catchment via the stream outlet. Therefore the main stem of the river network usually functions as a drainage structure unless the aquifers are heavily pumped or the climate is so dry that that large fractions of the water inputs evaporate and transpire before they reach the outlet. In fact I would argue that this drainage function can be seen in your data (Fig. 3), which for the major rivers in flat areas, even in dry climates such as in California (San Joaquin and Sacramento rivers) indicates gaining conditions (similarly for the larger river reaches in southeastern Texas and southern Mississippi in Fig. 3). This tendency, I would argue, would

be even more pronounced in flat landscapes with humid climates, which may be underrepresented in your data set as well as in the local studies of your meta-analysis, where there are no studies with very flat terrain and humid climate (left upper quadrant in Fig. 1).

The correlation between topography and frequency of losing conditions might be biased for two reasons: i) a general bias for drier climates in flat landscapes across the US (over the range of latitudes in the US), ii) the larger number of lower-order river reaches compared to higher-order reaches in the data base, which in dry climates will be more prone to become or to be losing than the higher order reaches. In summary, I think the conclusion that losing conditions are generally more common in flat landscapes needs to be revised or further specified.

The reviewer is correct that we can conceive of particular situations where some flat landscapes may have fewer losing streams, but the correlations and spatial patterns reported in Fig. 4 speak for themselves: there is, in fact, a strong tendency for flatter landscapes to have more losing streams. Our goal is to let the data speak. Furthermore, as we note in the text, this tendency persists even after any co-varying effects of aridity and groundwater pumping are factored out using multiple regression (the partial regression coefficient is $\beta = -0.22$, $p < 0.0001$). The reviewer points out that many of the highest-order reaches shown in Fig. 3 are identified as gaining reaches, and lie in relatively flat landscapes. The reviewer then says that the correlation in Fig. 4 "might be biased" due to "the larger number of lower-order river reaches compared to higher-order river reaches in the data base". There are vastly more lower-order river reaches than higher-order river reaches in the real world, not just in the database; the number of first-order streams is about four times the number of second-order streams, the number of second-order streams is in turn about four times the number of third-order streams, and so forth (the well-known "bifurcation ratio" of Horton-Strahler stream orders).

We likewise do not agree with the premise that there is "a general bias for drier climates in flat landscapes across the US". In comparing Fig. 4c (in which flat terrain is red) and Fig. 4b (in which dry climates are red), one can see that much of the flat terrain lies well outside the locations identified as arid. Conversely, much of the intermountain west is relatively arid, while very little of it is flat.

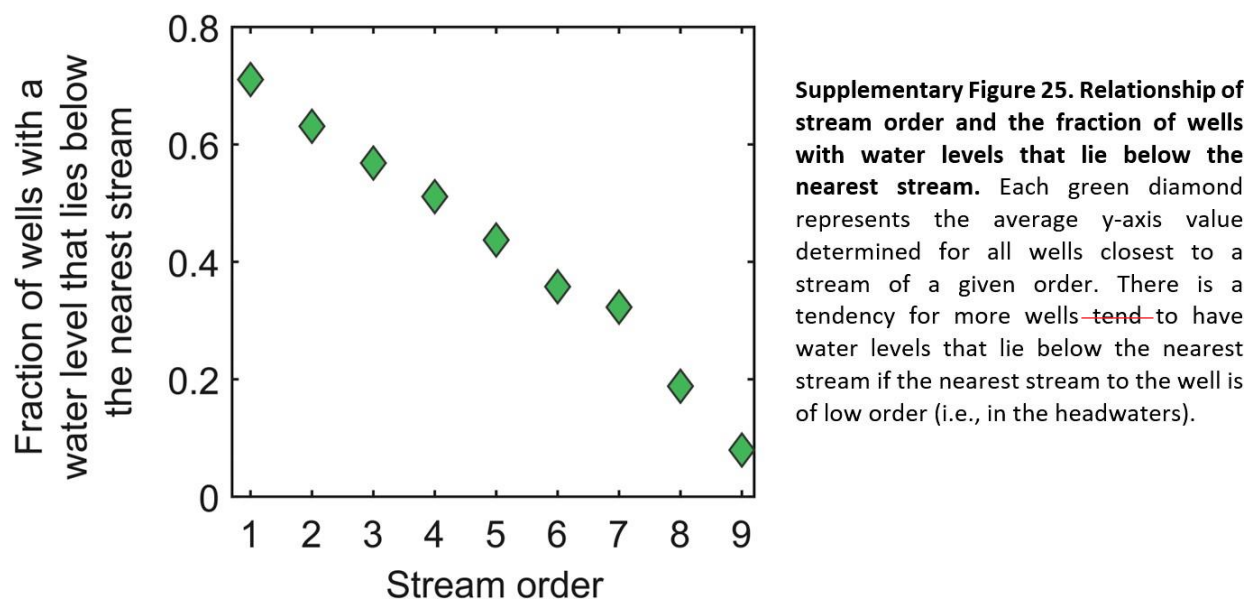
We recognize that although there is a statistically robust association between topographic slope and the prevalence of potentially losing rivers, not every low-gradient river is potentially losing and there is considerable unexplained variance in the statistical relationships (which may represent noise, or factors that have not been accounted for). We have added a new statement to our main text highlighting that our three explanatory variables may also leave much unexplained:

Lines 173-176: *“Although slope, pumping and aridity all correlate significantly ($p < 0.0001$) with the fraction of well water levels that lie below the nearest stream, the high fraction of unexplained variance highlights that other factors may also play important roles in river- aquifer exchanges (Supplementary Fig. 15).”*

Perhaps stream order and the relative frequency of reaches of different order in the data set could be additional valuable variables to further refine the analysis.

We have explored stream order in our analyses as shown in Fig. 3 (see lighter shades for lower order streams, darker shades for higher order streams). Thank you for your suggestion. We have made two additions to our manuscript in response to this helpful suggestion:

- (A) we have added a new Supplementary Figure showing how stream order and the prevalence of wells with water levels that lie below the nearest stream correspond:



- (B) we added a new sentence to our main text pointing readers to this finding:

Lines 345-346: *“Lower-order stream segments are more likely than higher-order stream segments to be potentially losing (Supplementary Fig. 25).”*

Line by line comments:

Line 37 to 38: Some earlier references on the aspect of groundwater contributions to stream flow and associated environmental flows would be Tabidian, M.A., and D.T. Pederson. 1995. Impact of irrigation wells on baseflow of the Big Blue River, Nebraska. Water Resources Bulletin 31, no. 2: 295–306 or Fleckenstein J.H., Anderson M, Fogg G.E., Mount J. (2004) Managing surface water-groundwater to restore fall flows in the Cosumnes River, Journal of

Water Resources Planning and Management, 130(4): 301-310 and Fleckenstein J.H., Niswonger R.G., Fogg G.E. (2006) River-aquifer interactions, geologic heterogeneity, and low flow management, Ground Water, 44(6): 837-852.

Thank you. We revised our manuscript to include references to all three of these suggested works. Our revised summary paragraph reads:

Lines 37-38: *"Identifying these 'losing' streams is important for managing water resources and sustaining environmental flows⁴⁻¹⁰,"*

References 4-6 are (as recommended):

"4 Tabidian, M. A., Pederson, D. T. Impact of irrigation wells on baseflow of the Big Blue River, Nebraska. Water Resources Bulletin 31, 295-306 (1995).

5 *Fleckenstein, J. H., Anderson, M, Fogg G. E., Mount, J. Managing surface water- groundwater to restore fall flows in the Cosumnes River. Journal of Water Resources Planning and Management 130, 301-310 (2004).*

6 *Fleckenstein, J. H., Niswonger, R. G., Fogg, G. E. River-aquifer interactions, geologic heterogeneity, and low flow management. Ground Water 44, 837-852 (2006)."*

Line 43: I would disagree with the general notion that losing conditions are more common in flat landscapes. Do you have references or data to back-up this part of the statement? In flat landscapes with humid climate streams and ditches are important structures to drain the landscape, while in steeper landscapes topographic and geologic structures can frequently cause local losing sections. The data you present in Fig. 1 for humid climates seems to show a prevalence of losing conditions in steeper terrain. The data for the drier climates shows the opposite. The latter might be related to the fact that flat landscapes are typically characterized by more extensive aquifers which in drier climates are typically more intensively pumped for water supply. This can cause spatially extensive hydraulic disconnection between aquifers and rivers. However, this is not due to the flat landscape but the more intensive pumping. Perhaps a better general statement with respect to flat landscapes would be that they are generally more susceptible to groundwater level declines caused by human activity and associated stream flow losses to groundwater.

We have revised the sentence to make it clearer that this is a result of our data analysis rather than a priori assumption. We note that Figure 1 shows that local scale studies in humid climates show general trends for gaining conditions and steeper terrain (as opposed to losing conditions and steeper terrain). We agree with Referee #3's point that there are numerous examples of local-scale studies in steep terrain that identify losing conditions. We specifically highlight a good example of such a case: Bonne Femme Creek in the karst landscapes of central Missouri published by Dogwiler et al. (2007; local-scale study code "L17" in our compilation presented in

Supplementary Table S2). Nevertheless, the general trend in the local scale studies, as depicted in Figure 1, indicate gaining conditions tend to be more common in steeper terrain; it is difficult to generalize from such individual examples (hence our nationwide analysis). Indeed, it is the trends shown by the local-scale studies that provided the foundation upon which we built our hypotheses pertaining to the prevalence of wells with water levels that lie below nearby streams and each of the explanatory variables. Our data confirm statistically robust results from our analysis of millions of wells across the US. That is, Line 43 is not a premise; it is a statistically robust result from our analysis of millions of wells across the US. The local-scale studies in Fig. 1 are too few, and too specific to their individual locations, to draw robust general conclusions.

That was the motivation for our analysis of millions of wells and our country-scale analysis of potential explanatory variables (see next paragraph). We do our best to explicitly note in the figure caption both general trends while also acknowledging that there are exceptions to these general trends:

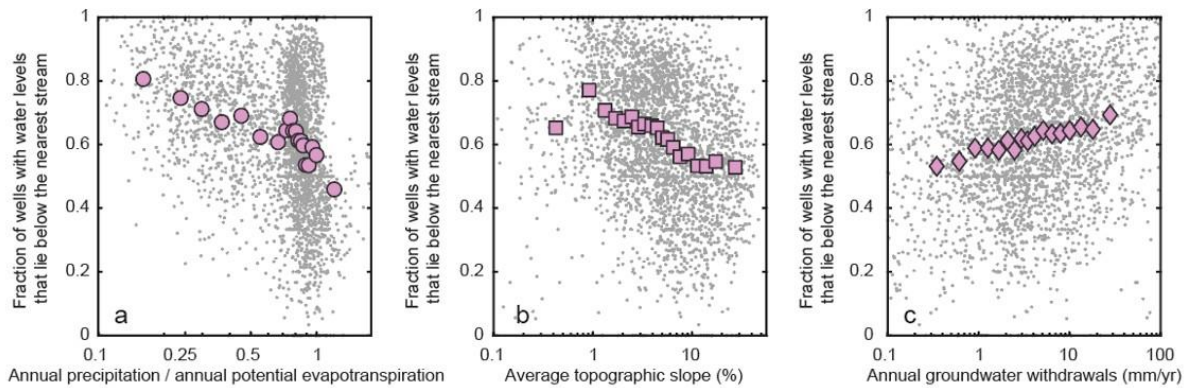
Lines 318-321: *“Losing reaches are more common in flat and arid regions and gaining reaches are more common in steep and humid areas, but both gaining and losing reaches are found at almost all combinations of aridity and topographic steepness.”*

That national-scale analysis of millions of wells shows a significant negative correlation between topographic slope and the fraction of wells with water levels that lie below the nearest stream ($\rho = -0.35$; $p < 0.0001$), indicates a tendency towards higher prevalence of losing rivers in flatter terrain. We of course recognize that this relationship could potentially arise spuriously if groundwater pumping tends to be greater in flatter areas or flatter places also tend to be more arid. Therefore, we took account of the interrelationships among the explanatory variables using multiple regression on the rank transforms of each variable, with the result that the partial regression coefficient for topographic slope (-0.22) remains highly significant ($p < 0.0001$). This result implies that losing streams are indeed more common in flatter landscapes than in steeper terrain, even after any confounding effects of climatic aridity or groundwater pumping are taken into account.

To improve clarity, we have added a new statement to our main text highlighting that our three explanatory variables may also leave much unexplained:

Lines 173-176: *“Although slope, pumping and aridity all correlate significantly ($p < 0.0001$) with the fraction of well water levels that lie below the nearest stream, the high fraction of unexplained variance highlights that other factors may also play important roles in river- aquifer exchanges (Supplementary Fig. 15).”*

In recognition that we should’ve done a better job of depicting how the data that we do have indicate that losing streams may be more prevalent in flatter areas, we added a new supplementary figure showing a scatterplot of our county scale data:



Supplementary Figure 24. Cross plots of each explanatory variable against the fraction of well water levels that lie below the nearest stream. Each grey point in each of the three plots represents one county. The large pink points represent the mean (average) fraction of wells in the county that have a water level that lies below the nearest stream elevation (i.e., average y-axis value) binned at intervals that capture 5% of the data as determined by the x-axis values (i.e., x-axis values have been binned at intervals designed to contain 5% of the grey points displayed on the plot). These plots suggest that the fraction of wells with water levels that lie below the nearest stream tends to be higher in (a) more arid counties (i.e., lower county-wide average values of annual precipitation divided by evapotranspiration), (b) counties that have lower average topographic slopes, and (c) counties where groundwater withdrawal rates are higher (i.e., higher groundwater pumping, expressed in a way that normalizes to the area of a county: units of mm of groundwater pumped per year).

Line 49: The words “near groundwater wells” are a significant constraint to the more general title of the manuscript, which seems to imply an even broader prevalence. Doesn’t the statement in this line also imply that human activities and their groundwater abstractions are one if not THE main reason for the observed prevalence? I think this would be a justifiable conclusion, which could be stressed more in the manuscript.

We cannot conclude that groundwater abstractions are the main cause of losing streams, because we use static water levels from constructed groundwater wells (intended to pump groundwater), which allow us to get such as dense set of data; unfortunately, this means that we have no data on aquifer levels where wells have not been drilled. The best we can do is to correlate the prevalence of losing streams with groundwater withdrawals, as we do in Fig. 4; this shows a substantial and robust relationship. We have done our best to make this important point that Referee #3 raises clear by stating:

Lines 128-130: “...our analysis inherently focuses on regions where wells pump groundwater, and we cannot quantify the prevalence of potentially losing rivers where wells are absent.”

While also stating elsewhere that our correlation analyses suggest a significant relationship (spatially) between areas where groundwater withdrawals are high and where our study wells have water levels that lie below the nearest stream:

Lines 190-191: *“Our analysis demonstrates the significant ($p < 0.0001$) relationship between groundwater pumping and losing streams across the USA (Fig. 4).”*

Lines 69 to 77: I agree with the statement that it is difficult to bridge between patchy local studies, but I think the assessment based on wells within 1km of the streams has its problems too. SW-GW exchange depends on many aspects in addition to just general hydraulic gradients, including topography and geology. Therefore exchange patterns can be locally quite diverse with nested flow regimes. Such patterns are difficult to generalize easily over larger scales using just local point information on GW head and stream stage elevations.

We agree that individual relationships can be complex, and that hydraulic gradients can only indicate the direction of possible groundwater flow (estimating fluxes would require permeability data, which is not available at the scales we would need). Although local flow patterns can be complex, we have no reason to believe that the complexities of these individual situations would substantially bias our main results, which concern the overall prevalence of losing vs. gaining streams, and how this varies with climate, physiography, and groundwater pumping.

To improve clarity, we more explicitly outlined the limitations of our analysis in a new Methods section, which reads:

Lines 258-269: *“Limitations to our analyses*

Because no locally relevant, high-resolution three-dimensional database of permeability exists at continental scale, only the potential for rivers to lose or gain (rather than the fluxes of actual losses or gains) can be calculated with confidence across the USA. Another limitation of our analysis arises from the lack of water level time series for most wells; because most of our wells have just one water level measurement, we cannot evaluate temporal variations in gaining versus losing conditions, nor can we assess the impact of seasonal fluctuations in climate and groundwater withdrawals on river-aquifer exchanges. Furthermore, imprecise well location data and the coarseness of the digital elevation data we analyze (30 m by 30 m, or 10 m by 10 m) will lead to uncertainty in land surface elevations and, therefore, in our estimate of hydraulic heads, although these random uncertainties should not substantially bias our assessment of the relative prevalence of losing and gaining conditions.”

We also added a new statement within our manuscript’s main text reinforcing the importance of locally relevant geologic data, and how the lack of it in 3D limits our ability to go beyond identifying potentially gaining and losing rivers to estimating fluxes:

Lines 107-112: *“Although these water level differences can identify potentially gaining and losing rivers, quantifying stream-aquifer fluxes would require high-resolution three-dimensional hydrogeologic data (e.g., permeability) that are presently unavailable at continental scale. The lack of high-resolution geologic data also means that we cannot resolve two-dimensional cross sections of groundwater flow fields, and thus cannot capture flow patterns in some complex aquifer systems.”*

Lines 86 to 87: As stated above I challenge the hypothesis that flat landscapes promote losing conditions.

We respect the Referee’s challenge to this hypothesis. We also agree that there are numerous examples of local-scale studies where flatter landscape and gaining conditions exist (e.g., blue upward-oriented triangles in the lower left quadrant of our main text Fig. 1). Our hypothesis that there is a tendency for losing rivers to be more prevalent (though far from the sole condition) in flatter terrain is motivated by 77 local-scale studies, as we state in our introduction section:

Lines 65-67: *“The 77 local-scale studies we reviewed (Supplementary Table 1) suggest that losing rivers may be more common in drier landscapes and flatter terrain, especially where groundwater pumping is extensive (Fig. 1b).”*

These local-scale studies suggest that losing conditions may be more common in flatter landscapes. We agree that there were some local-scale studies that do not support this hypothesis, which is one reason we were careful to add words like “more prevalent”. The data show that losing streams are more common in flatter landscapes than in steeper terrain, even after confounding effects of climatic aridity or groundwater pumping are taken into account. Perhaps we may be so bold to say that this is one of the values of the paper – it uses observed static water levels to give us continental scale insights into groundwater-surface water interactions, some of which transcend our expectations. Yet we restate the importance of including Referee’s comment as revisions to our manuscript; specifically, we highlight the new sentence we have added that clarifies that the correlations we present based on continent-wide analyses are imperfect:

Lines 173-176: *“Although slope, pumping and aridity all correlate significantly ($p < 0.0001$) with the fraction of well water levels that lie below the nearest stream, the high fraction of unexplained variance highlights that other factors may also play important roles in river-aquifer exchanges (Supplementary Fig. 15).”*

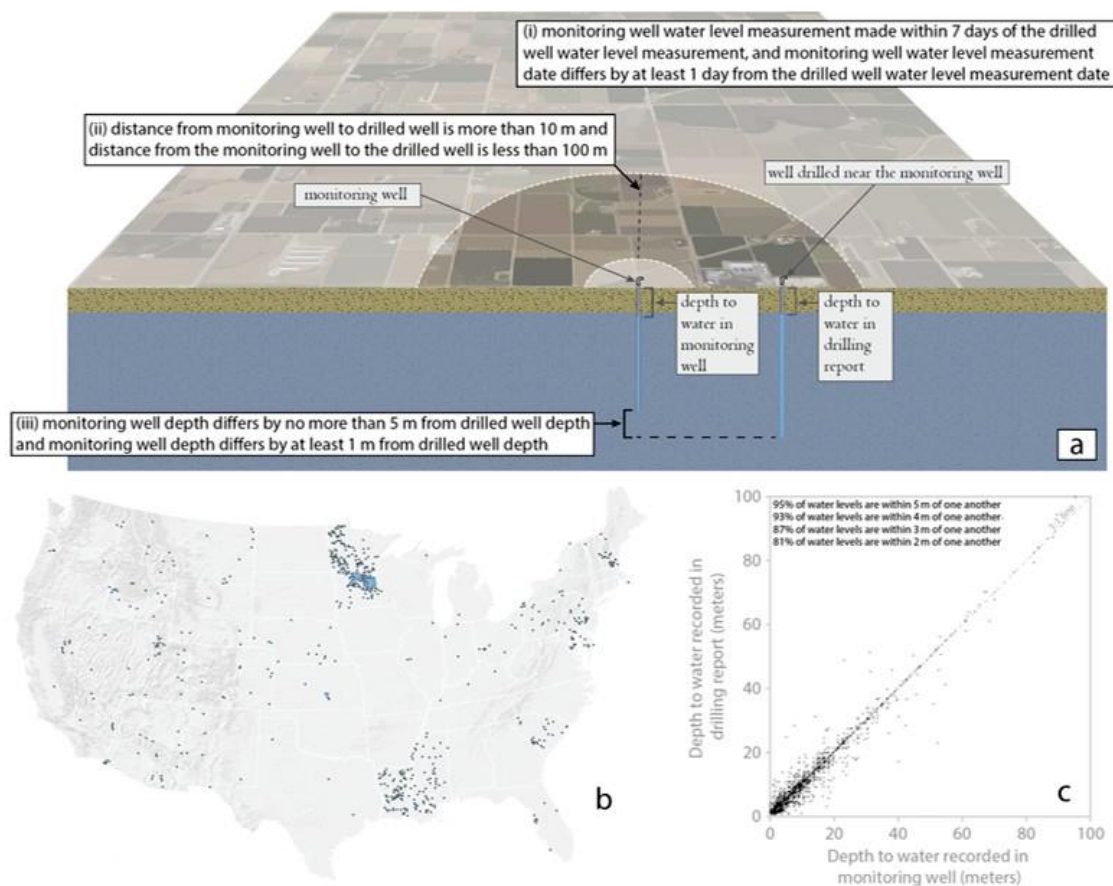
Lines 91 to 92: How could you ensure that that wells were not pumped at or shortly before the time of monitoring?

The water level measurements in the driller report databases we compiled (which represent the majority of all measurements in our dataset) are “static water levels”. Measurements of

static water levels that follow the National Groundwater Association guidelines (“ANSI/NGWA- 01-14 Water Well Construction Standard”) require “Determination of the static water level occurs when there is less than one (1) foot difference between two (2) consecutive water level measurements taken a minimum of sixty (60) minutes apart.” (quote from: http://www.gwpc.org/sites/default/files/event-sessions/mcCray_Keving_ANSI.pptx_.pdf).

While we accept that not all drilling companies may abide by these guidelines, the strong correspondence between static water levels (recorded in drilling reports) and nearby monitoring wells (with similar depths, with measurements made at similar times, with similar land surface elevations) lend confidence to our interpretation of these measurements as adequately capturing aquifer conditions (Jasechko et al., 2020 cited in paper – see “Note 3” in the SI of this paper; hyperlink is reproduced here for convenience https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-020-17038-2/MediaObjects/41467_2020_17038_MOESM1_ESM.pdf).

To do a better job of highlighting the value of driller report water levels and their consistency with nearby monitoring wells, we added a new Supplementary Figure that reinforces our finding in ref. 28 (that water levels recorded in well completion reports are consistent with those in nearby monitoring wells, implying that these data reflect prevailing groundwater levels at the time of observation):



Supplementary Figure 4. Comparison of depth to water measurements made in monitoring wells and nearby drilled wells. (a) Schematic depicting the comparison of depth to water measurements in a monitoring well and a nearby drilled well, with water level measurements made more than one day but less than one week from one another, well locations more than 10 m from one another but less than 100 m from one another, and depths that differ by at least one meter but less than five meters (minimum values are set to limit the potential that the record of drilling of the monitoring well (recorded in the well drilling database) is compared against water levels recorded in the monitoring well (recorded in the monitoring well database)). (b) Locations of pairs of monitoring wells and drilled wells meeting our criteria for analyses ($n=1979$ pairs of monitoring well and a drilled well; each black dot represents one well pair). (c) Comparison of depth to water measurements recorded in a monitoring well (y-axis values) versus a drilling report (x-axis values). The strong correspondence (e.g., 87% of comparisons have water levels the differ by less than 3 m; see the dashed 1:1 line) highlights that the driller report data capture similar water levels to monitoring wells when such wells are co-located, increasing our confidence that the water level data recorded in drilling reports captures actual hydrogeologic conditions. Not all measurements are shown, as the x- and y-axis limits are truncated at 100 m (98.5% of all well pairs are represented on this plot).

Are pumping cycles recorded in the well data base? How were seasonal patterns accounted for?

We share Referee #3's wish that pumping cycles could be included in these databases. Unfortunately, metering and reporting of actual groundwater pumping rates remains rare across much of the United States, and these data are not readily available within the well completion report databases we compiled. As pumping data remain unavailable, we could not

explicitly evaluate seasonal patterns with the data we analyzed. However, our well completion dates encompass a wide array of day-of-year values, meaning our dataset samples static water levels during all seasons of the year.

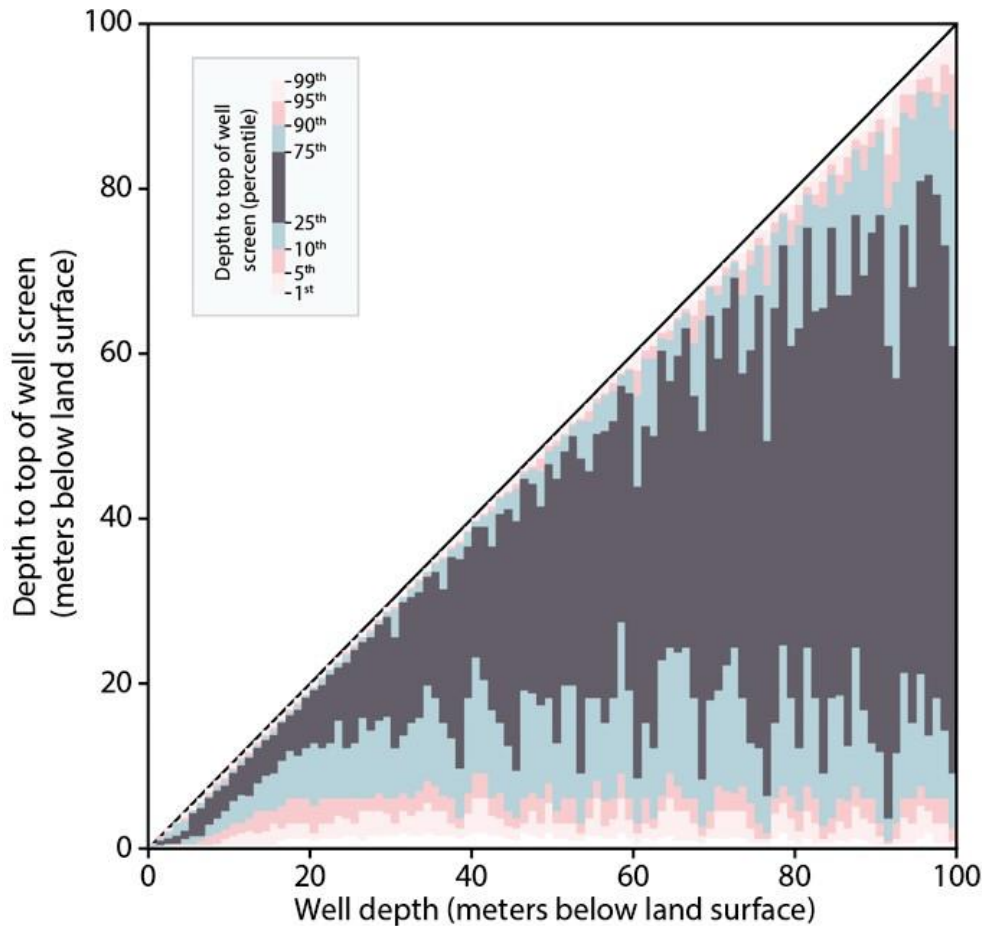
In order to ensure we modified our manuscript in response to Referee #3's comment, we have added to a new subsection entitled "Limitations to our analyses" the following statement pertaining to temporal variability:

Lines 261-265 *"Another limitation of our analysis arises from the lack of water level time series for most wells; because most of our wells have just one water level measurement, we cannot evaluate temporal variations in gaining versus losing conditions, nor can we assess the impact of seasonal fluctuations in climate and groundwater withdrawals on river- aquifer exchanges."*

Lines 89 to 106: How did you deal with the elevation of the well screens? If a groundwater well is only screened over a certain depth interval (which is usually the case) the head in this well and the head difference with respect to the stream may not tell you much about groundwater stream connectivity. In thick alluvial aquifer systems we typically have significant vertical hydraulic gradients, in particular close to the streams. Hence the head in a well may only give you limited information on the overall state of connectivity between groundwater and the stream. I think you need to provide more information on how the head data from the wells was assembled and what assumptions were made (depth of screens, time periods of measurements and seasonal variations etc.). I could not find sufficient information on those aspects, neither in the short methods section nor in the supplements.

Because our data derive from dozens of unique datasets, there is considerable variability in the availability of different fields, including perforated interval information. Some databases do not record well screen interval data at all. Other databases record information about the top of well screens and some have data for both the top and bottom of perforated interval(s). Yet even databases that report some well screen interval information are incomplete; a share of records in these databases do not record well screen information.

Using the well screen information that is available, however, we have added a new figure to the supplementary information comparing total well depths and the top of the shallowest recorded screen interval:



Supplementary Figure 9. Comparison of total recorded well depths (x-axis values; expressed in meters below land surface) and the recorded depth to the top of the perforated interval (y-axis values; expressed in meters below land surface). The diagonal black line marks a 1:1 relationship. Shaded areas mark the percentiles of depths to top of perforations (see legend in upper left corner): dark grey represents 25th-75th percentile range, light blue represents 10th-25th and 75th-90th percentile ranges, dark pink represents 5th-10th and 90th-95th percentile ranges, and light pink represents 1st-5th and 95th-99th percentile ranges (percentiles determined by binning the x-axis well depth data into 1 m intervals). We lack screen interval data for many well completion records, motivating us to use well depth data (which are recorded more widely than well screen interval data) in our sensitivity analyses described in section S4.2. Our comparison of the total depth of wells (x-axis) and the depth to the top of their perforations (y-axis) demonstrates that many wells (for which we have screen interval data) have perforated intervals that begin at considerably shallower depths (e.g., 75% of wells that have total depths of between 99m and 100m have a perforated interval that begins 9.1 m beneath the land surface). Therefore, the depth threshold presented in our sensitivity analyses in Supplementary Section 4.2 should be interpreted as the maximum (i.e., deepest) possible perforated interval.

We have also added a new statement in the main text directing readers to this new figure and its ramifications: that the total well depth represents a maximum value for the potential perforated interval, but that water levels recorded in deep wells can also capture conditions at shallower depths when their screened intervals lie above their bottoms.

Lines 275-276: “..for a comparison of well depths and the depths of the top of the shallowest screen interval see Supplementary Fig. 9)”

We have also added a new statement to our Supplementary Information (section S4.2) describing our rationale for using total well depth as a threshold in our analysis that reads:

Supplementary Information Page 78: “We analyze total well depth data rather than perforated interval data because the latter is unavailable for many of the wells in our dataset. We emphasize that the total well depth represents the maximum depth at which perforations exist for a given well, but that many production wells in our database have screened intervals far shallower than their bottom (see Supplementary Figure 9).”

Lines 121 to 124: As stated earlier your analysis is biased towards areas with human groundwater abstractions. Perhaps that should be stated earlier on in the manuscript to avoid false conclusions. For a regional assessment of the effects of human activities (e.g. groundwater abstractions) on groundwater-stream connectivity versus a baseline case with no pumping, even over a range of landscape types (e.g. flat versus steep), a modeling study would be better suited than the data driven approach presented here.

We thank Referee #3 for suggesting we make this point earlier in our manuscript. We have added a new sentence in the first paragraph of our section entitled “Millions of groundwater level observations” that reads:

Lines 96-98: “We emphasize that our dataset over-represents areas where hydrogeologic conditions have been disturbed by pumping, because we only have water level measurements where wells have been drilled.”

Lines 135 to 137: I would consider this a significant bias in the study, which should be addressed earlier in the manuscript. It also calls for some more information on how the head data from the wells was assembled and what assumptions and simplifications were made (see also one of my earlier comments).

We agree. Our analysis must overrepresent areas where pumping occurs, because we only have well water level measurements where wells exist (and these are the very same places where pumping is common). Our revised manuscript makes this point in numerous places:

- 1) We have now added a new sentence in our manuscript that reads:

Lines 96-98: “We emphasize that our dataset over-represents areas where hydrogeologic conditions have been disturbed by pumping, because we only have water level measurements where wells have been drilled.”

- 2) We also emphasize this point in our statement at another location in the main text of our manuscript:

Lines 126-130: *“Because we can only analyze river segments with at least one well nearby (~580,000 stream segments: 22% of all National Hydrography Database segments), our analysis inherently focuses on regions where wells pump groundwater, and we cannot quantify the prevalence of potentially losing rivers where wells are absent.”*

Lines 139 to 142: What do you mean with “not widely publicized”? Do you want to imply that groundwater depletion has not yet been fully recognized in those regions or that the problems that go along with it have not been seen as so severe?

To improve clarity, we revised this statement to state “considered” rather than “publicized”, as national-scale assessments of groundwater depletion led by the United States Geological Survey (e.g. Konikow 2014 <https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/gwat.12306>) suggest minimal groundwater depletion in these regions. Our revised text reads (modified text is in green):

Lines 143-146: *“Our analysis also reveals a substantial likelihood of losing rivers in regions that are **not widely considered to be** groundwater depletion hotspots, including Midwestern states near the Great Lakes, Idaho’s Snake River Plain, central Washington State, and Oregon’s Willamette Valley (Fig. 3).”*

Line 154: As said before, I do not believe that flatter topography alone promotes losing streams.

We agree that flatter topography alone does not promote losing streams. Line 154 (in our original manuscript) is summarizing findings from the literature, which highlights that losing rivers may be more common where topography is flatter:

Lines 154-156: *“Our meta-analysis of 77 local-scale studies (Supplementary Table 1) suggests that losing rivers may be more common where climate conditions are drier, topographic slopes are flatter, and groundwater withdrawals are greater (Fig. 1b).”*

The data used in our study support the local-scale analyses, as the data show a tendency for flatter landscapes to have more potentially losing streams, even after accounting for possible confounding by variations in climate or groundwater pumping.

Lines 162 to 164: I fully agree with this statement and it is important for interpreting your results. Even if the partial regressions imply a statistically significant correlation between topographic slopes and the frequency of losing conditions I would doubt a general mechanistic link between the two variables.

We agree that the correlations we find do not themselves imply a mechanistic link between flatter areas and a tendency for higher prevalence for losing conditions, and do our best to present little more than the data as they are (i.e., the regression coefficients) without commenting on mechanisms on this occasion. We did our best to convey the remaining

unexplained variance in our regression analyses by adding the following new statement to our main text:

Lines 173-176: *“Although slope, pumping and aridity all correlate significantly ($p < 0.0001$) with the fraction of well water levels that lie below the nearest stream, the high fraction of unexplained variance highlights that other factors may also play important roles in river- aquifer exchanges (Supplementary Fig. 15).”*

Line 178: I would avoid the term “water table” in this context. If stream and groundwater are hydraulically connected but losing, one would not refer to the groundwater head (or hydraulic potential) below the stream as the water table. Perhaps: “...by lowering the hydraulic head below the stream...”

We have revised our manuscript as suggested. It now reads:

Lines 179-181: *“Near-stream groundwater pumping can deplete streamflow either by withdrawing groundwater that would otherwise seep into the stream, or by lowering the hydraulic head below the stream surface, creating the potential for the stream itself to drain into the underlying aquifer^{1,3}.”*

Lines 181 to 184: I agree with this statement. Groundwater depletion, e.g in the Central Valley of California, has been going on for decades and has disconnected many streams from their underlying aquifer some time ago.

Thank you.

Lines 200 to 202: This is a common feature in rivers that transition from mountainous upper watersheds via alluvial fans in mountain forelands to lowlands towards the basin outlet (see also: Covino, T. P., and B. L. McGlynn (2007), Stream gains and losses across a mountain-to- valley transition: Impacts on watershed hydrology and stream water chemistry, *WRR*, 43, W10431, doi:10.1029/2006WR005544). Often these rivers become gaining again past the alluvial fans towards their watershed outlet in the deepest/flattest part of the basin (e.g. the Cosumnes River towards its confluence with the Mokelumne River and later the San Joaquin, Fleckenstein et al. *GW* 2006). I think this can be seen nicely in the close-up of California in Fig. 3c of your manuscript for the rivers draining the Sierra Nevada into the Central Valley).

We are pleased that the zoomed in maps agree with Referee #3’s expertise in these regional settings.

Lines 205 to 207: In California the notion that losing rivers are a main source for groundwater recharge in the agricultural areas of the Central Valley has been acknowledged for some time (e.g. Criss, R. E., and Davisson, M. L. (1996) Isotopic imaging of surface water/groundwater interactions, Sacramento Valley, California, *J. Hydrol.*, 178, 205–222) and it is embedded in

water management strategies including water banking (e.g. Kocis, T. N., & Dahlke, H. E. (2017). Availability of high-magnitude streamflow for groundwater banking in the Central Valley, California. *Environmental Research Letters* 12, 084009).

Thank you for recommending these references. We cite Prof. Dahlke's coauthored publication elsewhere in our manuscript, and now add a citation to Criss and Davisson at this location in our revised manuscript:

Lines 206-208: *"These examples suggest that losing rivers may recharge aquifers that in turn may support irrigated agriculture and groundwater-dependent ecosystems in lowland basins (e.g., ref.³⁸)."* (*note: ref. 38 points to the following citation: Criss, R. E., & Davisson, M. L. Isotopic imaging of surface water/groundwater interactions, Sacramento Valley, California, *Journal of Hydrology* 178, 205–222 (1996))

Figure 1: incorrect spelling of "Cosumnes River" in the figure (not "Consumnes")

Thank you for catching this typo. We have corrected it.

Supplements:

Section 1.2. How did you choose the radius of 1km? Is it sufficiently large to capture topographic features that can affect groundwater flow to the stream?

Thank you. The 1 km radius was selected to remain consistent with our methodology applied to selecting wells based on their proximity to the nearest river (i.e., Supplementary Information section S4.1. We also present maps of each local-scale study in Supplementary Table 2, and find the 1 km radius to be approximately consistent with the spatial scale considered in many of these local-scale studies (i.e., 1 km).

Figure 2: Why are there no studies in very flat humid landscapes (upper left quadrant of the figure)? Is that type of landscape rare in the US or not studied (perhaps because there are no problems of concern with respect to GW-SW interactions)?

Thank you for this observation. We note that there are some regional-scale studies along the Atlantic Coastal Plain (where slopes are generally flatter and P/PET ratios are 0.75-0.9; e.g., quote from <https://pubs.er.usgs.gov/publication/pp1773>: *"Additional recharge to deeper underlying aquifers occurs where the confined aquifers crop out along the Fall Line in the inner Coastal Plain and Sand Hills Physiographic Provinces. In this region, streams typically gain water from the aquifers."*). However, we could not immediately identify reach-scale studies directly applicable to our local-scale study compilation. With the exception of Belcher Canal (Florida, with slope close to 1% near the study site), we could not find any local-scale studies of humid- and-flat areas in the US. We do not know why such areas have not been studied in the US to date, though we are aware that some of the flatter areas in America also tend to be drier

because of the country's extratropical location where orographic and frontal air mass uplift tend to be dominant precipitation-forming mechanisms (more-so than, say, monsoonal regions influenced by the Intertropical Convergence Zone). The hypothesis posed by Referee #3 (that these regions have not been studied as widely because there are no problems of concern with respect to interactions between groundwaters and surface waters) may well be true.

Table 1, L49: "Cosumnes River" is incorrectly spelled (not "Consumnes")

Thank you; we fixed this typo.

Table 4: in the header of the right column it should be: "...wells have water levels that lie above the study stream (consistent with gaining conditions).", and not "lie below the stream"

This comment, which catches a typo in an SI table, exemplifies the thoroughness of this review. Thank you. We fixed this.

Figure 15: in the text boxes "fraction of" appears twice and one instance should be removed.

Thank you. We fixed this.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

I have reviewed a previous version of the manuscript and carefully read the revised version as well as the responses. They are very well articulated, but I still find several key points I raised previously remain unanswered. As highlighted previously, the amount of data compiled is impressive. But my most important concerns are that (1) the authors cannot demonstrate how their study is useful to water resources management and (2) From a scientific point of view, I maintain that all the factors that potentially influence gaining or losing conditions are already very well known, as also acknowledged by the authors.

I took note of the authors point that (quote from the reply) "Our goal is not to show that losing streams occur, but to shed light on how extensive and how common losing streams are». Without knowledge of fluxes, I fail to see how this information is useful, see also points below. I believe the authors missed the opportunity to demonstrate why this should be the case. Given that groundwater abstraction is ever-increasing in the US and most major aquifers are overpumped this is not a surprise in my opinion.

So, can the study be used to identify general patterns of gaining/losing conditions? The new supplementary figure 24 nicely shows that this is not possible, at least not with the very limited number of explanatory variables used. There is an extremely large scatter, which clearly indicates the complexity between environmental and anthropogenic factors and the presence of gaining/losing conditions. Many other factors will come into play, too, such as land-use, infiltration capacity of the soil, the average slope of the river or the drainage density, to list but a few of them. Even if on average there is somewhat of a tendency, the complexity and interdependence of the factors overshadow the general trend, thus making it impossible to infer gaining or losing conditions based on the knowledge of these factors.

Below some more specific comments that illustrate my points above.

- Again, essentially all the questions raised from line 54-58 require fluxes, as readily acknowledged by the authors in their response. The way the authors develop their manuscript suggests that this study can help in addressing these issues. But without fluxes, it does not. In their response the authors say (quote):

«Unfortunately, continental scale data are not available to calculate fluxes, especially fluxes in complex aquifer systems. We address this important point via revisions and new text added to our

revised paper: Lines 107-112: "Although these water level differences can identify potentially gaining and losing rivers, quantifying stream-aquifer fluxes would require high-resolution three-dimensional hydrogeologic data (e.g., permeability) that are presently unavailable at continental scale. The lack of high-resolution geologic data also means that we cannot resolve two-dimensional cross sections of groundwater flow fields, and thus cannot capture flow patterns in some complex aquifer systems.»

This is very true and I fully agree. What I am missing is a follow-up on the implications of this, specifically for the utility of this particular study. The rest of this reply is a discussion on policy. I do not think the authors have addressed (nor can address) my previous point.

In the same context (page 5), the authors additionally address this point (Quote):

«It would be wonderful to be able to quantify fluxes between rivers and aquifers at the millions of points where we have water table measurements, but no reliable data on streambed or aquifer permeability exist at continental scale. Our analysis nonetheless informs management by demonstrating that the potential for losing streams is widespread across the US, and by quantifying the influence of several causal factors.»

Again, the factors are all very well known, and the additional figure (Supplementary Figure 24) shows clearly the very wide scatter of gaining and losing conditions throughout all of the different parameters considered. A large-scale assessment does thus not help in any way in understanding the local conditions. The (unspecific) statement of the authors below is illustrating my point: "Thus, our analysis should help focus managers' attention where it is most clearly needed, even if we cannot quantify exactly which reaches of which streams are losing exactly how much " But how does this study help in this regard? Do losing rivers require more attention as opposed to gaining rivers? In the revised version the authors did not demonstrate how their paper can help any manager to identify where attention is required. And I am sure they cannot, because, again, the fluxes provide the relevant information. That they are hard to obtain does not increase the value of the knowledge of whether a river is gaining or losing. I noted that one reviewer is of a different opinion, he highlights that the study can have significant implications for water resources management. If this is the case, why arnt these points listed in the revised paper? Again, none of the points listed from line 54-58 can be addressed with the current study.

- I still maintain that many points are trivial. That surface water and groundwater need to be managed jointly is so well known by now, that I do not think the study increases its impact by making this point (see e.g. the last sentence in the abstract or section line 198). The same remains for the point that losing conditions are found in areas of significant groundwater abstraction. To avoid overselling the paper, I strongly suggest to significantly down-tune the relevance of the study by explicitly highlighting that the study can neither improve local water resources management nor provides any novelty in terms of process- understanding. The point that remains is how widespread losing conditions are. It is up to the editor to evaluate if this sufficient for publication in this journal.

Referee #2 (Remarks to the Author):

I am reviewing this manuscript for the 2nd time, and I find it to be much improved. In reviewing the entire set of first reviews of the paper, there were clearly some critical comments (constructive, mostly), and my assessment is that the authors did an excellent job responding to the range of comments and very much improving the manuscript. More than 1 reviewer was concerned about identifying gaining and losing streams/rivers/reaches without confirmation from the field, and it was not clear to me how the authors would deal with this without diluting the message. It is my opinion that their recrafting of their message throughout was excellent and that

this paper more precisely identifies the trends they find (even when they are counter to conventional wisdom as appears to be the case with 'flat' terrain), and truly "let's the data speak" as they state in their rebuttal comments. The revision process resulted in a very substantial revision of the text, which I believe is more on point and makes a sharper argument. In addition, it generated a larger supplemental information section, which is almost a paper on its own. It is full of additional analyses and summary tables. Together, these really make a strong point where the paper serves as the articulated synthesis of the 4 figures presented there, the millions of data points evaluated, and the extensive analyses in the supplemental materials, and the supplement really delves into the details that can address many skeptical comments (as was proven through the revision process).

The section 'Managing groundwaters and surface waters jointly' was only lightly revised (compared to other parts of the paper) and to me, this is really a key point. The authors' first citation is a USGS circular whose title is more or less the same message. That circular is something of a bible with respect to summarizing how surface water-groundwater interactions influence water quantity, quality, and ecosystems and their services to humans. It is largely a synthesis of findings from a large number of peer-reviewed papers that illustrates in many specific cases how these exchanges influence other aspects of natural systems. This paper comes to the same conclusion (i.e., this section heading) from a different perspective, one that is backed by very extensive data analyses. That is to say, this is the message that the data tell us when they speak and it's a critical message.

With all of this said, it is my opinion that indeed this is of broad enough interest and significance that it merits publication in Nature (as opposed to one of the other Nature topical journals). I conclude that the data overturned many of the arguments of the reviewers (i.e., I expect these are all specialists).

I am satisfied that the authors adequately addressed the simple and the more complicated reviewer comments. The message is straightforward and, given the data analysis, novel (in my opinion) because it is an outcome of these many many many observations (and their topographic/land use context of course).

Referee #3 (Remarks to the Author):

My main comments and concerns have been addressed and the manuscript has significantly improved. Refining the wording and the addition of a section on limitations have helped to put the presented data-driven analysis into a more balanced perspective. I acknowledge that data-driven analyses at this spatial scale are challenged by data availability issues and the need for simplifications. Although this naturally constrains the ability of such analyses to explain conditions at every individual location, I am generally in line with the authors that these analyses do have their merits and can reveal interesting general patterns. The identification of high potentials for losing conditions in areas that are usually not considered to be hotspots for groundwater overdraft is interesting and likely a sign of the growing pressures on groundwater resources. Admittedly the correlations between the main explanatory variables and the fraction of wells indicating losing conditions are not particularly strong (rank correlation coefficients <0.35), but considering the scale of the assessment I would still consider them robust for the evaluation of large-scale general patterns. I also appreciate that the authors added specific links to data in their data availability statement. This will be useful for future studies.

I have a few remaining points, which I would suggest to consider to further improve the manuscript.

42

You use the word "prevalence" from the beginning of the manuscript. To me this word implies some bias in the sense that one thing occurs more frequently than another. Therefore I would

avoid to use this term before you actually document this “prevalence”. For example in line 38 I would rather use an expression like “occurrence of losing streams”, which is poorly constrained by observations, because at this point a “prevalence” has not yet been shown. Although you state the 64% potentially losing river reaches in the next line of the abstract (which is more of a summary) I would also avoid the word “prevalence” in lines 72, 84 and 85. After you state the percentage of potentially losing reaches (64%) in the main text (line 121), which indeed indicates a “prevalence” for losing conditions the use of the term seems more appropriate.

On the impact of well screen elevations: I can follow the line of argument that deep wells tend to have longer screened intervals and that the hydraulic heads they show represent an averaged head. However, I disagree that this automatically collapses a 2D (or in fact a 3D) flow problem to a 1D flow problem as the authors state in their response to one of my comments. This may be fair to state for a more or less homogeneous aquifer with vertical gradients, but not for a structurally complex aquifer system with a sequence of aquifers and aquitards. But this leads back to the problem of locally more complex conditions, which cannot be resolved in an analyses at this scale. The newly added figure in the supplements on well screen locations for the wells, for which such data exists, is a good addition to the supplements and shows the large variation of well screen lengths that exist in wells (depending on the thickness of the vadose zone, the sequence of aquifers and aquitards and the targeted aquifer to pump from). However, the interpretation of the data in the figure in the figure captions contains an error. You state:

“Our comparison of the total depth of wells (x-axis) and the depth to the top of their perforations (y-axis) demonstrates that many wells (for which we have screen interval data) have perforated intervals that begin at considerably shallower depths (e.g., 75% of wells that have total depths of between 99m and 100m have a perforated interval that begins 9.1 m beneath the land surface)” The part in the brackets is not correct. Your figure shows that 75% of the wells with total depths between 99 and 100m have their screens start not deeper than about 60m, and only 25% have their wells screens start in the uppermost 10m from the land surface. In other words quite a few wells record heads in the deeper parts aquifers. You need to correct this, even if it probably does not change your general statement that unknown well screen locations may have minor impacts on the results of your analysis.

Line by line comments:

Line 132: I think to state that a potential for rivers to be losing is “the norm across much of the contiguous US” is quite a strong statement for only 64% of the well data indicating a potential for losing. I would suggest to scale down a bit, perhaps: “...well water levels that lie below nearby stream surfaces – indicative of losing streams – are surprisingly common across much of the contiguous United States....”

Line 261: I would replace the word “calculate” with “evaluate” or “assess”. You do not really “calculate” the potential for losing conditions, but rather “evaluate” it on the basis of the direction of a hydraulic gradient.

Line 268: “random uncertainties” sounds a bit strange. Perhaps rather: “...although these uncertainties, which will cause random errors in the data, should not substantially bias our assessment....”

Line 269: “...prevalence of losing versus gaining.” But perhaps my understanding of the word “prevalence” is too narrow?

Supplements

Section 4.3., page 87, third paragraph, first two lines: this should be 63% to 65% percent to be consistent with supplementary table 11.

Author Rebuttals to First Revision (please note that the authors quote the reviewers in grey and respond in blue):

Referee #1 (Remarks to the Author):

I have reviewed a previous version of the manuscript and carefully read the revised version as well as the responses. They are very well articulated, but I still find several key points I raised previously remain unanswered. As highlighted previously, the amount of data compiled is impressive. But my most important concerns are that (1) the authors cannot demonstrate how their study is useful to water resources management and (2) From a scientific point of view, I maintain that all the factors that potentially influence gaining or losing conditions are already very well known, as also acknowledged by the authors.

I took note of the authors point that (quote from the reply) “Our goal is not to show that losing streams occur, but to shed light on how extensive and how common losing streams are». Without knowledge of fluxes, I fail to see how this information is useful, see also points below. I believe the authors missed the opportunity to demonstrate why this should be the case. Given that groundwater abstraction is ever-increasing in the US and most major aquifers are overpumped this is not a surprise in my opinion.

It is useful, from the standpoint of framing large-scale policy, to understand how common losing streams are. It is also useful, again from the standpoint of framing large-scale policy, to understand how the prevalence of losing streams varies with key landscape, climatic, and anthropogenic drivers. Thus our analysis is useful in both of these senses, even if we cannot exactly determine the losing fluxes for individual stream reaches.

Of course it would be even more useful to have that flux information, but that will remain unavailable for the foreseeable future. But the fact remains that important public policy decisions are made, and must be made, without the perfect information that we would ideally like to have. Until now, we knew only that some particular streams had been shown to be losing, and that model simulations say that losing streams might become common sometime in the future. Our analysis, by contrast, shows that many streams across the continental US are likely to be losing *here and now*, at least where groundwater wells are nearby. That result could potentially shift the debate over groundwater policy.

A continental-scale analysis may influence *interest in and concern for* groundwater/surface-water interactions, and may do so at a broader scale that would support jurisdictional cooperation; this can have cascading implications on policy*, law, and management*.

While water managers make the day-to-day decisions that *may* require flux information, the legal powers available to managers are determined by broader policy initiatives that are more likely to depend on larger-scale analyses such as ours.

*Here, we define policy as the “strategies and plans adopted by stakeholders used to guide decision making processes”; management is the “day-to-day activities” or the “implementation of rules and measures” that aim to meet policy goals (quote Gleeson, T., Cuthbert, M., Ferguson, G., & Perrone, D. (2020). Global groundwater sustainability, resources, and systems in the Anthropocene. *Annual*

Review of Earth and Planetary Sciences **48**, 431-463). We have edited the paper to reflect these definitions, and, therefore, have removed the word management in a number of instances, as indicated below.

So, can the study be used to identify general patterns of gaining/losing conditions? The new supplementary figure 24 nicely shows that this is not possible, at least not with the very limited number of explanatory variables used. There is an extremely large scatter, which clearly indicates the complexity between environmental and anthropogenic factors and the presence of gaining/losing conditions. Many other factors will come into play, too, such as land-use, infiltration capacity of the soil, the average slope of the river or the drainage density, to list but a few of them. Even if on average there is somewhat of a tendency, the complexity and interdependence of the factors overshadow the general trend, thus making it impossible to infer gaining or losing conditions based on the knowledge of these factors.

We agree that it will be difficult to reliably infer gaining or losing conditions *for individual stream reaches* based on an analysis such as ours. But that is not the point of our analysis. The point of our analysis is to estimate *the overall prevalence of gaining and losing conditions*, and to estimate how this prevalence is likely to depend on several key factors. Of course there may be other factors as well, as the reviewer points out, but it is still useful to know that “on average there is somewhat of a tendency”, as the reviewer states. We also note that where better data are available (such as the state of Kansas, shown in Supplementary Figure 16), the correlations are much stronger than just “somewhat of a tendency”.

Below some more specific comments that illustrate my points above.

- Again, essentially all the questions raised from line 54-58 require fluxes, as readily acknowledged by the authors in their response. The way the authors develop their manuscript suggests that this study can help in addressing these issues. But without fluxes, it does not. In their response the authors say (quote):

«Unfortunately, continental scale data are not available to calculate fluxes, especially fluxes in complex aquifer systems. We address this important point via revisions and new text added to our revised paper: Lines 107-112: “Although these water level differences can identify potentially gaining and losing rivers, quantifying stream-aquifer fluxes would require high-resolution three-dimensional hydrogeologic data (e.g., permeability) that are presently unavailable at continental scale. The lack of high-resolution geologic data also means that we cannot resolve two-dimensional cross sections of groundwater flow fields, and thus cannot capture flow patterns in some complex aquifer systems.»

This is very true and I fully agree. What I am missing is a follow-up on the implications of this, specifically for the utility of this particular study. The rest of this reply is a discussion on policy. I do not think the authors have addressed (nor can address) my previous point.

45

In the same context (page 5), the authors additionally address this point (Quote):

It would be wonderful to be able to quantify fluxes between rivers and aquifers at the millions of points where we have water table measurements, but no reliable data on streambed or aquifer permeability exist at continental scale. Our analysis nonetheless informs management by demonstrating that the potential for losing streams is widespread across the US, and by quantifying the influence of several causal factors.»

Again, the factors are all very well known, and the additional figure (Supplementary Figure 24) shows clearly the very wide scatter of gaining and losing conditions throughout all of the different parameters considered. A large-scale assessment does thus not help in any way in understanding the local conditions. The (unspecific) statement of the authors below is illustrating my point: “Thus, our analysis should help focus managers' attention where it is most clearly needed, even if we cannot quantify exactly which reaches of which streams are losing exactly how much “ But how does this study help in this regard? Do losing rivers require more attention as opposed to gaining rivers? In the revised version the authors did not demonstrate how their paper can help any manager to identify where attention is required. And I am sure they cannot, because, again, the fluxes provide the relevant information. That they are hard to obtain does not increase the value of the knowledge of whether a river is gaining or losing. I noted that one reviewer is of a different opinion, he highlights that the study can have significant implications for water resources management. If this is the case, why arnt these points listed in the revised paper? Again, none of the points listed from line 54-58 can be addressed with the current study.

- I still maintain that many points are trivial. That surface water and groundwater need to be managed jointly is so well known by now, that I do not think the study increases its impact by making this point (see e.g. the last sentence in the abstract or section line 198). The same remains for the point that losing conditions are found in areas of significant groundwater abstraction. To avoid overselling the paper, I strongly suggest to significantly down-tune the relevance of the study by explicitly highlighting that the study can neither improve local water resources management nor provides any novelty in terms of process- understanding. The point that remains is how widespread losing conditions are. It is up to the editor to evaluate if this sufficient for publication in this journal.

We have revised our first (summary) paragraph to clarify that the magnitude of the fluxes towards or away from the stream cannot be derived from the hydraulic gradients alone:

“A lack of adequate permeability data prevents quantifying these flows”

We refrained from making a specific statement that “the study can neither improve local water resources management nor provides any novelty in terms of process- understanding”. As described above, we believe that our study results are likely to be highly relevant to water resources policy, even if they do not tell managers exactly how to manage individual river reaches. And our results are indeed novel. For example, although of course it has already been known that groundwater abstraction can result in losing conditions, there has previously been no large-scale assessment of how much groundwater abstraction actually does, ~~on~~ average, affect the prevalence of losing conditions.

Our analyses map areas where hydraulic gradients are conducive to losing (and gaining) rivers, allowing the assessment of possible stream water losses on the national scale. Moreover, we are aware of specific examples where distinguishing potentially losing rivers from potentially gaining rivers can be directly relevant to ecosystem characterization and management. For example, one of the groundwater basins designated as “Critically Overdrafted” under California’s Sustainable Groundwater Management Act (SGMA) submitted a Groundwater Sustainability Plan (GSP) in January-2020 stating:

“If there was evidence that the surface water was connected to groundwater (i.e., a gaining stream), that reach would be eligible for inclusion as a potential GDE. Because the vast majority of rivers and streams in the subbasin are not perennial and all are in a net-losing hydrological condition (i.e., losing water to the groundwater system), this criterion excluded most of the smaller river channels and associated terrestrial vegetation from consideration as GDEs” quoting: Madera Subbasin Joint Groundwater Sustainability Plan: Appendix 2. Plan area and basin setting: Technical Appendices 2.A. through 2.B (January-2020). 74 pp. https://www.maderacountywater.com/wp-content/uploads/2019/12/Appendix2A-B_FinalDraft_191201-1.pdf

This statement refers to a “GDE” – a groundwater dependent ecosystem, which are relevant to the Sustainable Groundwater Management Act (e.g., <https://www.scienceforconservation.org/assets/downloads/GDEsUnderSGMA.pdf>)

In an effort to include Referee #1’s recommendation, we revised the lead paragraph in our manuscript to avoid using the term ‘management’. Our revised manuscript reads:

“These ‘losing’ streams have important implications for water availability, riparian ecosystems, and environmental flows⁴⁻¹⁰, but the prevalence of losing streams remains poorly constrained by continent-wide in-situ observations.”

Additionally, we have added the following sentence to the main text:

“Although our analysis cannot be directly translated into management prescriptions for individual river reaches, it demonstrates that losing rivers are widespread and it identifies conditions under which they are more prevalent.”

Further, we also:

- deleted a sentence “Sustaining or restoring groundwaters and surface waters requires managing them jointly.”
- We deleted a section heading “Managing groundwaters and surface waters jointly”
- We deleted the statement “, highlighting the importance of managing groundwaters and surface waters jointly.”

Referee #2 (Remarks to the Author):

I am reviewing this manuscript for the 2nd time, and I find it to be much improved. In reviewing the entire set of first reviews of the paper, there were clearly some critical comments (constructive, mostly), and my assessment is that the authors did an excellent job responding to the range of comments and very much improving the manuscript. More than 1 reviewer was concerned about identifying gaining and losing streams/rivers/reaches without confirmation from the field, and it was not clear to me how the authors would deal with this without diluting the message. It is my opinion that their recrafting of their message throughout was excellent and that this paper more precisely identifies the trends they find (even when they are counter to conventional wisdom as appears to be the case with 'flat' terrain), and truly "let's the data speak" as they state in their rebuttal comments. The revision process resulted in a very substantial revision of the text, which I believe is more on point and makes a sharper argument. In addition, it generated a larger supplemental information section, which is almost a paper on its own. It is full of additional analyses and summary tables. Together, these really make a strong point where the paper serves as the articulated synthesis of the 4 figures presented there, the millions of data points evaluated, and the extensive analyses in the supplemental materials, and the supplement really delves into the details that can address many skeptical comments (as was proven through the revision process).

The section 'Managing groundwaters and surface waters jointly' was only lightly revised (compared to other parts of the paper) and to me, this is really a key point. The authors' first citation is a USGS circular whose title is more or less the same message. That circular is something of a bible with respect to summarizing how surface water-groundwater interactions influence water quantity, quality, and ecosystems and their services to humans. It is largely a synthesis of findings from a large number of peer-reviewed papers that illustrates in many specific cases how these exchanges influence other aspects of natural systems. This paper comes to the same conclusion (i.e., this section heading) from a different perspective, one that is backed by very extensive data analyses. That is to say, this is the message that the data tell us when they speak and it's a critical message.

With all of this said, it is my opinion that indeed this is of broad enough interest and significance that it merits publication in Nature (as opposed to one of the other Nature topical journals). I conclude that the data overturned many of the arguments of the reviewers (i.e., I expect these are all specialists).

I am satisfied that the authors adequately addressed the simple and the more complicated reviewer comments. The message is straightforward and, given the data analysis, novel (in my opinion) because it is an outcome of these many many many observations (and their topographic/land use context of course).

We are grateful to Referee #2 for their helpful reviews and comments that improved our manuscript.

Referee #3 (Remarks to the Author):

My main comments and concerns have been addressed and the manuscript has significantly improved. Refining the wording and the addition of a section on limitations have helped to put the presented data-driven analysis into a more balanced perspective. I acknowledge that data-driven analyses at this spatial scale are challenged by data availability issues and the need for simplifications. Although this naturally constrains the ability of such analyses to explain conditions at every individual location, I am generally in line with the authors that these analyses do have their merits and can reveal interesting general patterns. The identification of high potentials for losing conditions in areas that are usually not considered to be hotspots for groundwater overdraft is interesting and likely a sign of the growing pressures on groundwater resources. Admittedly the correlations between the main explanatory variables and the fraction of wells indicating losing conditions are not particularly strong (rank correlation coefficients <0.35), but considering the scale of the assessment I would still consider them robust for the evaluation of large-scale general patterns. I also appreciate that the authors added specific links to data in their data availability statement. This will be useful for future studies.

We thank Jan Fleckenstein for his helpful comments.

I have a few remaining points, which I would suggest to consider to further improve the manuscript.

You use the word “prevalence” from the beginning of the manuscript. To me this word implies some bias in the sense that one thing occurs more frequently than another. Therefore I would avoid to use this term before you actually document this “prevalence”. For example in line 38 I would rather use an expression like “occurrence of losing streams”, which is poorly constrained by observations, because at this point a “prevalence” has not yet been shown. Although you state the 64% potentially losing river reaches in the next line of the abstract (which is more of a summary) I would also avoid the word “prevalence” in lines 72, 84 and 85. After you state the percentage of potentially losing reaches (64%) in the main text (line 121), which indeed indicates a “prevalence” for losing conditions the use of the term seems more appropriate.

Thank you. The revisions we’ve made to our manuscript after considering each of these recommendations have strengthened our final product. On this specific occasion we would like to retain the term ‘prevalence’ in each of the lines identified by Referee #3, because we understand the word “prevalence” to mean “frequency of occurrence”, which is precisely what our work seeks to quantify. In line 38, it is not used as a statement that losing streams are indeed prevalent, but instead a statement that we seek to assess how prevalent they are.

On the impact of well screen elevations: I can follow the line of argument that deep wells tend to have longer screened intervals and that the hydraulic heads they show represent an averaged head. However, I disagree that this automatically collapses a 2D (or in fact a 3D) flow problem to a 1D flow problem as the authors state in their response to one of my comments. This may be fair to state for a more or less homogeneous aquifer with vertical gradients, but not for a structurally complex aquifer system with a sequence of aquifers and aquitards. But this leads back to the

problem of locally more complex conditions, which cannot be resolved in an analyses at this scale. The newly added figure in the supplements on well screen locations for the wells, for which such data exists, is a good addition to the supplements and shows the large variation of well screen lengths that exist in wells (depending on the thickness of the vadose zone, the sequence of aquifers and aquitards and the targeted aquifer to pump from). However, the interpretation of the data in the figure in the figure captions contains an error. You state: “Our comparison of the total depth of wells (x-axis) and the depth to the top of their perforations (y-axis) demonstrates that many wells (for which we have screen interval data) have perforated intervals that begin at considerably shallower depths (e.g., 75% of wells that have total depths of between 99m and 100m have a perforated interval that begins 9.1 m beneath the land surface)” The part in the brackets is not correct. Your figure shows that 75% of the wells with total depths between 99 and 100m have their screens start not deeper than about 60m, and only 25% have their wells screens start in the uppermost 10m from the land surface. In other words quite a few wells record heads in the deeper parts aquifers. You need to correct this, even if it probably does not change your general statement that unknown well screen locations may have minor impacts on the results of your analysis.

We thank Reviewer #3 for pointing this out. We agree that we could have been clearer in our figure caption and thank him for this comment. We have revised our figure caption in an effort to be clearer, and the revised version reads as follows:

“For example, half of wells that have total depths between 99 m and 100 m have tops of perforated intervals (or bottoms of casings) within 25 m of the land surface.”

Line by line comments:

Line 132: I think to state that a potential for rivers to be losing is “the norm across much of the contiguous US” is quite a strong statement for only 64% of the well data indicating a potential for losing. I would suggest to scale down a bit, perhaps: “...well water levels that lie below nearby stream surfaces – indicative of losing streams – are surprisingly common across much of the contiguous United States....”

Thank you for your suggestion; we changed “the norm” to, instead, read: “common”. We elected not to write “surprisingly” as we cannot be certain that all readers of this study will indeed be surprised.

Line 261: I would replace the word “calculate” with “evaluate” or “assess”. You do not really “calculate” the potential for losing conditions, but rather “evaluate” it on the basis of the direction of a hydraulic gradient.

Thanks, we replaced “calculated” with “evaluated”

Line 268: “random uncertainties” sounds a bit strange. Perhaps rather: “...although these uncertainties, which will cause random errors in the data, should not substantially bias our assessment....”

We deleted the word “random”

Line 269: "...prevalence of losing versus gaining." But perhaps my understanding of the word "prevalence" is too narrow?

We retained the word "prevalence" here. By "prevalence" we meant the frequency of occurrence of a phenomenon.

Supplements

Section 4.3., page 87, third paragraph, first two lines: this should be 63% to 65% percent to be consistent with supplementary table 11.

Thank you. We revised to: "The fractions of well water levels that lie below the nearest stream vary from 63% to 65% among the three studied time intervals (Supplementary Table 11)."