An investigation of anthropogenic influences on hydrologic connectivity using model stress tests

Amelie Herzog[1], Jost Hellwig[1], and Kerstin Stahl[1]

[1]University of Freiburg, Faculty of Environment and Natural Resources

Correspondence: Amelie Herzog (amelie.herzog@hydrology.uni-freiburg.de)

Abstract. Human influences threaten environmental flows directly or indirectly through groundwater abstraction. In alluvial geological settings these may affect the contributions from GW sustaining streamflow during drysummer months. The Dreisam River Valley in Southwest Germany represents a typical case where recurrent hydrological drought events between 2015 and 2022 have led to interruptions of longitudinal connectivity in the stream network. When and where vertical connectivity changes and streambeds are drying have therefore become important questions. To help answer them, zero flow level (ZWL) occurrences were previously measured on twenty locations in the river network during the drought year 2020, but they revealed high variability. This study therefore aimed to develop a methodology that allows assessing the connectivity along the entire stream network. The task required employing a numerical groundwater model to obtain the spatial distribution of the exchange flow between GW and SW along the river. A reference model simulation for the period 2010-2022 assumed near natural conditions. Stresstest scenario model runs then imposed either an altered recharge regime or a set of introduced groundwater abstraction wells or both. as To gain confidence in the model ZWL patterns are compared to observations of dry riverbed locations in 2020 and the model generally reproduces those relative drying patterns. Modelled exchange flows of the stress tests were then compared against the reference simulation. A set of specific metrics combining longitudinal and vertical connectivity is introduced for this task. The results of the stresstest model runs show stronger changes of vertical connectivity in response to GW abstraction than to the imposed recharge stress. Reaches are identified where the effects of the stresses are particularly strong. Nevertheless, these results have to be interpreted within the limits of model realism and uncertainty. For more model realism a number of improvements will be needed such as a higher resolution parametrization of the river bed's hydraulic conductivities and better coupling to contributions from hillslopes; for a quantification of the uncertainties a systematic sensitivity analysis would be required. The study introduces a framework for modelling stresstests and metrics for surface-groundwater interaction that can be easily transferred to similar studies. Theapplication also demonstrate that even if not all influences can be modelled such studies are useful locally to help inform a resilient management of water resources under multiple stresses.

1 Introduction

Expansion and contraction of non-perennial streams cause variations of hydrologic connectivity in space and time in three dimensions: longitudinal (upstream-downstream), vertical (surface-subsurface) and lateral (channel-floodplain) (Datry et al.,

2017; Allen et al., 2020; Godsey and Kirchner, 2014; Freeman et al., 2007). The degree of hydrologic connectivity and the direction of the exchange flow control streamflow intermittency and magnitude during dry spells, solute and contaminant transport and associated nutrient and carbon cycling, and consequently also affects water quality and the lotic ecosystem processes (i.e. aquatic biota) (Datry et al., 2017; Costigan et al., 2016; Pringle, 2001). When flow ceases, hydrological connectivity is interrupted and physical, chemical and biological processes are modified. Notable changes in the frequency and duration of such disruptions of hydrologic connectivity may lead to cascading changes in aquatic ecology. Fluctuations of hydrologic connectivity are particularly severe in non-perennial river systems, where the characterisation of spatial and temporal patterns of hydrologic connectivity is an important knowledge for water resources management as well as ecosystem conservation success. While regular or irregular connectivity alternations are the rule in non-perennial streams, it is often challenging to disentangle whether these occur naturally or whether they are due to anthropogenic influences. Water withdrawals from surface and groundwater are a major cause of hydrological alterations, such as lower flow volumes and prolongation of dry spells in rivers (Yildirim and Aksoy, 2022; AghaKouchak et al., 2021; Goodrich et al., 2018; Datry et al., 2017; Tijdeman et al., 2018). As water withdrawals create fluctuations of the groundwater level and the water stage in streams, they modify vertical connectivity and affect gaining or loosing conditions along a stream. These groundwater surface water interactions may alternate seasonally, depending on the hydrological regime and the hydrogeological setting. Particularly in snow dominated systems, vertical connectivity becomes more relevant in the summer season, as bank storage is less important (Huntington and Niswonger, 2012). In order to understand if flow alterations in dry phases are exacerbated in response to water withdrawals, it is therefore crucial to investigate the relationship of longitudinal and vertical connectivity.

Many studies have focused on one of the three dimensions of hydrologic connectivity, while fewer studies link the different dimensions of connectivity (Zimmer and McGlynn, 2018). Lateral connectivity studies mostly focus on hillslope connectivity (Zuecco et al., 2019; Blume and van Meerveld, 2015; M. Rinderer et al., 2019; Jencso et al., 2009) or connectivity of hillslopes to floodplains (Xu et al., 2020; Czuba et al., 2019; Gallardo et al., 2014). Recent studies on non-perennial river systems have predominantly analysed longitudinal connectivity in order to describe flow characteristics (magnitude, frequency, duration) for a characterisation of flow regime or to describe the spatio-temporal extent of non-perennial river systems (based on active drainage area, for example) (Price et al., 2021; Hammond et al., 2021; Belemtougri et al., 2021; Botter et al., 2021; Botter and Durighetto, 2020). This one-dimensional perspective with a focus on longitudinal connectivity has also been adopted to investigate whether anthropogenic activities cause hydrological alterations, e.g. expressed as "streamflow signatures" for regulated catchments and reservoirs (Salwey et al., 2023; Ruhi et al., 2022; Ferrazzi and Botter, 2019). Statistical models have been used (Jensen et al., 2018) to describe flow dynamics in non-perennial stream reaches but those models do not consider the physical processes underlying hydrologic connectivity.

Due to a lack of hydrometric data on non-perennial streams at necessary spatial resolution and over longer time periods, another prevalent research goal is the collection of streamflow and water level data at different scales, e.g. using field surveys (Zimmer and McGlynn, 2017), citizen science (Etter et al., 2020; Strobl et al., 2020) and new methods to measure dry phases in non-perennial streams (Herzog et al., 2022; Zanetti et al., 2021; Assendelft and van Meerveld, 2019; Jaeger and Olden, 2012). But without the necessary data, linking the different dimensions of hydrologic connectivity remains difficult (Meerveld et al.,

2020).

Specifically for an analysis of vertical connectivity, not only hydrometric data of the surface water (SW) system is required but also data for the groundwater (GW) system. Joint approaches considering GW and SW data are usually restricted by a low spatial data availability of GW heads but also by lacking measurements of transmission losses or GW-SW interaction itself,
which are difficult and only possible at particular locations along the stream. Measurement approaches for the quantification of GW-SW interaction are often based on tracers, e.g. heat (Angermann et al., 2012; Fleckenstein et al., 2010) or isotopes (Bertrand et al., 2014; Kalbus et al., 2006) which can be used to derive information on flow paths and residence times on very short timescales and small spatial scales only. Up to date, there is no reliable method for an up-scaling of GW data to large scales (Foster and Maxwell, 2018; Barthel and Banzhaf, 2016). Therefore, few data-driven studies on the spatio-temporal evo-
lution of vertical connectivity of non-perennial river systems are documented. Despite acknowledging the relevance of the GW storage term in the water balance (particularly in alluvial aquifers, where GW leakage is important) and the retroactive effects between headwater and lower catchment areas caused by GW leakage (Fan, 2019; Käser and Hunkeler, 2016; Covino and McGlynn, 2007), knowledge on the relative roles of the controls of GW-SW interactions remains restricted to few locations.

Given the lack of data, integrated models (IM) to-date have an important alternative role for the investigation of GW-SW-
dynamics at the catchment scale even though the possibility for calibration and validation of the simulated GW-SW interaction itself is limited as a result of missing data and required numerical effort (Barthel and Banzhaf, 2016). Uncertainties due to the influence of model resolution on model parameters and the complexity of the parameterisation are often discussed as limitations (Foster et al., 2020). IM have been successfully deployed to obtain a better understanding of the control factors of GW-SW-interactions, for example by means of sensitivity analysis in order to investigate the influence of specific parameters on model results (Herzog et al., 2021a; Foster and Maxwell, 2018) or hypothetic experiments to understand interaction of GW, SW and vegetation (Schilling et al., 2021). But, such experiments are computationally expensive. If obtaining a best-fit model might not be the primary aim and interest is more on general responses to changed conditions, targeted stress test model experiments may be a suitable alternative , e.g. to assess the response of groundwater drought and baseflow to changed antecedent recharge conditions (Hellwig et al., 2021). Stress test modeling has traditionally been used for management purposes to test the resilience of a system against specific stresses or even worst case scenarios and has recently been used in applied research in combination with sensitivity analysis. Most studies hereby inquire the effect of different climatic conditions (Hellwig et al., 2021; Stoelzle et al., 2014, 2020), while anthropogenic influences might be equally important in magnitude and relevant for management purposes.

In this study, we develop a modelling framework and specific assessment metrics to examine specifically the alteration of
(simulated) longitudinal and vertical connectivity along a stream in response to stresses. These stresses are formalized as a set of model stresstest scenarios that employ changes to climatically driven groundwater recharge and to groundwater abstraction and the combination of both. In addition, the aim of the study is to reassess findings from available observations of ZWL for a particularly dry year ((Herzog et al., 2022)). In this study the longitudinal connectivity was found to be highly variable with upstream to downstream drying in some and the reverse or more complex patterns in other tributaries, but finding clear patterns
was hampered by observations at few locations in the stream network. Modelling the entire stream network may enable further

insight into spatial and temporal drying patterns and the observations may help assess the model's ability to simulate stress responses. To analyse the response to the model stresstests on the SW-GW interaction, we introduce metrics that describe the relationship of vertical SW-GW interaction (leakage in modelling terminology) and longitudinal (zero water levels, ZWL) connectivity following general ideas form the hydrological alterations (HA) approach (Poff et al., 2010) . We test this stresstest approach in a meso-scale catchment in Southern Germany where discussions over stream and groundwater uses have emerged during recent dry years with reduced groundwater recharge. In this typical case of an alluvial valley aquifer and locally important river, approaches were needed to inform the debate with multiple water users. . The design of the model experiments and stress tests is chosen in order to answer local as well as more general research questions:

- Do the model experiments confirm the observed relative patterns and spatial distribution of longitudinal and vertical connectivity in the case study area?

- How strong are the responses of connectivity to impose stress from recharge deficit and GW withdrawals during a dry summer ?

- Can specific metrics distinguish the patterns of longitudinal and vertical connectivity and their responses to those stresses?

2 Methods

2.1 Study area and data availability

The study area is the Dreisam valley $25\,\mathrm{km}^2$, a sub-catchment of the Dreisam catchment in the federal state of Baden-Württemberg in Southern Germany $577\,\mathrm{km}^2$ (Figure 1 a)). All tributaries have their sources in the Black Forest Mountains and converge in the Dreisam valley which therefore has gentle slopes in the center and increasing slopes and altitudes towards the borders (Herzog et al., 2022). The geology is characterised by crystalline basement overlain by thick alluvial deposits. The uppermost alluvial materials belong to the so called Neuenburg formation, younger quartenary gravels with high hydraulic conducitivities ($3.2\,\mathrm{m\,d}^{-1}$ logarithmic). This alluvial filling reaches up to $25\,\mathrm{m}$–$40\,\mathrm{m}$ depth in the northern part whereas it decreases towards the southern part of the study area and contains the main aquifer (Wirsing and Luz, 2007). Older quarternary sediments with smaller grain sizes below this layer are less transmissive. Previous studies on runoff generation processes suggest, that groundwater exfiltration into streams (i.e. 'baseflow') is one of the main runoff processes contributing to streamflow during dry phases in the Dreisam valley (Ott and Uhlenbrook, 2004). However, the degree of the connectivity between groundwater and surface water varies and is difficult to quantify longitudinally.

The aquifer is used for about two thirds of the water supply of the city of Freiburg im Breisgau ($8.6 \times 10^6\,\mathrm{m}^3/\mathrm{yr}$). Based on data from government agencies on water withdrawals in the study area, it is estimated, that withdrawals for water supply account for $> 95\,\%$ of total withdrawals in the study area. Withdrawal data as well as GW level data are monitored and were provided for the period 2014-2022 by the largest regional water supplier (eight locations see Figure 1 a)). Other withdrawals were not available and could not be considered in the model experiment.

An dataset was available containing observed stream stages and ZWL measured at 20 locations in the study area in the summer of 2020 (Herzog et al., 2022). In addition, streamflow measurement campaigns in two years with contrasting climatic conditions, i.e. 2020 (dry) and 2021 (wet) were used to develop rating curves for the calculation of streamflow at these ZWL monitoring locations (Herzog et al., data prepared for publication). The letters in the station-IDs shown in the figure refer to the different main tributaries they belong to, i.e. Dreisam (D), Eschbach (E), Rotbach (RO), Wagensteigbach (W), Ibenbach (IB), Reichenbach (RE); Brugga (B), Krummbach (K), Zastlerbach (Z), out of which only those with ZWL in 2020 were used in this study.

2.2 The model concept

To represent both, the surface and the subsurface system we used a combination of the hydrological model RoGeR (Runoff Generation Research Model) (Steinbrich et al., 2021, 2016) and the GW model Modflow6 (Langevin et al., 2017) with the surface water routing (SFR-package) (Figure 1). RoGeR is an advanced rainfall-runoff model, which calculates runoff components (interflow, overland runoff, percolation) for unit areas of similar climatic, topographic and pedological properties (Version RoGeR WB 1D). Modflow6 solves the three-dimensional, transient GW flow equation (Darcy's law and continuity equation) for the simulation of GW heads. RoGeR and Modflow are offline coupled, i.e. the different runoff components modelled by RoGeR are used as input for Modflow's stress packages RCH (percolation) and SFR (fast runoff). Detailed information on
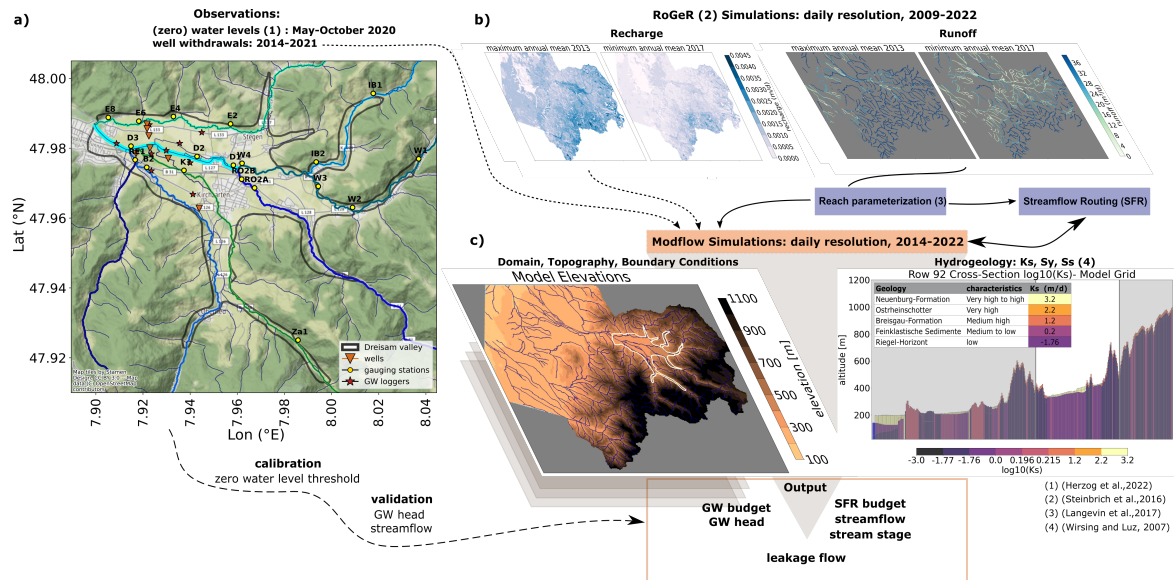
Figure 1. a) shows the Dreisam valley, the locations of the gauging stations from (Herzog et al., 2022) and the locations of the wells. b) RoGeR simulation output is converted to the model grid and added to the model as reach parameterization or boundary condition. c) The Modflow model uses data from a) and b) and additional hydrogeological data. Connections represented as dashed lines are input modified in the stress tests and bidirectional arrows indicate online coupling.

Modflow's modelling approach as well as the different packages are given in Langevin et al. (2017). Modflow uses a (block-centered) control-volume-finite-differences method in order to give an (iterative) approximation of the analytical solution of the partial-differential equation. For unconfined conditions (as we assume in this model approach), the transmissivity of a grid cell varies based on saturated thickness of the cell.

The model domain covers not only the river valley but the entire Dreisam catchment and the neighboring catchment Möhlin-Neumagen ($708.09\,\mathrm{km}^2$) with a river network length of $833\,\mathrm{km}$ and the spatial resolution of the GW model is $100 \times 100\,\mathrm{m}$ (Figure 1 c)). For the Modflow model set up spatially distributed parameters determining surface and subsurface flows are required. Elevation data were obtained from a 30 m DEM of the modelling domain. Four model layers are used. They were based on subsurface information from gridded data of aquifer thickness, hydraulic conductivity and storativity for each layer of alluvial valley fill according to (Wirsing and Luz, 2007) . The empirical equation of Marotz allows to determine specific yield as a function of hydraulic conductivity (Fuchs et al., 2017; Marotz, 1968).

River reaches are defined in Modflow's SFR-package as sections of streams in one model grid cell and further described by several parameters such as reach width, depth, slope, thickness of streambed sediments, Ks of streambed sediments. Details on parameterization requirements can be found in the Modflow6 documentation (Langevin et al., 2017). Parameters were derived from topography, river network data and governmental hydrographic surveying data.

Time-variable and spatially distributed input to the Modflow6 GW model is taken from high-resolution RoGeR simulation output and used for daily resolution modelling of the time period $2009 - 2022$. RoGeR's percolation component, which can also become negative in case of capillary rise, corresponds to recharge in Modflow. The sum of interflow and overland runoff components corresponds to runoff directly contributing to streamflow. Additionally, transient GW extractions are added to Modflow by means of the WEL-package. The regional water supplier provided the required daily pumping rates.

In general, one model grid cell can contain several river reach units and thus, simulated GW head in one model grid cell is used to calculate the exchange flow (or leakage) of all the streams corresponding to this particular grid cell. The sign of leakage reflects the groundwater terminology with exfiltrating conditions (negative leakage) describing gaining streams and infiltrating conditions (positive leakage) describing losing streams. The leakage between the aquifer and the riverbed depends on the simulated hydraulic gradient between surface water and GW. For GW heads above the streambed, the gradient is the difference of stream stage and GW head whereas for GW heads below the streambed leakage becomes independent of GW head and the gradient is solely the surface water stage above streambed. Leakage is then calculated as a product of hydraulic conductivity of streambed sediments, the streambed area of the reach and the hydraulic gradient divided by the thickness of streambed sediments(Langevin et al., 2017).

Streamflow for each river reach is obtained based on the principle of continuity, considering that source terms (inflow from up-stream reach, direct overland runoff and GW leakage to a reach) equal the sink terms (outflow to downstream reach, diversions from another reach, leakage to the aquifer). Based on this flow, stream stage is calculated for every reach using Manning's

equation. As leakage flow and the overall water budget are dependent on the stream stage, they cannot be directly calculated but the equations need to be solved iteratively.

2.3   Reference simulation and stresstest scenario modelling

A reference simulation was needed as a benchmark to compare the stress test simulation results against. Groundwater recharge from the ROGER model was available for the period 2009-2022, which was used as input for a transient Modflow model run. A near-natural system scenario with this input was chosen as the reference simulation (Figure 2 (a)) that neglects all water uses in the study area. The case study area has a long history of SW and GW use for a range of purposes and most of these are not quantitatively known. Only the groundwater withdrawal rates from a main drinking water supplier were available. Other infrastructures include private and smaller groundwater wells, river abstractions for small scale irrigation, weirs and small scale hydropower, trained and stabilized sections of the river for erosion control etc.. Therefore, the model cannot be calibrated or validated to a known natural system state and the assumption of a natural system scenario is uncertain. Nevertheless this reference simulation can serve as a benchmark to analyse relative changes in stream connectivity in response to specifically applied model stress tests.

As stresstest scenarios we defined stresses due to the impact of climatic conditions, i.e. by applying altered recharge conditions to the reference model (Figure 2 c) and e), left side) and by adding stresses due to the anthropogenic impact in the form of known GW withdrawals (Figure 2 b),d) and f), right side) in the valley bottom near the main river. GW recharge is the quantity of water percolating from the surface-water system into the GW system and is directly linked to both, hydro-meteorological conditions (precipitation and evapotranspiration) and soil properties (infiltration capacity and runoff pathways). According to the ROGER simulation the most 'normal (average)' water year in terms of groundwater recharge was the year 2012, the lowest annual GW recharge was 2017, and the winter season with the lowest GW recharge was 2019 (Figures A6 and A7) (a water year starts in November with the summer season starting in May). Synthetic, gridded recharge stress was generated from altering the order of simulated recharge output of the RoGeR model (2009 − 2022) to create two scenarios: one with changed recharge magnitude (Figure 2 (c)) and one with changed recharge seasonality (Figure 2 (e)). Because the dry summer of 2020 was chosen as the main year of analysis, the recharge stresstests focus on changing the recharge conditions preceding 2020. The magnitude scenario focuses on drought stress and therefore replaced the real recharge time series with alternating normal years (2012) and drought years (2020). The seasonality scenario expresses a system change with a tendency for winters to be drier; to simulate this, the driest winter (2019) was repeated again before the summer of 2020. Stress tests of GW water withdrawals are based on real daily abstractions for drinking water that were available for the time period 2014-2022. The withdrawal stress was added to the reference simulation (for simplicity called "wells" scenario) as well as to the magnitude (mag+wells) and seasonal (seas+wells) recharge scenario simulations allowing a comparison of the system response to altered groundwater recharge, to GW abstractions, and to the combination of all stresses.

2.4 Assessment of SW-GW interaction: approach and metrics

A direct validation of modeled SW-GW interaction is not possible for a number of reasons. None of the model runs includes all human influences on surface and groundwater in the valley as these have a long history and quantitative data is unavailable. Therefore there is no simulation that would correspond exactly to the real conditions in the field. Also, exchange rates along the stream cannot be measured continuously in the field. Model grid cells with $100\,\mathrm{m}$ resolution will not resolve many smaller stream reach variations that might influence stages and zero flows measured at certain points. Nevertheless, one aim was to test the usefulness of the experimental dataset that exists for the summer of 2020 to build on the insight from a previous experimental study. Therefore the approach is to compare relative patterns and changes of longitudinal and vertical connectivity with specifically developed metrics. For that we were guided by work on the hydrologic regime of non-perennial rivers (Magand et al., 2020; Costigan et al., 2016; Gallart et al., 2012). From an ecohydrological perspective, three (hydrological) aquatic phases (dry, standing, flowing) relate to five (ecological) aquatic states (Meerveld et al., 2020; Datry et al., 2017; Gallart et al., 2017,
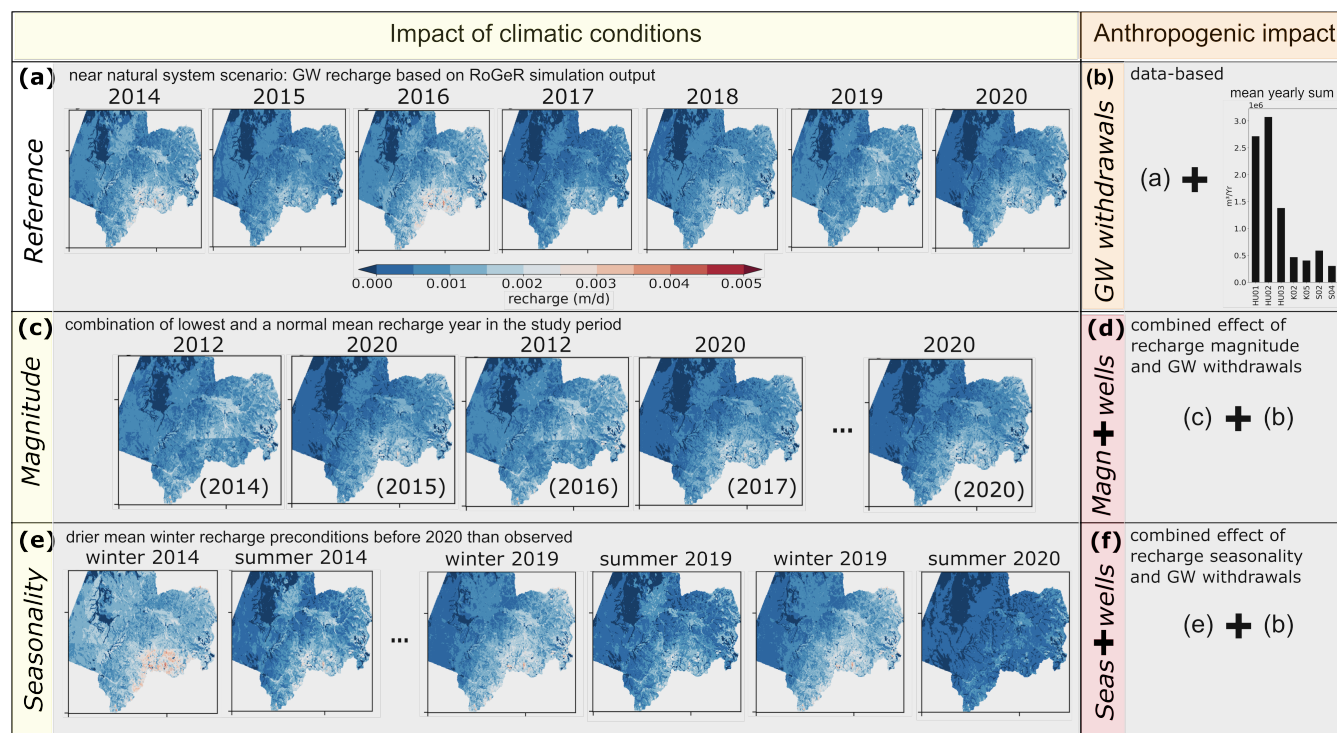
Figure 2. Concept showing all model experiments with the recharge scenarios representing climate stress in the left column with a) the reference simulation with real recharge input from 2014-2020 and c) the magnitude stress and e) the seasonality stress; and anthropogenic impact in the right column with b) expressing the reference with abstractions and d) and e) the combined stress of altered recharge and abstractions . For climate impact, the annual mean recharge of the water year or the season are shown. For anthropogenic impact, the mean annual GW withdrawal rates for the period at the wells are shown.

2012). The available water level dataset of (Herzog et al., 2022) contains derived ZWL occurrences and thus, the dry phase. The occurrence of ZWL is therefore an indicator for a loss of longitudinal connectivity. Indicators for vertical connectivity are the direction and amount of GW-SW exchange flow (leakage, in Modflow terminology).

Some prepossessing and selection of suitable locations out of the available 20 observational records was necessary to derive ZWL and leakage indicators. For this study the raw dataset of observed ZWL occurences (15 min resolution) was converted to ZWL-days respecting a maximum duration ($<24\,$h) between two ZWL occurrences. In the simulations, negative outliers of simulated stream stages may occur for simulated GW heads far below the surface. We first determined a range of possible zero water level thresholds T for each observational location x within the sum of minimum water level and either 5 % (or 1 %) depending on the variability) of the mean and the 90 % quantile (equation 1) of water levels h at this station.

$$T_x = [\min(h) + 0.05 * (\frac{1}{n}\sum_{i=i}^{n} h_i)...Q_{90}] \tag{1}$$

We then chose the threshold with the smallest difference in total number of observed and simulated ZWL days. For the locations where the so 'calibrated' ZWL-days are comparable with the observations, we assume that the drying phases are well represented by the model and select this subset of 'trusted' locations for further metrics and comparisons among the different model runs. The direction (and quantity) of the modelled vertical connectivity of a stream reach can change with time depending on the flow conditions in the stream and the GW head. We assume that the connection between GW and the streambed is lost at first when GW leakage changes to zero. Thus, we extracted direction changes to zero leakage and compare simulated zero leakage and the measured ZWL days. As for ZWL, the definition of zero leakage also requires the use of a location-specific threshold T_{leak}. We define T_{leak} as 10% of the maximum leakage simulated at each location.

In order to analyse the effect of the imposed model stresses we developed questions regarding vertical connectivity (leakage) and longitudinal connectivity (ZWL). These questions relate to duration, timing and frequency of the different phases of vertical and longitudinal connectivity (Table 1). For each set of questions we developed a corresponding metric that allow a quantitative comparison between model and observation, between reference and stresstests and among the different stresstests.

3 Results

3.1 At-site comparison of zero water levels

ZWL occurrences were measured at 20 locations in the study area (Herzog et al., 2022). (Table 2) shows the ZWL for the wells simulation (reference+withdrawals) derived with the help of calibrating the thresholds T. For about 50% of the locations (E4, W4, RO2A, E6, E8, E2) agreement is good, but for the rest of the locations differences in ZWL percentage are greater than 15 % . In general, the simulated percentage of ZWL days is higher than for the observations, in particular for locations further upstream as indicated by the distances from the main gauging station at the downstream outlet of the Dreisam catchment. An exception are the locations E2 and E6. A comparison of mean simulated water levels against observed water levels

Table 1. Metrics for the assessment of connectivity and their underlying questions

Characteristic	ZWL-day variable	Leakage variable	Metric combined	Question
Duration	t_{zwl}: longest duration of ZWL days	t_{0leak}: longest duration of zero leakage days	$\delta t = \frac{(t_{0leak} - t_{zwl})}{365}$	How different is the duration of the longest ZWL-period and the longest period with zero leakage?
Timing	t_0: time of the first ZWL occurence	t_{ch}: time of the change to zero leakage prior to t_0	$\Delta t = t_0 - t_{ch}$	Is there a delay of the first ZWL after leakage has ceased and how long is it?
Frequency	$N0_d$ and $N0_w$: mean duration of ZWL events for the drying (d) phase and rewetting (w) phase	$N0l_d$ and $N0l_w$ mean duration of zero leakage for the two phases	$\omega = \frac{(N0l_{d,w} - N0_{d,w})}{D_{d,w}}$ $D_{d,w}$: number of days in each phase	How long are periods of ZWL and no leakage during during phases of drying and rewetting?

250 indicates a general underestimation of water levels in the model despite more heterogeneity in the derived streamflow and an overestimation of mean GW levels (Figures A2, A1). This disparate bias suggests, that the underestimation of water levels are likely due to the streambed parameterisation. While the wells model simulation should be closest to today's real situation as explained before, it only contains the most important abstractions in the valley. Regarding general relative patterns of ZWL along particular tributaries with three observation locations, the model runs confirm the successively increasing percentage of

255 ZWL from upstream to downstream for the location along the Eschbach (E) while the more complex decreasing-increasing-decreasing pattern in the Wagensteigbach is simulated as consistent decrease. Nevertheless, the two different overall directions are captured by the model run.

Table 2. Drived ZWL-day percentages observed and modelled for the study period for two simulations (ordered from upstream to downstream by distance to catchment outlet gauging station, with grey shaded rows with bold numbers indicating best correspondences with percentage differences $< 15\%$)

Location	T_{wells}	Observed ZWL (%)	Simulated ZWL (%) by wells scenario	Location from catchment outlet (km)
E8	**0,0**	**74**	**81**	**0,5**
D3	5,3	15	38	1,5
E6	**0,0**	**43**	**31**	**1,7**
RE1	0,9	3	28	1,9
E4	**0,0**	**24**	**29**	**3,0**
D2	3,8	8	36	3,7
W4	**5,3**	**31**	**36**	**5,3**
E2	**-0,2**	**13**	**1**	**5,4**
RO2A	**0,7**	**55**	**57**	**6,2**
IB2	1,4	7	47	8,4
W3	5,0	7	51	8,7
W2	5,8	19	63	10,4
IB1	0,2	13	58	12,1

3.2 Stresstest effects on GW leakage

3.2.1 Spatial distribution of GW leakage in the study area during different GW conditions

260 The spatial distribution of leakage flow (L) differs for the driest GW conditions (days with highest mean depth to GW) and the wettest GW conditions (days with lowest, simulated mean depth to GW) in the simulation period (Figure 3). For dry GW conditions the leakage approaches zero especially in the northern part of the catchment (E tributary) (Figure 3 a) and b)). In general, depth to GW is lower in the north eastern part of the catchment. Strong topographic gradients towards the border of the catchment lead to strong hydraulic gradients which are problematic to represent as an average for one grid cell and

265 therefore depth to GW is less reliable in these areas. (Figure A3). Including the abstraction wells in the simulation leads to stronger decrease in leakage in the downstream part of the E tributary (for all stress tests with wells). The recharge scenarios affect the timing of GW drought. The dry GW conditions occur earlier in summer and wet GW conditions earlier in January (at least for the magnitude recharge scenario) but the leakage pattern itself for dry conditions and wet conditions does not differ significantly from reference/near-natural conditions (only if wells are included additionally). For wet GW conditions, the river

270 system is entirely connected to the GW system in the study area as there is almost no place with zero leakage.
The difference in mean leakage (Δ L) of the reference conditions versus stress test conditions shows, that main differences in the downstream part of the catchment are caused by the implementation of wells in the simulation (Figure 3 c)). Concerning the mangitude and seasonality recharge scenarios, the Δ L is slightly higher (leakage in natural system is greater) in the

upper part of the E tributary and in the downstream part of the eastern tributaries (towards W and R tributaries). As in Fig.
3 a), the effect of wells and modified recharge adds up when both are combined in one stress test. The standard deviation of
leakage is highest along the E river, indicating that the variability of leakage is particularly high along this tributary (Figure 3
d)). Standard deviation is particularly low at the W and SW borders where slopes are increasing. However, due to the abrupt
change of simulated GW head at the catchment borders (Figure A3), the simulated leakage is highly uncertain for areas with
increasing slopes. Nevertheless, the results give a general overview where in the study area leakage variability and dynamics
are particularly strong and more specifically how these changes in leakage are related to physiographic characteristics (such as
slopes, topography or whether the location is situated upstream or downstream).
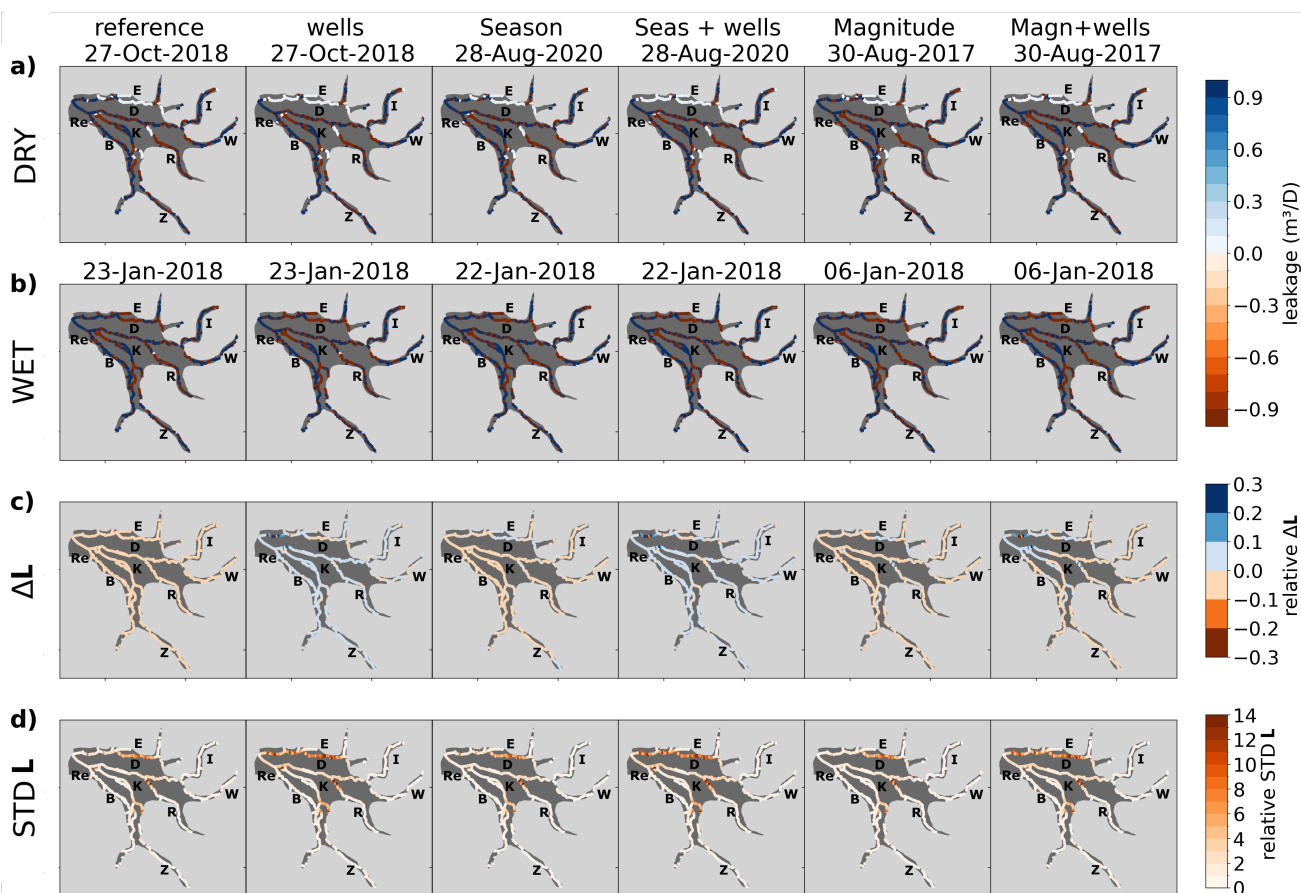
Figure 3. Spatial distribution of leakage for the reference and all stresstest simulations a) for wet conditions of mean maximum depth to GW,
b) for dry conditions with minimum depth to GW, c) the relative difference between the stress test simulation and reference simulation d) the
relative standard deviation of leakage

3.2.2 Longitudinal variation of leakage flows

As leakage quantities cannot be validated but depend on simulated streamstage/streamflow, we analyse the longitudinal variation of GW leakage in detail for the different measurement locations in the following. We divided leakage flow $(\mathrm{m^3\,d^{-1}})$
285 by the simulated leakage flow area in $\mathrm{m^2}$ to obtain specific leakage $(\mathrm{mm\,d^{-1}})$. For better visualisation and comparison, we normalized the leakage flow of 2020 (setting the minimum to zero. Minima for the reference simulation and the wells scenario are close). Quantities of leakage show strong variations in a longitudinal direction along the river bed (Figure 4) and a tendency for decreasing leakage variation with distance from the outlet (IB2, W3, W2, Za1). This pattern however has to be interpreted with caution, particularly due to the overestimation of ZWL days at the stations located far from the outlet. Characteristic for
290 most locations in the main Dreisam river (D3, D2, D1) are loosing conditions (from a river perspective). Even though the main gauging station near the outlet (PE) is also located in the main river, it shows gaining conditions. Gaining conditions are
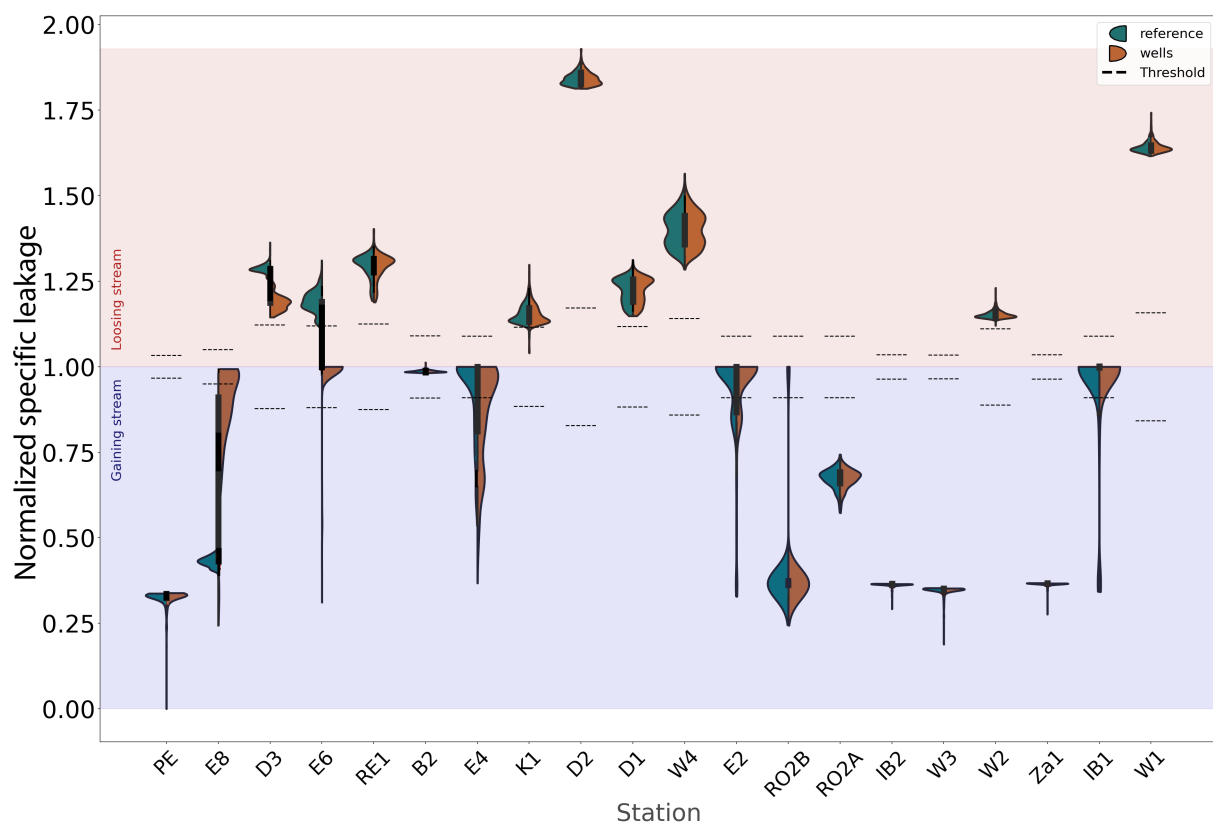
Figure 4. Distribution of normalized specific leakages for the reference scenario and the wells scenario for all gauging station locations in 2020. From left to right the distance from the catchment outlet (station PE) is increasing. Values below 1 (above 1) indicate, that the water flow is directed towards the river (the aquifer).Note: Thresholds are location-specific as described in (see section 2.4).

otherwise primarily found along the E tributary (E8, E4, E2), except for E6 in the reference simulation. At most locations, there is no profound difference between the reference and the wells stress test. The most obvious exceptions are locations in the downstream E tributary (E6, E8 and E4) as well as D3 in the D river.

295   3.2.3   Comparison of leakage direction changes

In order to understand which stream reaches experience changes from gaining conditions to no connection or loosing conditions, we first evaluate the temporal evolution of direction changes at the specific locations for the different stress tests. Three different GW leakage conditions may occur: zero leakage, meaning that there is no exchange between the river and the aquifer, positive leakage (from the perspective of the GW body), meaning that the river experiences losing conditions and negative
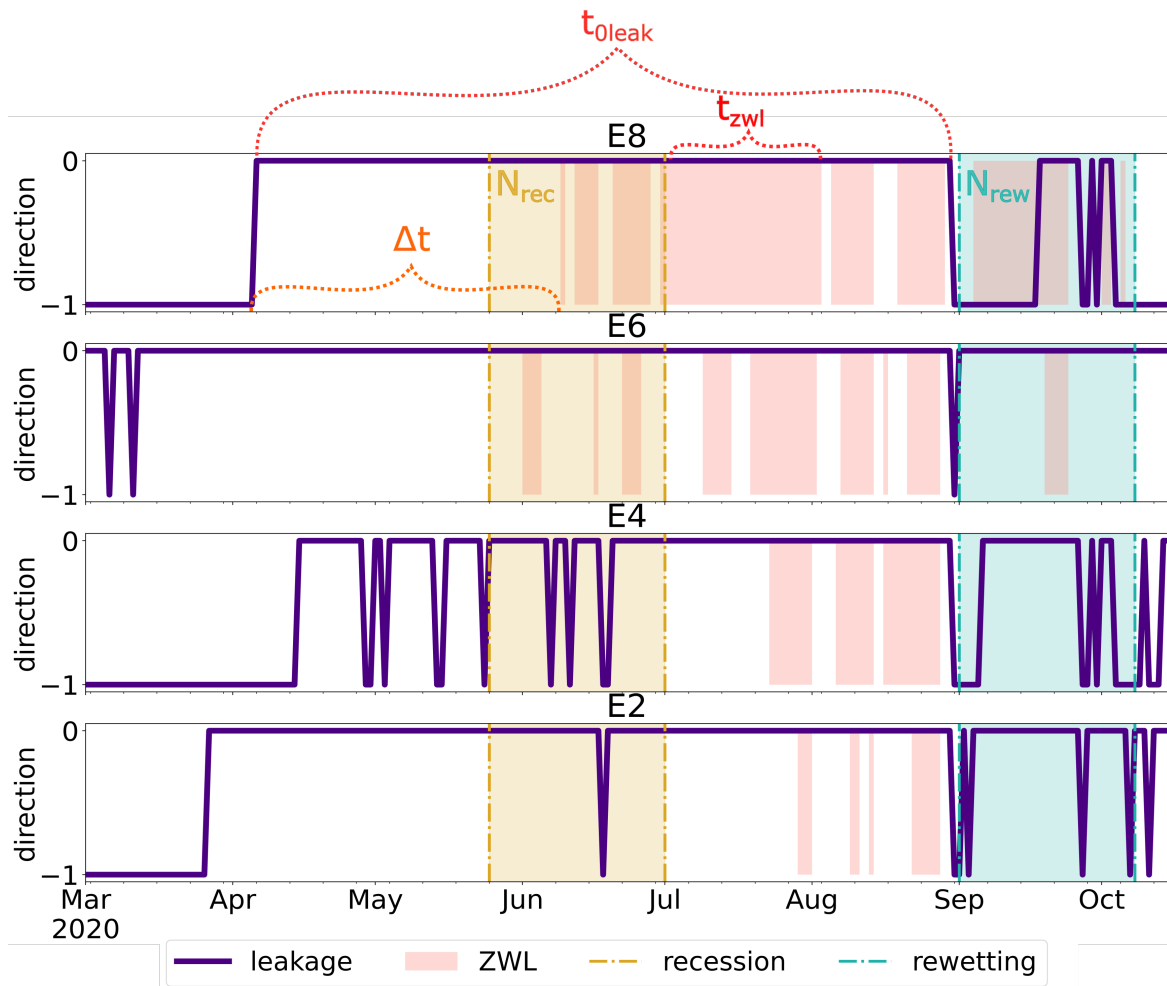
Figure 5. An example of leakage direction changes from gaining conditions (-1) to zero leakage (0) and ZWL days for the wells scenario at the measurement locations along the E tributary.

300 leakage, meaning that the river experiences gaining conditions. For normalized leakage, we define zero leakage within a range of $[1 - T_{leak}, 1 + T_{leak}]$, where T_{leak} is the location specific threshold (as illustrated by the dashed lines in Fig. 4). For the following analysis we make a choice of locations based on the occurrence of leakage direction changes first. Note however, that there are differences in how well ZWL days have been simulated at these locations (Table 2). The characteristics of the temporal evolution of zero leakage and ZWL days form the metrics as described in section 2.4 (Figure 5). While E8, E4 and

305 E2 experience only direction changes from gaining conditions to zero exchange, E6 experiences direction changes from losing conditions (for natural conditions) to mostly zero exchange (Figure A5). However, E6 stands out because it is the location closest to the well with the highest abstraction rates and the flow direction changes differ strongly between stress tests with and without wells. As already noted, the upstream stations along the E tributary experience less ZWL days than downstream locations.
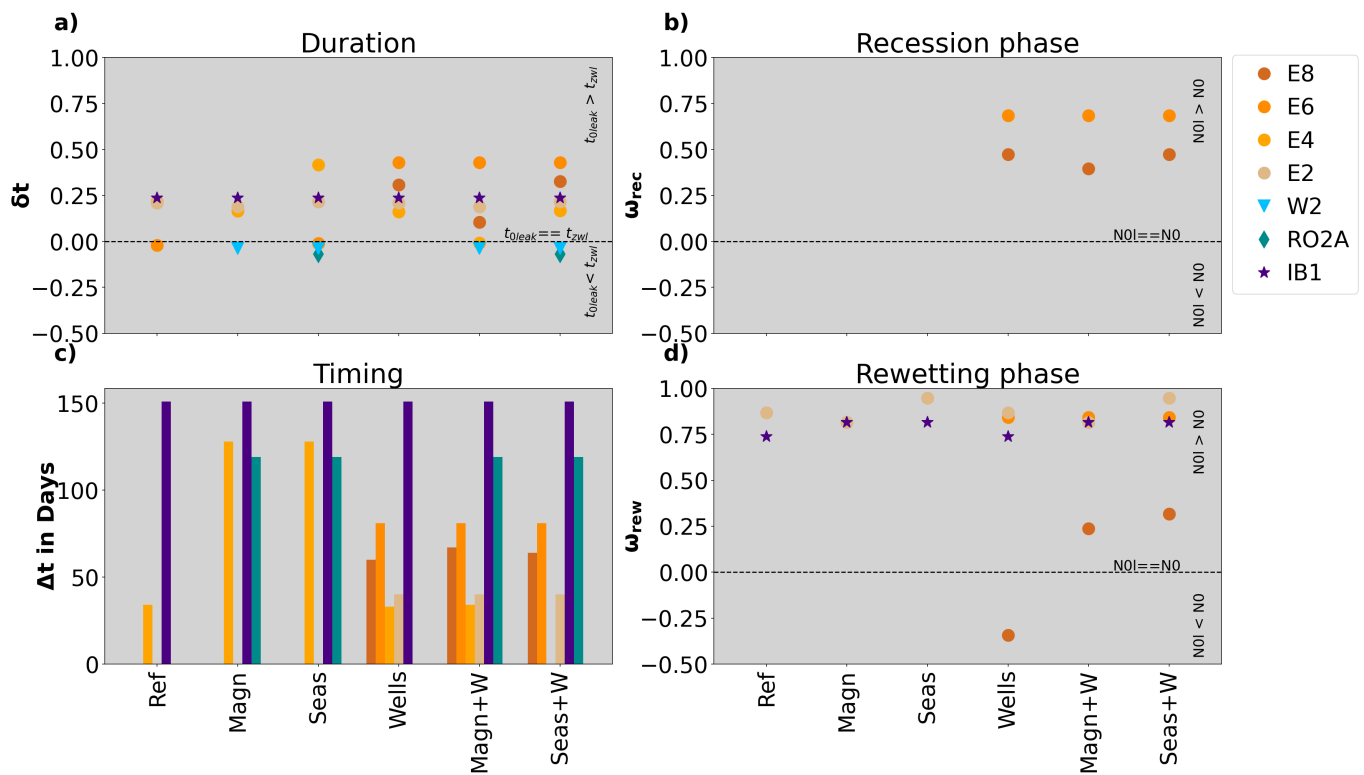
Figure 6. Results of connectivity metrics for the different stress tests at all measurement locations with zero leakage and ZWL days, with a) duration, b) and d) frequency and c) timing. For a detailed description of the metrics we refer to Table 1. Stress tests are displayed in Figure 2. Dashed lines denote, where zero leakage and ZWL days are equally important.

3.3 The relationship of vertical and longitudinal connectivity

For the duration metric δt, values above 0 (below 0) indicate, that the duration of the longest zero leakage event t_{0leak} is greater (smaller) than the duration of the longest zero water level event t_{zwl}. Mostly, t_{0leak} is greater than t_{zwl} (Figure 6 a)). For stress-tests without wells, the ratio is closer to 0 for E6 and W2. For E8, no zero leakage was simulated in these scenarios. In all stress tests with wells δt increases with zero leakage becoming more important at E8 and E6. At E4, δt is more heterogeneous, less influenced by well withdrawals but without a clear response to recharge scenarios either. However, at E2, δt does not vary significantly among scenarios. E2 is also the last station towards the upstream end of the tributary. At IB1, no difference is observed in comparison to the stress tests without wells. δt at W2 and RO2A is more responsive to the recharge stress tests as zero leakage does not occur at these locations in the reference simulation.

The frequency metric ω differ only slightly for the recession and rewetting phases (Figure 6 b) and d)). For stress tests without wells, almost no zero leakage occurs during the recession and rewetting phase. Hereby, it should be noted that the observations don't show ZWL days in the recession phase for some locations. However, at locations such as E8 and E6, where ZWL days occur in the recession phase (Figure 5), zero leakage already occurs much earlier and thus, no ω can be defined. Nevertheless, some differences can be analysed. At E6 ω has the same magnitude for the rewetting and recession phase. At E8, ω is significantly lower with wells included for the rewetting phase in comparison to the recession phase pointing towards a stronger importance of ZWL days in this phase (as leakage might switch back to gaining conditions earlier). However, the other stress tests with wells show a higher ω during rewetting than during recession. The timing metric, i.e. time delay between first zero leakage and first ZWL day, provides further information about whether ZWL occurs directly in response to zero leakage or not (Figure 6 c)). For scenarios without wells, no zero leakage occurred before ZWL in 2020 at all stations except E4 and IB1 in the reference simulation with Δt being particular high at IB1. Zero leakage occurs before ZWL at RO2A in the reference and wells simulations. In the scenarios with wells Δt increases for almost all stations (excluding W2). The presence of wells therefore seems to result in a time shift of zero leakage appearing earlier than in the reference simulation and also earlier than ZWL. Interestingly, Δt does not change at E4 for all the stress tests with wells, but it increases for the recharge stress tests. The latter points towards zero leakage occurring earlier under dry recharge conditions. As for δt, Δt at RO2A is only changed by recharge stress tests but the magnitude of Δt does not change among them. At W2 no result is found for Δt as there is rarely any zero leakage (Figure A5).

4 Discussion

4.1 Model evaluation and uncertainties

The application of the complex distributed groundwater-model in this study comes along with some substantial sources of uncertainty. Firstly, the model has multiple parameters and processes implemented that can influence the same hydrological variables. Hence, equifinality prevents finding a single 'realistic' parametrisation, but multiple parameter sets will be able to depict observations reasonably well. Secondly, observational data used for model evaluation are only partially comparable to

model output. While data on groundwater heads or surface water levels are point measurements, model outputs integrate flows and processes for the area of the model grid cell. Particularly for high model resolutions and or strong (hydraulic) gradients in one cell, 'correct' model outputs can differ substantially from 'correct' observational data taken in the area of that cell. Finally, none of the simulations incorporates all human influences in the stream network and in the groundwater, as noted earlier. As often for complex models such as that used in this study, computation time hampers an automated formal calibration. Similarly, a sensitivity analysis that might address the parameter uncertainty and ideally would be based on the full parameter space, has so far not been possible, in part also because if invested, it would also have to include an online coupling between the recharge model ROGER and the Modflow model, which again demands resources. Reduced model run times are the main advantage of offline-coupled models (Condon et al., 2021).

As a result of the limitations differences between observed and simulated ZWL days are substantial at some locations. These differences reflect the simplifications made by the models. Most important to name are the conversion of GW heads from the scale of a grid cell to a river reach unit and the uncertainties in riverbed parameterisation. Other studies have found such difficulties before. (Brunner et al., 2010) showed, that coarse resolutions of the Modflow GW model can lead to underestimation of infiltration in losing streams, which results in lower GW head and thus also influences streamflow results. Increasing model resolution may reduce this problem (Mehl and Hill, 2010). Strong topographic gradients and coarse resolution do not only impact GW heads but also result in distortion of hydrogeologic properties (Foster et al., 2020; Fleckenstein et al., 2006). This might affect simulated hillslope contributions, which this study did not focus on with high detail. Hillslope contributions that are not correctly represented might explain the poorly modeled ZWL at some of the upstream locations (for example IB1). Additional observations on hillslope connectivity would be necessary to identify where the model misses such inflows into the main tributaries from the headwaters or the hillslopes. In general, future studies my need to assess the influence of spatial resolution on modeled water levels and streamflow (e.g. by decreasing the GW grid cell size) in order to use this model for a quantification of surface water availability. Also, the riverbed parameterisation depends on the available raster data and human structures such as weirs and bridges have not been considered in the parameterisation, which may lead to individual affected locations not being well represented. Overall, the spatial discretisation of stream reaches likely does not represent the small-scale heterogeneity of streambed. In order to estimate which streambed parameter has the largest influence on the model results, a targeted sensitivity analysis might be helpful. Another option to reduce parameterisation errors would be to calibrate hydrodynamic parameters based on hydrodynamic models (Quan et al., 2020) or conceptual models (Meert et al., 2018; Vermuyten et al., 2018) instead of relying on available gridded data. Such additions to the modelling concept were beyond the scope of this study, but may present useful extensions in future work.

Model outputs need to be interpreted accordingly, i.e. with those limitations in mind. Simulations at specific locations can be used for relative comparisons but should not directly be compared to real world observations similar to a validation. The first research question addressed to what degree the model setup would be able to simulate the dynamics or patterns of connectivity and consequentially the river bed drying in the Dreisam and its tributaries. The question was raised by the experimental zero level data from the dry summer of 2020 that had revealed a high variability in longitudinal connectivity ((Herzog et al., 2022)). Overall, the model did simulate ZWL days well for locations in the valley bottom and it distinguished between upstream to

downstream increase or decrease in ZWL days in different tributaries. An analysis of longitudinal and vertical connectivity based on the model results allowed to distinguish between predominantly gaining and losing stream reaches in general to explain these patterns. With the application of stress tests as it is done in this study the more general and relative system
380 response could be analysed which is more robust than absolute numeric outputs. The second objective of this study was to obtain responses of the spatio-temporal dynamics of GW-SW-interaction to different stresses. Despite the uncertainties with respect to simulated system the model setup provided confidence for the representation of the correct relative responses to changed input or conditions.

4.2 Response of groundwater-surface water interactions to stress tests

385 The second research question addressed the response and changes of modelled connectivities in response to the applied stress test scenarios. Following the findings of Hellwig et al. (2021), who found that baseflow reacts on shorter timescales to intensified drought events (especially in fast reacting systems), one would have expected drier recharge preconditions to modify leakage flow. In the stress test simulations, recharge stress mimicking the exacerbation of climatic situations, did not respond very strongly, neither for dry nor for wet GW conditions compared to the reference simulation (Figure 3). Leakage responses
390 were highest for dry GW conditions and the stress tests with direct water abstractions - although with variability among locations. Nevertheless, this suggests, that anthropogenic activities exert large influences on GW-SW-interaction on such short time scales that may exceed climatic conditions. Locally, this concerns the occurrence of zero leakage in specific parts of the stream network. Hereby, the critical distance to the wells should be further investigated, for example by means of using stream proximity criteria (Zipper et al., 2019; Li et al., 2022). In our simulations, the well locations and depths were based on observations.
395 However, synthetic stress tests with equally distributed wells and different withdrawal rates could also be envisioned to obtain a general understanding of the impact of water withdrawals. This could help to answer practical questions, such as where GW contributions may buffer the effects of well withdrawals or for the definition of critical withdrawal rates. Critical withdrawal rates could also be investigated, taking seasonality of water withdrawals into account, for example to compare stress tests with more and no water withdrawals in summer.
400 The stronger impact of well withdrawals on vertical connectivity in comparison to recharge stress may not be valid for longer timescales as particularly GW is a slow reacting system (Cuthbert et al., 2019). One has to note also, that we designed the stress tests to account for current climatic trends of dry recharge years and dry recharge winters to occur more often and not for climatic extremes. We also did not evaluate the effect of longer durations (more than two years) of dry recharge conditions, which have been shown to increase the durations of streamflow droughts in rivers from catchments with porous aquifers (Stoel-
405 zle et al., 2014). For the two recharge stress scenarios (mag and seas), the interpretation of changes induced by the magnitude scenario is more complex in comparison to the seasonality stress test because the whole recharge sequence was modified.

By evaluating leakage only for the driest and the wettest GW conditions in the study period, we obtain an overview on how leakage differs in the extreme (GW) situation. While this may be particularly relevant for water management, the understanding of how vertical connectivity evolves in response to GW could be improved by an additional event-based analysis (e.g. looking
410 at whole periods of relatively deep (or low) GW head) or an analysis of the temporal evolution of leakage during changing

GW conditions. These synergies between connectivity changes and drought imply furthermore, that research on IRES and on drought should not be independent, as elaborated e.g. by (Yildirim and Aksoy, 2022).

While the influence of recharge stress tests on leakage was low, they did have an impact on the timing of the occurrence of dry GW conditions. With dry winter preconditions in 2020, the driest GW conditions occurred in late august 2020 and with every
415 second year being the driest modeled recharge year between 2009 and 2020, driest GW conditions occurred in late August 2017, as compared to late August 2018 in the reference simulation. On an interannual basis, GW heads were generally lowest in the end of summer, which is in agreement with trends found in other studies based on climate scenarios (Dams et al., 2012). This shift of dry GW conditions due to climatic stress could also lead to a shift in timing of zero leakage in general. However, the timing of dry GW conditions could also be caused by the interplay of different hydrologic variables in specific years, lead-
420 ing to more or less resilience of the GW system to climatic influences. Assessing this further was not the central interest in this study that focused on 2020 due to the available observations. A broadened understanding on vertical connectivity changes due to climatic stress can likely only be achieved through a multi year analysis of the effects of possible more severe climatic stress tests.

4.3 Appraisal of connectivity metrics

425 The third research question addressed the use of metrics that may help assess vertical and longitudinal connectivities - in models or observations. The metrics as described in Table 1 allow a joint analysis of longitudinal and vertical connectivity that for example confirms the impact of water withdrawals in a specific part of the area as in Fig. 3 but they also allow to demarcate locations that are likely more responsive to climatic (recharge) impacts. The investigation of δ t shows, that zero leakage persists for longer durations than ZWL for almost all stations (except the ones experiencing very few zero leakage),
430 supporting that ZWL days do not occur independently of zero leakage. δ t, Δ t and ω along the downstream E tributary stations are influenced by the presence of wells. RO2A and IB1 are most influenced by climatic preconditions, which was shown to have an impact on δ t (indicating that zero leakage appears) and Δ t (only for RO2A) but not on ω on the other hand. This different behavior underlines that the relationship of vertical and longitudinal connectivity is very site specific. W2 and IB1 are both relatively far from the outlet but IB2 is experiencing a lot and W2 very few zero leakage. These differences can only be
435 expressed looking at multiple temporal characteristics, which highlights the value of the use of such metrics. In this study, zero leakage was mostly found at stations characterised by predominantly gaining conditions (see also Figure A5). The findings of this study are therefore not necessarily valid for the spatio-temporal relationship between vertical and longitudinal connectivity in losing stream sections. Differences among losing and gaining streams in their connectivity relationship therefore requires more research in different catchment context and with larger samples.

440 Additionally, one also has to be aware of the limits of the metrics due to their definition. First of all, the metrics can only be used for streams experiencing dry spells and zero leakage. The metric for timing could not be calculated if zero leakage does not occur in the same year (stations without bars in Fig. 6 c)). Considering longer time spans, infinite or very large values would appear for location where the time delay between changes in vertical connectivity and longitudinal connectivity is large. But such large values are difficult to display and imply that there is no link between vertical and longitudinal connectivity changes

445   anyway. Furthermore, the frequency metric is difficult to assess when looking at one year only. For a long-term analysis of the evolution of connectivity changes, the frequency metric would serve to compare the recession and rewetting phase in different years. However, this is problematic for analysing intraseasonal differences. The definition of recession and rewetting phase depends on available data (May-November 2020) and same number of days in each phase but could also be defined differently to actually take into account, that each station shows a different drying and rewetting pattern, which is not always happening at

450   the same time during the year. A possibility to adjust the frequency metric to intraseasonal timescales would be to extract the actual recession period at each station based on the hydrograph.

In general, the focus of this study was on understanding the spatio-temporal relationship of leakage flow and ZWL. It could be expanded further for example to seasonality if the focus wasn't only on the summer season. Further metrics might also use leakage quantities (magnitude, rate of change) if sufficient confidence existist in simulated leakage quantities without leakage

455   measurements for validation. An analysis of leakage flow quantities will also require more work to obtain the best simulation result.

5   Conclusions

This study presents a model based approach for assessing the connectivity along the entire stream network and tested metrics on combined longitudinal and vertical connectivity. The Dreisam River Valley in Southwest Germany that the approach was

460   tested on represents a typical case where recurrent hydrological drought events between 2015 and 2022 have led to interruptions of longitudinal connectivity in the stream network. Employing a numerical groundwater model allowed to confirm and fill in the knowledge gaps on surface water groundwater interaction along the river network that was previously based on a few observations of zero flows. .The evaluation of leakage flow for dry and wet GW conditions identifies parts of the catchment, with particularly strong variability of GW-SW-interaction and where GW-SW interaction ceases during dry conditions. This was

465   found for all the model simulations. However it has to be kept in mind that 'natural conditions' with no abstractions and an untrained river system have not existed in this catchment for more than 50 years. Hence, the reference cannot be validated and the closest integration of human influences only considers a portion of them, i.e. those that are available quantitively..While confidence was obtained in the model's abaility to simulate relative patterns, uncertainties in modeled leakage flows are still high. In order to enhance the soundness of such models we suggest a broader sensitivity analysis, which includes stress tests, parameter

470   uncertainty and model resolution as a follow-up to our work. However, an analysis of metrics describing the relationship of zero leakage and ZWL (duration, timing and frequency) helped to disentangle the spatio-temporal relationship of zero leakage and ZWL at specific locations. Combined analysis of longitudinal and vertical connectivity is therefore a promising approach. To verify if such metrics are transferable to other contexts, additional case studies in various hydrogeological settings as well as on different spatial scales will be needed. A reference model simulation assumed near natural conditions. Stresstest scenario

475   model runs then imposed either an altered recharge regime or a set of introduced groundwater abstraction wells or both.The stress tests showed, that GW withdrawals affect leakage possibly more stongly than the recharge stresses employed. In our study area, well withdrawals influence the intraseasonal relationship of longitudinal and vertical connectivity (duration, timing

and frequency), in a way that zero leakage becomes more dominant and thus, vertical connectivity decreases. Seasonality of the drying may be different for other catchments, which calls for additional studies considering different timescales, especially not only seasonal but also annual drying dynamics. Future analysis should also focus on zero flow in addition to ZWL and not only on zero leakage but on all types of leakage directions and on multi-annual datasets. The magnitude of the decrease in vertical connectivity is possibly influenced by the distance to the wells but in general, our analysis showed that connectivity fluctuations increase and are exacerbated during the recession and rewetting phase in a specific part of the stream network. However, the combined analysis of vertical and longitudinal metrics reveals, that the distance from the wells is not the only factor leading to zero leakage. For some locations (particularly upstream ones), climatic preconditions were the primary influencing factor. Overall, this shows the potential of the stress tests to disentangle climatic and human impact. Furthermore, the findings indicate, that changes of connectivity patterns in response to different types of stresses might differ depending on the location (upstream or downstream). This could be a starting point for future analysis of such connectivity differences between upstream and downstream locations. Apart from GW withdrawals, other human activities, such as urbanisation, soil sealing, land use changes, or water withdrawals due to irrigation, can also influence recharge, groundwater heads and interaction of groundwater and surface water. To assess other factors, model stress test approaches, such as presented in this study, can be adapted accordingly in the future. For example, the Modflow6 drain package can also be used to implement agricultural drains and other stresses, which potentially modify the GW head.The study introduces a framework for modelling stresstests and metrics for surface-groundwater interactionthat can be easily transferred to similar studies. The application also demonstrate that even if not all influences can be modelled such studies are useful locally to help inform a resilient management of water resources under multiple stresses.

Code and data availability. The Zero water level dataset is publically available (Herzog et al., 2021b). The streamflow dataset will be published.

Author contributions. Amelie Herzog was responsible for running the simulations, data treatment of the model outputs, implemented the design and analysis of the study as well as all figures and writing of the manuscript. She also provided data for validation of the model (zero water levels and streamflow). Jost Hellwig gathered the input data for the modflow model and performed initial runs with a prior Modflow5 version of the final Modflow6 model used in this study. Jost Hellwig and Kerstin Stahl contributed to the study design, revised drafts and helped to finalise the manuscript.

Competing interests. The authors declare no conflict of interest.

Financial support. This work was supported by the Badenova Fund for Innovation

Acknowledgements. We acknowledge the work of Max Schmit (and Hannes Leistert) for performing the RoGeR model runs used as input data for the Modflow6 model.

Appendix A
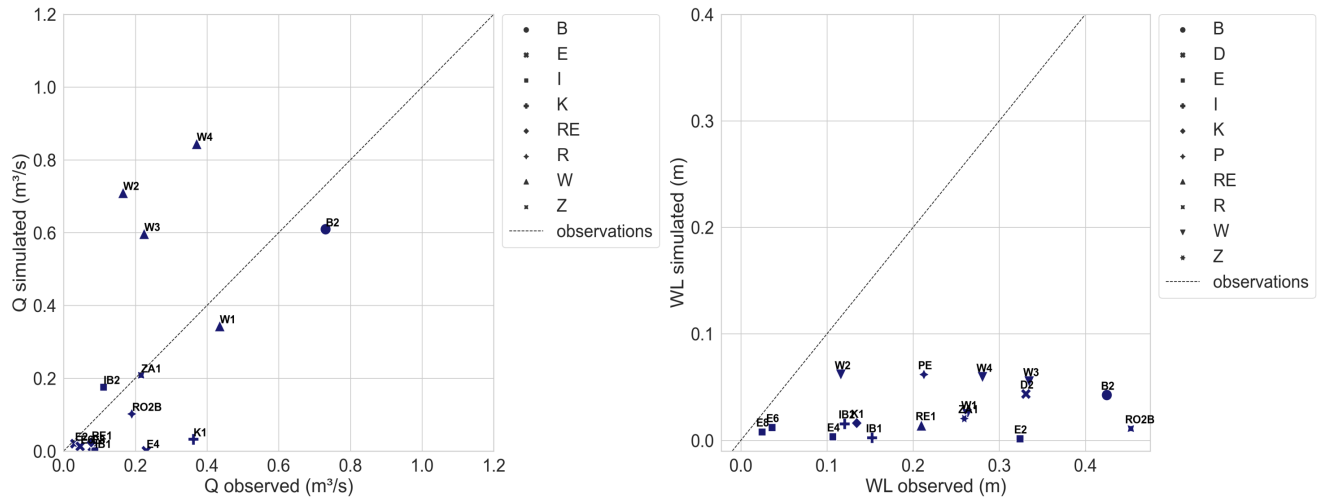
Figure A1. The simulated and observed streamflow and water levels in summer 2020
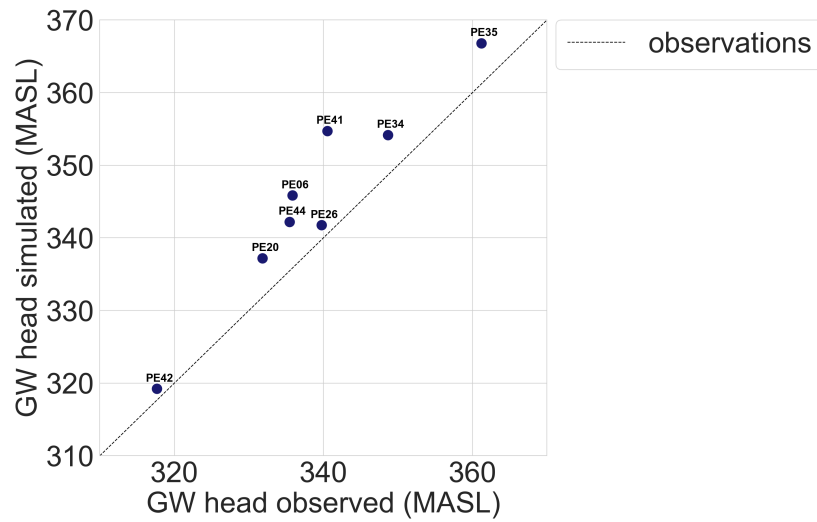
Figure A2. The simulated and observed GW head in summer 2020
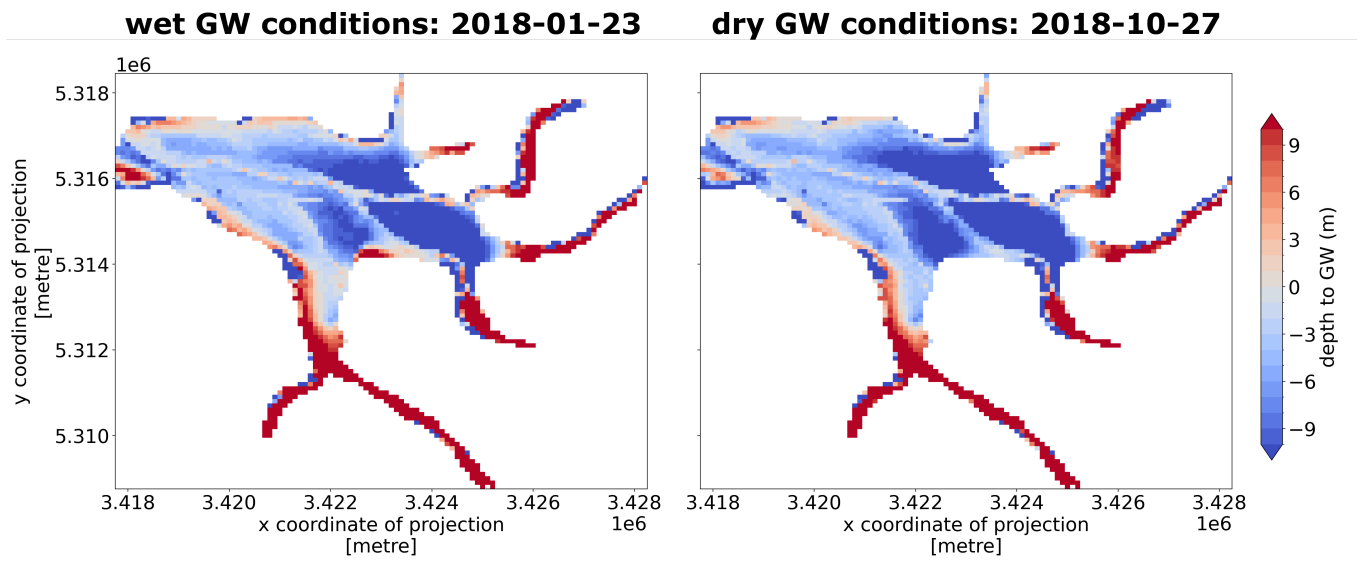
Figure A3. The simulated depth to GW for wettest GW conditions and driest GW conditions in the study period
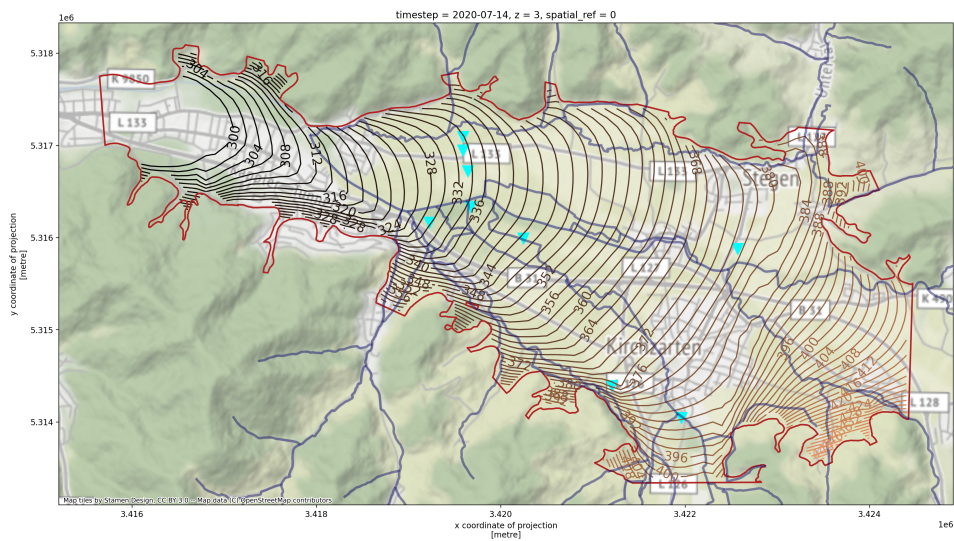
Figure A4. Simulated GW contour lines for the reference simulation for 14th of july 2020.
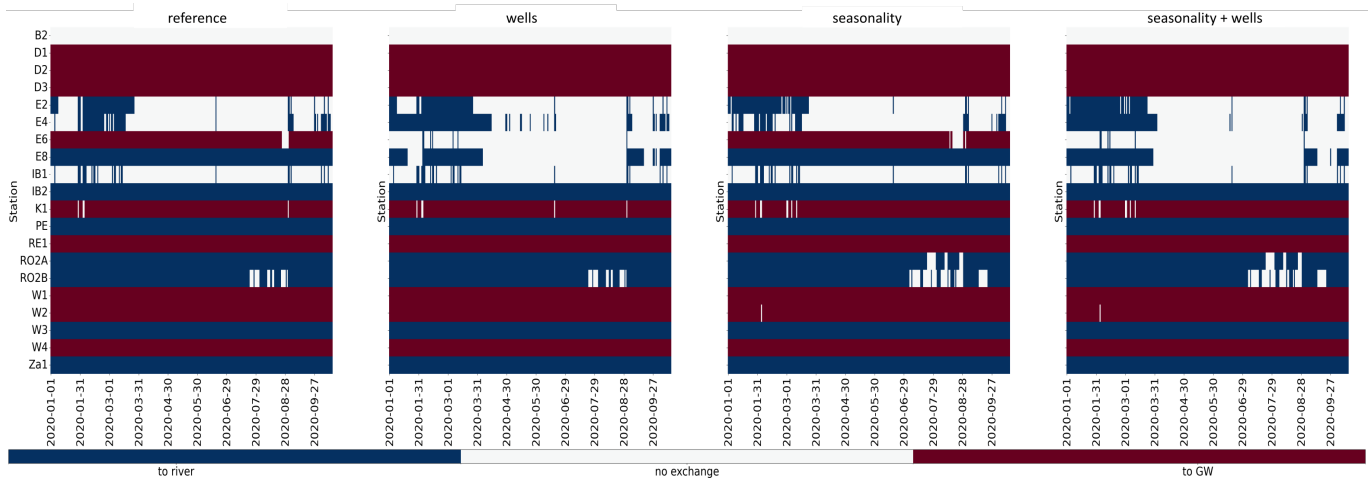
Figure A5. The extracted, normalized direction changes in 2020.
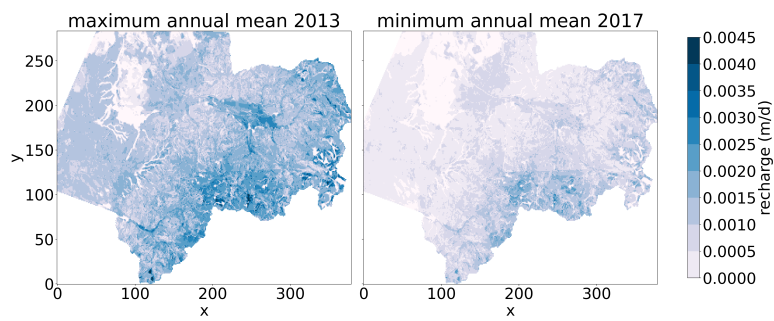
Figure A6. Mean annual minimum and maximum recharge water years for the period between 2009 and 2020 (RoGeR simulations)
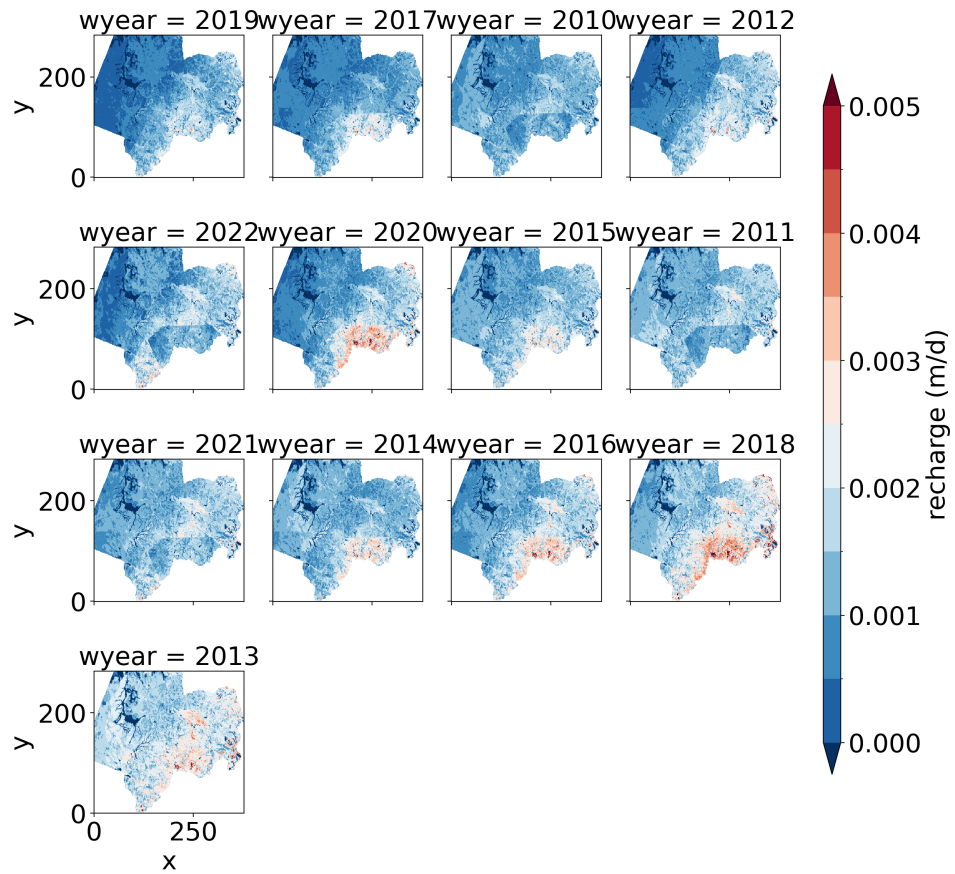
Figure A7. Mean annual recharge for the winter season between 2010 and 2020 (RoGeR simulations)

References

510 AghaKouchak, A., Mirchi, A., Madani, K., Di Baldassarre, G., Nazemi, A., Alborzi, A., Anjileli, H., Azarderakhsh, M., Chiang, F., Hassanzadeh, E., Huning, L. S., Mallakpour, I., Martinez, A., Mazdiyasni, O., Moftakhari, H., Norouzi, H., Sadegh, M., Sadeqi, D., van Loon, A. F., and Wanders, N.: Anthropogenic Drought: Definition, Challenges, and Opportunities, Reviews of Geophysics, 59, https://doi.org/10.1029/2019 RG000683, 2021.

Allen, D. C., Datry, T., Boersma, K. S., BOGAN, M. T., Boulton, A. J., Bruno, D., Busch, M. H., Costigan, K. H., Dodds, W. K., Fritz, K. M.,
515 Godsey, S. E., Jones, J. B., Kaletova, T., Kampf, S. K., Mims, M. C., Neeson, T. M., OLDEN, J. D., Pastor, A. V., POFF, N. L., Ruddell, B. L., Ruhi, A., Singer, G., Vezza, P., Ward, A. S., and Zimmer, M.: River ecosystem conceptual models and non–perennial rivers: A critical review, Wiley Interdisciplinary Reviews: Water, 7, e1473, https://doi.org/10.1002/wat2.1473, 2020.

Angermann, L., Krause, S., and Lewandowski, J.: Application of heat pulse injections for investigating shallow hyporheic flow in a lowland river, Water Resources Research, 48, https://doi.org/10.1029/2012 WR012564, 2012.

520 Assendelft, R. and van Meerveld, H.: A Low-Cost, Multi-Sensor System to Monitor Temporary Stream Dynamics in Mountainous Headwater Catchments, Sensors (Basel, Switzerland), 19, https://doi.org/10.3390/s19214645, 2019.

Barthel, R. and Banzhaf, S.: Groundwater and Surface Water Interaction at the Regional-scale – A Review with Focus on Regional Integrated Models, Water Resources Management, 30, 1–32, https://doi.org/10.1007/s11269-015-1163-z, 2016.

Belemtougri, Axel P. andDucharne, A., Tazen, F., Oudin, L., and Karambiri, H.: Understanding key factors controlling the duration of river
525 flow intermittency: Case of Burkina Faso in West Africa, Journal of Hydrology: Regional Studies, 37, 100 908, https://doi.org/10.1016/j.ejrh.2021.100908, 2021.

Bertrand, G., Siergieiev, D., Ala-Aho, P., and Rossi, P. M.: Environmental tracers and indicators bringing together groundwater, surface water and groundwater-dependent ecosystems: importance of scale in choosing relevant tools, Environmental Earth Sciences, 72, 813–827, https://doi.org/10.1007/s12665-013-3005-8, 2014.

530 Blume, T. and van Meerveld, H. I.: From hillslope to stream: methods to investigate subsurface connectivity, Wiley Interdisciplinary Reviews: Water, 2, 177–198, https://doi.org/10.1002/wat2.1071, 2015.

Botter, G. and Durighetto, N.: The Stream Length Duration Curve: A Tool for Characterizing the Time Variability of the Flowing Stream Length, Water Resources Research, 56, e2020WR027 282, https://doi.org/10.1029/2020 WR027282, 2020.

Botter, G., Vingiani, F., Senatore, A., Jensen, C., Weiler, M., McGuire, K., Mendicino, G., and Durighetto, N.: Hierarchical climate-driven
535 dynamics of the active channel length in temporary streams, Scientific Reports, 11, 21 503, https://doi.org/10.1038/s41598-021-00922-2, 2021.

Brunner, P., Simmons, C. T., Cook, P. G., and Therrien, R.: Modeling surface water-groundwater interaction with MODFLOW: some considerations, Groundwater, 48, 174–180, https://doi.org/10.1111/j.1745-6584.2009.00644.x, https://ngwa.onlinelibrary.wiley.com/doi/full/10.1111/j.1745-6584.2009.00644.x, 2010.

540 Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H.-J. H., Verhoef, A., van Loon, A. F., Sulis, M., and Abesser, C.: Global Groundwater Modeling and Monitoring: Opportunities and Challenges, Water Resources Research, 57, https://doi.org/10.1029/2020 WR029500, 2021.

Costigan, K. H., Jaeger, K. L., Goss, C. W., Fritz, K. M., and Goebel, P. C.: Understanding controls on flow permanence in intermittent rivers to aid ecological research: integrating meteorology, geology and land cover, Ecohydrology, 9, 1141–1153,
545 https://doi.org/10.1002/eco.1712, 2016.

Covino, T. P. and McGlynn, B. L.: Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry, Water Resources Research, 43, https://doi.org/10.1029/2006 WR005544, 2007.

Cuthbert, M. O., Gleeson, T., Moosdorf, N., Befus, K. M., Schneider, A., Hartmann, J., and Lehner, B.: Global patterns and dynamics of climate–groundwater interactions, Nature Climate Change, 9, 137–141, https://doi.org/10.1038/s41558-018-0386-4, 2019.

550   Czuba, J. A., David, S. R., Edmonds, D. A., and Ward, A. S.: Dynamics of Surface–Water Connectivity in a Low–Gradient Meandering River Floodplain, Water Resources Research, 55, 1849–1870, https://doi.org/10.1029/2018 WR023527, 2019.

Dams, J., Salvadore, E., van Daele, T., Ntegeka, V., Willems, P., and Batelaan, O.: Spatio-temporal impact of climate change on the groundwater system, Hydrology and Earth System Sciences, 16, 1517–1531, https://doi.org/10.5194/hess-16-1517-2012, 2012.

Datry, T., Bonada, N., and Boulton, A. J.: Intermittent rivers and ephemeral streams: Ecology and management, Academic Press, an imprint

555   of Elsevier, London, United Kingdom, 2017.

Etter, S., Strobl, B., van Meerveld, I., and Seibert, J. S.: Quality and timing of crowd–based water level class observations, Hydrological Processes, 34, 4365–4378, https://doi.org/10.1002/hyp.13864, 2020.

Fan, Y.: Are catchments leaky?, WIREs Water, 6, https://doi.org/10.1002/wat2.1386, 2019.

Ferrazzi, M. and Botter, G.: Contrasting signatures of distinct human water uses in regulated flow regimes, Environmental Research Com-

560   munications, 1, 071 003, https://doi.org/10.1088/2515-7620/ab3324, 2019.

Fleckenstein, J. H., Niswonger, R. G., and Fogg, G. E.: River-aquifer interactions, geologic heterogeneity, and low-flow management, Ground water, 44, 837–852, https://doi.org/10.1111/j.1745-6584.2006.00190.x, 2006.

Fleckenstein, J. H., Krause, S., Hannah, D. M., and Boano, F.: Groundwater-surface water interactions: New methods and models to improve understanding of processes and dynamics, Advances in Water Resources, 33, 1291–1295,

565   https://doi.org/10.1016/j.advwatres.2010.09.011, 2010.

Foster, L. M. and Maxwell, R. M.: Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions, Hydrological Processes, 33, 332–349, https://doi.org/10.1002/HYP.13327, 2018.

Foster, L. M., Williams, K. H., and Maxwell, R. M.: Resolution matters when modeling climate change in headwaters of the Colorado River, Environmental Research Letters, 15, 104 031, https://doi.org/10.1088/1748-9326/ aba77f, 2020.

570   Freeman, M. C., Pringle, C. M., and Jackson, C. R.: Hydrologic Connectivity and the Contribution of Stream Headwaters to Ecological Integrity at Regional Scales1, JAWRA Journal of the American Water Resources Association, 43, 5–14, https://doi.org/10.1111/j.1752-1688.2007.00002.x, 2007.

Fuchs, S., Ziesche, M., and Nillert, P.: Empirische Verfahren zur Ableitung verschiedener Porositätsarten aus Durchlässigkeitsbeiwert und Ungleichkörnigkeitszahl – ein Überblick, Grundwasser, 22, 83–101, https://doi.org/10.1007/s00767-017-0355-2, 2017.

575   Gallardo, B., Dolédec, S., Paillex, A., Arscott, D. B., Sheldon, F., Zilli, F., Mérigoux, S., Castella, E., and Comín, F. A.: Response of benthic macroinvertebrates to gradients in hydrological connectivity: a comparison of temperate, subtropical, Mediterranean and semiarid river floodplains, Freshwater Biology, 59, 630–648, https://doi.org/10.1111/fwb.12292, 2014.

Gallart, F., Prat, N., García-Roger, E. M., Latron, J., Rieradevall, M., Llorens, P., Barberá, G. G., Brito, D., de Girolamo, A. M., Lo Porto, A., Buffagni, A., Erba, S., Neves, R., Nikolaidis, N. P., Perrin, J. L., Querner, E. P., Quiñonero, J. M., Tournoud, M. G., Tzoraki, O.,

580   Skoulikidis, N., Gómez, R., Sánchez-Montoya, M. M., and Froebrich, J.: A novel approach to analysing the regimes of temporary streams in relation to their controls on the composition and structure of aquatic biota, Hydrology and Earth System Sciences, 16, 3165–3182, https://doi.org/10.5194/hess-16-3165-2012, 2012.

Gallart, F., Cid, N., Latron, J., Llorens, P., Bonada, N., Jeuffroy, J., Jiménez-Argudo, S.-M., Vega, R.-M., Solà, C., Soria, M., Bardina, M., Hernández-Casahuga, A.-J., Fidalgo, A., Estrela, T., Munné, A., and Prat, N.: TREHS: An open-access software tool for investigating and evaluating temporary river regimes as a first step for their ecological status assessment, The Science of the total environment, 607-608, 519–540, https://doi.org/10.1016/j. scitotenv.2017.06.209, 2017.

Godsey, S. E. and Kirchner, J. W.: Dynamic, discontinuous stream networks: hydrologically driven variations in active drainage density, flowing channels and stream order, Hydrological Processes, 28, 5791–5803, https://doi.org/10.1002/hyp.10310, 2014.

Goodrich, D. C., Kepner, W. G., Levick, L. R., and Wigington, P. J.: Southwestern Intermittent and Ephemeral Stream Connectivity, JAWRA Journal of the American Water Resources Association, 54, 400–422, https://doi.org/10.1111/1752-1688.12636, 2018.

Hammond, J. C., Zimmer, M., Shanafield, M., Kaiser, K., Godsey, S. E., Mims, M. C., Zipper, S. C., Burrows, R. M., Kampf, S. K., Dodds, W., Jones, C. N., Krabbenhoft, C. A., Boersma, K. S., Datry, T., OLDEN, J. D., Allen, G. H., Price, A. N., Costigan, K., Hale, R., Ward, A. S., and Allen, D. C.: Spatial Patterns and Drivers of Nonperennial Flow Regimes in the Contiguous United States, Geophysical Research Letters, 48, https://doi.org/10.1029/2020 GL090794, 2021.

Hellwig, J., Stoelzle, M., and Stahl, K.: Groundwater and baseflow drought responses to synthetic recharge stress tests, Hydrology and Earth System Sciences, 25, 1053–1068, https://doi.org/10.5194/hess-25-1053-2021, 2021.

Herzog, A., Hector, B., Cohard, J.-M., Lawson, F. M. A., Peugeot, C., and de Graaf, I.: A parametric sensitivity analysis for prioritizing knowledge needs for modeling water transfers in the West African critical zone, Vadose Zone Journal, https://doi.org/10.1002/vzj2.20163, 2021a.

Herzog, A., Stahl, K., Blauhut, V., and Weiler, M.: Water Level Dataset Dreisam Valley V1.0, https://doi.org/10.6094/UNIFR/228702, 2021b.

Herzog, A., Stahl, K., Blauhut, V., and Weiler, M.: Measuring zero water level in stream reaches: A comparison of an image–based versus a conventional method, Hydrological Processes, 36, https://doi.org/10.1002/hyp.14658, 2022.

Huntington, J. L. and Niswonger, R. G.: Role of surface-water and groundwater interactions on projected summertime streamflow in snow dominated regions: An integrated modeling approach, Water Resources Research, 48, 303, https://doi.org/10.1029/2012 WR012319, 2012.

Jaeger, K. L. and Olden, J. D.: Electrical resistance sensor arrays as a means to quantify longitudinal connectivity of rivers, River Research and Applications, 28, 1843–1852, https://doi.org/10.1002/rra.1554, 2012.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, Water Resources Research, 45, https://doi.org/10.1029/2008 WR007225, 2009.

Jensen, C. K., McGuire, K. J., Shao, Y., and Andrew Dolloff, C.: Modeling wet headwater stream networks across multiple flow conditions in the Appalachian Highlands, Earth Surface Processes and Landforms, 43, 2762–2778, https://doi.org/10.1002/esp.4431, 2018.

Kalbus, E., Reinstorf, F., and Schirmer, M.: Measuring methods for groundwater – surface water interactions: a review, Hydrology and Earth System Sciences, 10, 873–887, https://doi.org/10.5194/hess-10-873-2006, 2006.

Käser, D. and Hunkeler, D.: Contribution of alluvial groundwater to the outflow of mountainous catchments, Water Resources Research, 52, 680–697, https://doi.org/10.1002/2014 WR016730, 2016.

Langevin, C. D., Hughes, J. D., Banta, E. R., Niswonger, R. G., Panday, S., and Provost, A. M.: Documentation for the MODFLOW 6 Groundwater Flow Model, https://doi.org/10.3133/tm6A55, 2017.

Li, Q., Gleeson, T., Zipper, S. C., and Kerr, B.: Too Many Streams and Not Enough Time or Money? Analytical Depletion Functions for Streamflow Depletion Estimates, Ground water, 60, 145–155, https://doi.org/10.1111/gwat.13124, 2022.

620   M. Rinderer, H. J. van Meerveld, and B. L. McGlynn: From Points to Patterns: Using Groundwater Time Series Clustering to In-
       vestigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics, Water Resources Research, 55, 5784–5806,
       https://doi.org/10.1029/2018 WR023886, 2019.

       Magand, C., Alves, M. H., Calleja, E., Datry, T., Dörflinger, G., England, J., Gallart, F., Gomez, R., Jorda-Capdevila, D., Marti, E., Munne,
       A., Pastor, A. V., Stubbington, R., Tziortzis, I., and von Schiller, D.: Intermittent rivers and ephemeral streams: what water managers need
625   to know, Zenodo, https://doi.org/10.5281/ZENODO.3888474, 2020.

       Marotz, G.: Technische Grundlagen einer Wasserspeicherung im natürlichen Untergrund, Schriftenreihe des KWH, p. 228 S., 1968.

       Meert, P., Pereira, F., and Willems, P.: Surrogate modeling-based calibration of hydrodynamic river model parameters, Journal of Hydro-
       environment Research, 19, 56–67, https://doi.org/10.1016/j. jher.2018.02.003, 2018.

       Meerveld, H. J. I., Sauquet, E., Gallart, F., Sefton, C., Seibert, J., and Bishop, K.: Aqua temporaria incognita, Hydrological Processes, 34,
630   5704–5711, https://doi.org/10.1002/hyp.13979, 2020.

       Mehl, S. and Hill, M. C.: Grid-size dependence of Cauchy boundary conditions used to simulate stream–aquifer interactions, Advances in
       Water Resources, 33, 430–442, https://doi.org/10.1016/j. advwatres.2010.01.008, 2010.

       Ott, B. and Uhlenbrook, S.: Quantifying the impact of land-use changes at the event and seasonal time scale using a process-oriented
       catchment model, Hydrology and Earth System Sciences, 8, 62–78, https://doi.org/10.5194/hess-8-62-2004, 2004.

635   Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M. C.,
       Henriksen, J., Jacobson, R. B., Kennen, J. G., Merritt, D. M., Oâkeeffe, J. H., Olden, J. D., Rogers, K., Tharme, R. E., and Warner, A.: The
       ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards, Freshwater
       Biology, 55, 147–170, https://doi.org/10.1111/j.1365-2427.2009.02204.x, 2010.

       Price, A. N., Jones, C. N., Hammond, J. C., Zimmer, M. A., and Zipper, S. C.: The Drying Regimes of Non–Perennial Rivers and Streams,
640   Geophysical Research Letters, 48, https://doi.org/10.1029/2021 GL093298, 2021.

       Pringle, C. M.: Hydrologic connectivity and the management of biological reserves: A global perspective, Ecological Applications, 11,
       981–998, https://doi.org/10.1890/1051-0761(2001)011[0981:HCATMO]2.0.CO;2, 2001.

       Quan, T. Q., Meert, P., Huysmans, M., and Willems, P.: On the importance of river hydrodynamics in simulating groundwater levels and
       baseflows, Hydrological Processes, 34, 1754–1767, https://doi.org/10.1002/hyp.13667, 2020.

645   Ruhi, A., Hwang, J., Devineni, N., Mukhopadhyay, S., Kumar, H., Comte, L., Worland, S., and Sankarasubramanian, A.: How Does Flow
       Alteration Propagate Across a Large, Highly Regulated Basin? Dam Attributes, Network Context, and Implications for Biodiversity,
       Earth's Future, 10, https://doi.org/10.1029/2021EF002490, 2022.

       Salwey, S., Coxon, G., Pianosi, F., Singer, M. B., and Hutton, C.: National–Scale Detection of Reservoir Impacts Through Hydrological Sig-
       natures, Water Resources Research, 59, e2022WR033 893, https://doi.org/10.1029/2022WR033893, https://agupubs.onlinelibrary.wiley.
650   com/doi/full/10.1029/2022WR033893, 2023.

       Schilling, O. S., Cook, P. G., Grierson, P. F., Dogramaci, S., and Simmons, C. T.: Controls on Interactions Between Surface Water, Ground-
       water, and Riverine Vegetation Along Intermittent Rivers and Ephemeral Streams in Arid Regions, Water Resources Research, 57,
       e2020WR028 429, https://doi.org/10.1029/2020 WR028429, 2021.

       Steinbrich, A., Leistert, H., and Weiler, M.: Model-based quantification of runoff generation processes at high spatial and temporal resolution,
655   Environmental Earth Sciences, 75, https://doi.org/10.1007/s12665-016-6234-9, 2016.

Steinbrich, A., Leistert, H., and Weiler, M.: RoGeR – ein bodenhydrologisches Modell für die Beantwortung einer Vielzahl hydrologischer Fragen, RoGeR – ein bodenhydrologisches Modell für die Beantwortung einer Vielzahl hydrologischer Fragen, 2021, 94–101, https://doi.org/10.3243/kwe2021.02.004, 2021.

Stoelzle, M., Stahl, K., Morhard, A., and Weiler, M.: Streamflow sensitivity to drought scenarios in catchments with different geology, Geophysical Research Letters, 41, 6174–6183, https://doi.org/10.1002/2014 GL061344, 2014.

Stoelzle, M., Staudinger, M., Stahl, K., and Weiler, M.: Stress testing as complement to climate scenarios: recharge scenarios to quantify streamflow drought sensitivity, Proceedings of the International Association of Hydrological Sciences, 383, 43–50, https://doi.org/10.5194/piahs-383-43-2020, 2020.

Strobl, B., Etter, S., van Meerveld, I., and Seibert, J.: Accuracy of crowdsourced streamflow and stream level class estimates, Hydrological Sciences Journal, 65, 823–841, https://doi.org/10.1080/02626667.2019.1578966, 2020.

Tijdeman, E., Hannaford, J., and Stahl, K.: Human influences on streamflow drought characteristics in England and Wales, Hydrology and Earth System Sciences, 22, 1051–1064, https://doi.org/10.5194/hess-22-1051-2018, 2018.

Vermuyten, E., Meert, P., Wolfs, V., and Willems, P.: Combining Model Predictive Control with a Reduced Genetic Algorithm for Real-Time Flood Control, Journal of Water Resources Planning and Management, 144, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000859, 2018.

Wirsing, G. and Luz, A.: Hydrogeologischer Bau und Aquifereigenschaften der Lockergesteine im Oberrheingraben (Baden-Württemberg), LGRB Informationen, 19, 1–130, 2007.

Xu, H., van der Steeg, S., Sullivan, J., Shelley, D., Cely, J. E., Viparelli, E., Lakshmi, V., and Torres, R.: Intermittent Channel Systems of a Low–Relief, Low–Gradient Floodplain: Comparison of Automatic Extraction Methods, Water Resources Research, 56, https://doi.org/10.1029/2020 WR027603, 2020.

Yildirim, I. and Aksoy, H.: Intermittency as an indicator of drought in streamflow and groundwater, Hydrological Processes, 36, https://doi.org/10.1002/hyp.14615, 2022.

Zanetti, F., Durighetto, N., Vingiani, F., and Botter, G.: Analysing river network dynamics and active length - discharge relationship using water presence sensors, Hydrology and Earth System Sciences Discussions, pp. 1–27, https://doi.org/10.5194/hess-2021-103, 2021.

Zimmer, M. A. and McGlynn, B. L.: Ephemeral and intermittent runoff generation processes in a low relief, highly weathered catchment, Water Resources Research, 53, 7055–7077, https://doi.org/10.1002/2016 WR019742, 2017.

Zimmer, M. A. and McGlynn, B. L.: Lateral, Vertical, and Longitudinal Source Area Connectivity Drive Runoff and Carbon Export Across Watershed Scales, Water Resources Research, 54, 1576–1598, https://doi.org/10.1002/2017 WR021718, 2018.

Zipper, S. C., Gleeson, T., Kerr, B., Howard, J. K., Rohde, M. M., Carah, J., and Zimmerman, J.: Rapid and Accurate Estimates of Streamflow Depletion Caused by Groundwater Pumping Using Analytical Depletion Functions, Water Resources Research, 55, 5807–5829, https://doi.org/10.1029/2018 WR024403, 2019.

Zuecco, G., Rinderer, M., Penna, D., Borga, M., and van Meerveld, H. J.: Quantification of subsurface hydrologic connectivity in four headwater catchments using graph theory, The Science of the total environment, 646, 1265–1280, https://doi.org/10.1016/j. scitotenv.2018.07.269, 2019.