The impact of model choice on the quantification of seasonal hyporheic exchange depths using a multi-model approach on a long time series

Enter authors here: Lara-Maria Schmitgen[1], Reinhard Bierl[1], and Tobias Schuetz[1]

[1] Department of Hydrology, Regional and Environmental Sciences, University of Trier, Trier, 54296, Germany

Corresponding author: Lara-Maria Schmitgen (lara.schmitgen@gmail.com)

Key Points:

- Temperature-based models result in higher magnitudes of seasonality and higher proportions of upward flux conditions than Darcy flux
- Seasonal variations of hyporheic exchange fluxes and depths have a larger extent than episodic variations
- Distinctive impact of model choice on hyporheic exchange flux modelling regarding parameter sensitivity and model performance

Abstract

Direction and depths of hyporheic exchange fluxes at the groundwater surface water interface are a driver of biogeochemical processes influencing e. g. nutrient supply and water quality. Here, we quantify seasonal and episodic variations of hyporheic exchange fluxes and hyporheic exchange depths with methods of heat tracing. Numerically (FLUX-BOT) and analytically (VFLUX) working program scripts were used to solve the one-dimensional conduction-advection-dispersion equation and compute hyporheic flux rates from vertical sediment water temperature profiles recorded continuously in a small low mountain creek between 2011 and 2017. Based on vertical hydraulic gradients Darcy exchange fluxes were used as a benchmark. Volumetric heat capacity and thermal conductivity of the sediment-water system were identified as the most influential thermal properties for both modelling approaches and thus chosen for a detailed analysis. By comparing the behavior of the two different water temperature-based modelling approaches, dissimilarities in sensitivity to sediment thermal properties and deviating optima for both parameters were found. In general, using the analytical model achieved higher Kling-Gupta-Efficiencies with regard to similarity with the benchmark flux. The differences in parameter responsivity can explain deviating performances of the models under several boundary conditions. We show that the extension of hyporheic exchange depth has a distinctive temporal variability, which is strongly influenced by seasonal effects but also by the chosen model. Furthermore, surface water levels, groundwater levels and stream discharges have a significant effect on flux direction and hyporheic zone extension, whereas model performance for both model types depended on air temperatures as well.

1 Introduction

The hyporheic zone as the interface between groundwater and surface waters (Orghidan, 1959) has a strong impact on the overall water quality (Winter et al., 1999). This active ecotone harbors many different biogeochemical processes from sediment to catchment scale (Boulton et al., 1998) such as carbon, energy and nutrient cycling as well as contaminant transport and removal by the vast community of organisms residing there (Buss et al., 2009). The mixing between waters of different origin and with different chemical composition in the hyporheic zone enhances chemical reactions and microbial diversity and activity and thus contributes largely to the self-purification of river systems (Boulton et al., 2008; Fanelli & Lautz, 2008). This importance is limited by its vertical extension, which can vary seasonally (Boano et al., 2014; Wondzell & Swanson, 1996), and is depending on the intensity and direction of hyporheic exchange fluxes (Boulton et al., 2010). Hyporheic exchange flux is the ensemble of flows entering and exiting the hyporheic zone. Boano et al. (2014) name two main drivers for these fluxes: hydrostatically influenced flow and hydrodynamically driven flow. According to Boano et al. (2014), the first is caused by pressure head gradients between surface water and groundwater and has its largest influence on the spatial scale of riffle-pool-sequences, steps, cascades or meandering banks, whereas the latter is controlled by momentum transfer dependent on the roughness of the river bed and has therefore greater effect on finer scales smaller than stream depth. Hyporheic zone depths are highly variable over time and space (Wondzell & Swanson, 1996; Wondzell & Swanson, 1999) and values in the literature range from only a few centimeters (Harvey & Fuller, 1998; Hill & Lymburner, 1998; Singh et al., 2019) over variable depths between 10 cm and 30 cm (Boano et al., 2008) up to half a meter (Palmer et al., 1992; Gariglio et al., 2013). The vertical extension of the hyporheic zone in combination with sediment particle size distributions and the velocity of occurring exchange fluxes strongly influences ongoing matter turnover and retention processes (Boulton et al., 1998). A match between hyporheic retention time and the appropriate reaction time for various biogeochemical process can optimize contaminant attenuation (Herzog et al., 2018; Grant et al. 2014; Zarnetske et al., 2011).

Seasonal shifting (Allander. 2003) as well as episodic transition (McCallum & Shanafield, 2016; Trauth & Fleckenstein, 2017) between gaining and loosing conditions is often observable. However, many studies focus on short time periods spanning from several days (Wang et al., 2017; Lu et al., 2017; Bhaskar et al., 2012) over months (Briggs et al., 2012) to one full year (Gariglio et al., 2013; Birkel et al., 2016), where it is possible to detect episodic (flood pulse driven) events, but periodic (evapotranspiration driven) behavior, especially seasonal fluctuations stay

rather invisible. Hence, the impact of either episodic events or periodic behavior over a longer time series on direction and intensity of hyporheic exchange fluxes as well as on the depth extent of the hyporheic zone needs further research.

There are several methods to quantify groundwater – surface water interactions, e.g. direct measurement methods using seepage meters, methods based on Darcy's Law, mass balance approaches and temperature/heat tracing methods (an overview on available methods is given in the review paper of Kalbus et al., 2006). In contrast to methods using Darcy's Law to calculate water flow through porous media, heat-tracing methods must take conductive and convective transport into account (Stallman, 1965; Rau et al., 2014). The range of occurring variations in thermal conductivities is narrower than the one of hydraulic conductivity and it is not dependent on grain size distribution but on material characteristics alone (Stonestrom & Constantz, 2003), while the latter is, which gives heat tracing methods a great advantage over hydrometric methods (Rau et al., 2014). Heat-tracing methods are based on the interdependency of sediment water temperature depths profiles and water infiltration rates (Suzuki, 1960). The conductive and convective transport of heat in water-sediment-systems is described in the one-dimensional conduction-advection-dispersion equation, which can be used to quantify percolation rates in the surficial zone near the sediment water interface (Stallman, 1965).

There are numerous different numerical (Lapham, 1989; Munz & Schmidt, 2017; Rau et al., 2015; Silliman et al., 1993; Voytek et al., 2014) and analytical solutions (Hatch et al., 2006; Keery et al., 2007; Kurylyk & Irvine, 2016; Luce et al., 2013; McCallum et al., 2012) solving the conduction-advection-dispersion equation. While the analytical solutions need a sinusoidal temperature boundary signal, some numerical solutions can also incorporate arbitrary temperature boundaries (Munz & Schmidt, 2017; Silliman et al., 1993). However, numerical and analytical models both need, in addition to a time series of water temperature depth profiles, information on thermal and physical sediment properties. These physical properties of the sediment-water-system are seldomly quantified, e.g. in laboratory experiments (e.g. Wang et al., 2017). Commonly, they are estimated (coupled with uncertainty analysis) in many studies (e.g. Gariglio et al. 2013; Briggs et al., 2012; Lu et al., 2017). The uncertainty and the spatial heterogeneity of sediment thermal parameters remain as challenges (Rau et al. 2014) in any attempt to model processes in or the hyporheic zone itself. Due to the heterogeneity of streambed sediment, information on the system properties is not always directly inferable. The estimation of streambed properties with the use of literature values might lead to wrong choices and thus a biased simulation of hyporheic exchange fluxes (Hatch et al., 2006; Munz & Schmidt, 2017; Rau et al., 2010).

Here, we follow the recommendation by Boulton et al. (2010) to study the long-term variability of hydrological conditions occurring during several years, to represent seasonal variation and episodic hydrological extremes and their effect on hyporheic exchange depth under varying hydro-meteorological conditions. Parameter sensitivity analysis of a recently released numerical model (Munz & Schmidt, 2017) in comparison to a well-known analytical model approach (Hatch et al., 2006) is applied to:

I. identify the effect of estimated thermal properties on the results by evaluating parameter sensitivities for two contrasting exchange flux modelling approaches,
II. identify seasonal and short-term dynamics of hyporheic flow direction and intensity,
III. identify hyporheic exchange depth and its variation due to episodic boundary conditions or seasonal effects.

2 Materials and Methods

2.1 Study site and data

The Olewiger Bach is a tributary of the river Mosel and is located south of the city of Trier (Rhineland-Palatinate; Germany) with a catchment area of about 35 km² and a total length of 14 km (Krein & Schorer, 2000). The altitude difference between headwater and mouth is about 300 m. The study site is located at the lower reach at a height of approximately 170 m above sea level. The river has a pluvial regime with a mean discharge of 245 l s^{-1}. The study site includes a riffle-pool-sequence at a 50 m long straight part of the river with a mean width of 3 m. The sediment layer at the study site is made of coarse-grained low metamorphic schist and quartzite materials. The underlying aquifers are consisting of Devonian schist and quartzite deposits. Banzhaf and Scheytt (2009) reported a filtration coefficient (Kf) of $8*10^{-4}$ m s^{-1} using the grain curve method. Considering the method related procedures this value was converted, according to German National Association for Water Management, Waste Water and Waste Management (DWA) set of rules A 138 (DWA, 2005), to a Kf of $1.6*10^{-4}$ m s^{-1}.

Three temperature lances were positioned in succession on the riffle head with 7 m distance between the first and second lance and 5 m distance between the second and third. Hyporeic exchange fluxes and exchange depths were calculated for all lances, while lance 2 was chosen for an in-depth model evaluation and sensitivity analysis, to

exclude effects of spatial heterogeneity in sediment parameters and decimate computation time. Continuous measurements of sediment water temperature depth profiles were carried out using temperature lances (Umwelt- und Ingenieurtechnik (UIT) GmbH Dresden), with a resolution of 0.032 °C and an accuracy of 0.1 °C. Sensors were installed at depths of 2 cm, 5 cm, 10 cm, 15 cm, 25 cm, 45 cm and 65 cm on the 67 cm long temperature lance. The use of a temperature lance prevented sensors from varying their spatial location relative to each other, as this can result in large errors (Munz & Schmidt, 2017; Sebok et al., 2017). The temporal resolution of temperature measurements varied during the seven years between 1 min and 10 min. To receive an evenly spaced water temperature time series with 10 min intervals the original time series has been decimated. Surface water and groundwater levels were measured 50 m downstream using an Eijkelkamp CTD + BaroDiver (standard temporal resolution: 10 min) in a groundwater observation well (filter depth of 2 m) and a SEBAPLUS radar measurement system mounted on a bridge (standard temporal resolution: 15 min (2011-2014) or 5 min (2015-2017)). Groundwater level data showed a slight declining trend which was attributed to several subsidence events at the measurement location (which were caused by freezing and thawing) and was corrected by using linear regression. Surface water levels and river discharges were continuously measured at a radar gauging station 1 km downstream. Between the continuously measured water levels at the gauging station and the surface water levels observed regularly (weekly to monthly) right next to the groundwater observation well, an empirical linear relationship (R^2: 0.58, n: 150, p-value < 0.001) was established. Daily mean air temperature and daily precipitation sum were obtained from the nearby Petrisberg weather monitoring station (operated by: German Weather Service (DWD), published under DWDs Climate Data Center).

2.2 Modelling tools

The temperature-based modelling tools as well as the Darcy flux calculation all result in an estimation of vertical hyporheic flux (q) in m s^{-1}, which is distinct from the vertical fluid velocity (v_f) in m s^{-1}. While the first is a measure of volume per area per time, the second incorporates the porosity (n) and is a measure of distance per time (Gordon et al. 2012). Upward flux is denoted with a negative sign. If not mentioned otherwise calculated hyporheic exchange flux and depth values are given for lance 2.

2.2.1 Temperature-based hyporheic exchange flux modelling

Two different modelling approaches were used to simulate vertical hyporheic exchange fluxes based on water temperature time series observed in seven depths. Both modelling approaches incorporate the one-dimensional conduction-advection-dispersion equation (modified after Stallman, 1968):

$$\frac{\partial T}{\partial t} = K_e \frac{\partial^2 T}{\partial z^2} - q \frac{C_w}{C} * \frac{\partial T}{\partial z} \quad \textbf{(1)}$$

where t is time in s, T is water temperature in °C, z is depth in m, and C_w is volumetric heat capacity of water in J m^{-3} $°C^{-1}$. In addition to that, C is the volumetric heat capacity of the system in J m^{-3} $°C^{-1}$, which can be described by (modified after Stallman, 1968):

$$C = n * C_w + (1 - n) * C_s \quad \textbf{(2)}$$

where C_s is the volumetric heat capacity of the sediment in J m^{-3} $°C^{-1}$.
K_e is the effective thermal diffusivity in m^2 $s^{—1}$:

$$K_e = \frac{K}{C} \quad \textbf{(3)}$$

where K is the thermal conductivity of the saturated sediment in J s^{-1} m^{-1} $°C^{-1}$

2.2.1.1 Numerical model

A numerical solution to Stallman's equation is implemented in the inverse numerical computer program FLUX-BOT published by Munz and Schmidt (2017), where Equation 1 is solved under steady-state conditions by Bredehoeft and Papadopulos (1965):

$$\frac{T_z - T_0}{T_L - T_0} = \frac{exp^{\left(\frac{C_W}{K} * q * z\right)} - 1}{exp^{\left(\frac{C_W}{K} * q * L\right)} - 1} \quad \textbf{(4)}$$

where T_z is the water temperature at depth z in °C, T_0 is the water temperature at the uppermost sensor in °C, T_L is the water temperature at the lowermost sensor in °C, and L is the length of the vertical section in m.
Equation 1 is solved numerically using a finite difference method developed by Crank and Nicolson (1996) followed by the derivative-free Nelder-Mead simplex optimization method described by Lagarias et al. (1998) to minimize the error between calculated water temperature and measured water temperature (Schmidt et al., 2006, Munz & Schmidt, 2017). As input, a water temperature time series with depth information is needed as well as the thermal parameters: K, C and C_w. A time window of 1 day was chosen to exclude the diurnal temperature variation and artificial effects coming along with it (Munz & Schmidt, 2017). With this setting the model computes mean daily hyporheic exchange fluxes. Vertical flux is calculated between the uppermost and lowest temperature sensor depth given, using all the sensors in between for the numerical optimization process. At least three temperature measuring depths are needed for the model to run (Munz & Schmidt, 2017). In theory it should be possible to calculate 15 different vertical hyporheic flux time series using all available sensor combinations as boundaries and a varying number of sensors. However, due to a limitation in the numerical code, namely excluding combinations where the distance between the uppermost and lowermost sensor is smaller than the mean depth, only 14 combinations were realized.

2.2.1.2 Analytical model

Six different approaches to solve Stallman's conduction-advection-dispersion equation analytically using phase shift and amplitude dampening are incorporated in the MATLAB® toolbox VFLUX (Gordon et al., 2012; Irvine et al., 2015). A method based on amplitude dampening alone was chosen, since this approach leads to the most accurate results under transient flow conditions (Irvine et al., 2015). The method by Hatch et al. (2006) was chosen, since it offers the opportunity to perform sensitivity analysis (SA) on thermal dispersivity (β). Hatch et al (2006) used the following equation to compute vertical hyporheic flux:

$$q = \frac{C}{C_w} * \left(\frac{2K_e}{\Delta z} \ln A_r + \sqrt{\frac{a + v^2}{2}} \right) \quad \textbf{(5)}$$

where A_r is the amplitude ratio of corresponding water temperature measurements at different depths, v is thermal front velocity in m s^{-1} and a is defined as:

$$a = \sqrt{v^4 + \left(\frac{8\pi K_e}{P}\right)^2} \quad \textbf{(6)}$$

where P is the period of the sine temperature wave in s.
The original input variables for the amplitude method described by Hatch et al. (2006) in VFLUX are the thermal parameters: n, K, β, C_s and C_w; as well as a water temperature time series with corresponding depths (Gordon et al., 2012). Considering, that there was no depth integrative information on n or C_s available and that C_w can be assumed as a constant, C was used as a direct input parameter solving Equation 5 instead of calculating it using Equation 2, as it is normally implemented in VFLUX(2). This procedure allows parameter estimation with a reduced number of free parameters and results in a higher comparability between the numerical and the analytical modelling approach, as they now have identical parameter inputs. Only for one-at-time (OAT) sensitivity analysis (SA) all original input variables, described in (Gordon et al., 2012), were used to also assess the effect of variation in these parameters. The periodicity of the temperature signal was always set to 1 day to filter the diurnal signal. To improve filtering and prevent oversampling a sample size of 12 samples per day was applied, following the recommendations of Gordon et al. (2012). The dynamic harmonic regression function implemented in the Captain toolbox (Taylor et al., 2007) has been used to isolate the diurnal signal and extract the amplitude information to calculate A_r. For the calculation of one flux time series, two water temperature time series from different depths are needed as upper and lower boundaries. Vertical water flux was calculated between a sensor pair defined by a sensor spacing window. With

seven measurement depths, it is possible to compute 21 different hyporheic exchange flux time series with the analytical model for a single temperature lance using different temperature sensor depths. With the chosen sample size, vertical flux was calculated in 2-hour intervals and was resampled using linear interpolation to daily values to fit the numerically derived flux time series.

2.2.2 Darcy flux calculation

Applying Darcy's equation (1856):

$$q = Kf * i \quad \textbf{(7)}$$

with the hydraulic gradient i, which describes in this case the gradient between the surface water level and the groundwater level in the observation well, q was calculated using a 2 m depth integrating hydrometric method as reference to the temperature-based methods. The vertical Darcy flux time series was then resampled to a period of 1 day using linear interpolation to match the time steps of the analytically and the numerically simulated hyporheic exchange fluxes. Even if there is a high probability of deviations between the signals of vertical hydraulic heads and temperature-based water flux calculations (e. g. Krause et al., 2012) the results of the Darcy flux model are considered here as a benchmark (not as the true flux) which is used to compare the functioning and the results of two different modelling approaches and their sensitivity to parameter uncertainty.

2.3 Model evaluation and sensitivity analysis

To bypass errors in hyporheic flux calculation originating in uncertain thermal parameter assignment (Hatch et al., 2006; Munz & Schmidt, 2017; Rau et al., 2010) , OAT and all-at-a-time (AAT) SA, as classified by Pianosi et al. (2016), were applied with the time series of the sensor combinations with the highest similarity to the benchmark hyporheic flux. In OAT SA only one parameter is varied, while all other parameters are kept at a base value. A global and a local OAT SA were performed using the input parameter variances depicted in Table 1. AAT SA was done using Monte-Carlo-Analysis (MCA) (see below). Regional Sensitivity analysis (RSA) performed with the SAFE-toolbox by Pianosi et al. (2015), was used to identify the sensitivity of the output metric (KGE) to the parameters K and C. Therefore, the output metrics were split into a behavioral (KGE above 0) and a non-behavioral group (Saltelli et al. 2007). To assess the robustness bootstrapping with 1000 resamples and a significance level of 0.05 was performed. Afterwards the sensitivity was assessed by comparing the difference in cumulative distribution functions (CDFs) between the behavioral and non-behavioral data sets, using Kolmogorov-Smirnov statistics with their 95 % confidence interval. Reported, as a measure of sensitivity is the mean d-stat of this test, where a high value is associated with a higher sensitivity to this parameter.

MCA with 1000 model realizations (as recommended by Pianosi et al., 2016) was performed for each of the 21 possible depth combinations of all three lances for the analytical model and of lance 2 for the 14 combinations of the numerical model using randomly distributed parameter values for K and C according to Table 1. The parameter space was defined broader than the parameter ranges suggested by Lapham (1989) or Gordon et al (2012), to include all possible parameter specifications. Since C_w is quasi a constant and the effect of β is neglectable (see results of OAT SA), both were kept constant for the further analysis (Table 1). With regard to the Darcy reference flux, the fluxes with highest similarity were identified for all possible depth combinations per model. This was done using the Kling-Gupta-Efficiency (KGE) (Gupta et al., 2009) as a quality criterion. As described in detail by Gupta et al. (2009), the KGE is an improved version of the Nash-Sutcliff-Efficiency (NSE), which incorporates values for correlation, bias and variability to describe how well a modelled data set describes a measured data set. The bias is the ratio between the mean of the modelled data set and the mean of the measured data set, whereas the variability is described by the ratio of the standard deviations of the modelled data set and the measured data set. A KGE of 1 would be a perfect fit of the data. In this study, model performance is defined by the KGE between the modelled hyporheic exchange flux and the benchmark, meaning a high KGE value represents a good model performance, whereas a low or negative KGE is attributed with a poor model performance.

To assess the influence of K and C under specific boundary conditions, KGE values were calculated for selected characteristic short time periods in addition to the performance evaluation over the whole time series: namely a period of 6 months with a comparably good model performance including downward and upward vertical hyporheic flux behavior, a 3 month period with mainly downward vertical hyporheic flux behavior, a 3 month period with mainly upward vertical hyporheic flux behavior and short-term upward vertical hyporheic flux events (< 5 days)).

To quantify directly the influence of variation in C and K, on hyporheic flux intensity, 12 distinct values for K (set between the boundaries given in Table 1) were applied to calculate vertical hyporheic fluxes for a given day with

upward flux behavior and a given day with downward flux behavior, while all other inputs were kept constant. The same procedure was used for 12 distinct values for C, with all other parameters kept constant. The calculated vertical hyporheic exchange fluxes with differing values for either K or C were compared to the Darcy fluxes of these days as a benchmark.
Additionally, the hyporheic exchange flux simulations of both temperature-based approaches were compared against each other, to assess the similarity between the two best-fit fluxes as well as the similarity between the models if the same parameter values for K and C were used.

2.4 Process analysis

2.4.1 Identification of hyporheic exchange depths

The extent of the hyporheic zone can be described by the zero-flux plane as a boundary between upward and downward fluxes (Khalil et al. 2003), where vertical fluxes are closest to zero. Based on the works by Kim et al. (2013) and González-Pinzón et al. (2015) we define the approximated depth of the hyporheic zone, where the downwelling fluxes calculated between a given sensor combination turn into upwelling fluxes at the following lower sensor combination as the extent of the hyporheic zone. This so-called zero-flux method was applied to estimate hyporheic exchange depths by comparing the numerically and analytically generated results of hyporheic exchange flux calculations using 21 (analytical model at 3 lances) or 14 (analytical and numerical model at one lance) depth combinations. The parameter combinations of the model runs with the highest overall similarity to the benchmark were applied to exclude effects introduced by parameter choice.
Alternatively, mean hyporheic exchange depths were identified using a logistic regression model, to combine the depth integrated information from the Darcy benchmark flux with the depth resolved simulation results of the temperature-based models. Here the individual best fit to the benchmark for each depth combination was used. Obtained fluxes were categorized into a binary system; where 1 is marking upward flux behavior and 0 is representing downward flux. Very small values lower than (-)10^{-7} m s^{-1} are characterized as not assignable, because the method's errors induced by filtering are inside this boundary (Hatch et al. 2006). A binomial logistic model with either river discharges, surface water or groundwater levels as predictors (X) and the binary flux direction as response variable (Y) was fitted:

$$P(Y) = \frac{1}{1 + e^{-(b+\alpha X)}} \quad \textbf{(8)}$$

with the probability for upward flux P(Y), the predictor variable X and slope α as well as intercept b as regression coefficients.
The resulting shape of the model was compared to the reference logistic model, which was set up with flow directions from the Darcy flux time series as a depth integrated information. The lowest depth, where the slope α has the same direction as the reference model slope α, was identified as the mean exchange depth.

2.4.2 Temporal variation of hyporheic exchange fluxes and depths

Statistical location parameters (mean, median) and variation parameters (standard deviation, percent standard deviation, variance, range) were calculated to describe the time series of fluxes and depths. Furthermore, skewness and kurtosis were determined to further assess the distributions. Daily means (dm) and daily standard deviations (dstd) over the 7 years were calculated. The seasonal amplitude of these averaged years was used as an indicator for seasonal variation, whereas the range of the daily standard deviation suits as indicators for episodic variation. We define the episodic variation index as:

$$episodic\ variation\ index = 1 - \frac{\max(dm) - \min(dm)}{2 * mean\ (dstd)} \quad \textbf{(9)}$$

Following equation 9 episodic variation is higher than seasonal variation if the episodic variation parameter is positive.
In addition, the datasets were split into summer (April until September) and winter (October until March) periods to compare hyporheic exchange depths in different seasons. As calculated hyporheic exchange depth and flux data is

not in its entirety normally distributed (confirmed by a one sample Kolmogorov-Smirnov test (significance level alpha of 0.001)), non-parametric statistical tests were used to assess significant differences between the seasons and models. For comparison of two samples (e.g. summer and winter season) two sample Kolmogorov-Smirnov test and Wilcoxon-Mann-Whitney test were used, whereas Kruskal-Wallis test combined with Dunn's posthoc test was used to compare more than two samples (e.g. different models).

All time series were downsampled to monthly mean values for additional time series analysis. Additive time series decomposition in trend, seasonal (frequency of 12 month) and random component using local regression (loess) as described by Cleveland et al. (1990) implemented in the R function stl_plus (Hafen, 2016) was performed on the downsampled time series. The strength of the seasonal component was calculated according to Wang et al. 2006 using the R package tsfeatures (Hyndman et al., 2019). Seasonal strength is given as a value between 0 (no seasonal component) and 1 (very strong seasonal component). The strength of the seasonal component was estimated for calculated fluxes and resulting depths using the parameter optimum range (Figure 1) for each model, to assess the effect of slightly uncertain parameters on the modelling results. In addition, the same was done using the whole parameter range used in MCA (Table 1) to assess the influence of high parameter uncertainties.

To investigate the effects of water levels, discharge, precipitation and daily mean air temperature on hyporheic zone extension we categorized the hyporheic exchange depths into three expressions (< 0.2 m, 0.2 m to 0.4 m, > 0.4 m) and used them in a multinomial logistic regression (MNR) model for ordinal responses (Long & Freese, 2001; Dobson, 2002) with the named boundary conditions as predictors, to investigate the effects of these parameters on hyporheic zone extension. The effects of those boundary conditions on flux direction were identified using the binary logistic approach, due to the binary nature (upwards and downwards) of the predicted dependent variable.

3 Results

3.1 Model evaluation and sensitivity analysis

The effects of OAT SA are presented in Table 1. Global OAT SA indicates that in the analytical model C has a positive effect on KGE; higher C values result in an increased similarity between simulated hyporheic flux and the benchmark. A negative relationship appears when K is increased; KGE drops below 0 with high values of K. C_w has almost no effect on KGE. A slight increase in β also has no effect; only with very high β values a negative effect on KGE can be observed. Therefore, in the AAT SA C_w and β were kept at base value. Although, local OAT SA generally confirms these results (Table 1), there are only very small changes in KGE.

Global OAT SA for the numerical model resulted in unstable model runs with either very low values for K or very high C values. In contrast to the analytical model, high K values as well as low C values increase KGE. C_w has hardly any effect. Local OAT SA confirms these results (Table 1).

There is a large difference between both models regarding the optimum range for C and K (Figure 1). The analytical model has its optimum for parameter C starting at $3.1*10^6$ J m^{-3} $°C^{-1}$ and stopping moderately before the modelling boundary at $3.68*10^6$ J m-3 $°C^{-1}$. At the contrary, the optimum for C in numerically modelling is much lower starting at $1.76*10^6$ J m^{-3} $°C^{-1}$ and reaching $2.49*10^6$ J m^{-3} $°C^{-1}$. The optimum of K for analytical modelling starts at 0.84 J s^{-1} m^{-1} $°C^{-1}$ and reaches to 1.45 J s^{-1} m^{-1} $°C^{-1}$, whereas the optimum for the numerical model starts at 1.47 J s^{-1} m^{-1} $°C^{-1}$ and ends at 2.85 J s^{-1} m^{-1} $°C^{-1}$.

The benchmarking procedure highlighted the parameter combinations resulting in the highest similarity between the modelled hyporheic fluxes and the benchmark flux. Using values for K of 1.16 J s^{-1} m^{-1} $°C^{-1}$ and 2.61 J s^{-1} m^{-1} $°C^{-1}$ as well as values for C of $3.68*10^6$ J m-3 $°C^{-1}$ and $1.66*10^6$ J m-3 $°C^{-1}$ for the analytical model and the numerical model, respectively, achieve the best model performances. With the analytical model, using shallower depth combinations to estimate vertical hyporheic flux lead to better model performances (Table 2). The highest KGE of 0.36 is computed for a mean flux depth of 0.125 m using sensors at 0.10 m and 0.15 m. Using the numerical model, only one depth combination results in a positive KGE between the reference benchmark and simulated hyporheic fluxes. A KGE of 0.2 is achieved using a sensor combination of depths 0.05 m and 0.45 m, incorporating five sensors in the model. Shallower depth combinations do not lead to higher KGE values, but increasing the number of sensors used has a slightly positive effect on KGE in most cases. However, including the uppermost and lowest sensor results in lower similarity than including only the five sensors in the middle of the depth profile (Table 2).

Table 1: Parameter values for sensitivity analysis (n: numerical Model, a: analytical Model) and resulting KGEs

Parameter	Base value	KGEn	KGEa	**Low value**	**KGEn**	**KGEa**	**High value**	**KGEn**	**KGEa**

Global SA	K [J s^{-1} m^{-1} °C^{-1}]	1.675	-1.82	0.07	0.837	< -3	0.26	3.349	0.08	-0.9
	C [kJ m^{-3} °C^{-1}]	2 904			1 506	0.09	-0.39	3 682	< -3	0.23
	C_w [kJ m^{-3} °C^{-1}]	4 187			4 142	-1.83	0.07	4 226	-1.81	0.07
	β	0	n.a.		0.001	n.a.	0.06	0.1	n.a.	<-3
Local SA n	K [J s^{-1} m^{-1} °C^{-1}]	1.840	0.18		1.7	-1.47		1.9	0.18	
	C [kJ m^{-3} °C^{-1}]	2 160			2 000	0.18		2 300	-1.52	
	C_w [kJ m^{-3} °C^{-1}]	4 187			4 142	0.18		4 226	0.18	
Local SA a	K [J s^{-1} m^{-1} °C^{-1}]	1.160		0.37	1.1		0.38	1.2		0.37
	C [kJ m^{-3} °C^{-1}]	3 681			3 600		0.37	3 700		0.36
	C_w [kJ m^{-3} °C^{-1}]	4 187			4 142		0.37	4 226		0.37
	β	0			0.001		0.37	0.01		0.3

Table 2: Kling-Gupta-Efficiencies of time series analysis and slope α of the logistic regression model with groundwater level as predictor for the analytical model and the numerical model (* indicates significance on an alpha of 0.001)

	Analytical model						**Numerical model**		
Depth [m]	**Lance 1**		**Lance 2**		**Lance 3**		**Lance 2**		
	KGE	**Slope α**	**KGE**	**Slope α**	**KGE**	**Slope α**	**No. of sensors**	**KGE**	**Slope α**
0.035	0.06	2.21*	0.07	1.09	-0.07	1.07			
0.06	0.02	-0.37	0.13	5.85*	0.12	4.58*	3	< -3	-6.02*
0.075	-0.07	-0.9	0.28	11.75*	0.09	6.61*			
0.085	-0.07	-0.6	0.27	11.37*	0.06	3.39*	4	-0.34	5.47*
0.1	-0.15	-1.90*	0.35	15.94*	0.06	4.16*	3	< -3	0
0.125	-0.07	-1.32	0.36	14.09*	0.03	2.31*			
0.135	-0.21	-3.12*	0.27	9.6*	-0.84	-11.53*	5	-0.86	10.87*
0.15	-0.26	-4.85*	0.22	10.41*	-0.83	-11.42*	4	< -3	-1.31-
0.175	-0.24	-5.45*	0.21	8.69*	-0.97	-10.58*	3	< -3	0
0.2	-0.19	-5.32*	0.08	5.65*	-1.42	-9.96*			
0.235	-0.17	-5.66*	0.03	1.70	-0.88	-11.46*	6	-0.7	8.71*
0.25	-0.25	-7.06*	0.03	1.93*	-0.87	-10.56*	5	0.2	10.01*
0.275	-0.25	-7.69*	-0.023	0.55	-0.87	-9.60*	4	< -3	4.32*
0.3	-0.19	-8.37*	-0.09	-1.25	-0.95	-8.65*	3	< -3	8.17*
0.335	-0.15	0	-0.17	1.77	-0.65	-8.70	7	-0.25	13.74*
0.35	-0.14	-6.61*	-0.27	-2.22	-0.46	1.42			
0.35	-0.19	0	-0.20	1.59	-0.62	-6.68	6	-0.96	10.63*
0.375	-0.17	0	-0.24	0.38	-0.59	-12.33*	5	<-3	1.84*
0.4	-0.18	0	-0.28	-1.91	-0.55	-12.72*	4	< -3	5.16*
0.45	-0.18	-9.93	-0.38	-3.47	-0.19	-0.55	3	n.a.	n.a.
0.55	-0.32	-13.83*	-0.26	-9.87*	-3.42	0			

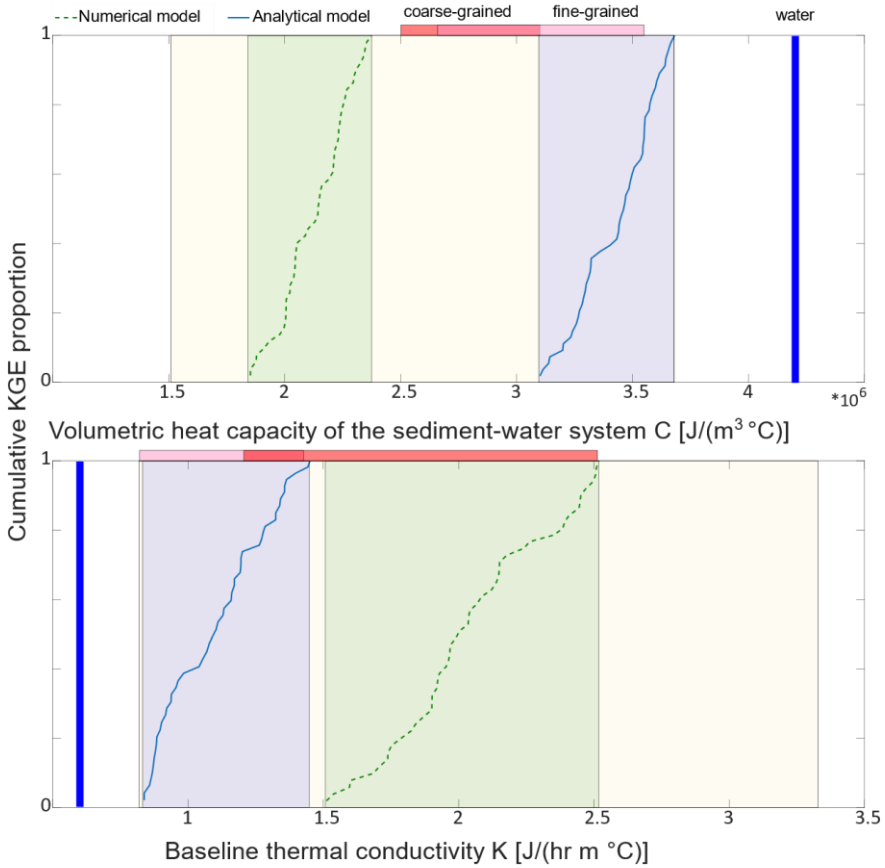

Figure 1: 5 percent best fit cumulative KGE proportion of volumetric heat capacity of the sediment-water system and thermal conductivity for both models with modelling boundaries (slight yellows) and literature values (reddish colors, blue bar) given by Lapham (1989)

Comparing the multiplicative relation between the model parameters (Figure 2), the combined optimum of both parameters appears to be in opposing areas of the parameter ranges under all tested conditions. In general, higher KGEs are achieved under downwelling conditions rather than upwelling conditions. Regional sensitivity analysis confirms a strong influence of both parameters on the output metric KGE, but a significantly higher sensitivity is attributed to the thermal parameter K for the analytical model under all conditions (e.g. d-stats of 0.72 for K and 0.47 for C for the whole time series). The results are not as clear for the numerical model, where the sensitivity of KGE to the thermal parameters is strongly influenced by the chosen boundary conditions. Using the whole time series both parameters have nearly the same level of sensitivity (d-stats of 0.5 for C and 0.54 for K). However, under upward conditions the sensitivity to K (d-stat of 0.89) increases, whereas under downward conditions the output metric KGE is clearly more sensitive to C (d-stat of 0.82). Comparing both models with each other, a common optimum for both parameters is identifiable only for downwelling conditions with comparably low expressions of KGE.

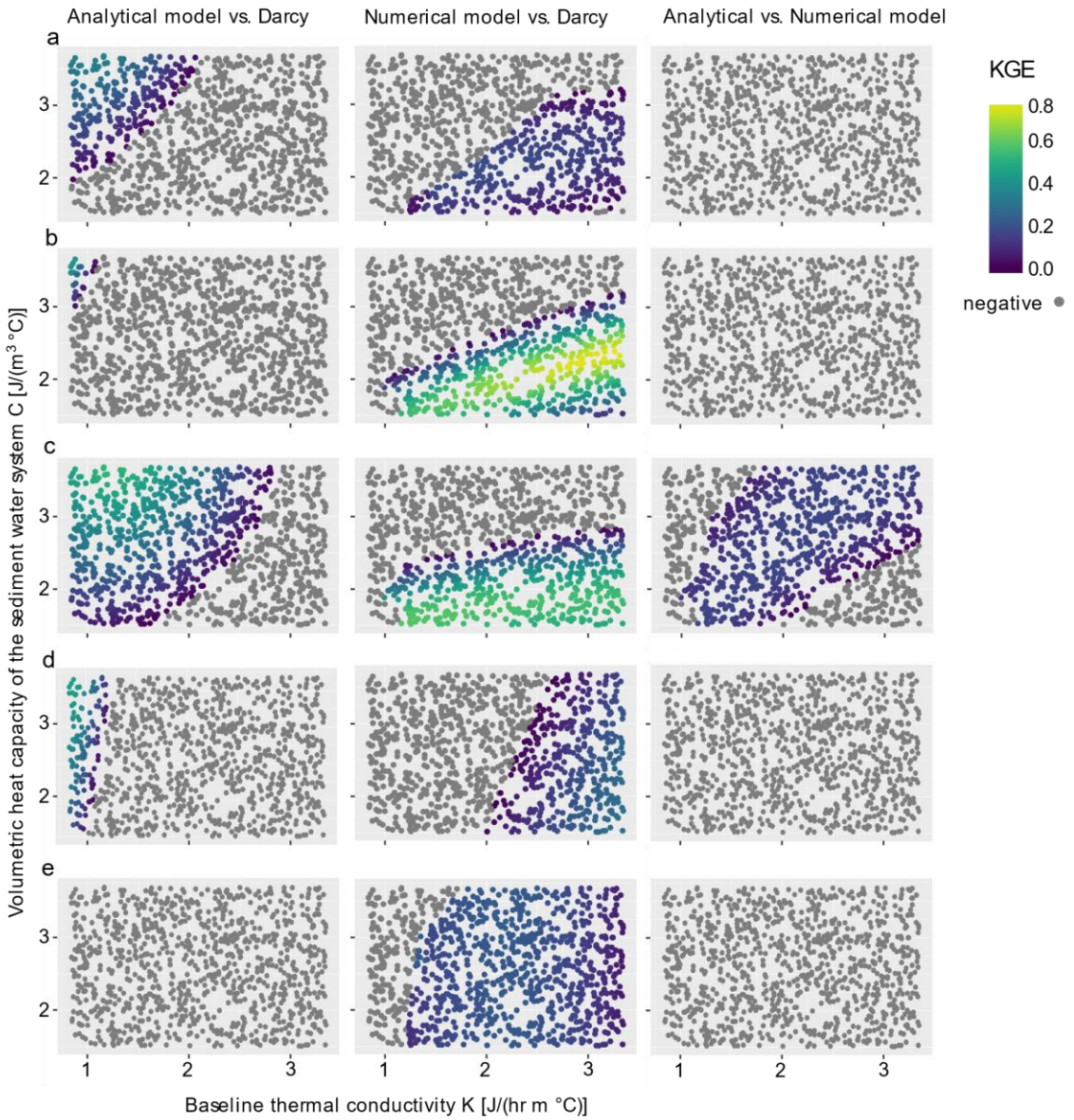

Figure 2: Colored scatter plots of parameter influence on the similarity (as KGE) between both temperature-based models and Darcy flux (a: whole time series, b: good fit period of 6 months, c: dominantly downward period of 3 months, d: dominantly upward period of 3 months, e: short-term upward events)

Looking at Darcy fluxes as well as numerically and analytically modelled fluxes in parallel, for specific days under upwelling or downwelling conditions, the effect of parameter variation on flux intensity becomes clear (Figure 3). A change in only one thermal parameter has the same effect on the resulting flux in both models. However, there is a difference in flux intensity between both models at the same parameter value. In general, an increase in C leads to higher flux values under downwelling conditions for both models, whereas an increase in K results in lower flux values for both models under down- and upwelling conditions. C has near to zero effect on the flux calculated with the numerical model under upwelling conditions and again a positive effect on the flux calculated with the analytical model. As a consequence of the mentioned offset, Darcy flux is matched by both temperature-derived fluxes with different thermal parameter values. This is confirmed in the comparison between fluxes calculated with different models using the same parameter value and the respective best fitting depth combination, which results in mostly negative KGE. Only under dominant downward flux behavior positive KGEs are calculated (Figure 2), but even here only with a maximum of 0.17. However, comparing the best fit model runs (highest similarity between

modelled flux and benchmark) of both models ignoring the use of different parameter combinations, results in a KGE of 0.55 between the simulated fluxes.

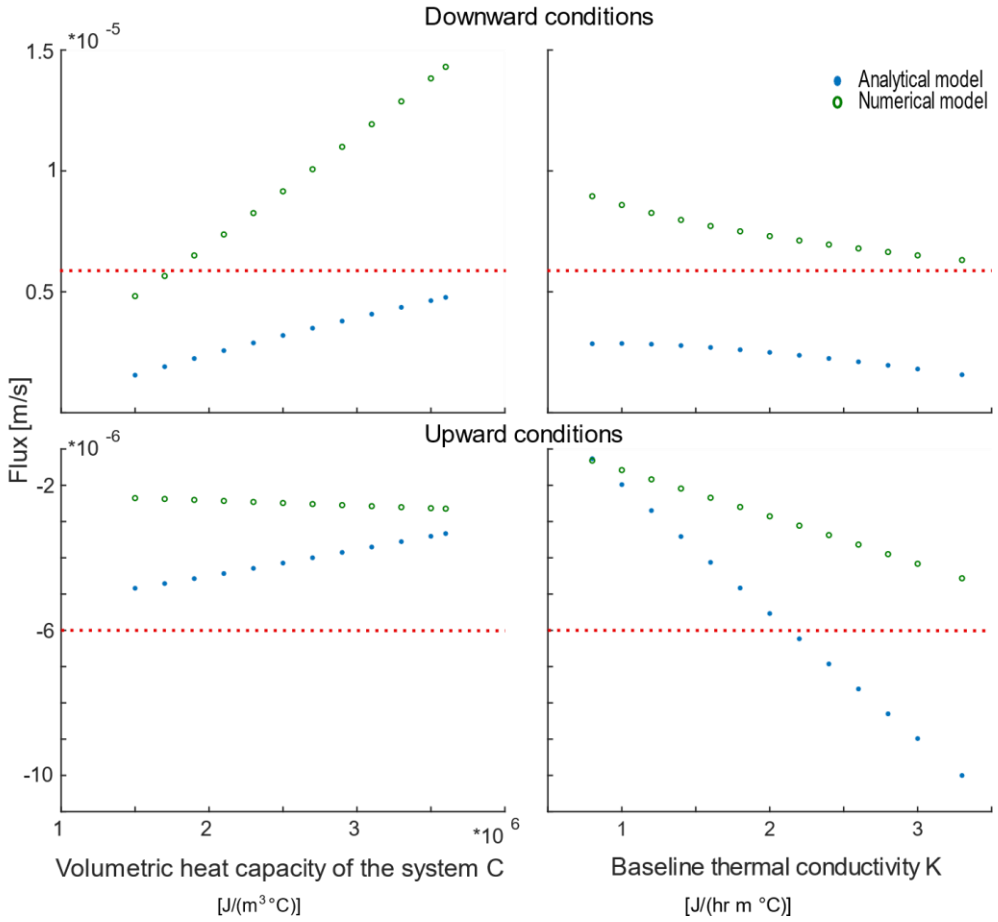

Figure 3: Effect of one-at-a-time variation of thermal properties K and C on flux intensity under downward or upward conditions for both temperature-based models. Darcy flux marked in red.

3.2 Process analysis

3.2.1 Temporal variation of hyporheic exchange fluxes

Temporal variation of hyporheic exchange fluxes was analyzed using the best-fit fluxes of both model approaches (Figure 4). The median values of the temperature-based modelled fluxes ($2.69*10^{-6}$ m s^{-1} -analytical model) and ($1.23*10^{-6}$ m s^{-1} -numerical model) are smaller than the Darcy flux with $4.08*10^{-6}$ m s^{-1}. Mean values of the temperature-based modelled fluxes are smaller than the mean of Darcy flux, too. The standard deviations of the analytical model (171 %) and the numerical model (213 %) are higher than of the Darcy flux (130 %). Distributions are with a skewness of 0.15 (analytical model) and 2.21 (numerical model) right skewed whereas Darcy flux distribution is slightly left skewed (-0.06). All fluxes show a leptokurtic distribution.

Darcy flux has with 0.11 a smaller seasonality strength than the temperature-based modelled fluxes with 0.37. The stronger seasonality is visible in higher downward fluxes in summer and a tendency for upward fluxes in winter (Figure 4). Kolmogorov-Smirnov test and Wilcoxon-Mann-Whitney-U test both show a significant difference (alpha of 0.001) between summer and winter period for all fluxes. However, the calculated p-value using Darcy flux is smaller than the ones using the temperature-based models. The episodic variation indices are 0.12 (Darcy flux), -0.3 (analytical model) and -0.07 (numerical model). Episodicity and seasonality of hyporheic exchange fluxes have a comparable magnitude at the other two lances.

The reference logistic model using Darcy flux directions has a positive slope α of 24.67 and a turning point at 168.68 m a.s.l.. The depth combinations resulting in the most similar hyporheic exchange fluxes to the benchmark (namely 0.1 m to 0.15 for the analytical model and 0.05 to 0.45 for the numerical model) follow the same shape as the logistic model using directions from Darcy flux (Figure 5). With low groundwater tables, the probability for upward flux is near zero and increases with increasing groundwater levels. These models were highly significant (alpha of 0.001). All show a positive slope α, which was with 14.09 (analytically derived flux direction) and 10.01 (numerically derived flux direction) fairly lower than the slope α of the logistic Darcy flux model. With increasing depth the logistic model based on analytically derived flux direction becomes insignificant, before in the lowest depth it had a significant but negative slope α (Figure 5 and Table 2). The logistic model based on the numerically derived flux directions largely depends on the number of considered sensors. If only few sensors are applied the model tends to be insignificant or changes slope α direction as well (Table 2). Using discharge or surface water levels as predictor variable does not change the shape of the results.

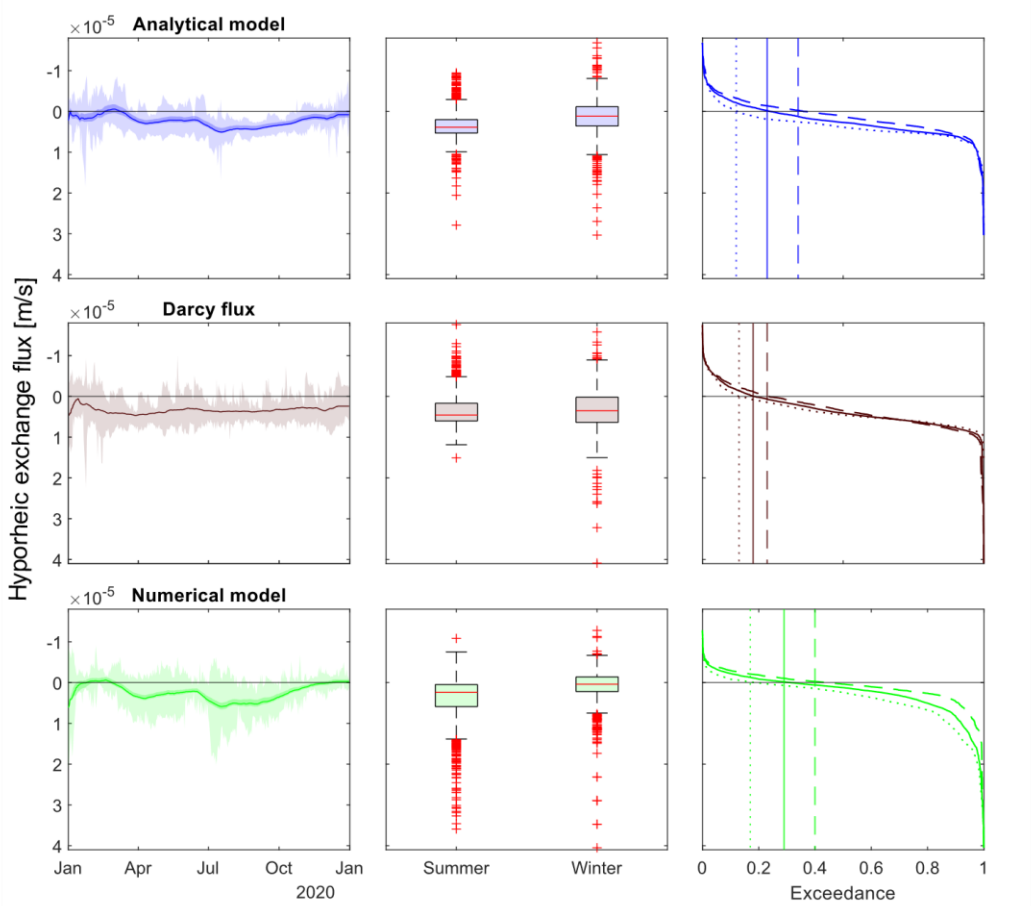

Figure 4: Left panel: seven year mean daily hyporheic exchange fluxes (line: 30 days moving average, dark shaded area: 5 % best fit interval (5 % of fluxes from MCA with the highest similarity to the benchmark) as 30 days moving average, light shaded: seven year daily standard deviation). Middle panel: boxplots of hyporheic exchange fluxes for summer and winter periods (red line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, red cross: outlier). Right panel: flow duration curves with vertical lines marking the change between upwards and downwards flux for the whole time series (full line), summer (dotted line) and winter (dashed line).

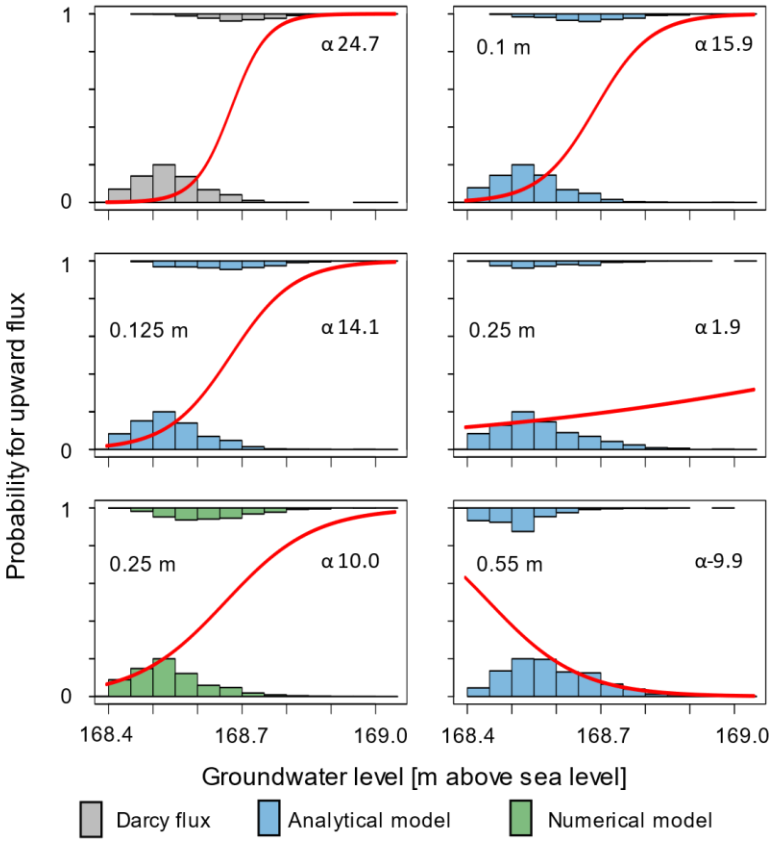

Figure 5: Results of the logistic model approach (p-value < 0.001) for the best fitting depth of the different model approaches (left) and different depth of the analytical model (right) with slope α

The model-fit of the temperature-based hyporheic flux in accordance with the gradient derived benchmark is better in summer months (April till September) than in the winter period (November to March). KGE improves drastically from 0.12 in winter up to 0.54 in summer for the analytical model, compared to the -0.04 to 0.09 for the numerical model.

In comparison to the benchmark flux both temperature-based models result in longer times of dominating upward fluxes (Figure 4). Darcy flux calculation proposes upward flux in 18 % of the cases, whereas the analytical and numerical model estimate a proportion of 23 % and 29 %, respectively. About 80 % of the fluxes vary between $-1*10^{-5}$ m s^{-1} and $1*10^{-5}$ m s^{-1}. Both temperature–based models estimate slightly smaller downward flux intensities than the Darcy flux calculations. These differences are more prominent in winter than in the summer period. In winter Darcy flux proposes upward flux behavior in 23 % of the cases, whereas the analytical and numerical model estimated the proportion with 34 % and 40 %, respectively. In summer the estimation of upward flux behavior is with values of 12 % (analytical model) and 17 % (numerical model) much closer to the amount of cases calculated with the Darcy flux, which is 13 %. Slight parameter uncertainties result in maximum ranges of $2.72*10^{-6}$ m s^{-1} (analytical model) and $4.24*10^{-6}$ m s^{-1} (numerical model)). In addition, seasonality strengths vary slightly between 0.37 and 0.38 for small parameter changes. Considering the whole range of parameter variability, seasonality strengths ranges from 0.36 to 0.4 for the analytical model, and 0.3 to 0.45 for the numerical model.

3.2.2 Quantification and temporal variation of hyporheic exchange depths

The zero-flux method was applied on analytically and numerically modelled fluxes with 21 or 14 depths combinations as input data sets. Using the 21 depth combinations with the analytical model, the zero-flux method results in a mean extension of the hyporheic exchange depth of 0.27 m (with a standard deviation of 0.18 m). Including only 14 depth combinations with the analytical model results in an extension of 0.24 m (with a standard

deviation of 0.11 m). In contrast, the numerical model (14 depth combinations) results in a mean hyporheic extension of 0.26 m (with a standard deviation of 0.08 m). Kruskal-Wallis combined with Dunn's posthoc test reveals significant (alpha of 0.001) differences in depths calculated using the different model approaches.

Using the logistic regression approach results in a comparable hyporheic zone extension of 0.25 m based on the combination of the analytical model and the Darcy benchmark. The logistic approach to estimate mean hyporheic extension coupled with numerical data could not be realized, since the performance of the numerical model is strongly influenced by the number of sensors used.

Time series decomposition reveals a seasonal component with a strength of 0.31 in hyporheic exchange depth using the numerical model with 14 depth combinations. Seasonal strength is slightly higher with 0.36 (AM 21) and 0.37 (AM 14) for the analytical model. The episodical variation indices are -0.09 (AM 21), 0.01 (AM 14) and -0.2 (NM 14). The absolute vertical range of episodic variation is 0.54 m (AM 21), 0.34 m (AM 14) and 0.27 m (NM 14). At the other two lances seasonal variation is slightly more dominant.

The analytical model shows a deeper extent of the hyporheic zone in summer, whereas it is the opposite using the numerical model (Figure 6). Kolmogorow-Smirnow tests (alpha of 0.001) confirm significant differences between the seasons regardless of the chosen model or the number of depth combinations. However, the difference is small when the numerical model is used. In addition, strong variations during the whole year are observable.

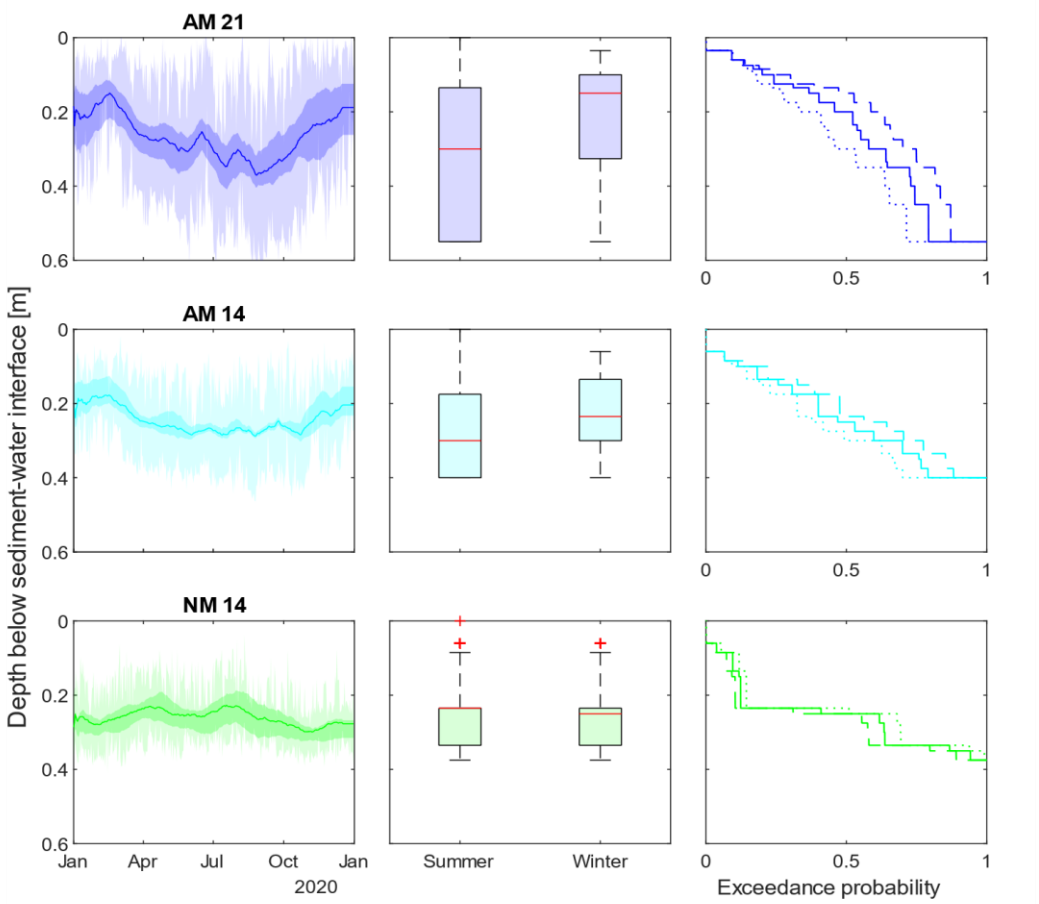

Figure 6: Left panel: seven year mean daily hyporheic exchange depth (line: 30 days moving average, dark shaded area: 5 % best fit interval (5 % of fluxes from MCA with the highest similarity to the benchmark) as 30 days moving average, light shaded: seven year daily standard deviation). Middle panel: boxplots of hyporheic zone extension for summer and winter periods (red line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, red cross: outlier). Right panel: exceedance probability curves of hyporheic zone extension for the whole time series (full line), summer (dotted line) and winter (dashed line).

MNR results in significant effects (alpha of 0.001) of water levels and river discharges on hyporheic exchange depth for all models, while air temperature is a significant predictor using the numerical model (Figure 7). The effect of water levels and river discharges is clearly visible for both analytical models. No statistical relationship has been detected between precipitation and hyporheic zone extension. In addition, air temperature shows no effects when the analytical model is used. However, an increase of hyporheic zone extension with increasing daily mean air temperatures is observable and statistically significant when the numerical model is used. Parameter uncertainties result in maximum differences of 0.29 m (AM 21), 0.17 m (AM 14) and 0.16 m (NM 14) (Figure 6). In addition, seasonality strengths range from 0.31 to 0.38 (AM 21), 0.37 to 0.46 (AM 14) and 0.27 to 0.36 (NM 14) if there is a small parameter variation. High parameter uncertainties lead to ranges of seasonality strengths from 0.2 to 0.43 (AM 21), 0.11 to 0.46 (AM 14) and 0.23 to 0.52 (NM 14).

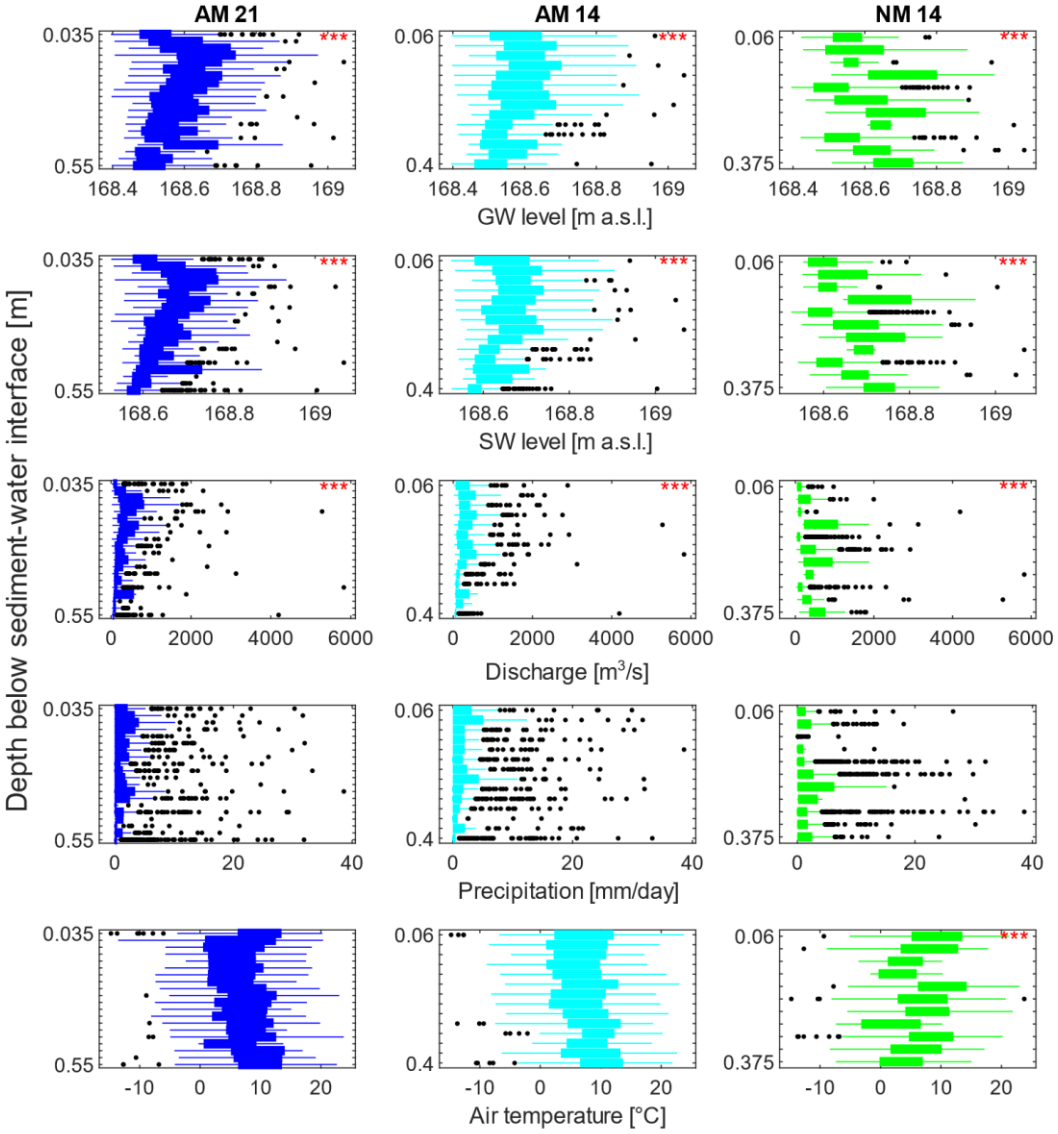

Figure 7: Whisker-Box-Plots depicting the relationship between identified hyporheic exchange depth (shown on the y-axes as zero flux plane) and hydro-meteorological boundary conditions (groundwater (GW) level, surface water (SW) level, discharge, precipitation and mean air temperature) ; * marks significant effects in MNR.)**

Comparing the relationships between hyporheic exchange flux intensities end exchange depths we find significant (alpha of 0.001) correlations between both for all three model setups. While hyporheic exchange flux intensities end

exchange depths for the analytical model setups with 21 and 14 sensor depths, respectively, are positively correlated the numerical model applications result in a negative correlation between both. However, the explanatory power of these correlations is limited (R^2 029, 0.21, 0.12). Looking at the relationships between the daily standard deviations of hyporheic exchange flux intensities end exchange depths the observed behavior of the correlations was contrary: all signs were conversed and the explanatory powers were even reduced (R^2 0.09, 0.0, 0.06). The relationship for the analytical model with 14 sensor depths was not significant (alpha of 0.001).

4 Discussion

4.1 Model evaluation and sensitivity analysis

Deviations between temperature-based flux and Darcy flux can be explained with the spatial distance between the GW well and the temperature lances (e. g. Krause et al., 2012) as well as with the depth integrative nature of the hydraulic method (2 m filter depth), while the temperature-based approaches quantify fluxes on smaller spatial scales. Therefore, we consider Darcy flux as a benchmark for model evaluation and sensitivity analysis but acknowledge that we cannot identify it as the true flux.

This study shows that for the analytical model a shallower sensor depth and a small sensor spacing window as well as a high number of sensors used for the numerical model led to the highest similarity with the used benchmark. This is in line with Gordon et al. (2012) who show an increasing modelling uncertainty with increasing sensor spacing. However, minimizing sensor spacing is limited by the amplitude ratio reaching unity, which would result in improbable high flux estimations (Irvine et al., 2017). Regarding absolute sensor depth, we observed two quite common phenomena: using a very shallow sensor deployment, the topmost sensor at 0.02 m depth was uncovered and covered through erosion and deposition processes, which resulted in unstable measurements. A very deep sensor depth (here 0.65) partially did not show any diurnal signal. This extinction depth can already be in very shallow grounds of 0.2 m under upwelling flux conditions (Briggs et al., 2014). The mean depth of 0.125 m, with the strongest similarity between analytical model and Benchmark flux is way above the extinction depth but deep enough to be not affected by erosion. Munz and Schmidt (2017) stated that an increase in the number of sensors used in their numerical model led to a higher accuracy. In this study, model runs using all seven available sensors gave results that did not meet these expectations. We attribute the poor model performance to the unstable measurements of the uppermost sensor (0.02 m). However, the exclusion of the shallowest and the deepest sensor drastically improved model performance (e. g. KGE increase from -0.25 to 0.2 for the whole time series).

A positive correlation between daily mean air temperature and model performance is shown for the analytical model (Figure 8). These observations coincide with the seasonal differences of model performances. The performance of the numerical model in relation to the daily mean air temperature follows an optimum curve: it was best performing at a medium daily mean air temperatures and similarity with the benchmark decreased drastically at air temperatures below 6 °C or above 16 °C (Figure 8, lower panel). The reduced performances of both models at lower air temperatures can be explained with smaller temperature amplitudes measured in winter due to smaller temperature ranges and reduced solar radiation (Onderka et al., 2013).

The impact of C_w is neglectable considering the fact that most literature references agree about a value between $4.184*10^6$ J m^{-3} $°C^{-1}$ and $4.187*10^6$ J m^{-3} $°C^{-1}$ for C_w (e. g. Gordon et al., 2012; Munz & Schmidt, 2017) and that even a substantial variation in OAT SA did only show negligible influence. The influence of β is subject of great discussion in the scientific community over the last years (Rau et al., 2012). As described in Rau et al. (2014) the influence of thermal dispersivity is much lower than that of solute dispersion. It could be shown that the analytical model was insensitive to β. The numerical model by Munz and Schmidt (2017) does not incorporate a β term.

Considering the results of the SA: it can be stated that there was always a difference in calculated flux between the two models even under uniform conditions and with constant parameter settings as also reported by Swanson and Cardenas in 2011. Using two contrasting modelling approaches both incorporating the same temperature data can therefore result in different thermal parameter estimation. Uncertain thermal properties combined with two different model approaches can lead to miscellaneous results of some magnitude (Figure 3). Applying the best fit parameters for the analytical and numerical modelling approaches resulted in the highest similarity between both simulated exchange flux time series with a KGE even higher than that of the best fits to the benchmark flux.

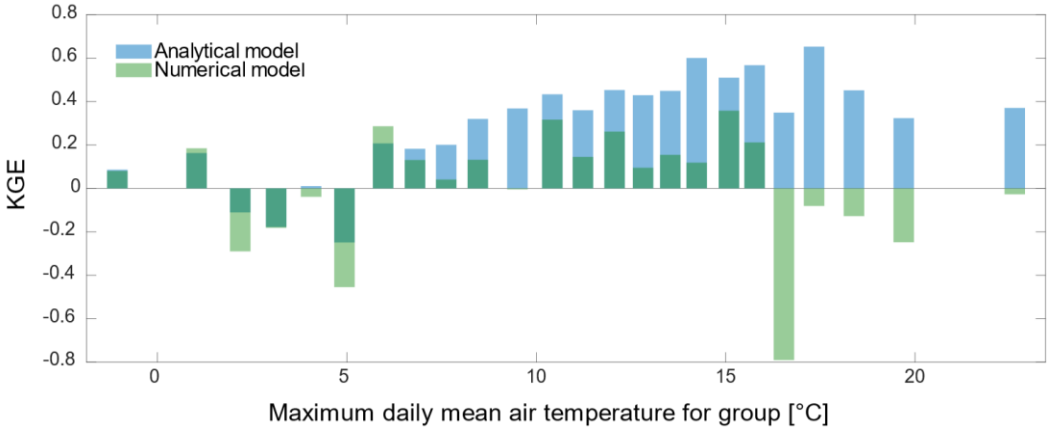

Figure 8: Influence of daily mean temperature on the model performance (KGE) of the analytical and numerical model (data is classified in 24 groups according to maximum daily mean air temperature to enable KGE calculation per group).

The identification of C and K as sensitive parameters for the numerical model and the higher effect of C, confirms findings by Munz and Schmidt (2017). Yet, the slightly higher effect of C reported by Munz and Schmidt (2017) was not reproduced under dominating upward flow conditions. Furthermore, it could also be validated that there is a correlated effect of both parameters (Munz & Schmidt, 2017). However, especially the very low values for C ($1.76*10^6$ J m^{-3} $°C^{-1}$), falling outside the boundaries set by Lapham (1989) identified by the numerical model, seem questionable for real sediments, although there is some literature reporting comparable low values for C (Goto & Matsubayashi, 2009).

The high sensitivity of the analytical model to K is also demonstrated in other studies (i.e. Gordon et al., 2012). As second key parameter n is identified and its impact directly attributed to its influence on C (Gordon et al., 2012), which in the end results in a high sensitivity to the parameter C. Shanafield et al. (2011) varied n, K and C_s in the analytical model of Hatch et al. (2006) using MCA. They only report an effect of K_e, for which they conclude that uncertainty in this parameter estimation is disruptive especially under gaining conditions.

4.2 Process analysis

4.2.1 Temporal variation of hyporheic exchange fluxes

Gariglio et al. (2013) suggest to use different model approaches for winter, as there is no sinusoidal signal in winter. In this study, the sinusoidal signal was always detectable even in winter. Nevertheless, deviations between the modelled fluxes and the gradient derived fluxes are smaller in the summer period than during winter. In addition to that, an increased upward flux behavior was simulated in the winter period by the temperature-based models, whereas the calculated percentages of upward flux were more similar in summer. An explanation for this could be the diminished diurnal signal or the significant change in hyporheic zone extension during the year, an effect also reported by Boano et al. (2014).

The flux time series generated by the analytical model was similar to the benchmark in regard of distribution shape. However, statistical location and variation parameters reveal that lower average flux intensities with a higher temporal variation are simulated by both temperature-based models. The analytical model achieves again a higher similarity to the benchmark.

Even though Birkel et al. (2016) reported no clear relationship between hyporheic flux and river discharge, it could be shown here that there is at least a binary logistic relationship between flux direction and river discharge or water level in the study area. Groundwater level, surface water level and surface water discharge were all positively correlated (correlation coefficient > 0.8) and logistic regression always described a higher tendency for upward flux with an increase in these parameters. Su et al. (2015) question that hyporheic flux is solely dependent on surface water level, but they also confirm the influence observed here. A direct influence of precipitation as suggested by Sophocleus (2002) could not be validated. Probably the resolution of daily data was not suited to find clear connections here. It cannot be excluded that precipitation might have an indirect influence by affecting discharge and water level, which are both influential conditions.

In several studies (e.g. McCallum & Shanafield, 2016; Trauth & Fleckenstein, 2017) a shift from upwelling to downwelling conditions is often observed during a hydrological event. However, at our study site an increase in upwards flow came along with an increasing discharge in most of the cases. This could be attributed to the temporal resolution of our modelled hyporheic fluxes (daily) being sometimes larger than the temporal resolution of occurring hydrological events. Another reason could be a fast reaction of the near surface groundwater, observable in a fast raise of the groundwater table during event flow. With only a short time shift high discharge and surface water level caused by high precipitation led to an increase in near-surface groundwater, which in turn resulted in upward flux behavior. Yet, there are single events that show a more intense downward flux during high discharges. Probably, this phenomenon is also dependent on the connectivity and the shape of the whole hyporheic corridor (Stanford & Ward, 1993) e.g. the overall antecedent moisture content of the whole area.

4.2.2 Identification of hyporheic exchange depths and its temporal variation

The absolute values of hyporheic exchange depth calculated at our study site range from 0.035 m to 0.55 m and are in accordance with values reported elsewhere (e. g. Kim et al., 2014; Harvey & Fuller, 1998; Hill & Lymburner, 1998; Boano et al., 2008). The mean hyporheic exchange depth estimated with the zero-flux plane method based on the analytical and numerical approaches deviated with 0.03 m only slightly. The time-weighted average of hyporheic exchange depth calculated by the logistic regression method based on the depth integrated information of the Darcy benchmark flux matches the mean extension simulated using the zero-flux plane method based on spatially and temporally resolved water temperature data.

The hyporheic zone extension is subject to great variation during the whole year as reported in previous studies (e.g. Wondzell & Swanson, 1996; Wondzell & Swanson, 1999). There is a significant difference between the summer and winter period, confirmed by all modelling approaches. The strength of the seasonal component is slightly higher when the analytical model is used to calculated exchange depth. This is confirmed by the small but still significant difference between the seasons by applying the numerical model. We identified several seasonal drivers influencing hyporheic zone extension. Using the zero-flux method very high standard deviations were calculated throughout the year. Hence, there must be short time drivers influencing hyporheic exchange depth as well. Although seasonal variation is comparable high or higher (episodic variation index $\approx< 0$), MNR could identify short time drivers like discharge and water. Calculated episodic variation was comparable between the models, but the range of the absolute depth values was twice as large using the analytical model. The reduced ability of the numerical model to detect seasonal and episodical variations in hyporheic zone extension can be explained with the limited number of possible depth combinations in comparison to the analytical model.

As expected, a higher groundwater level came along with a shallower hyporheic zone extension as also described by Boano et al. (2008). An increase of hyporheic zone extension under flood peaks as reported by other authors (Bhaskar et al., 2012; Singh et al., 2019; Wu et al., 2018), was not observed. This was supported by the results of the correlation analysis between hyporheic exchange flux intensities end exchange depths for the modelling results and the seven years daily standard deviations: the correlations were ambiguous regarding the signs of analytical and numerical modelling approaches. However, the low explanatory power of all correlations suggests that other processes and boundary conditions might contribute to the observed combinations of hyporheic exchange flux intensities end exchange depths as well. E. g. a diverging behavior (smaller hyporheic zone extension under peak flow) can be explained with the fast reaction of the groundwater table coming along with flooding observed in the study area. Watson et al., (2018) show that even water temperature differences between flood peak and and groundwater might even lead to differences in temperature distributions and thus simulated hyporheic exchange depths.

Since hyporheic zone extension and residence times are linked, a larger hyporheic zone increases reaction times for biogeochemical processes (Boano et al. 2014). Considering the impact of diel water temperature variations on hyporheic nutrient cycling (e.g. Zheng and Cardenas, 2018), the integration of observed seasonal and episodic variations of hyporheic exchange depths and exchange flux intensities into the continuous and long-term prediction of travel time related nutrient turnover in the hyporheic zone might improve the abilities of our modelling tools even further.

5 Conclusions

Simulated hyporheic exchange flux time series are characterized by an underlying seasonal behavior, tending to upwelling conditions in winter and downwelling conditions in summer. Both temperature-based models tend to estimate an increased upward flux behavior and seasonal variation as well as lower flux intensities than depth

integrated Darcy flux calculations. Parameter uncertainty has a small influence on the seasonality of hyporheic exchange fluxes using the analytical model. The numerical model is more sensitive to parameter variation regarding absolute flux values and time series features. Given that daily mean air temperatures affect model accuracies the use air temperatures as quality criteria is therefore recommended to access times of low model reliability.

The simulated seasonality of hyporheic exchange zone depth shows a differing behavior for each model type. While the analytical model revealed a larger hyporheic zone extension in summer than in winter, the results of the numerical model were indistinctive. The smaller number of possible sensor depth combinations using the numerical model in comparison to the analytical model results in a reduced ability to detect seasonal variation. Short-term effects resulted in a strong variability of hyporheic zone extension during the whole year. These short-term effects could be attributed to changes in groundwater and surface water levels and river discharges. A rise in groundwater or surface water level as well as discharge showed a positive correlation with an increasing probability for upward flux and a reduced hyporheic exchange depth.

Finally, we show that thermal methods have to be applied carefully as different algorithms lead to diverging flux calculations and the influence of thermal properties should in no case be neglected. In addition, we present new insights into temporal variation of hyporheic zone extension and flux intensities, which might contribute to fully understand the functioning of river self-purification processes in the hyporheic zone of anthropogenically impacted rivers.

Acknowledgments, Samples, and Data

This study was supported by the German Research Foundation (DFG) – Project-No.: SY 12/27-2.
We thank Bernhard Fink and Andreas Kurtenbach for their efforts during the sensor installation and the data aquisition. As of today the data are submitted for open access publication under Pangea.de (we will receive a DOI within the next weeks).

References

Allander, K. K. (2003): Trout Creek—Evaluating ground-water and surface water exchange along an alpine stream, Lake Tahoe, California. In: David Arthur Stonestrom und Jim Constantz (Hg.): Heat as a tool for studying the movement of ground water near streams. Reston, Va.: U.S. Geological Survey Information Services (Circular / U. S. Department of the Interior, U.S. Geological Survey, 1260), S. 35–45.

Banzhaf, S., & Scheytt, T. (2009). Auswirkung einer künstlichen Hochwasserwelle auf den fließgewässernahen Grundwasserleiter. *Grundwasser*, *14*(4), 265–275. https://doi.org/10.1007/s00767-009-0109-x

Bhaskar, A. S., Harvey, J. W., & Henry, E. J. (2012). Resolving hyporheic and groundwater components of streambed water flux using heat as a tracer. *Water Resources Research*, *48*(8), 951. https://doi.org/10.1029/2011WR011784

Birkel, C., Soulsby, C., Irvine, D. J., Malcolm, I., Lautz, L. K., & Tetzlaff, D. (2016). Heat-based hyporheic flux calculations in heterogeneous salmon spawning gravels. *Aquatic Sciences*, *78*(2), 203–213. https://doi.org/10.1007/s00027-015-0417-4

Boano, F., Revelli, R., & Ridolfi, L. (2008). Reduction of the hyporheic zone volume due to the stream-aquifer interaction. *Geophysical Research Letters*, *35*(9), 995. https://doi.org/10.1029/2008GL033554

Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, *52*(4), 603–679. https://doi.org/10.1002/2012RG000417

Boulton, A. J., Findlay, S., Marmonier, P., Stanley, E. H., & Valett, H. M. (1998). The functional significance of the hyporheic zone in streams and rivers. *Annual Review of Ecology and Systematics*, *29*(1), 59–81. https://doi.org/10.1146/annurev.ecolsys.29.1.59

Boulton, A. J., Fenwick, G. D., Hancock, P. J., & Harvey, M. S. (2008). Biodiversity, functional roles and ecosystem services of groundwater invertebrates. *Invertebrate Systematics*, *22*(2), 103. https://doi.org/10.1071/IS07024

Boulton, A. J., Datry, T., Kasahara, T., Mutz, M., & Stanford, J. A. (2010). Ecology and management of the hyporheic zone: Stream–groundwater interactions of running waters and their floodplains. *Journal of the North American Benthological Society*, *29*(1), 26–40. https://doi.org/10.1899/08-017.1

Bredehoeft, J. D., & Papadopulos, I. S. (1965). Rates of vertical groundwater movement estimated from the Earth's thermal profile. *Water Resources Research*, *1*(2), 325–328. https://doi.org/10.1029/WR001i002p00325

Briggs, M. A., Lautz, L. K., McKenzie, J. M., Gordon, R. P., & Hare, D. K. (2012). Using high-resolution distributed temperature sensing to quantify spatial and temporal variability in vertical hyporheic flux. *Water Resources Research*, *48*(2), 951. https://doi.org/10.1029/2011WR011227

Briggs, M. A., Lautz, L. K., Buckley, S. F., & Lane, J. W. (2014). Practical limitations on the use of diurnal temperature signals to quantify groundwater upwelling. *Journal of Hydrology*, *519*, 1739–1751. https://doi.org/10.1016/j.jhydrol.2014.09.030

Buss, S., Cai, Z., Cardenas, B., Fleckenstein, J., Hannah, D., Heppell, K., Hulme, P., Ibrahim, T., Kaeser, D., Krause, S., Lawler, D., Lerner, D., Mant, J., Malcom, I., Old, G., Parkin, G., Pickup, R., Pinay, G., Porter, J., Rhodes, G., Richie, A., Riley, J., Robertson, A., Sear, D., Shields, B., Smith, J., Tellam, J., Wood, P. (2009). *The hyporheic handbook: A handbook on the groundwater-surface water interface and hyporheic zone for environment managers*. *Science report: SC050070*. Bristol: Environment Agency.

Cleveland, Robert W. (1990): STL: a seasonal-trend decomposition procedure based on loess. In: *Journal of official statistics : JOS : an international quarterly*.

Crank, J., & Nicolson, P. (1996). A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Advances in Computational Mathematics*, *6*(1), 207–226. https://doi.org/10.1007/BF02127704

Darcy, H. (1856). *Les fontaines publiques de la ville de Dijon: Exposition et application des principes a suivre et des formules a employer dans les questions de distribution d'eau; ouvrage terminé par un appendice relatif aux fournitures d'eau de plusieurs villes au filtrage des eaux à la fabrication des tuyaux de fonte, de plomb, de tole et de bitume*. Paris: Dalmont.

Dobson, A. J. (2002). *An introduction to generalized linear models* (2. ed.). *Chapman & Hall/CRC texts in statistical science series: Vol. 51*. Boca Raton, Fla.: Chapman & Hall/CRC. Retrieved from http://www.loc.gov/catdir/enhancements/fy0646/2001047417-d.html

DWA. (2005). *Arbeitsblatt. DWA-A 138 - Regelwerk*. Hennef: Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall.

Fanelli, R. M., & Lautz, L. K. (2008). Patterns of water, heat, and solute flux through streambeds around small dams. *Ground Water*, *46*(5), 671–687. https://doi.org/10.1111/j.1745-6584.2008.00461.x

Gariglio, F. P., Tonina, D., & Luce, C. H. (2013). Spatiotemporal variability of hyporheic exchange through a pool-riffle-pool sequence. *Water Resources Research*, *49*(11), 7185–7204. https://doi.org/10.1002/wrcr.20419

González-Pinzón, R., Ward, A. S., Hatch, C. E., Wlostowski, A. N., Singha, K., Gooseff, M. N., Hagerty, R., Harvey, J. W., Cirpka, O. A., Brock, J. T. (2015). A field comparison of multiple techniques to quantify groundwater–surface-water interactions. *Freshwater Science*, *34*(1), 139–160. https://doi.org/10.1086/679738

Gordon, R. P., Lautz, L. K., Briggs, M. A., & McKenzie, J. M. (2012). Automated calculation of vertical pore-water flux from field temperature time series using the VFLUX method and computer program. *Journal of Hydrology*, *420-421*, 142–158. https://doi.org/10.1016/j.jhydrol.2011.11.053

Goto, S., & Matsubayashi, O. (2009). Relations between the thermal properties and porosity of sediments in the eastern flank of the Juan de Fuca Ridge. *Earth, Planets and Space*, *61*(7), 863–870. https://doi.org/10.1186/BF03353197

Grant, S. B., Stolzenbach, K., Azizian, M., Stewardson, M. J., Boano, F., & Bardini, L. (2014). First-order contaminant removal in the hyporheic zone of streams: Physical insights from a simple analytical model. *Environmental Science & Technology*, *48*(19), 11369–11378.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hafen, Ryan P. (2016): Enhances Seasonal Decomposition of Time Series by LOESS. stlplus R package. Version 0.5.1.

Harvey, J. W., & Fuller, C. C. (1998). Effect of enhanced manganese oxidation in the hyporheic zone on basin-scale geochemical mass balance. *Water Resources Research*, *34*(4), 623–636. https://doi.org/10.1029/97WR03606

Hatch, C. E., Fisher, A. T., Revenaugh, J. S., Constantz, J., & Ruehl, C. (2006). Quantifying surface water-groundwater interactions using time series analysis of streambed thermal records: Method development. *Water Resources Research*, *42*(10), 72. https://doi.org/10.1029/2005WR004787

Herzog, S. P., Higgins, C. P., Singha, K., & McCray, J. E. (2018). Performance of Engineered Streambeds for Inducing Hyporheic Transient Storage and Attenuation of Resazurin. *Environmental Science & Technology*, *52*(18), 10627–10636. https://doi.org/10.1021/acs.est.8b01145

Hill, A. R., & Lymburner, D. J. (1998). Hyporheic zone chemistry and stream-subsurface exchange in two groundwater-fed streams. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(2), 495–506. https://doi.org/10.1139/f97-250

Hyndman, Rob; Kang, Yanfei; Montero-Manso, Pablo; Talagala, Thiyanga; Wang, Earo; Yang, Yangzhuoran et al. (2019): Time series feature extraction. tsfeatures R package. Version 0.5.1.

Irvine, D. J., Lautz, L. K., Briggs, M. A., Gordon, R. P., & McKenzie, J. M. (2015). Experimental evaluation of the applicability of phase, amplitude, and combined methods to determine water flux and thermal diffusivity from temperature time series using VFLUX 2. *Journal of Hydrology*, *531*, 728–737. https://doi.org/10.1016/j.jhydrol.2015.10.054

Irvine, D. J., Briggs, M. A., Lautz, L. K., Gordon, R. P., McKenzie, J. M., & Cartwright, I. (2017). Using Diurnal Temperature Signals to Infer Vertical Groundwater-Surface Water Exchange. *Groundwater*, *55*(1), 10–26. https://doi.org/10.1111/gwat.12459

Kalbus, E., Reinstorf, F., & Schirmer, M. (2006). Measuring methods for groundwater - surface water interactions: a review. *Hydrology and Earth System Sciences*, *10*(6), 873–887. https://doi.org/10.5194/hess-10-873-2006

Keery, J., Binley, A., Crook, N., & Smith, J. W.N. (2007). Temporal and spatial variability of groundwater–surface water fluxes: Development and application of an analytical method using temperature time series. *Journal of Hydrology*, *336*(1-2), 1–16. https://doi.org/10.1016/j.jhydrol.2006.12.003

Khalil, M., Sakai, M., Mizoguchi, M., & Miyazaki, T. (2003). Current and prospective applications of zero flux plane (ZFP) method. *Soil Physical Conditions and Plant Growth (Japan)*, *95*, 75–90.

Kim, H., Lee, K.-K., & Lee, J.-Y. (2014). Numerical verification of hyporheic zone depth estimation using streambed temperature. *Journal of Hydrology*, *511*, 861–869. https://doi.org/10.1016/j.jhydrol.2014.02.052

Krause, S., Blume, T., & Cassidy, N. J. (2012). Investigating patterns and controls of groundwater up-welling in a lowland river by combining Fibre-optic Distributed Temperature Sensing with observations of vertical hydraulic gradients. *Hydrology and Earth System Sciences*, *16*(6), 1775–1792. https://doi.org/10.5194/hess-16-1775-2012

Krein, A., & Schorer, M. (2000). Road runoff pollution by polycyclic aromatic hydrocarbons and its contribution to river sediments. *Water Research*, *34*(16), 4110–4115. https://doi.org/10.1016/S0043-1354(00)00156-1

Kurylyk, B. L., & Irvine, D. J. (2016). Analytical solution and computer program (FAST) to estimate fluid fluxes from subsurface temperature profiles. *Water Resources Research*, *52*(2), 725–733. https://doi.org/10.1002/2015WR017990

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147. https://doi.org/10.1137/S1052623496303470

Lapham, W. W. (1989). *Use of temperature profiles beneath streams to determine rates of vertical ground-water flow and vertical hydraulic conductivity*. *US Geological Survey water supply paper: Vol. 2337*. Denver, CO.

Long, J. S., & Freese, J. (2001). *Regression models for categorical dependent variables using stata*. College Station, Tex.: Stata Press.

Lu, C., Chen, S., Zhang, Y., Su, X., & Chen, G. (2017). Heat tracing to determine spatial patterns of hyporheic exchange across a river transect. *Hydrogeology Journal*, *43*, 951. https://doi.org/10.1007/s10040-017-1553-9

765 Luce, C. H., Tonina, D., Gariglio, F., & Applebee, R. (2013). Solutions for the diurnally forced advection-diffusion equation to
766 estimate bulk fluid velocity and diffusivity in streambeds from temperature time series. *Water Resources Research*, *49*(1),
767 488–506. https://doi.org/10.1029/2012WR012380

768 McCallum, A. M., Andersen, M. S., Rau, G. C., & Acworth, R. I. (2012). A 1-D analytical method for estimating surface water-
769 groundwater interactions and effective thermal diffusivity using temperature time series. *Water Resources Research*, *48*(11),
770 951. https://doi.org/10.1029/2012WR012007

771 McCallum, J. L., & Shanafield, M. (2016). Residence times of stream-groundwater exchanges due to transient stream stage
772 fluctuations. *Water Resources Research*, *52*(3), 2059–2073. https://doi.org/10.1002/2015WR017441

773 Munz, M., & Schmidt, C. (2017). Estimation of vertical water fluxes from temperature time series by the inverse numerical
774 computer program FLUX-BOT. *Hydrological Processes*, *31*(15), 2713–2724. https://doi.org/10.1002/hyp.11198

775 Onderka, M., Banzhaf, S., Scheytt, T., & Krein, A. (2013). Seepage velocities derived from thermal records using wavelet
776 analysis. *Journal of Hydrology*, *479*, 64–74. https://doi.org/10.1016/j.jhydrol.2012.11.022

777 Orghidan, T. (1959). Un nou domeniu de viata acvatica subterana 'Biotopul hipreic'. *Buletin Stiintific sectia de Biologie si stiinte*
778 *Agricole si sectia de Geologie si Geografie*, *7*(3), 657–676.

779 Palmer, M. A., Bely, A. E., & Berg, K. E. (1992). Response of invertebrates to lotic disturbance: A test of the hyporheic refuge
780 hypothesis. *Oecologia*, *89*(2), 182–194. https://doi.org/10.1007/BF00317217

781 Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling &*
782 *Software*, *70*, 80–85. https://doi.org/10.1016/j.envsoft.2015.04.009

783 Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of
784 environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232.
785 https://doi.org/10.1016/j.envsoft.2016.02.008

786 Rau, G. C., Andersen, M. S., McCallum, A. M., & Acworth, R. I. (2010). Analytical methods that use natural heat as a tracer to
787 quantify surface water groundwater exchange, evaluated using field temperature records. *Hydrogeology Journal*, *18*(5),
788 1093–1110. https://doi.org/10.1007/s10040-010-0586-0

789 Rau, G. C., Andersen, M. S., & Acworth, R. I. (2012). Experimental investigation of the thermal dispersivity term and its
790 significance in the heat transport equation for flow in sediments. *Water Resources Research*, *48*(3), 951.
791 https://doi.org/10.1029/2011WR011038

792 Rau, G. C., Andersen, M. S., McCallum, A. M., Roshan, H., & Acworth, R. I. (2014). Heat as a tracer to quantify water flow in
793 near-surface sediments. *Earth-Science Reviews*, *129*, 40–58. https://doi.org/10.1016/j.earscirev.2013.10.015

794 Rau, G. C., Cuthbert, M. O., McCallum, A. M., Halloran, L. J. S., & Andersen, M. S. (2015). Assessing the accuracy of 1-D
795 analytical heat tracing for estimating near-surface sediment thermal diffusivity and water flux under transient conditions.
796 *Journal of Geophysical Research: Earth Surface*, *120*(8), 1551–1573. https://doi.org/10.1002/2015JF003466

797 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. (2007). *Sensitivity*
798 *analysis of scientific models*. Hoboken, N.J., Chichester: Wiley; John Wiley [distributor].

799 Schmidt, C., Bayer-Raich, M., & Schirmer, M. (2006). Characterization of spatial heterogeneity of groundwater-stream water
800 interactions using multiple depth streambed temperature measurements at the reach scale. *Hydrology and Earth System*
801 *Sciences Discussions*, *3*(4), 1419–1446.

802 Sebok, E., Engesgaard, P., & Duque, C. (2017). Long-term monitoring of streambed sedimentation and scour in a dynamic
803 stream based on streambed temperature time series. *Environmental Monitoring and Assessment*, *189*(9), 469.
804 https://doi.org/10.1007/s10661-017-6194-x

805 Shanafield, M. A., Hatch, C. E., & Pohll, G. (2011). Uncertainty in thermal time series analysis estimates of streambed water
806 flux. *Water Resources Research*, *47*(3), 951. https://doi.org/10.1029/2010WR009574

807 Silliman, S. E., & Booth, D. F. (1993). Analysis of time-series measurements of sediment temperature for identification of
808 gaining vs. losing portions of Juday Creek, Indiana. *Journal of Hydrology*, *146*, 131–148. https://doi.org/10.1016/0022-
809 1694(93)90273-C

810 Singh, T., Wu, L., Gomez-Velez, J. d., Lewandowski, J., Hannah, D. M., & Krause, S. (2019). Dynamic Hyporheic Zones:
811 Exploring the Role of Peak Flow Events on Bedform-Induced Hyporheic Exchange. *Water Resources Research*, *55*(1), 218–
812 235. https://doi.org/10.1029/2018WR022993

813 Sophocleous, M. (2002). Interactions between groundwater and surface water: The state of the science. *Hydrogeology Journal*,
814 *10*(1), 52–67. https://doi.org/10.1007/s10040-001-0170-8

815 Stallman, R. W. (1965). Steady one-dimensional fluid flow in a semi-infinite porous medium with sinusoidal surface
816 temperature. *Journal of Geophysical Research*, *70*(12), 2821–2827. https://doi.org/10.1029/JZ070i012p02821

817 Stanford, J. A., & Ward, J. V. (1993). An Ecosystem Perspective of Alluvial Rivers: Connectivity and the Hyporheic Corridor.
818 *Journal of the North American Benthological Society*, *12*(1), 48–60. https://doi.org/10.2307/1467685

Stonestrom, D. A., & Constantz, J. (Eds.). (2003). *Circular / U. S. Department of the Interior, U.S. Geological Survey: Vol. 1260*. *Heat as a tool for studying the movement of ground water near streams*. Reston, Va.: U.S. Geological Survey Information Services.

Su, X., Shu, L., Li, W., Lu, C., Zhu, J., Wu, G., Wang, X., Wang, G. (2015). Monitoring temporal patterns of vertical hyporheic flux via distributed temperature sensors. *Proceedings of the International Association of Hydrological Sciences*, *368*, 299–304. https://doi.org/10.5194/piahs-368-299-2015

Suzuki, S. (1960). Percolation measurements based on heat flow through soil with special reference to paddy fields. *Journal of Geophysical Research*, *65*(9), 2883–2885. https://doi.org/10.1029/JZ065i009p02883

Swanson, T. E., & Cardenas, M.B. (2011). Ex-Stream: A MATLAB program for calculating fluid flux through sediment–water interfaces based on steady and transient temperature profiles. *Computers & Geosciences*, *37*(10), 1664–1669. https://doi.org/10.1016/j.cageo.2010.12.001

Taylor, C., Pedegral, D., Young, P., & Tych, W. (2007). Environmental time series analysis and forecasting with the Captain toolbox. *Environmental Modelling & Software*, *22*(6), 797–814. https://doi.org/10.1016/j.envsoft.2006.03.002

Trauth, N., & Fleckenstein, J. H. (2017). Single discharge events increase reactive efficiency of the hyporheic zone. *Water Resources Research*, *53*(1), 779–798. https://doi.org/10.1002/2016WR019488

Voytek, E. B., Drenkelfuss, A., Day-Lewis, F. D., Healy, R., Lane, J. W., & Werkema, D. (2014). 1dtemppro: Analyzing temperature profiles for groundwater/surface-water exchange. *Ground Water*, *52*(2), 298–302. https://doi.org/10.1111/gwat.12051

Wang, L., Jiang, W., Song, J., Dou, X., Guo, H., Xu, S., Zhang, G., Wen, M., Long, Y., Li, Q. (2017). Investigating spatial variability of vertical water fluxes through the streambed in distinctive stream morphologies using temperature and head data. *Hydrogeology Journal*, *25*(5), 1283–1299. https://doi.org/10.1007/s10040-017-1539-7

Wang, Xiaozhe; Smith, Kate; Hyndman, Rob (2006): Characteristic-Based Clustering for Time Series Data. In: *Data Min Knowl Disc* 13 (3), S. 335–364. DOI: 10.1007/s10618-005-0039-x.

Watson, J. A., Cardenas, M. B., Ferencz, S. B., Knappett, P. S.K., & Neilson, B. T. (2018). The effects of floods on the temperature of riparian groundwater. *Hydrological Processes*, *32*(9), 1267–1281. https://doi.org/10.1002/hyp.11504

Winter, T. C., Harvey, J. W., Franke, O. L., & Alley, W. M. (1999). *Ground water and surface water: A single resource* (Repr. (with rev.)). *U.S. Geological Survey circular: Vol. 1139*. Denver, Colo.: U.S. Geological Survey.

Wondzell, S. M., & Swanson, F. J. (1996). Seasonal and Storm Dynamics of the Hyporheic Zone of a 4th-Order Mountain Stream. I: Hydrologic Processes. *Journal of the North American Benthological Society*, *15*(1), 3–19. https://doi.org/10.2307/1467429

Wondzell, S. M., & Swanson, F. J. (1999). Floods, channel change, and the hyporheic zone. *Water Resources Research*, *35*(2), 555–567. https://doi.org/10.1029/1998WR900047

Wu, L., Singh, T., Gomez-Velez, J., Nützmann, G., Wörman, A., Krause, S., & Lewandowski, J. (2018). Impact of Dynamically Changing Discharge on Hyporheic Exchange Processes Under Gaining and Losing Groundwater Conditions. *Water Resources Research*, *53*(1), 3941. https://doi.org/10.1029/2018WR023185

Zarnetske, J. P., Haggerty, R., Wondzell, S. M., & Baker, M. A. (2011). Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone. *Journal of Geophysical Research*, *116*(G1), 758. https://doi.org/10.1029/2010JG001356

Zheng, L., & Bayani Cardenas, M. (2018). Diel Stream Temperature Effects on Nitrogen Cycling in Hyporheic Zones. *Journal of Geophysical Research: Biogeosciences*, *123*(9), 2743–2760. https://doi.org/10.1029/2018JG004412

Captions

Table 1: Parameter values for sensitivity analysis (n: numerical Model, a: analytical Model) and resulting KGEs

Table 2: Kling-Gupta-Efficiencies of time series analysis and slope α of the logistic regression model with groundwater level as predictor for the analytical model and the numerical model (* indicates significance on an alpha of 0.001)

Figure 1: 5 percent best fit cumulative KGE proportion of volumetric heat capacity of the sediment-water system and thermal conductivity for both models with modelling boundaries (slight yellows) and literature values (reddish colors, blue bar) given by Lapham (1989)

Figure 2: Colored scatter plots of parameter influence on the similarity (as KGE) between both temperature-based models and Darcy flux (a: whole time series, b: good fit period of 6 months, c: dominantly downward period of 3 months, d: dominantly upward period of 3 months, e: short-term upward events)

Figure 3: Effect of one-at-a-time variation of thermal properties K and C on flux intensity under downward or upward conditions for both temperature-based models. Darcy flux marked in red.

Figure 4: Left panel: seven year mean daily hyporheic exchange fluxes (line: 30 days moving average, dark shaded area: 5 % best fit interval (5 % of fluxes from MCA with the highest similarity to the benchmark) as 30 days moving average, light shaded: seven year daily standard deviation). Middle panel: boxplots of hyporheic exchange fluxes for summer and winter periods (red line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, red cross: outlier). Right panel: flow duration curves with vertical lines marking the change between upwards and downwards flux for the whole time series (full line), summer (dotted line) and winter (dashed line).

Figure 5: Results of the logistic model approach (p-value < 0.001) for the best fitting depth of the different model approaches (left) and different depth of the analytical model (right) with slope α

Figure 6: Left panel: seven year mean daily hyporheic exchange depth (line: 30 days moving average, dark shaded area: 5 % best fit interval (5 % of fluxes from MCA with the highest similarity to the benchmark) as 30 days moving average, light shaded: seven year daily standard deviation). Middle panel: boxplots of hyporheic zone extension for summer and winter periods (red line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, red cross: outlier). Right panel: exceedance probability curves of hyporheic zone extension for the whole time series (full line), summer (dotted line) and winter (dashed line).

Figure 7: Whisker-Box-Plots depicting the relationship between identified hyporheic exchange depth (shown on the y-axes as zero flux plane) and hydro-meteorological boundary conditions (groundwater (GW) level, surface water (SW) level, discharge, precipitation and mean air temperature) ; *** marks significant effects in MNR.)

Figure 8: Influence of daily mean temperature on the model performance (KGE) of the analytical and numerical model (data is classified in 24 groups according to maximum daily mean air temperature to enable KGE calculation per group).