Model choice impacts the quantification of seasonal hyporheic exchange depths and fluxes

Lara-Maria Schmitgen[1], Reinhard Bierl[1], and Tobias Schuetz[1]

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

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

Key Points:

- Seven-year data analysis reveals stronger seasonal variations of hyporheic exchange fluxes than episodic variations
- Hyporheic exchange depths were stronger influenced by episodic variation than hyporheic exchange fluxes
- The impact of model choice is stronger for quantifying hyporheic exchange depths than for hyporheic exchange flux intensities

Abstract

Direction and depths of hyporheic exchange fluxes at the groundwater - surface water interface are drivers of biogeochemical processes influencing nutrient cycling and water quality. Model concepts on the dynamic relationship between hyporheic exchange fluxes and exchange depth are typically based on the assumption of a linear relationship between both measures. 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; method based on temperature amplitude dampening developed by Hatch et al., 2006) working program scripts were used to solve the one-dimensional conduction-advection-dispersion equation and compute hyporheic flux rates from three vertical sediment water temperature profiles recorded continuously in a small low mountain creek between 2011 and 2017. By comparing the behavior of two differing water temperature-based modelling approaches, dissimilarities in the sensitivity to sediment thermal properties were identified. These differences in parameter responsivity explain deviating behavior of the models regarding exchange flux and depth calculations. We show that the vertical extension of hyporheic exchange depth has a distinctive seasonal pattern over seven years, which differs between the chosen models. Surface water levels, groundwater levels and stream discharges show significant correlations with both flux direction and hyporheic zone extension. In contrast to the numerical modelling approach, analytically derived flux data allowed for establishing a significant relationship between the hydraulic gradient observed at a nearby groundwater well and simulated hyporheic exchange depths.

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 depends on the intensity and direction of hyporheic exchange fluxes (Boulton et al., 2010). Although, the impact is reduced in bigger streams with higher discharge (Wondzell, 2011), studies show that smaller first- and second-order streams constitute up to 80 % of the length of river networks and are of great importance to the whole stream network (Wohl, 2017; Downing et al. 2012). Hyporheic exchange flux is defined as 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) or even deeper. Furthermore, there is not only a relationship between the river's width and the hyporheic zone extension (Boano et al., 2008, Kim et al., 2014), but river bed topography, stream curvature and the ambient groundwater gradient affect the dimension as well (Cardenas et al., 2004). The thickness of the hyporheic zone is positively correlated with the ambient hydraulic head (Fox et al., 2014; Marzadri et al., 2016). 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 processes 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 inter-annual variations of seasonal fluctuations stay rather invisible if short time periods of days or weeks are used. 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 estimate 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). This 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-sedimentsystems 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. Although, there are also analytical solutions using point in time temperature-depth profiles instead of a time series (Irvine et al., 2019), but these are not considered further in this work. The physical properties of the sediment-water-system are either quantified, e.g. in laboratory experiments (e.g. Wang et al., 2017) or they are estimated (coupled with uncertainty analysis) in other 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 many attempts to model processes in the hyporheic zone itself using temperature-based methods. 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 identify the effect of seasonal variation and episodic hydrological extremes on hyporheic exchange depth under varying hydro-meteorological conditions. By applying parameter sensitivity analysis with the results of a recently released numerical model (Munz & Schmidt, 2017) in comparison to those of a well-known analytical model approach (Hatch et al., 2006) we

I. identify the effect of estimated thermal properties on the resulting exchange fluxes,
II. identify seasonal and short-term dynamics of hyporheic flow direction and intensity,
III. determine hyporheic exchange depths variations in relation to hydraulic boundary conditions.

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} between 2011 - 2017. 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 material stemming from the underlying aquifers, which consist of low metamorphic slate and quartzite deposits. Banzhaf and Scheytt (2009) reported a hydraulic conductivity (K_f) of $8*10^{-4}$ m s^{-1} for the underlying aquifer using the grain curve method. To consider method-based

bias a correction factor of 2 was applied (German National Association for Water Management, Waste Water and Waste Management (DWA) set of rules A 138, 2005) and the resulting K_f of $1.6*10^{-4}$ m s^{-1}used further on. Continuous measurements of sediment water temperature- depth profiles were carried out using three temperature lances (Umwelt- und Ingenieurtechnik (UIT) GmbH Dresden), with a resolution of 0.032 °C and an accuracy of 0.1 °C. The temperature lances were positioned in succession in a riffle-pool sequence with 7 m distance between the first and second lance and 5 m distance between the second and third (Figure 1). 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 resampled using linear interpolation. 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). 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 using a SEBAPLUS radar measurement system mounted on a bridge (standard temporal resolution: 15 min (2011-2014) or 5 min (2015-2017)). 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, to receive a continuous time series at the measurement site. Daily mean air temperature and daily precipitation sum were obtained from the nearby (~2 km) Petrisberg weather monitoring station (operated by: German Weather Service (DWD), published under DWDs Climate Data Center).

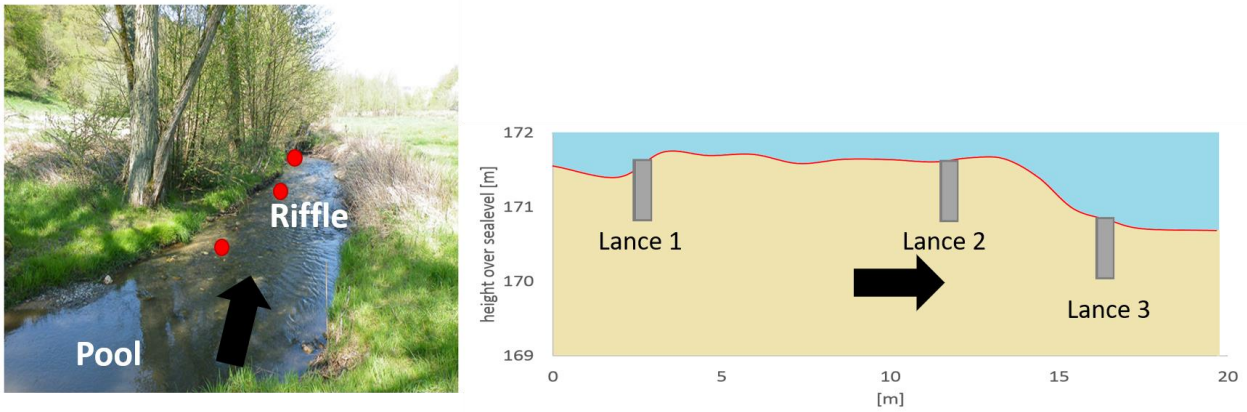


Figure 1: Left panel: Photograph showing the location of the three lances (red dots) in the riffle-pool-sequence. Right panel: Simplified profile of the riffle-pool- sequence with the positions of the three lances. Black arrow denotes the flow direction.

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.

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 are based on the one-dimensional conduction-advection-dispersion equation (modified after Stallman, 1968):

$$\frac{\partial T}{\partial t} = K_e \frac{\partial^2 T}{\partial z^2} - q \frac{C_w}{C} * \frac{\partial T}{\partial z}$$

(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 sediment water system in J m^{-3} $°C^{-1}$, which can be described by:

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

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 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, which is not able to run if the distance between the uppermost and lowermost sensor is smaller than the mean depth between those sensors, only 14 combinations were realized. Due to numerical instability some runs produced unreasonable flux calculations for some timesteps. We eliminated fluxes with an intensity higher $5*10^{-5}$ m/s in both directions and calculated for each iteration the percentage of days where a reasonable flux calculation was possible (Table **1**).

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). An extension described by Irvine (2017b) makes it even possible to estimate K_e from a reliable time period where a combined approach (phase shift and amplitude dampening) is used and fluxes are recalculated with the additional information. However, a method based on amplitude dampening alone was chosen, to exclude the uncertainty introduced by a shifting phase lag (Irvine et al., 2015) or the necessity to identify a reliable time period for K_e estimation (Irvine et al., 2017b). 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:

$$q = \frac{C}{C_w} * \left(\frac{2K_e}{\Delta z} \ln A_r + \sqrt{\frac{a + v^2}{2}} \right) \quad \textbf{(4)}$$

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:

$$a = \sqrt{v^4 + \left(\frac{8\pi K_e}{P}\right)^2} \quad \textbf{(5)}$$

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 was assumed as a constant, C was used as a direct input parameter solving Equation 4 instead of calculating it using Equation 2, as it is normally implemented in VFLUX. This procedure allows for both models 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. 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 potentially possible to compute 21 different hyporheic exchange flux time series with the analytical model for a single temperature lance. Within this analysis, we considered only those depth combinations, which could be realized using the numerical model. Therefore, the combinations where a sensor and its directly neighboring sensor are used are excluded, as the numerical model always needs a sensor in between. With the chosen sample size, vertical fluxes were calculated in 2-hour intervals and were resampled using linear interpolation to daily values fitting the numerically derived flux time series.

2.2.2 Darcy flux calculation

Applying Darcy's equation (1856):

$$q = Kf * i \quad \textbf{(6)}$$

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 as an external reference, which is used in addition to the inter-model comparison evaluating the overall temporal distribution of up-welling and down-welling water fluxes at the study site.

2.3 Model evaluation and sensitivity analysis

To bypass errors in hyporheic flux calculation originating in uncertain thermal parameter assignment, one-at-a-time (OAT) and all-at-a-time (AAT) SA as Monte Carlo Analysis (MCA, after Pianosi et al. (2016)) were applied. In OAT SA only one parameter is varied, while all other parameters are kept at a base value. A global OAT SA was performed using the change in mean flux and standard deviation to assess the effect of the varied parameters individually. To quantify directly the influence of variation in C and K on hyporheic flux intensity under different conditions, distinct values for K (set between the boundaries used for MCA) 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 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 reference.
In the following the Kling-Gupta-Efficiency (KGE) is used 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 analytically modelled hyporheic exchange flux and the numerically modelled hyporheic exchange flux, meaning a high KGE value represents a good fit between the models, whereas a low or negative KGE is attributed with a poor model fit.

MCA with 1000 model realizations (as recommended by Pianosi et al., 2016) was performed for each of the 14 depth combinations for all three lances using randomly distributed parameter values for K (0.837 J s^{-1} m^{-1} °C^{-1} – 3.349 J s^{-1} m^{-1} °C^{-1}) and C (1.506*10^6 J m^{-3} °C^{-1} – 3.682*10^6 J m^{-3} °C^{-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 and test both models under the extreme condition that there is no information on thermal parameters available. 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. The fluxes with the highest similarity between both models were identified for all possible depth combinations per model calculating the KGE.

To assess the influence of K and C under specific boundary conditions, KGE values were calculated for the days with downward flux and the days with upward flux in addition to the performance evaluation over the whole time series..

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.

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 neighboring 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 14 depth combinations at each of the three lances. For each depth the individual best fit between both models was chosen.

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. 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 served as indicators for episodic variation. We define the episodic variation index as:

$$episodic\ variation\ index = 1 - \frac{\max(dm) - \min(dm)}{2 * mean\ (dstd)} \quad \textbf{(7)}$$

Following equation 7 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) the two sample Kolmogorov-Smirnov test and Wilcoxon-Mann-Whitney test were used, whereas the Kruskal-Wallis test combined with Dunn's posthoc test was used to compare more than two samples (e.g. different lances).

Monthly mean values were calculated for all time-series enabling 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) was performed on the down-sampled time series. The r function stl_plus (Hafen, 2016) was used for this. 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 3) 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 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 (< 0.15 m, 0.15 m to 0.3 m, > 0.3 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. The effects of those boundary conditions on flux direction were identified using mnr with flux direction as ordinal response.

3 Results

3.1 Model evaluation and sensitivity analysis

The highest similarity between both models with a KGE of 0.46 is computed for a mean flux depth of 0.135 m at lance 2, using values for K (thermal conductivity) of 3.29 J s^{-1} m^{-1} $°C^{-1}$ and for C (volumetric heat capacity) of 1.70*10^6 J m-3 $°C^{-1}$. The best fit between the models at lance 1 (KGE of 0.31) was found using the sensor pairs 0.05 m and 0.45 m. However, here only 46 % of the time series were realized in the numerical model, so that the second best fit (KGE of 0.15) at 0.135 m was chosen for lance 1 with a K of 2.9 J s^{-1} m^{-1} $°C^{-1}$ and a C of 1.55*10^6 J m-3 $°C^{-1}$. At lance 3 the fit between the models was not as good (KGE of 0.07) at a mean depth of 0.085 m with a K of 1.326 J s^{-1} m^{-1} $°C^{-1}$ and a C of 3.48*10^6 J m-3 $°C^{-1}$. In general, the inter-model comparison shows that using the same parameter combinations in both models results in substantial differences in most depths (Table **1**).
To evaluate the models individually, the effect that a varied parameter has on the mean flux and its standard deviation was assessed. In the models C has a positive effect on mean flux, which is especially pronounced in the analytical model. A negative relationship appears when K is increased. This effect on the mean is especially strong for the numerical model. C_w has almost no effect. A slight increase in β (only applicable in the analytical model) also has no effect; only with very high β values the flux is drastically reduced. Therefore, in the AAT SA C_w and β were kept at base value. The standard deviations of hyporheic exchange flux increase with an increase in C for both models and this effect is again more pronounced when using the analytical model. Standard deviations also increase with an increase in K. However, the increase in the numerical model is more dominant here.
Considering the numerically and analytically modelled fluxes in parallel for specific days under upwelling or downwelling conditions, respectively, the influence of parameter variation on flux intensity becomes clear (Figure 2**Error! Reference source not found.**). A change in only one thermal parameter has the same effect direction 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, although the effect is much stronger for the analytical model. An increase in K results in higher flux intensities for both models under down- and upwelling conditions. C has a slightly decreasing effect on the flux intensity 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 by comparing the fluxes calculated with different models using the same parameter values, resulting in mostly negative KGE (Table **1** and Figure 4).
Comparing the models with each other (Figure 3), the calculated optima for both thermal parameters appear to be outside the realistic boundaries (Lapham, 1989) for lance 2. The optima for C at lance 1 and 3 have a quite similar range, whereas the optima for K at these lances differ again significantly. No clear trend between an increase in mean flux depth (chosen sensors) and location and dispersion parameters describing the optima for K and C are visible.
Comparing the multiplicative relation between the model parameters (Figure 4), it can be seen that the combined optimum of the two parameters is distributed differently among the three lances. Especially between lance 1 and 3 optimum parameter combinations appear to be in opposing areas of the parameter range. Lance 2 is covering both areas but a higher model similarity is achieved using parameter combinations with a high value for K and a low value for C. The difference between times with differing flux direction is also pronounced. However, the number of runs which result in a positive KGE decreases when only downward fluxes are considered at lance 2 and increases when only upward fluxes are considered at lance 1. In general RSA confirms the visual results. Considering the whole time series C and K appear to have a fairly equal effect, with a higher effect of C at lance 1 and 2 and a higher effect of K at lance 3. Under upward conditions, K has a higher effect (d-stats of 0.63 and 0.6 for K and 0.17 and 0.2 for C) for lance 1 and 2. There are not enough behavioral results to calculate d-stats under specific conditions at lance3. C is more influential with downward flux behavior at lance 1 and 2.

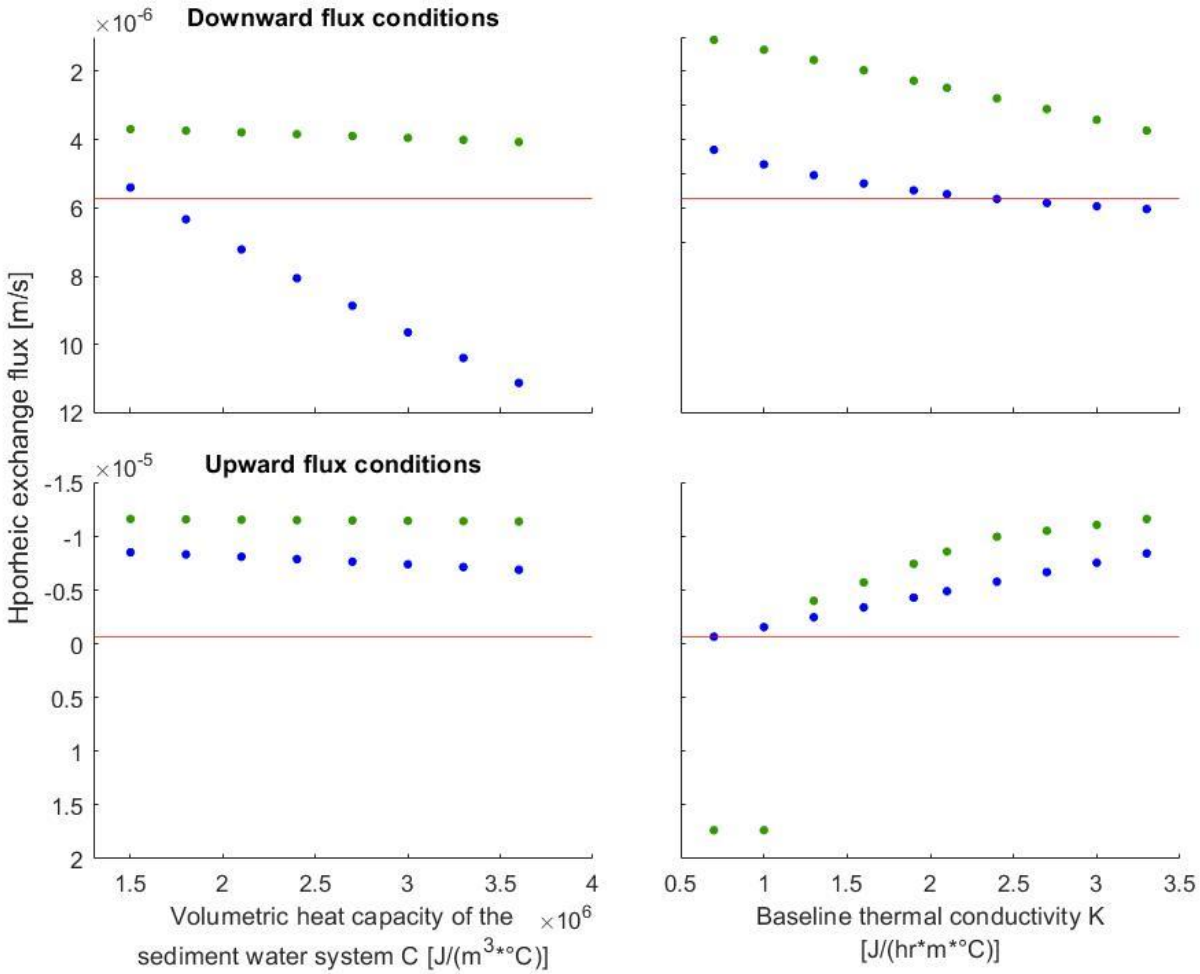


Figure 2: Effect of one-at-a-time variation of thermal properties K and C on flux intensity under downward or upward conditions for both temperature-based models. Darcy flux marked in red, analytical model in blue and numerical model in green.

Table **1**: Kling-Gupta-Efficiencies (KGE) **of time series analysis between the analytical model and the numerical model (AvN) including the percentage of the time series which could be realized using the numerical model (% real) for AvN as well as the maximum percentage which could be achieved using 1000 parameter combinations (max poss. % real). The number of sensors used in the numerical model is given behind the respective depth.**

	Lance 1			Lance 2			Lance 3		
Depth [m]	**AvN**		**max poss.**	**AvN**		**max poss.**	**AvN**		**max poss.**
	KGE	**% real**	**% real**	**KGE**	**% real**	**% real**	**KGE**	**% real**	**% real**
0.06 (3)	-0.27	92	92	-0.14	56	100	-0.07	96	100
0.085 (4)	-0.09	1	1	0.22	98	100	0.07	100	100
0.1 (3)	-0.49	2	4	-0.11	0	4	-0.89	5	5
0.135 (5)	0.15	98	99	0.46	100	100	-3.56	76	99
0.15 (4)	0.15	42	84	0.21	71	71	-0.63	46	93
0.175 (3)	-0.21	1	5	-0.11	2	6	-1.71	2	3

0.235 (6)	-0.33	97	98	0.23	100	100	-3.11	100	100
0.25 (5)	0.31	46	95	0.01	99	100	-0.73	32	100
0.275 (4)	-0.09	21	52	-0.31	64	68	-0.71	21	59
0.3 (3)	-0.04	0	9	-0.10	3	31	-0.66	6	24
0.335 (7)	-0.09	96	96	-0.12	99	100	-0.03	100	100
0.35 (6)	-0.33	46	98	-0.52	97	100	-0.52	53	100
0.375 (5)	-0.49	6	22	-0.44	3	85	-0.59	15	30
0.4 (4)	-0.22	2	12	-0.35	1	28	-0.49	7	23

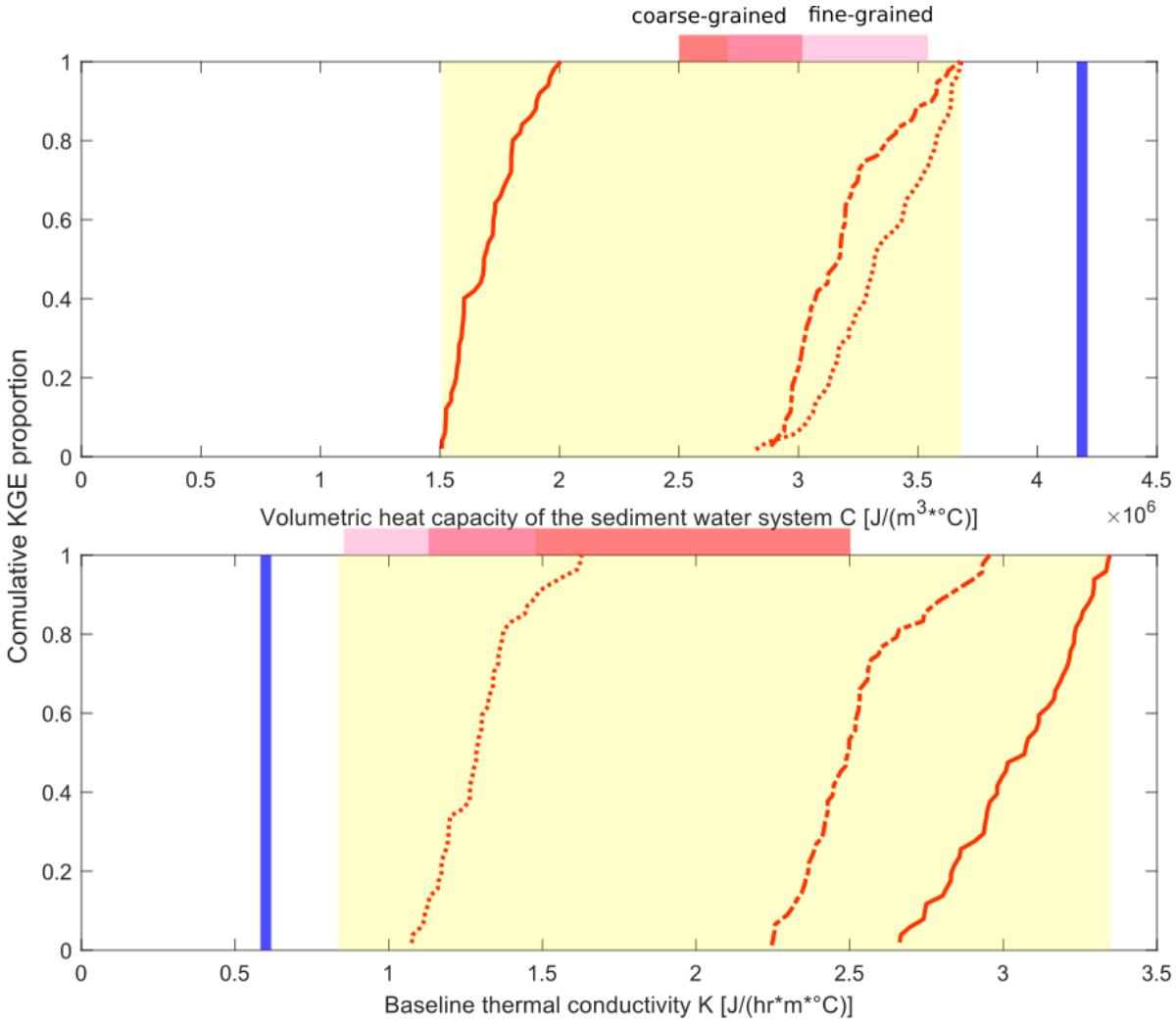


Figure 3: 5 percent best fit cumulative KGE proportion of volumetric heat capacity of the sediment-water system and thermal conductivity for both models with modelling boundaries (slight yellows) and literature values (reddish colors (light: fine-grained sediment, darker: coarse-grained sediment), blue bar: water) given by Lapham (1989) (dashed line: lance 1, line: lance 2, dotted line: lance 3).

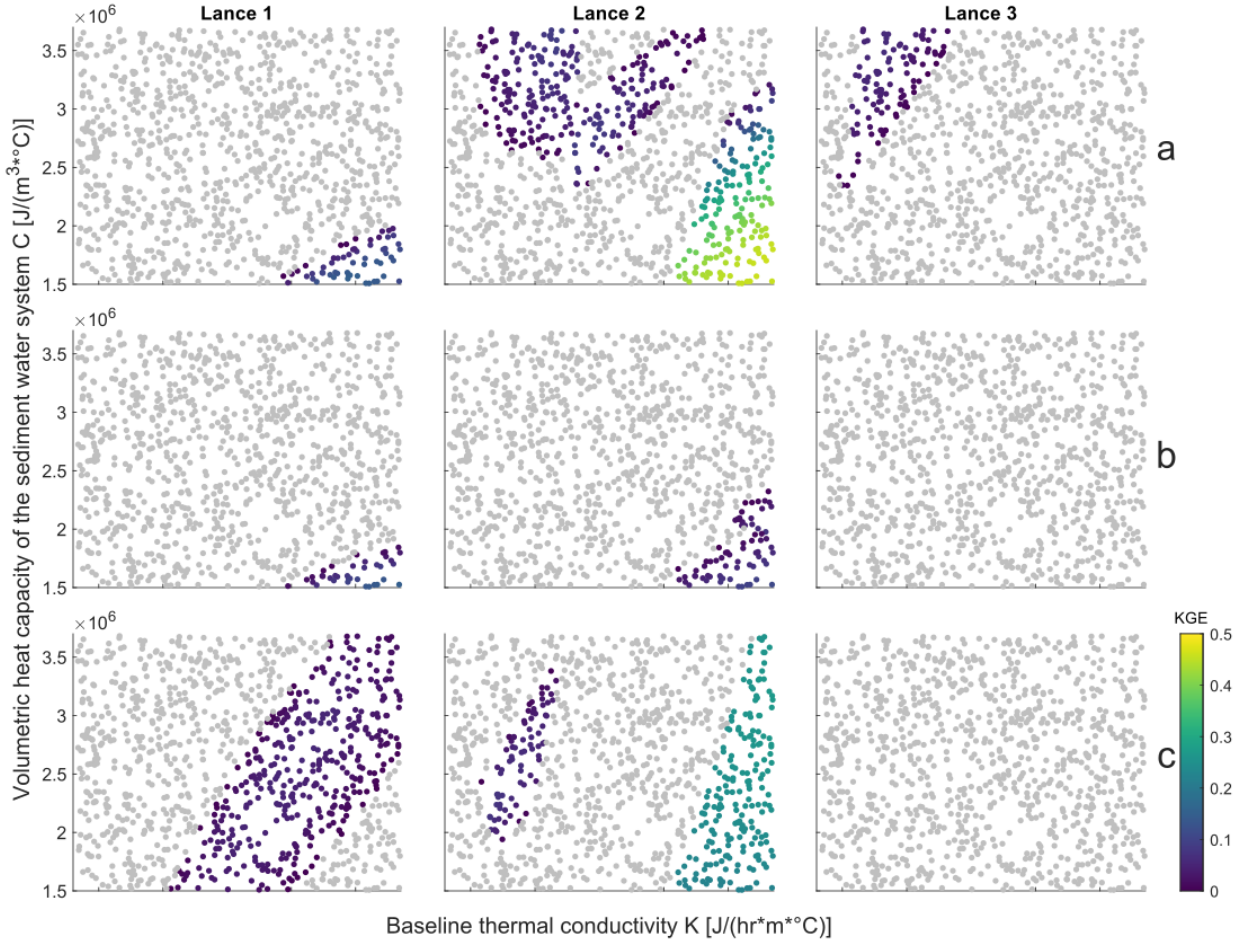


Figure 4: Colored scatter plots of parameter influence on the similarity (as KGE) between both temperature-based models (a: whole time series, b: downward flux, c: upward flux)

3.2 Process analysis

3.2.1 Temporal variation of hyporheic exchange fluxes

Temporal variation of hyporheic exchange fluxes was determined by using the best-fit fluxes between both model approaches (Figure 5). The mean values of the temperature-based modelled fluxes calculated at lance 3 are comparable to the mean Darcy fluxes with $3.33*10^{-6}$ m s^{-1}. At lance 1 the temperature-based models propose smaller flux intensities, while at lance 2 the mean direction is changed (Table 2). The standard deviations of the temperature based modelled fluxes are up to 4 times higher than of the Darcy flux (130 %).

Darcy flux has a smaller seasonal strength than the temperature-based hyporheic exchange fluxes (Table 2). The stronger seasonality is expressed with stronger downward fluxes in summer and a stronger tendency for upward fluxes in winter at lance 2 and 3 (Figure 5). This is contradicted at lance 1, where both models propose more intense downward flux in winter and upward fluxes in summer. 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 regardless which model or lance was chosen. The variation in seasonal strength is small in both models when the thermal parameters are varied only around the optimum, but when they are varied over the entire range, there is a large variation in seasonality for the numerical model. The episodic variation index is with 0.12 positive for Darcy flux, but negative for all temperature based modelled fluxes.

In comparison to the Darcy flux time series both temperature-based models result in longer times of dominating upward fluxes (Figure 5). Darcy flux calculations propose upward fluxes in 18 % of the cases, whereas the analytical

and numerical model estimate distinctively larger proportions (Table 2). All models propose a higher proportion of upward flux behavior in winter than in the summer period for lance 2 and 3, but not at lance 1.
Regarding the effect of the boundary conditions on the flux direction tested using mnr, there are large differences between the lances. An increase in discharge increased the probability of upward flux at lance 2 significantly, while it reduced the probability at lance 1 significantly when the analytical model was used. At lance 3 the same tendency as at lance 2 can be observed but is not significant. A rise in surface water level has a significant positive effect on the chance for upward flux at lance 1, while it is significantly negative at lance 2 and 3 (significant here only for AM). Groundwater level promotes exactly the opposite effects. The reaction to precipitation is mostly indifferent.

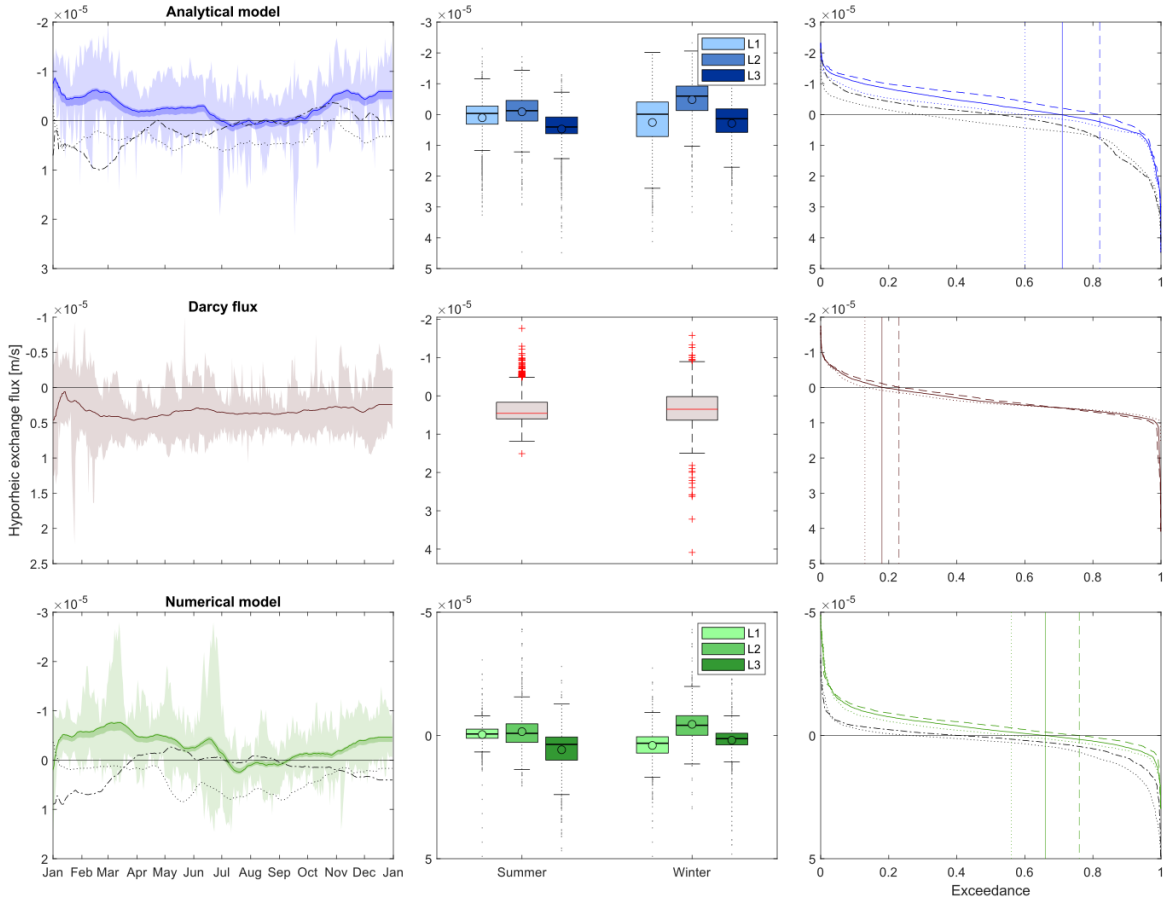


Figure 5: Left panel: seven year mean daily hyporheic exchange fluxes (line: 30 days moving average, dark shaded area: 5 % best fit interval (5 % of fluxes from MCA with the highest similarity to the benchmark) as 30 days moving average, light shaded: seven year daily standard deviation) for lance 2. Seven year mean daily hyporheic exchange fluxes as 30 days moving average are added in black for lance 1 (dashed line) and lance 3 (dotted line) Middle panel: boxplots of hyporheic exchange fluxes for summer and winter periods (black line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, point: mean, grey dot: outlier). Right panel: flow duration curves with vertical lines marking the change between upwards and downwards flux for the whole time series (coloured full line), summer (coloured dotted line) and winter (coloured dashed line) for lance 2 and for the whole time series of lance 1 (black dashed line) and lance 3 (black dotted line).

Table 2: Summary of seasonal and episodic variations of hyporheic exchange fluxes

		Mean flux [m/s]	Seasonal strength	Episodic variation	Percentage upward flux [%]
Lance 1	AM	$1.78*10^{-6}$	0.45	-0.61	52
	NM	$1.80*10^{-6}$	0.53	-0.81	40

Lance 2	AM	$-2.82*10^{-6}$	0.36	-0.41	71
	NM	$-3.01*10^{-6}$	0.36	-0.37	66
Lance 3	AM	$3.84*10^{-6}$	0.23	-0.31	29
	NM	$3.97*10^{-6}$	0.27	-0.37	26
Darcy		$3.33*10^{-6}$	0.11	0.12	18

406

407 3.2.2 Quantification and temporal variation of hyporheic exchange depths

408 The zero-flux method was applied on analytically and numerically modelled fluxes with 14 depths combinations as
409 input data sets. Using the numerical model results in smaller extent of hyporheic exchange depths for lance 2 and 3
410 in comparison to the depth derived from the analytical model (Table 3). Kolmogorow-Smirnow tests and Wilcoxon-
411 Mann-Whitney tests reveal significant (alpha of 0.001) differences in depths calculated using the different model
412 approaches. Furthermore, Kruskal-Wallis combined with Dunn's posthoc test (alpha of 0.001) confirms significant
413 differences between lance 3 and the other two lances when the analytical model and between lance 1 and the other
414 two lances when the numerical model is used.
415 Seasonal strength is always higher for the analytical modelled time series (Table 3). The episodical variation indices
416 are again all negative. The seasonal strength is strongly changed for both models if parameters are varied even slightly.
417 However, this time the analytical model seems to be more influenced by parameter variation in contrast to the
418 observations made for the effect on seasonality when considering hyporheic flux.
419 The analytical model shows a deeper extent of the hyporheic zone in summer, whereas it is the opposite using the
420 numerical model (Figure 6). Kolmogorow-Smirnow tests (alpha of 0.001) confirm significant differences between the
421 seasons for all lances except for lance 3 if the numerical model is used. In addition, strong variations during the whole
422 year are observable.
423 MNR results in significant effects (alpha of 0.01) of surface water levels and river discharges on hyporheic exchange
424 depth for nearly all cases (Figure 7). However, the effect is contrary in both models. Groundwater levels are significant
425 predictors for the hyporheic exchange depth calculated by the numerical model at all lances, while they are only
426 significant at lance 3 for the analytical model. Again, the effect is contrary between both temperature-based models.
427 Air temperature is a significant predictor at lance 1 (for both models) and using the analytical model at lance 3. No
428 statistical relationship has been detected between precipitation and hyporheic zone extension.
429 Comparing the relationships between Darcy derived hyporheic exchange flux and exchange depths, for each of the
430 lances we find a significant (alpha of 0.001) correlation with the analytical model results (Figure 8). Although the
431 explanatory power of these correlations is limited a clear positive relationship between increasing exchange depth
432 with increased downwards flux intensity is visible. The relationship between exchange flux and depth is indifferent
433 with even lower explanatory power for the numerical model.
434

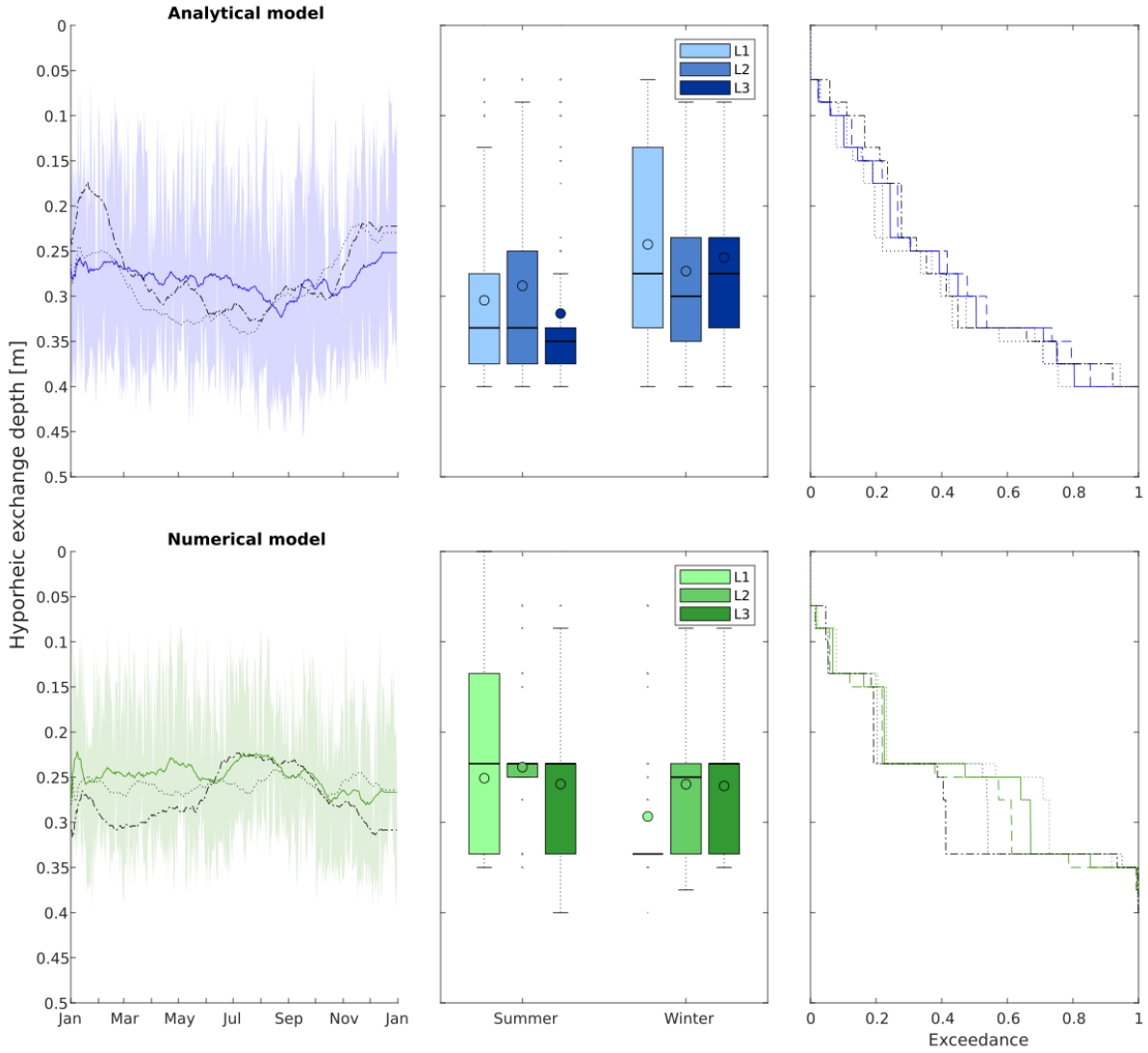


Figure 6: Left panel: seven year mean daily hyporheic exchange depth (line: 30 days moving average, light shaded: seven year daily standard deviation) for lance 2. Seven year mean daily hyporheic exchange fluxes as 30 days moving average are added in black for lance 1 (dashed line) and lance 3 (dotted line). Middle panel: boxplots of hyporheic zone extension for summer and winter periods (black line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, point: mean, grey dot: outlier). Right panel: exceedance probability curves of hyporheic zone extension for the whole time series (coloured full line), summer (coloured dotted line) and winter (coloured dashed line) for lance 2 and for the whole time series of lance 1 (black dashed line) and lance 3 (black dotted line).

Table 3: Parameters describing seasonal and episodic variation of hyporheic exchange depth

		Mean depth [m]	Standard deviation [%]	Seasonal strength	Episodic variation
Lance 1	AM	0.24	40	0.63	-0.34
	NM	0.27	32	0.43	-0.28
Lance 2	AM	0.28	36	0.3	-0.01
	NM	0.25	33	0.23	-0.14
Lance 3	AM	0.29	36	0.55	-0.21
	NM	0.26	32	0.27	-0.03

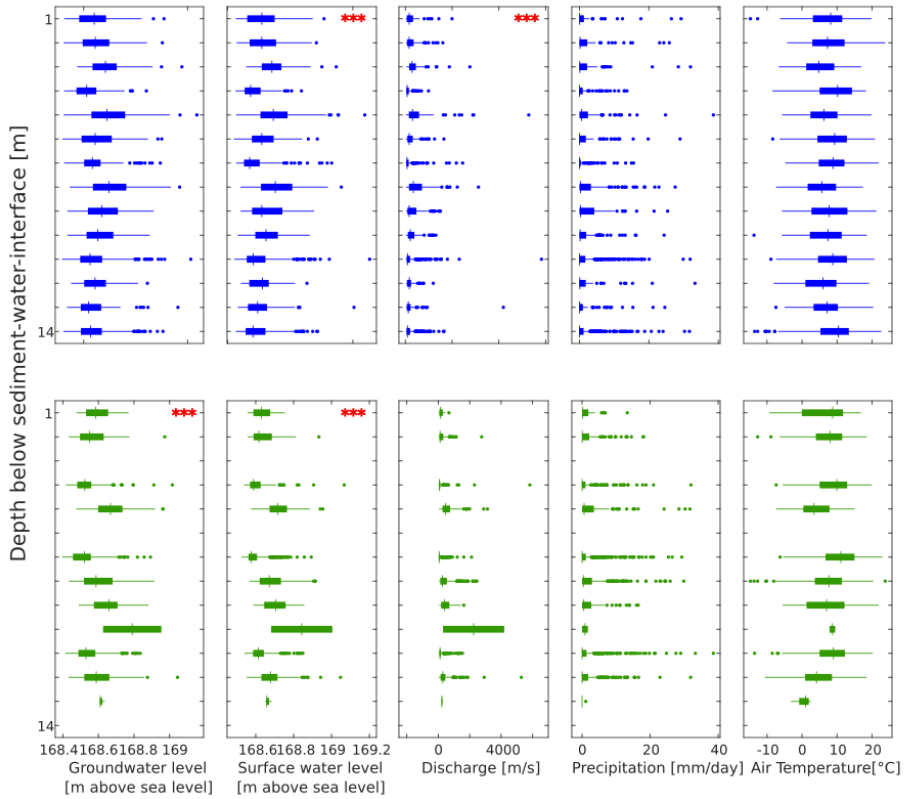


Figure 7: Whisker-Box-Plots depicting the relationship between identified hyporheic exchange depth (shown on the y-axes as zero flux plane) and hydro-meteorological boundary conditions (groundwater (GW) level, surface water (SW) level, discharge, precipitation and mean air temperature). Depth values from 1 to 14 give increasing depth steps according to Table 1.; * marks significant effects in MNR.)**

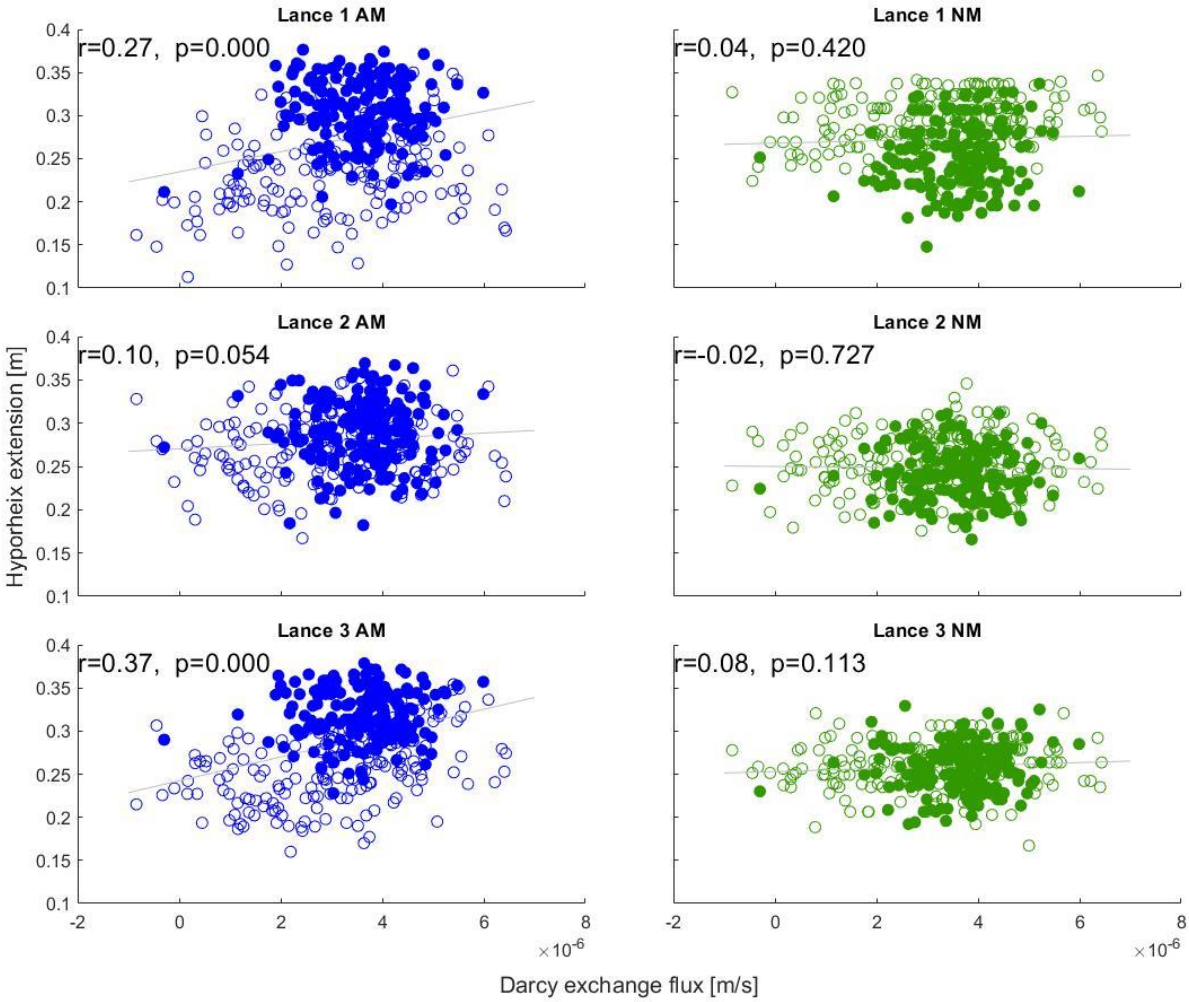


Figure 8: Relationship between seven year daily mean hyporheic exchange flux and seven year daily mean hyporheic exchange depth (open circles: winter period, filled circles: summer period).

4 Discussion

4.1 Model evaluation and sensitivity analysis

We have identified sensor combinations where the analytical and numerical models have a higher similarity. For both models there are different factors mentioned in the literature which increase model uncertainty. Gordon et al. (2012) show an increasing modelling uncertainty with increasing sensor spacing for the analytical model approach. However, minimizing sensor spacing is limited by the amplitude ratio reaching unity, which would result in improbable high flux estimations (Irvine et al., 2017). Munz and Schmidt (2017) stated that an increase in the number of sensors used in their numerical model led to a higher accuracy. This would naturally come along with a wider sensor spacing window between the two sensors used as boundaries. These contradicting implications for reducing uncertainty can explain some of the differences between the models. Regarding absolute sensor depth as a factor, it is important to consider effects of erosion and sediment remobilization. Although the uppermost sensor (0.02 m) at lance 2 was temporarily affected by this, the inclusion of it still led to the most similar results between the models. Our data gives the impression that the extinction depth, at which the diurnal signal is no longer detectable is of greater concern. A very deep sensor depth (here 0.65) partially did not show any diurnal signal. Far more shallow extinction depths have been documented as well, e.g. 0.2 m under upwelling flux conditions (Briggs et al., 2014).

The impact of variations in 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.
The identification of C and K as sensitive parameters for the numerical model and a slightly higher sensitivity of C, especially under downward flux conditions, confirms findings by Munz and Schmidt (2017). Yet, as expected, K has a stronger effect under exclusively upward flux conditions. Furthermore, it could also be validated that there is a correlated effect of both parameters (Munz & Schmidt, 2017). The high sensitivity of the analytical model to K is also demonstrated in other studies (i.e. Gordon et al., 2012). A second key parameter n is identified and its impact directly attributed to its influence on C (Gordon et al., 2012), which results in a high sensitivity of the parameter C. Shanafield et al. (2011) varied n, K and C_s in the analytical model of Hatch et al. (2006) using MCA and report an effect of K_e, for which they conclude that uncertainty in this parameter estimation is biased especially under gaining conditions.
Considering the effect of the thermal parameters on absolute flux values, it appears worth mentioning that the parameter C has a very strong influence on the analytical model, especially under downward flux conditions (Figure 2), but not on the numerical approach, leading to drastic deviations in calculated fluxes, although the same input is used. Under upward flux conditions, the response to a change in C does not result in such divergent behavior. In general, an increase in K increases flux intensity autonomously from flux direction. The numerical model seems to be stronger affected by a variation in K, which again intensifies under upward flux conditions. This can generally be explained by the divergent directions of upward advection along with the flow direction and downward propagation of the temperature signal through conductive transport. However, this is not a sufficient explanation why this effect seems to be stronger for the numerical model.
We attribute the main differences between the fluxes calculated with different modeling approaches to a different sensitivity to variation in thermal parameters, which is probably due to the nature of the mathematical solutions used. The main difference between the two approaches is that the analytical method depends on the relationship between the temperature amplitudes caused by a daily temperature cycle and its propagation through the sediment, whereas the numerical model explicitly excludes these sinusoidal signals and is instead based on absolute temperature differences. It needs to be noted, that some of the calculated combined optimum ranges (lance 2) for the thermal parameters are outside the boundaries set by Lapham (1989) and seem questionable for real sediments, although there is some literature reporting comparable low values for C and high values for K (Goto & Matsubayashi, 2009). However, this is not influencing the main message that both temperature-based methods result in different flux results, even with the same input parameter values.
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 or significantly different flux calculations even if thermal parameters are known. In conclusion, uncertain thermal properties combined with two different model approaches can cause differences of some magnitude (Figure 2**Error! Reference source not found.**).

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 the sinusoidal signal can be strongly attenuated in winter. In this study, the sinusoidal signal was always detectable even in winter. Hence, we could not find a clear difference in model performance between the seasons, which holds true for all lances. However, for lances 1 and 2, similarities between both models in use where stronger in summer. As the analytical model directly depends on distinctive amplitude differences, which can be minimal in winter, we advise, to use the here presented temperature-based methods, especially the analytical one depending on amplitude dampening carefully at study sites with pronounced cold winters.
An increased upward flux behavior was simulated for the winter periods by the temperature-based models at lance 2 and 3 in correspondence with increased discharges and groundwater levels as well as catchment saturation during the rainy season. An additional effect could be attributed to the significant change in hyporheic zone extension during the year, an effect also reported by Boano et al. (2014) or the diminished diurnal signal of water temperatures.
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. Furthermore our 1-D modelling approaches are not able to detect multidirectional flow patterns, which can also lead

to a deviation between the benchmark and the modelled fluxes (Briggs et al. 2012). However, it is still worthy to use Darcy flux as a comparison especially for the longterm seasonality, as it is a convenient and cost-efficient method to estimate groundwater surface water exchange.

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. We observed combined increase in downward flux intensities and stream flow only at lance 1, whereas at lance 2 and 3 increasing stream flow coincided with increasing upward flux intensities. This can be explained with the location along the riffle-pool-sequence, where a higher discharge would induce increased downward fluxes at elevated positions, where lance 1 is situated (Figure 1). Another reason could be the reaction of the near surface groundwater table, observable as a fast raise of the groundwater table during event flow, which in turn resulted in upward flux behavior. Probably, this phenomenon is also dependent on the connectivity and the shape of the whole hyporheic corridor (Stanford & Ward, 1993) and the overall antecedent moisture content of the whole catchment.

In general, the temperature-based models propose a distinctively higher seasonal variation than the Darcy fluxes who were endemically dominated by episodic events. A reason might be the direct reaction of gradient-based Darcy fluxes to velocity and celerity of increased lateral water input occurring during flood-pulse-driven events (McDonnell & Beven, 2014), whereas the temperature-based models are estimated based on energy transport velocities alone.

4.2.2 Identification of hyporheic exchange depths and its temporal variation

The absolute values of hyporheic exchange depths calculated at our study site range from 0.06 m to 0.4 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 a maximum of 0.03 m only slightly between the models. However, the calculated depths differed significantly between the models on a daily basis. Depending on the model choice the differences between the three lances changed. The numerical model proposes on average a deeper extension at the head of the riffle, than on the crest or in the following pool, whereas the analytical model proposes a more extended zone on the crest and in the pool. The difference between the lances can of course be explained with the location and the typical flow patterns through a riffle pool sequence (e.g. Cranswick et al., 2014, Gariglio et al., 2013), but it is striking that both models propose a different ranking between the lances regarding the hyporheic zone extension.

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. However, both models contradict each other: while the analytical model is proposing a smaller hyporheic zone extension in winter, the numerical model proposes a deeper extension in winter (especially visible at lance 1). Although significant differences in calculated hyporheic exchange depth were observed, we did not anticipate such a diverging behavior between the models regarding the seasonal behavior as observed at lance 1. Therefore, it is very important to consider small scale riverbed formations in positioning of measurement equipment and record changes as it seems that they can influence different temperature-based model approaches in different ways. We identified a strong seasonal component in hyporheic zone extension. The strength of the seasonal component is always slightly higher when the analytical model is used to calculate exchange depth. Using the zero-flux method very high standard deviations were calculated throughout the year. Hence, there must be short timescale drivers influencing hyporheic exchange depth as well. Although seasonal variation is comparable high or higher (episodic variation index $\approx<$ 0), MNR could identify short timescale drivers like discharge and water levels. Calculated episodic variation was comparable between the models.

As expected, higher groundwater levels came along with a shallower hyporheic zone extension as also described by Boano et al. (2008). However, this effect was only observed using the analytical model, while the numerical model proposed the contrary. 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 on the other hand only observed using the numerical model. The positive effect of an increasing ambient hydraulic head on hyporheic zone extension as mentioned in other studies (Fox et al., 2014; Marzadri et al., 2016) could be reproduced by using the analytical model approach to identify hyporheic exchange depths. However, the still low to moderate explanatory power of all correlations suggests that other processes and boundary conditions might contribute to the observed combinations of hyporheic exchange flux intensities and 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 groundwater might lead to differences in temperature distributions and thus simulated hyporheic exchange depths.

Since the extent of the hyporheic zone 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, regardless of the positioning of each lance. Both temperature-based models tend to simulate 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, but a strong impact on the simulated exchange depths. The simulated seasonality of hyporheic exchange zone depths shows a characteristic behavior for each model type. While the analytical model revealed a larger extent of the hyporheic zone in summer than in winter, the results of the numerical model were contradicting. Episodic event dynamics resulted in a strong variability of hyporheic zone extension during the whole year. These short-term effects could be attributed to combined effects of changes in groundwater and surface water levels and river discharges. However, the models again propose a different nature in correlations.

Finally, we show that thermal methods must be applied carefully because different algorithms lead to divergent flux calculations, resulting in different estimates for the extent of the hyporheic zone, even when the same input data are used. The influence of thermal properties should in no case be neglected. Hence, these results might contribute in future to an improved understanding of the functioning of river self-purification processes in the hyporheic zone of anthropogenically impacted rivers.

601

Acknowledgments, Samples, and Data

603 This study was supported by the German Research Foundation (DFG) – Project-No.: SY 12/27-2.
604 We thank Bernhard Fink and Andreas Kurtenbach for their efforts during the sensor installation and the data
605 acquisition. Data is published as open access publication under Pangaea.de (doi: 10.1594/PANGAEA.922956)
606
607

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.

Cardenas, M. B., Wilson, J. L., & Zlotnik, V. A. (2004). Impact of heterogeneity, bed forms, and stream curvature on subchannel hyporheic exchange. *Water Resources Research*, *40*(8). https://doi.org/10.1029/2004WR003008

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

Cranswick, R. H., Cook, P. G., Shanafield, M., & Lamontagne, S. (2014). The vertical variability of hyporheic fluxes inferred from riverbed temperature data. *Water Resources Research*, *50*(5), 3994–4010. https://doi.org/10.1002/2013WR014410

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

Downing, J. A., Cole, J. J., Duarte, C. M., Middelburg, J. J., Melack, J. M., Prairie, Y. T., Kortelainen, P., Striegl, R. G.; McDowell, W. H., Tranvik, L. J. (2012). Global abundance and size distribution of streams and rivers. *Inland Waters*, *2*(4), 229–236. https://doi.org/10.5268/IW-2.4.502

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

Fox, A., Boano, F., & Arnon, S. (2014). Impact of losing and gaining streamflow conditions on hyporheic exchange fluxes induced by dune-shaped bed forms. *Water Resources Research*, *50*(3), 1895–1907. https://doi.org/10.1002/2013WR014668

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, R,; Kang, Y.,; Montero-Manso, P.,; Talagala, Th., Wang, E., Yang, Yet 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. (2017a). Using diurnal temperature sgnals to infer vertical groundwater-surface water exchange. *Groundwater*, *55*(1), 10–26. https://doi.org/10.1111/gwat.12459

Irvine, D. J., Briggs, M. A., Cartwright, I, Scruggs, C. R., & Lautz. L. K. (2017b). Improved Vertical Streambed Flux Estimation Using Multile Diurnal Temperature Methods in Series. *Groundwater*, *55*(1), 73–80. https://doi.org/10.1111/gwat.12436

Irvine, D. J., Kurylyk, B. L., Briggs, M. A. (2019). Quantitative guidance for efficient vertical flow measurements at the sediment-water interface using temperature-depth profiles. Hydrological Processes, 34(3), 649-661. https://doi.org/10.1002/hyp.13614

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

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

Marzadri, A., Tonina, D., Bellin, A., & Valli, A. (2016). Mixing interfaces, fluxes, residence times and redox conditions of the hyporheic zones induced by dune-like bedforms and ambient groundwater flow. *Advances in Water Resources*, *88*, 139–151. https://doi.org/10.1016/j.advwatres.2015.12.014

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

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

McDonnell, J. J., & Beven, K. (2014). Debates-The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resources Research*, *50*(6), 5342–5350. https://doi.org/10.1002/2013WR015141

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Stanford, J. A., & Ward, J. V. (1993). An Ecosystem Perspective of Alluvial Rivers: Connectivity and the Hyporheic Corridor. *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.

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, X.,; Smith, K., Hyndman, R. (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.

Wohl, E. (2017). The significance of small streams. *Frontiers of Earth Science*, *11*(3), 447–456. https://doi.org/10.1007/s11707-017-0647-y

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

Wondzell, S. M. (2011). The role of the hyporheic zone across stream networks. *Hydrological Processes*, 25(22), 3525-3532. https://doi.org/10.1002/hyp.8119

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

Error! Reference source not found.
Table **1**: Kling-Gupta-Efficiencies (KGE) **of time series analysis between the analytical model and the numerical model (AvN)**

Table 2: Summary of seasonal and episodic variations of hyporheic exchange fluxes

Table 3: Parameters describing seasonal and episodic variation of hyporheic exchange depth

Figure 1: Left panel: Photograph showing the location of the three lances (red dots) in the riffle-pool-sequence. Right panel: Simplified profile of the riffle-pool- sequence with the positions of the three lances. Black arrow denotes the flow direction.

Figure 2: 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, analytical model in blue and numerical model in green.

Figure 3: 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 (light: fine-grained sediment, darker: coarse-grained sediment), blue bar: water) given by Lapham (1989)

Figure 4: Colored scatter plots of parameter influence on the similarity (as KGE) between both temperature-based models (a: whole time series, b: downward flux, c: upward flux

Error! Reference source not found.
Figure 5: 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) for lance 2. Seven year mean daily hyporheic exchange fluxes as 30 days moving average are added in black for lance 1 (dashed line) and lance 3 (dotted line) Middle panel: boxplots of hyporheic exchange fluxes for summer and winter periods (black line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, point: mean, grey dot: outlier). Right panel: flow duration curves with vertical lines marking the change between upwards and downwards flux for the whole time series (coloured full line), summer (coloured dotted line) and winter (coloured dashed line) for lance 2 and for the whole time series of lance 1 (black dashed line) and lance 3 (black dotted line).

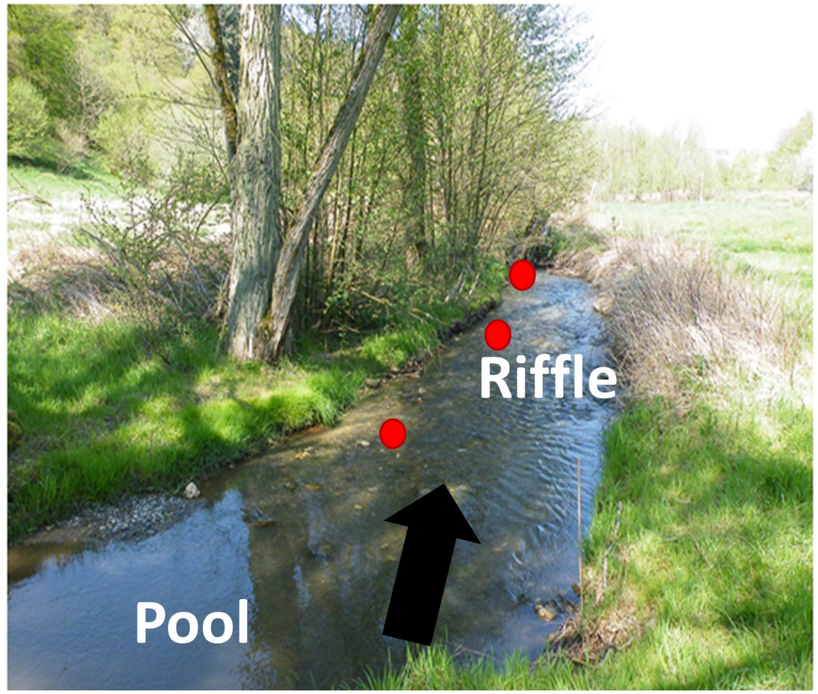
Error! Reference source not found. Figure 6: Left panel: seven year mean daily hyporheic exchange depth (line: 30 days moving average, light shaded: seven year daily standard deviation) for lance 2. Seven year mean daily hyporheic exchange fluxes as 30 days moving average are added in black for lance 1 (dashed line) and lance 3 (dotted line). Middle panel: boxplots of hyporheic zone extension for summer and winter periods (black line: median, black box: first and third quartile, upper and lower whisker: 99 % coverage, point: mean, grey dot: outlier). Right panel: exceedance probability curves of hyporheic zone extension for the whole time series (coloured full line), summer (coloured dotted line) and winter (coloured dashed line) for lance 2 and for the whole time series of lance 1 (black dashed line) and lance 3 (black dotted 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). Depth values from 1 to 14 give increasing depth steps according to Table **1**.; *** marks significant effects in MNR.)

Figure 8: Relationship between seven year daily mean hyporheic exchange flux and seven year daily mean hyporheic exchange depth (open circles: winter period, filled circles: summer period).

Error! Reference source not found.

Figure 1.


Riffle
Pool

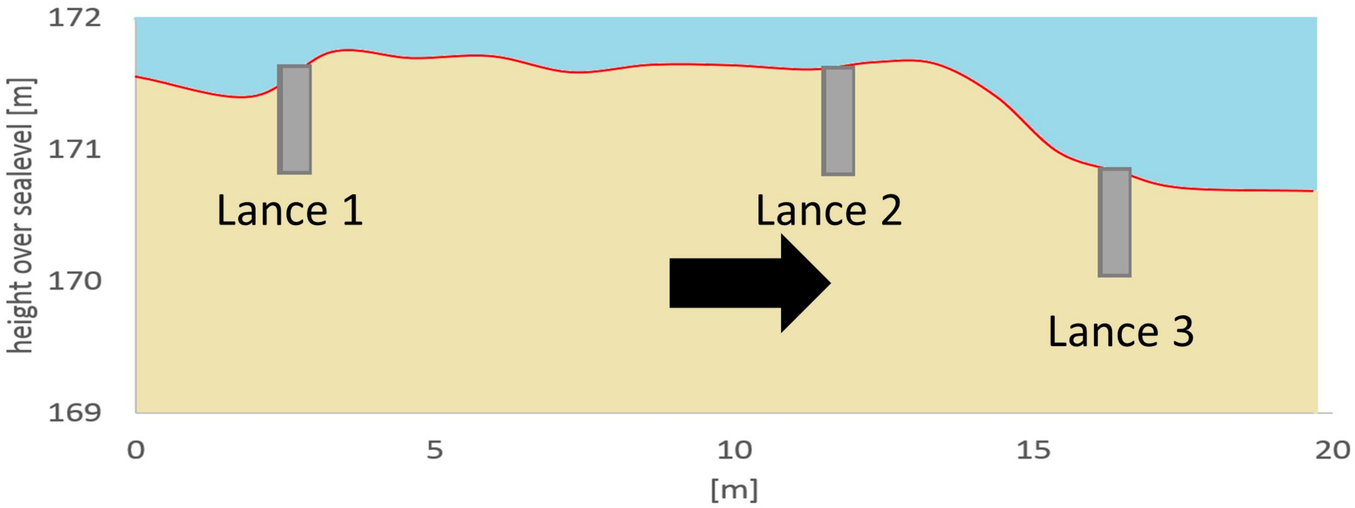

172
171
170
169
height over sealevel [m]
Lance 1
Lance 2
Lance 3
0
5
10
15
20
[m]

Figure 2.

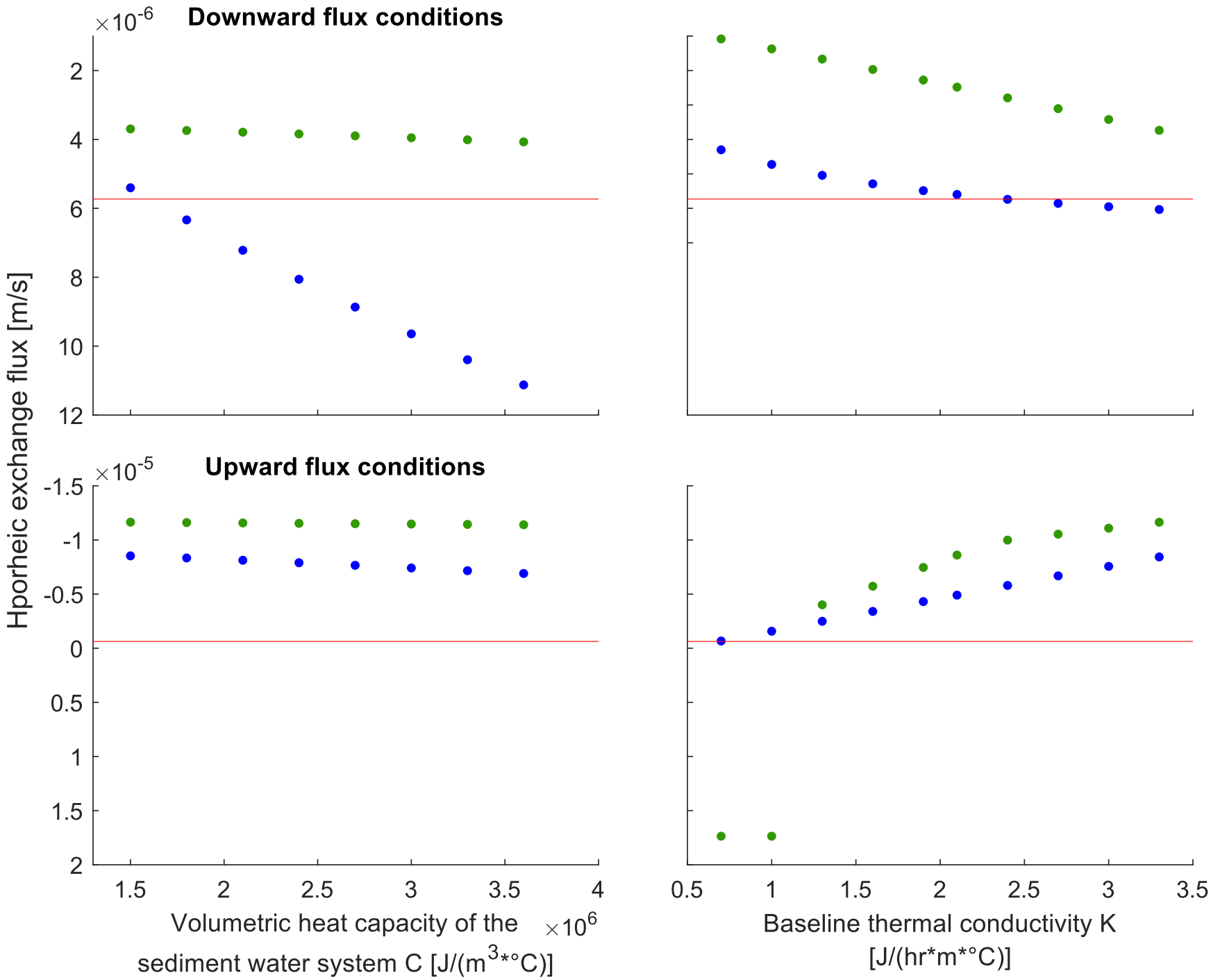

Downward flux conditions
Upward flux conditions
Hporheic exchange flux [m/s]
×10^-6
×10^-5
2
4
6
8
10
12
-1.5
-1
-0.5
0
0.5
1
1.5
2
1.5
2
2.5
3
3.5
4
×10^6
Volumetric heat capacity of the sediment water system C [J/(m^3*°C)]
0.5
1
1.5
2
2.5
3
3.5
Baseline thermal conductivity K [J/(hr*m*°C)]

Figure 3.

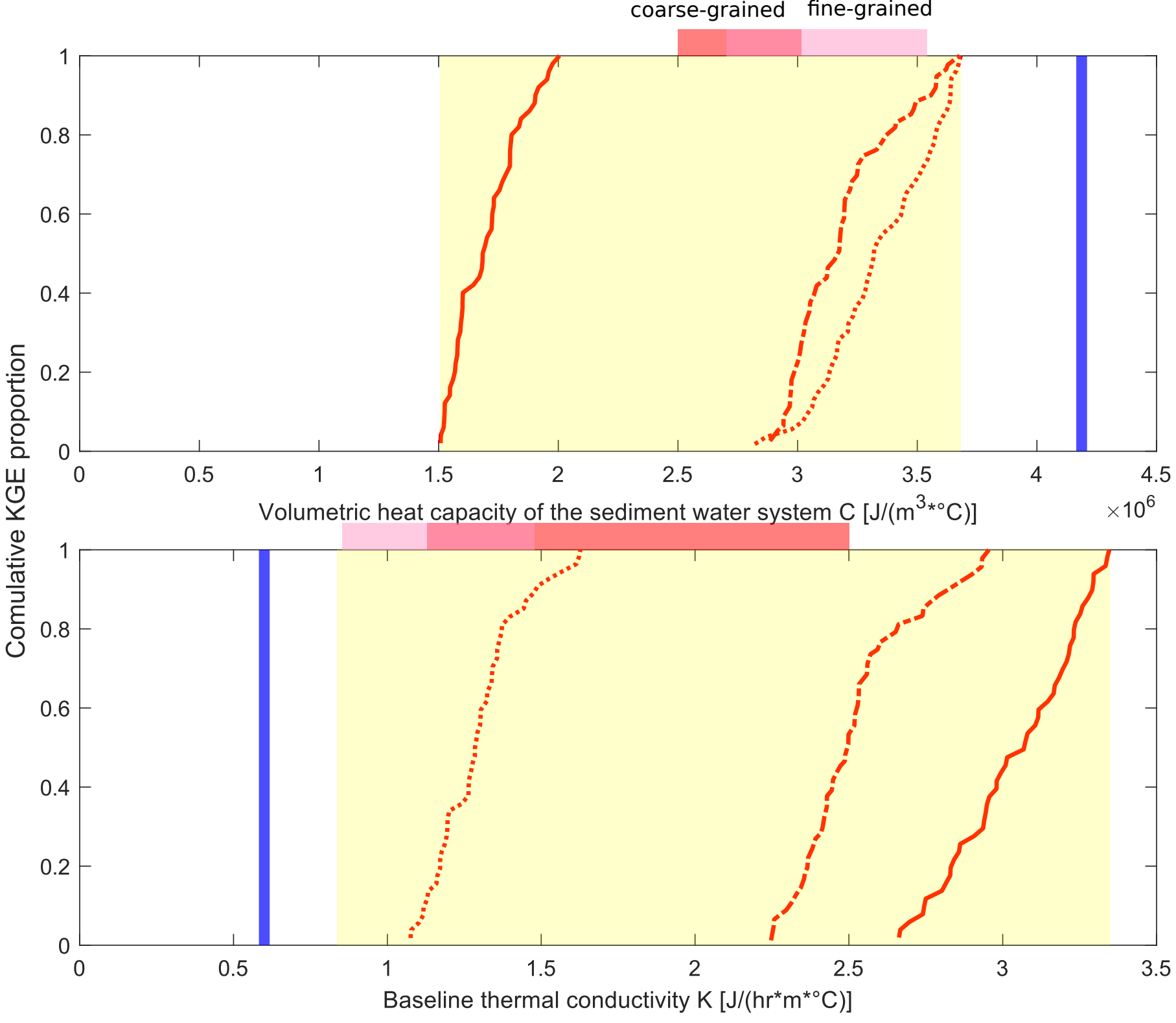

coarse-grained
fine-grained
Comulative KGE proportion
Volumetric heat capacity of the sediment water system C [J/(m3*°C)]
×10^6
Baseline thermal conductivity K [J/(hr*m*°C)]

Figure 4.

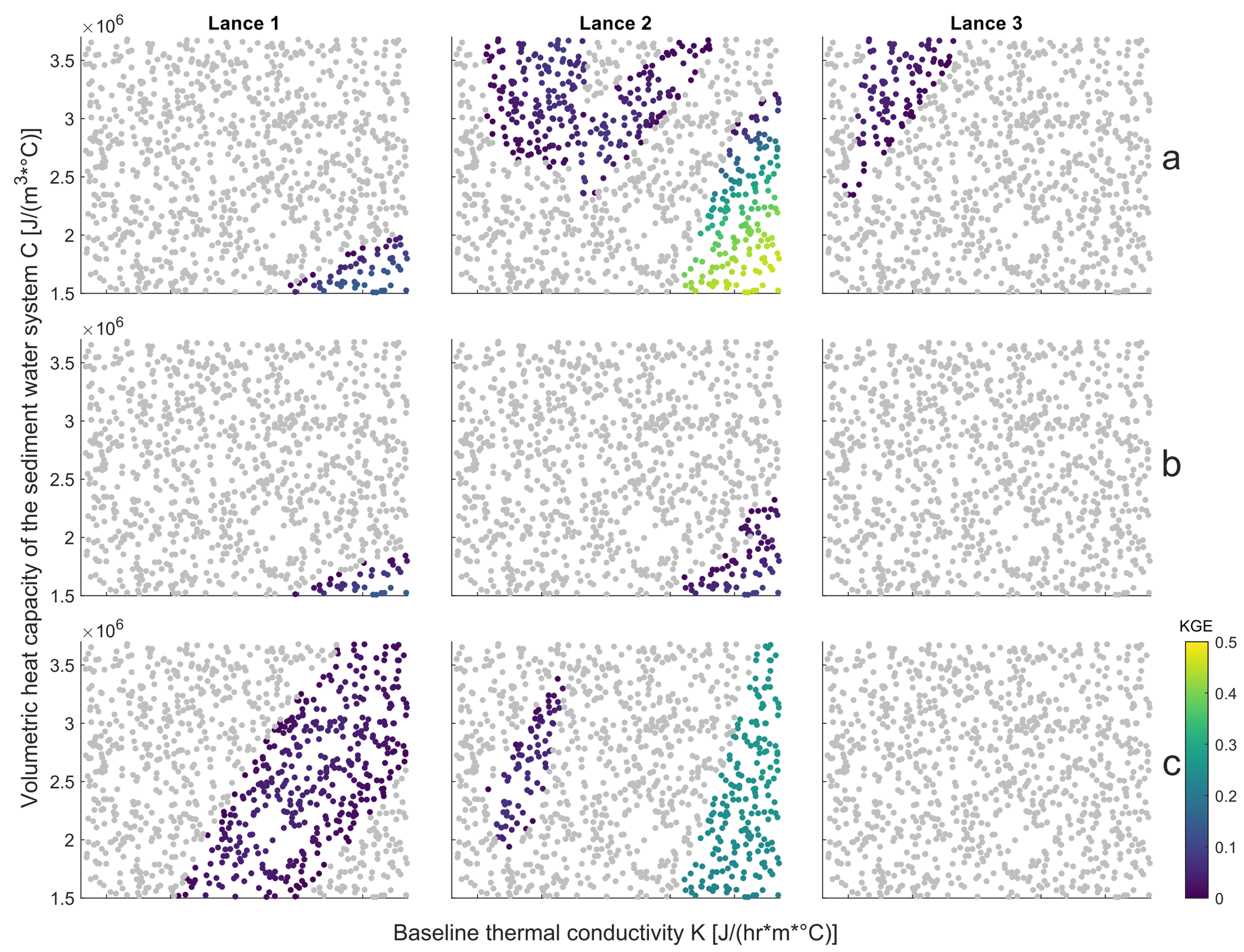

Lance 1
Lance 2
Lance 3
a
b
c
×10^6
3.5
3
2.5
2
1.5
Volumetric heat capacity of the sediment water system C [J/(m^3*°C)]
Baseline thermal conductivity K [J/(hr*m*°C)]
KGE
0.5
0.4
0.3
0.2
0.1
0

Figure 5.

Analytical model

Darcy flux

Numerical model

Hyporheic exchange flux [m/s]

L1
L2
L3

Summer
Winter

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan

Exceedance

Figure 6.

Analytical model

Numerical model

Hyporheic exchange depth [m]

0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Jan Mar May Jul Sep Nov Jan

L1
L2
L3

Summer Winter

0 0.2 0.4 0.6 0.8 1

Exceedance

Figure 7.

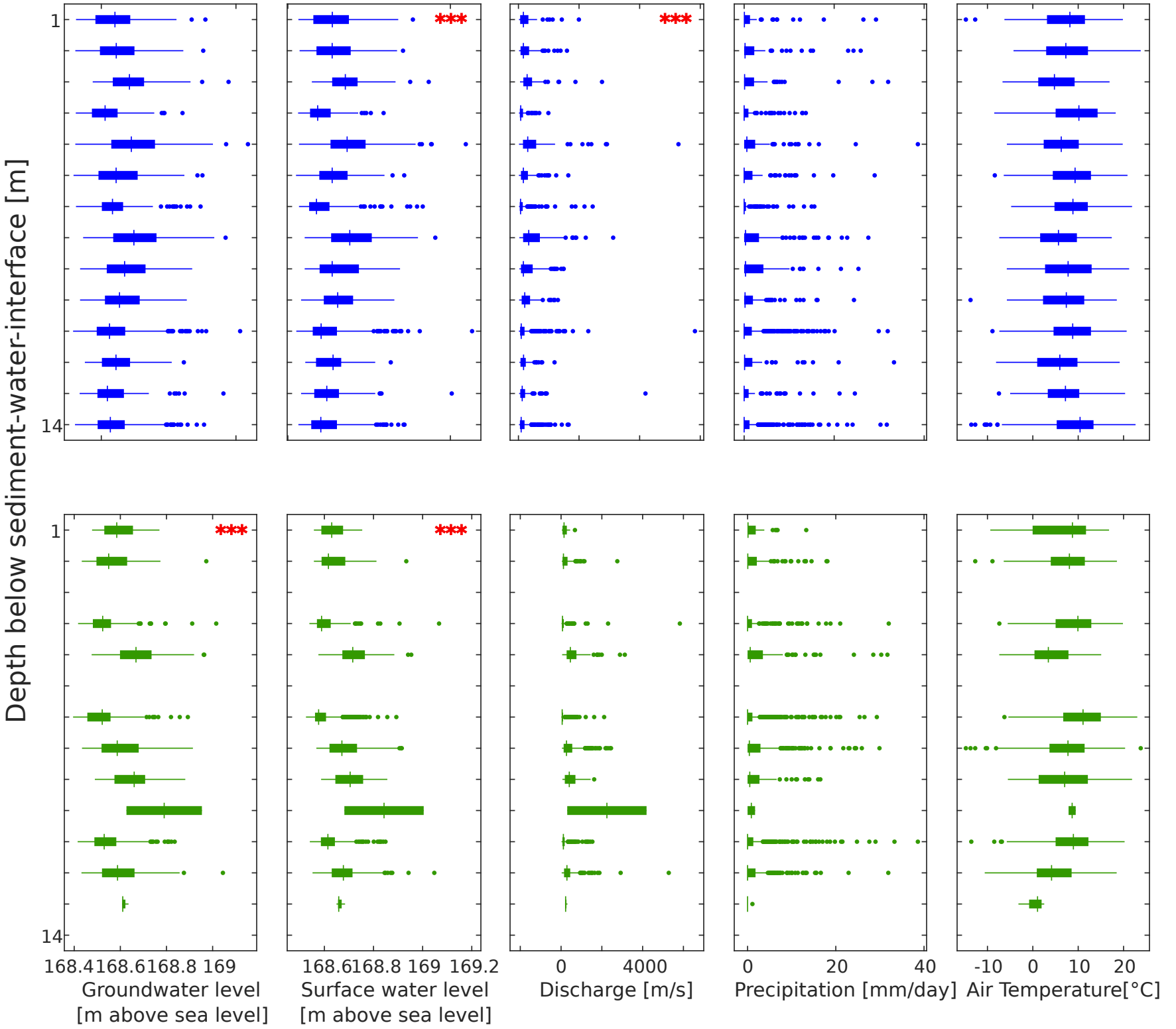
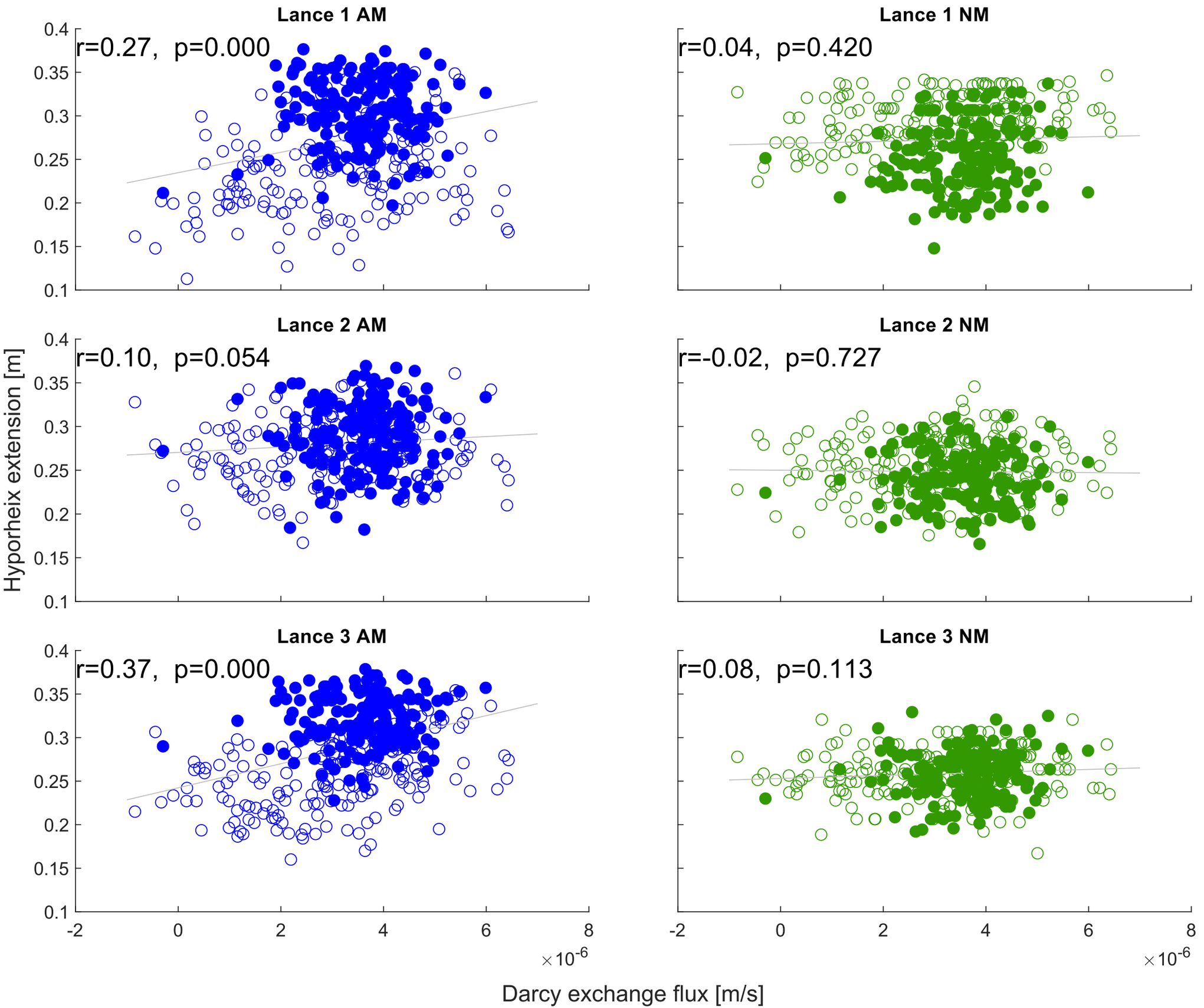

Depth below sediment-water-interface [m]
1
14
***
***
***
***
168.4 168.6 168.8 169
Groundwater level
[m above sea level]
168.6 168.8 169 169.2
Surface water level
[m above sea level]
0 4000
Discharge [m/s]
0 20 40
Precipitation [mm/day]
-10 0 10 20
Air Temperature[°C]

Figure 8.


Lance 1 AM
r=0.27, p=0.000
Lance 1 NM
r=0.04, p=0.420
Lance 2 AM
r=0.10, p=0.054
Lance 2 NM
r=-0.02, p=0.727
Lance 3 AM
r=0.37, p=0.000
Lance 3 NM
r=0.08, p=0.113
Hyporheix extension [m]
0.4
0.35
0.3
0.25
0.2
0.15
0.1
-2
0
2
4
6
8
×10-6
Darcy exchange flux [m/s]