**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**

Dylan J. Irvine ^{a, b, *}, Laura K. Lautz ^a, Martin. A. Briggs ^c, Ryan P. Gordon ^a, Jeffrey M.
McKenzie ^d

^a Syracuse University, Department of Earth Sciences, 204 Heroy Geology Laboratory,
Syracuse, NY 13244, USA.

^b School of Earth, Atmosphere and Environment, Monash University, Clayton, VIC 3800,
Australia.

^c U.S. Geological Survey, Office of Groundwater, Branch of Geophysics, 11 Sherman Place,
Unit 5015 Storrs CT 06269

^d McGill University, Earth and Planetary Sciences, 3450 University Street, Montreal, Quebec,
Canada H3A 2A7.

* Corresponding Author:

Email dylan.irvine@monash.edu

For Journal of Hydrology

21 **Abstract**

22 Vertical fluid exchange between surface water and groundwater can be estimated using
23 diurnal signals from temperature time series methods based on amplitude ratios (A_r), phase
24 shifts ($\Delta\phi$), or combined use of both ($A_r\Delta\phi$). The A_r , $\Delta\phi$, and $A_r\Delta\phi$ methods are typically
25 applied in conditions where one or more of their underlying assumptions are violated, and the
26 reliability of the various methods in response to non-ideal conditions is unclear. Additionally,
27 $A_r\Delta\phi$ methods offer the ability to estimate thermal diffusivity (κ_e) without assuming any
28 thermal parameters, although the value of such output has not been broadly tested. The A_r , $\Delta\phi$,
29 and $A_r\Delta\phi$ methods are tested under non-steady, 1D flows in sand column experiments, and
30 multi-dimensional flows in heterogeneous media in numerical modeling experiments. Results
31 show that, in non-steady flow conditions, estimated κ_e values outside of a plausible range for
32 streambed materials (0.028 to $0.180 \text{ m}^2 \text{ d}^{-1}$) coincide with time periods with erroneous flux
33 estimates. In heterogeneous media, sudden changes of κ_e with depth also coincide with
34 erroneous flux estimates. When (known) fluxes are variable in time, poor identification of $\Delta\phi$
35 leads to poor flux estimates from $\Delta\phi$ and $A_r\Delta\phi$ methods. However, when fluxes are steady, or
36 near zero, $A_r\Delta\phi$ methods provide the most accurate flux estimates. This comparison of A_r , $\Delta\phi$
37 and $A_r\Delta\phi$ methods under non-ideal conditions provides guidance on their use. In this study,
38 $A_r\Delta\phi$ methods have been coded into a new version of VFLUX, allowing users easy access to
39 recent advances in heat tracing.

40

41 **1. Introduction**

42 Anderson (2005), in her review of heat as a groundwater tracer, called for groundwater
43 temperature to be measured and interpreted more widely. Since the Anderson (2005), review,
44 the study of surface water-groundwater interactions has been an area that has seen a large

45 increase in the use of water temperature data. Following on from seminal work by Stallman
46 (1965) and important steps from theory to application by Lapham (1989) and Goto et al.
47 (2005), diurnal temperature signals in particular have become widely used. Temperature time
48 series approaches, based on variations in diurnal signals, have been used to investigate fluid
49 exchange between streams and aquifers in a range of settings (Lapham 1989; Stonestrom and
50 Constantz 2003; Hatch et al., 2006; Rau et al., 2010; Briggs et al., 2013; Briggs et al., 2014;
51 McCallum et al., 2014; Birkel et al., 2015), to determine changes in streambed hydraulic
52 conductivity with time (Hatch et al., 2010), and to produce spatial maps of exchange fluxes
53 (Lautz and Ribaud, 2012; Gordon et al., 2013; Irvine and Lautz, 2015). The use of heat as a
54 tracer to determine the fluid exchange between surface water and groundwater is not
55 restricted to stream settings. For example, the use of heat as a tracer has also been applied in
56 wetlands (e.g. Hunt et al. 1996; Bravo et al. 2002), and in lakes (e.g. Krabbenhoft and
57 Babiarz 1998). Similarly, the use of temperature signals is not restricted to analysis of the
58 diurnal signal, as annual signals contain similar fluid flux information but are expected to
59 propagate to deeper depths (e.g. Lapham, 1989), and thermal profiles without periodic signals
60 can be numerically simulated (e.g. Koch et al. 2015).

61

62 Temperature time series can be collected at discrete points along a vertical profile (e.g. Hatch
63 et al., 2006), or near-continuous temperature with depth profiles using fiber-optic high-
64 resolution distributed temperature sensing (e.g. Vogt et al., 2010; Briggs et al., 2012). While
65 temperature time series are straightforward to collect in streambed sediments, the calculation
66 of fluxes can involve complex steps, including the isolation of the diurnal component of the
67 temperature signal from other signals and stochastic variation. There are also multiple
68 analytical solutions based on diurnal signals presented in the literature available to solve for

69 fluid flux (e.g. Hatch et al., 2006; Keery et al., 2007; McCallum et al., 2012; Luce et al.,
70 2013).

71

72 The VFLUX MATLAB routines (Gordon et al., 2012, <http://hydrology.syr.edu/vflux>) address
73 many of the complexities in processing temperature time series data by integrating data
74 management, signal processing, fluid flux modeling, and post-analysis based on straight-
75 forward user input. VFLUX (version 1.2.5 and earlier versions, from herein referred to as
76 VFLUX 1) is an all-in-one modelling system that synchronizes multiple temperature time
77 series, isolates and extracts the amplitude and phase of the signals for the period of interest
78 using advanced signal processing techniques, calculates fluxes at high spatial and temporal
79 resolution, and is able to quantify the influence of uncertainty in thermal parameters. VFLUX
80 1 includes analytical solutions by Hatch et al. (2006) and Keery et al. (2007), both of which
81 provide approaches to calculate time series of flux using either the amplitude ratio (A_r) or
82 phase shift ($\Delta\phi$) of diurnal signals between two temperature time series.

83

84 The assumptions of a number of 1D analytical methods (e.g. Lapham, 1989; Hatch et al.
85 2006; Keery et al. 2007) include that fluid flow is vertical and one-dimensional, that the
86 temperature signal is sinusoidal, that there is no mean thermal gradient in the streambed with
87 depth, and that streambed properties are homogeneous. The influence of non-ideal field
88 conditions (where one or more of the abovementioned assumptions are not met) on the use of
89 the Hatch et al. (2006) methods have been widely investigated. Lautz (2010), Roshan et al.
90 (2012), and Cuthbert and Mackay (2013) investigated the influence of multi-dimensional
91 flow fields on flux estimates in homogeneous porous media. Lautz (2010) identified that
92 multi-dimensional flow was a greater source of error compared to the presence of thermal

93 gradients in the streambed or non-sinusoidal temperature signals. Roshan et al. (2012)
94 showed that errors in flux estimates can be reduced by locating sensors as close to the
95 streambed surface as possible, and suggested that multiple arrays of sensors should be
96 deployed horizontally to evaluate the presence of horizontal flow. Cuthbert and Mackay
97 (2013) suggested that differences in fluxes estimated on the basis of A_r or $\Delta\phi$ may indicate
98 the degree of complexity in the flow field. Lautz (2010) showed that the presence of a mean
99 thermal gradient in the streambed was not a significant source of error in downwelling zones;
100 this was recently confirmed for upwelling zones by Briggs et al. (2014). Shanafield et al.
101 (2011) and Soto-López et al. (2011) investigated the influence of uncertainties in thermal
102 parameters and sensor precision on flux estimates. Shanafield et al. (2011) identified that
103 uncertainties in thermal parameters and noise in temperature measurements (due to sensor
104 accuracy) were particularly an issue for estimating upwelling fluxes. Soto-López et al. (2011)
105 identified that sensor resolution has a greater influence on the accuracy of flux estimates
106 based on A_r compared to $\Delta\phi$. Irvine et al. (2015) investigated the effect of streambed
107 heterogeneity on flux estimates under downwelling conditions, and showed that on average,
108 the A_r methods produced lower errors in flux estimates relative to $\Delta\phi$ in heterogeneous media.

109

110 Clearly there is somewhat conflicting evidence regarding which diurnal signal-based method
111 may be most reliable for field conditions, which is complicated by the fact that fluxes
112 estimated using A_r and $\Delta\phi$ alone rarely agree (e.g. Hatch et al., 2006; Lautz, 2012; Briggs et
113 al., 2012; Roshan et al., 2012; Cuthbert and Mackay, 2013; Irvine et al., 2015). More recently
114 presented analytical solutions to the 1D heat transport equations by McCallum et al. (2012),
115 and Luce et al. (2013) simultaneously use both A_r and $\Delta\phi$ to calculate flux, thereby
116 eliminating the possibility of obtaining different flux estimates when using A_r and $\Delta\phi$ in
117 isolation. Benefits of the McCallum et al. (2012) and Luce et al. (2013) combined methods

($A_r\Delta\phi$) include the fact that no estimate of thermal conductivity is required for the estimation of flux (thereby reducing the uncertainty due to uncertainty in thermal parameters), and both methods can also produce an estimated time series of thermal diffusivity (κ_e , $\text{m}^2 \text{s}^{-1}$). The Luce et al. (2013) methods can also be used to calculate vertical sediment thickness. Beyond their intrinsic utility to heat transport modeling, κ_e estimates shows promise as an indicator of poor flux estimates because it is a parameter expected to be relatively homogeneous in time and space; large spatial or temporal changes in κ_e from the $A_r\Delta\phi$ methods could suggest errors in flux estimates. Using field data, McCallum et al. (2012, in their figure 4) suggest that variations in κ_e in time can be used to identify time periods where flux estimates are unreliable.

The influence of non-ideal field conditions (e.g. 3D heterogeneity in sediments and flux) on $A_r\Delta\phi$ model-based flux estimates has not been investigated in great detail. Irvine et al. (2015) compared the use of the Hatch et al. (2006), McCallum et al. (2012), and Luce et al. (2013) methods in 2D numerical simulations with heterogeneous streambed properties. The Irvine et al. (2015) study was restricted to estimates of downwelling, and they did not explore the utility of κ_e output from the McCallum et al. (2012) and Luce et al. (2013) methods. They showed that (for downwelling conditions), on average, the $A_r\Delta\phi$ and A_r methods produce fluxes with smaller errors than the $\Delta\phi$ method. The influence of non-sinusoidal temperature signals, uncertainty in thermal parameters, and non-steady flows on $A_r\Delta\phi$ flux estimates was investigated by Rau et al. (2015). Using synthetic 1D data sets, Rau et al. (2015) showed that errors in flux and κ_e estimates during non-steady flow periods can be attributed to limitations in filtering methods.

It is important to note that all field sites are ‘non-ideal’, in that one or more of the above-mentioned assumptions of 1D analytical methods will likely not be met. It is therefore important to identify the impact of these non-ideal conditions on flux estimates from using known (numerically simulated and controlled laboratory) data. For this study, we investigate how flux estimates are influenced by non-sinusoidal temperature signals, non-steady flows, multi-dimensional flows, streambed heterogeneity and uncertainty in thermal parameters. The Hatch et al. (2006), Keery et al. (2007), McCallum et al. (2012) and Luce et al. (2013) methods were used to determine fluxes from temperature time series. All data are assessed with an updated version of VFLUX (version 2.0.0, from herein referred to as VFLUX 2) which includes the McCallum et al. (2012) and Luce et al. (2013) $A_r\Delta\phi$ methods to determine time series of flux, κ_e and sediment thickness.

2. Methods

The investigation is divided into two sets of experiments. First, temperature time series from the Lautz (2012) sand column experiments are re-examined. The analysis of the Lautz (2012) experiments allows for the determination of the influence of a non-sinusoidal temperature boundary condition, and non-steady flow rates on known flux estimates from real temperature data. Second, the influence of multi-dimensional flows and the influence of streambed heterogeneity on flux estimates are demonstrated using a 3D numerical model simulation.

In the sections below, first we present the equations of heat and fluid flow (Section 2.1), followed by the analytical solutions included in VFLUX 1 (Section 2.2) and VFLUX 2

(Section 2.3). Finally, we outline the experimental set up of the sand column (Section 2.4) and numerical model simulation (Section 2.5).

2.1 1D Equation of heat and fluid flow

The Hatch et al. (2006), Keery et al. (2007), McCallum et al. (2012) and Luce et al. (2013) methods are analytical solutions to the one-dimensional heat transport equation (e.g. Stallman, 1965):

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

where T is temperature ($^{\circ}\text{C}$), t is time (s), q is fluid flux (m s^{-1}) and C_w is volumetric heat capacity of water ($\text{J m}^{-3} ^{\circ}\text{C}^{-1}$), C is volumetric heat capacity of saturated sediment ($\text{J m}^{-3} ^{\circ}\text{C}^{-1}$), and z is depth (m). κ_e is calculated by:

$$\kappa_e = \frac{\lambda_0}{C} + \beta \left(\frac{C_w |q|}{C} \right) = \frac{\lambda_w^n \lambda_s^{(1-n)}}{nc_w \rho_w + (1-n)c_s \rho_s} + \beta \left(\frac{c_w \rho_w |q|}{nc_w \rho_w + (1-n)c_s \rho_s} \right), \quad (2)$$

where λ_0 , λ_w and λ_s are the thermal conductivities of the bulk saturated medium, water and the solid phase ($\text{W m}^{-1} ^{\circ}\text{C}^{-1}$), β is thermal dispersivity (m), n is porosity (–), c_w and c_s are specific heat capacities of water and the solid phase ($\text{J kg}^{-1} ^{\circ}\text{C}^{-1}$), and ρ_w and ρ_s are densities of water and the solid phase. Rau et al. (2012) show that the influence of β is negligible over the range

of q (e.g. $\pm 3.0 \text{ m d}^{-1}$), and sensor spacing (Δz , e.g. up to $\sim 1.0 \text{ m}$) considered in a groundwater–surface water interaction context, and so the second term on the right hand side of Eqn. 2 is often neglected.

Upper and lower bounds of κ_e that could realistically be expected from streambed materials can be determined from the range of parameter values that make up κ_e . Typical ranges of parameter values are shown in Table 1.

Thermal parameters (including n) of porous media can vary outside of the ranges presented in Table 1. For example, Conant (2004) measured n up to 0.72, or c_s can exceed $1100 \text{ J kg}^{-1} \text{ }^\circ\text{C}^{-1}$ for some minerals (Waples and Waples, 2004), and organic-rich sediments may contrast strongly with mineral grains (Stonestrom and Constantz 2003). However, the parameters in Table 1 provide a reasonable parameter set to identify the limits of κ_e that could be expected at a field site. Other ranges of thermal properties can be found in Lapham (1989) or Stonestrom and Constantz (2003), although it should be noted that Lapham (1989), presents bulk thermal properties (i.e. λ_0 not λ_s), and thermal properties of the solid and liquid phases are required for heat capacities (i.e. $C_s = \rho_s c_s$, $C_w = \rho_w c_w$) are required for the VFLUX inputs, with C calculated internally (i.e. $C = (1 - n)C_s + nC_w$). The range of possible κ_e values calculated from Table 1 is $0.028\text{--}0.180 \text{ m}^2 \text{ d}^{-1}$ (calculated at $20 \text{ }^\circ\text{C}$). This range could be reduced by measuring any of the parameters included in κ_e .

The methods described in Sections 2.2 and 2.3 calculate the thermal front velocity (or the “advective thermal velocity”, Luce et al. 2013), v_t (m s^{-1}), using two temperature time series

from a vertical profile. Estimates of q are obtained by multiplying v_t by C/C_w . All q calculations are generally considered representative of the mid-point between two temperature sensors. Estimates of q are based on either A_r (i.e. $A_r = A_d/A_s$, where subscript d and s represent the deep and shallow sensor, respectively), $\Delta\phi$ (i.e. $\Delta\phi = \phi_s - \phi_d$), or both, as shown below. VFLUX (1 and 2) uses Dynamic Harmonic Regression (DHR) from the Captain toolbox (Young et al., 1999; Young et al., 2010) to determine A_r and $\Delta\phi$, although other methods are available (e.g. see Hatch et al., 2006; Rau et al., 2010; Swanson and Cardenas, 2011; McCallum et al., 2012). Benefits of the DHR method include the fact that variable fluxes can be calculated on a sub-daily time step (e.g. Lautz, 2012).

2.2 Existing methods in VFLUX 1

VFLUX 1 includes the Hatch et al. (2006) and Keery et al. (2007) models. The Hatch et al. (2006) model for A_r is:

$$q = \frac{C}{C_w} \left(\frac{2\kappa_e}{\Delta z} \ln A_r + \sqrt{\frac{\alpha + v_t^2}{2}} \right), \quad (3)$$

and the $\Delta\phi$ model is:

$$|q| = \frac{C}{C_w} \sqrt{\alpha - 2 \left(\frac{4\pi\Delta\phi\kappa_e + v_t^2}{P\Delta z} \right)^2}, \quad (4)$$

228 where P is the period of the temperature signal (s), Δz is the sensor spacing (m), and α is
 229 calculated by $\alpha = \sqrt{v_t^4 + 8\pi\kappa_e/P}$. Eqn. 4 allows the magnitude of flux to be determined, but
 230 not the direction of flow because phase shift is reduced with water movement in either
 231 direction (Hatch et al., 2006; Briggs et al., 2014).

232

233 Keery et al. (2007) also provides individual solutions based on A_r (Eqn. 5) and $\Delta\phi$ (Eqn. 6).

234 The Keery et al. (2007) model for A_r is:

235

$$236 \quad \left(\frac{H^3 \ln A_r}{4\Delta z} \right) q^3 - \left(\frac{5H^2 \ln^2 A_r}{4\Delta z^2} \right) q^2 + \left(\frac{2H \ln^3 A_r}{z^3} \right) q + \left(\frac{\pi C}{\lambda_0 P} \right)^2 - \frac{\ln^4 A_r}{z^4} = 0, \quad (5)$$

237

238 where $H = C_w/\lambda_0$. As Eqn. 5 is presented, both here, and in Keery et al., (2007, their
 239 equation 9), λ_0 is included (in the H term), but the influence of β is not.

240

241 The $\Delta\phi$ model from Keery et al. (2007) is:

242

$$243 \quad |q| = \sqrt{\left(\frac{C\Delta z}{\Delta\phi C_w} \right)^2 - \left(\frac{4\pi\Delta\phi\lambda_0}{P\Delta z C_w} \right)^2}. \quad (6)$$

244

245 As with Eqn. 4, Eqn. 6 allows the magnitude of flux to be calculated but not the flow
 246 direction.

247

248 When $\beta = 0$ m, calculated fluxes using A_r (Eqns. 3 and 5) or $\Delta\phi$ (Eqns. 4 and 6) are identical
 249 (Irvine et al., 2015). Here, we consider $\beta = 0$ m for all analyses. Throughout this paper, q
 250 calculated from either A_r method (Eqn. 3 or 5) are presented as q_{A_r} , and q calculated from
 251 either $\Delta\phi$ method (Eqn. 4 or 6) are presented as $q_{\Delta\phi}$.

252

253 The additions to VFLUX 2 are outlined below. No additional user input is required to run the
 254 main routine of VFLUX 2 relative to previous releases.

255

256 **2.3 Additions in VFLUX 2**

257 McCallum et al. (2012) produced direct solutions to the 1D heat transport equation that use
 258 both A_r and $\Delta\phi$ simultaneously, where q can be determined from:

259

$$260 \quad q = -\frac{C}{C_w} \left(\frac{\Delta z (P^2 \ln^2 A_r - 4\pi^2 \Delta\phi^2)}{\Delta\phi \sqrt{(16\pi^4 \Delta\phi^4 + 8P^2 \pi^2 \Delta\phi^2 \ln^2 A_r + P^4 \ln^4 A_r)}} \right). \quad (7)$$

261

262 Note that Eqn. 7 does not require a value of κ_e to determine q (hence it does not require a
 263 *priori* knowledge of λ_0), which offers advantages over the Hatch et al. (2006) and Keery et al.
 264 (2007) models. The negative sign in Eqn. 7 is due to McCallum et al. (2012) adopting the

convention of positive flow for upwelling. The implementation of the Hatch et al. (2006), Keery et al. (2007), McCallum et al. (2012) and Luce et al. (2013) methods in all VFLUX releases use the convention of a negative sign for upwelling.

Another key benefit of the methods outlined by McCallum et al. (2012) is that it is possible to calculate a time series of κ_e from:

$$\kappa_e = \frac{\Delta z^2 P^2 \ln A_r (4\pi^2 \Delta \phi^2 - P^2 \ln^2 A_r)}{\Delta \phi (P^2 \ln^2 A_r + 4\pi^2 \Delta \phi^2) (P^2 \ln^2 A_r - 4\pi^2 \Delta \phi^2)}. \quad (8)$$

Luce et al. (2013) also present approaches that simultaneously use A_r and $\Delta \phi$ to determine a time series of q (Eqn. 9), and κ_e (Eqn. 10):

$$q = \frac{C}{C_w} \left(\frac{\omega \Delta z}{\Delta \phi_r} \left(\frac{1 - \eta^2}{1 + \eta^2} \right) \right), \text{ and} \quad (9)$$

$$\kappa_e = \frac{\omega \Delta z^2}{\Delta \phi_r^2 \left(\frac{1}{\eta} + \eta \right)}, \quad (10)$$

where $\Delta \phi_r$ is the phase shift in radians, $\eta = -\ln A_r / \Delta \phi_r$, and $\omega = 2\pi/P$.

Eqns. 7, 8, 9 and 10 have been added to VFLUX 2. Irvine et al. (2015) show that the McCallum et al. (2012) and Luce et al. (2013) equations for q produce equivalent results, and hence all q calculated by simultaneously using both A_r and $\Delta\phi$ are presented as $q_{A_r\Delta\phi}$. Estimates of κ_e also are identical for McCallum et al. (2012), and Luce et al. (2013), and hence they are presented as $\kappa_{eA_r\Delta\phi}$.

Other changes in VFLUX 2 (e.g. the ability to determine streambed scour using a reformulation of Equation 10 outlined in Luce et al. (2013), and the ability to refine the range of plausible κ_e for a given field site) are described in the documentation provided with the computer code. The VFLUX 2 program files are included in the Supplementary materials, available with the online version of this article. The most up to date versions are available at the following website: <http://hydrology.syr.edu/vflux>.

2.4 Sand column experiments

Lautz (2012) performed laboratory experiments to investigate the influence of non-steady q (a step change and gradual reduction in q), and the influence of stream temperature time series with variable amplitudes under a constant q . The Hatch et al. (2006) methods were applied to vertical temperature time-series collected during controlled sand column experiments (Fig. 1). For all experiments, actual flow rates were determined using a tipping bucket mechanism, and temperatures were measured in the column using Omega[®] “t-type” stainless steel thermocouples which were located at depths of 0.025 m, 0.05 m, 0.10 m, 0.15 m, 0.25 m and 0.35 m (Fig. 1). The sand column was insulated with a 2.5 cm thick Grainger[®] insulating jacket. Temperature time series were generated with $P = 0.25$ d. In all experiments

reanalyzed here, flows are vertical, one-dimensional, and downwelling (positive); for full details on the experiments, see Lautz (2012).

Temperature time series from the Lautz (2012) sand column study are reanalyzed to investigate the estimates of q from the combined $A_r\Delta\phi$ methods, and to explore the benefits of the additional ability to compute κ_e time series. The `vfluxmc.m` (Monte Carlo) script from VFLUX 2 was used to estimate q values as the mean of 200 realizations, using the thermal parameter values and uncertainties (standard deviations) shown in Table 2. The `vfluxmc.m` program assumes normally distributed parameter distributions defined by the user defined mean and standard deviation. The benefit of the use of the `vfluxmc.m` program is that both the mean q , as well its uncertainty based on estimated parameter values are obtained. κ_e estimates do not require thermal parameters, and hence the source of uncertainty in κ_e comes from the uncertainty in values of A_r and $\Delta\phi$ (which is not accounted for in `vfluxmc.m`). Estimates of q and κ_e were output 12 times for each temperature cycle using sensors located at depths of 0.05 and 0.10 m (i.e. $\Delta z = 0.05$ m). Using data from other pairs of sensors did not significantly alter the results.

2.5 Numerical modeling

To investigate the influence of multi-dimensional flows and streambed heterogeneity (in hydraulic conductivity, $K \text{ m s}^{-1}$), synthetic daily streambed temperature time series were generated using HydroGeoSphere, HGS (Therrien et al., 2006), using the model domain from Irvine and Lautz (2015). The numerical model domain (Fig. 2) was 10 m long (x), 4 m wide (y) and 3 m deep (z). Model discretization was uniform in x and y , with $\Delta x = \Delta y = 0.05$ m.

Vertical discretization was variable, with $\Delta z = 0.02$ m for the upper 2 m of the domain, and 0.05 m for the lower 1 m of the domain. This discretization produced a grid with 1.92 million elements.

The hydraulic boundary conditions of the model were no flow boundaries on all faces, except the top boundary, where the hydraulic head (h) was a function of x , with a gradient of 0.1 m m^{-1} . These boundary conditions cause a flow-through hyporheic flow cell with the (constant in time) vertical component of flux (q_z) ranging between 8.83 to -6.18 m d^{-1} (with downwards flow denoted as positive). This range may seem high, however it is important to note that these extreme values occur at the edges of the model ($x = 0$ and 10 m), and do not feature in any of the profiles used for analysis. The thermal boundary condition at the bottom of the domain was a constant temperature of $15 \text{ }^\circ\text{C}$, and the top thermal boundary condition was calculated using Eqn. 11:

$$T_{(t)} = T_0 + \Delta T \cdot \sin\left(\frac{2\pi(t - P/4)}{P}\right), \quad (11)$$

where $T_{(t)}$ is stream temperature ($^\circ\text{C}$) as a function of time, T_0 is the mean stream temperature ($17.5 \text{ }^\circ\text{C}$), and ΔT is the amplitude of the stream temperature signal ($2.5 \text{ }^\circ\text{C}$). Eqn. 11 produced a sinusoidal temperature variation with a period of 1 day, and temperature varying between 15 and $20 \text{ }^\circ\text{C}$. As per Irvine and Lautz (2015), the initial temperature conditions were established by first running the simulation for 14 days. The temperature distribution after 14 days was then used as the initial temperature distribution for the simulations of a

further 14 days, which produced time series that were analyzed to calculate q from Eqns. 3, 4, 5, 6, 7 and 9, and κ_e from Eqns. 8 and 10.

Observation points were horizontally located at $x = 0.5, 2.0, 5.0, 8.0$ and 9.5 m to record temperature profiles that span the range from strong downwelling to strong upwelling (Fig.2). In the y direction (across the model), observation points were located at $y = 1.0, 2.0$ and 3.0 m. Observation points were vertically located every 0.1 m, from $z = 0.0$ m to 0.7 m. A vertical observation point spacing (Δz) of 0.1 m provided up to seven flux estimates with depth at each vertical profile.

The heterogeneous streambed K distribution was generated using the Sequential Gaussian Simulator from the Geostatistical Software Library (Deutsch and Journel, 1998). The mean K was $1 \times 10^{-4} \text{ m s}^{-1}$, and correlation lengths were 1.0 m, 1.0 m, and 0.2 m in the x , y , and z directions respectively, and the variance of $\ln(K)$ was 0.5 . For the purposes of this study, only one realization of the $\ln(K)$ field was considered, because the goal was to compare the results of the A_r , $\Delta\phi$, and $A_r \Delta\phi$ methods in heterogeneous (rather than homogeneous) media, rather than a detailed analysis of the influence of varied K distribution. It is important to note that Gaussian simulators produce relatively smoothly varying fields, and more sophisticated methods, such as Multiple-Point Statistics (e.g. Mariethoz and Caers, 2014) can be used to produce fields with more connected preferential flow paths.

All thermal properties of the HGS model are as presented in Table 2. As with the Lautz (2012) sand column data, all HGS model output were analyzed with the `vfluxmc.m` script,

using 200 realizations, with 12 q and κ_e outputs per cycle (in this instance, per day). As numerical models can produce output (e.g. temperature) with far higher precision than can be obtained from field equipment, all temperature time series from HGS were adjusted before being processed. All temperature time series from the HGS simulation were rounded to 2 decimal places before being read into vfluxmc.m.

The use of 3D, heterogeneous K -fields provides the opportunity to assess the performance of temperature time series solutions with highly complex flow paths, where previous investigations have either used homogeneous K distributions (e.g. Lautz, 2010; Roshan et al., 2012; Cuthbert and Mackay, 2013) or 2D model simulations (Irvine et al., 2015).

3. Results

The results below are presented as follows: first, the influence of non-steady flows (Section 3.1), and non-sinusoidal temperature signals (Section 3.2) are assessed for the sand column experiments. Second, the influence of multi-dimensional flows in heterogeneous media (Section 3.3) are assessed using the numerical model output.

3.1. Influence of non-steady flows

Fig. 3 shows the results from the Lautz (2012) ‘step experiment’, where three step changes in measured q (denoted as q_{meas}) occur after approximately 4.5, 12.7 and 21.5 temperature cycles. Fig. 3 shows the output from the A_r , $\Delta\phi$, and $A_r\Delta\phi$ methods. The gray box in Fig. 3c represents the plausible range of κ_e calculated from Table 1.

396

397 In Fig. 3b, the temperature dependency of hydraulic conductivity causes the cyclic variability
398 in q_{meas} (e.g., between cycles 4.5 and 12.7). The greatest temperature induced variation in
399 q_{meas} occurs when there is the largest amplitude in the temperature signal (e.g., compare q_{meas}
400 from cycle 21.5 to 25.9 with cycles 12.7 to 21.5). Interestingly, all three methods (q_{A_r} , $q_{\Delta\phi}$,
401 and $q_{A_r\Delta\phi}$) produce flux estimates with these cyclic variations; however, modeled flux
402 changes are out of phase with measured changes, further indicating the 1D analytical models
403 lag real step changes in flux as observed by Lautz (2012).

404

405 To analyze the performance of the temperature time series models, the mean from the Monte
406 Carlo simulation (for each time step, i.e. each 1/12th cycle) is compared to the known q_{meas} .
407 Errors are calculated as $\bar{q}_{A_r} - \bar{q}_{meas}$, $\bar{q}_{\Delta\phi} - \bar{q}_{meas}$, or $\bar{q}_{A_r\Delta\phi} - \bar{q}_{meas}$ (a positive error denotes
408 estimates of stronger downwelling than $\bar{q}_{meas}$). In order to present the results in a more
409 meaningful way, the average error is presented between cycles 4.5 and 12.7 (labelled as time
410 period A), cycles 12.7 and 21.5 (time period B), and from cycle 21.5 to the end of the
411 experiment (time period C). The estimates and errors for each method (A_r , $\Delta\phi$, and $A_r\Delta\phi$) are
412 presented in Table 3 below.

413

414 It is clear from Table 3 that the $q_{\Delta\phi}$ method was the poorest performing method across all
415 time periods considered. During time periods A and B, the q_{A_r} method outperforms the
416 $q_{A_r\Delta\phi}$ method, with the $q_{A_r\Delta\phi}$ method producing estimates that, on average, were slightly lower
417 than the q_{A_r} method for period C. However, Fig. 3 shows that the errors from the A_r method
418 are larger than the $A_r\Delta\phi$ method immediately after the large step change in q_{meas} after 21.5

cycles. If period C was calculated between cycle 22.5 and the end of the experiment, the mean errors are -0.46 , 0.87 and 0.59 m d^{-1} for q_{A_r} , $q_{A\phi}$, and $q_{A_r\Delta\phi}$ respectively (i.e. smaller errors for q_{A_r} relative to $q_{A_r\Delta\phi}$).

The time series of $\kappa_{e A_r\Delta\phi}$ shows cyclic variations as was the case for each of the q estimation methods. However, the $\kappa_{e A_r\Delta\phi}$ time series also shows rapid changes that coincide with the step changes in q_{meas} . The step change in q_{meas} that occurs at cycle 4.5 produces a rapid increase in $\kappa_{e A_r\Delta\phi}$, which takes the value outside of the reasonable range. The $\kappa_{e A_r\Delta\phi}$ estimates are also often outside of the theoretical range from cycle 21.7 to the end of the experiment. In both cases, the time periods where unreasonable $\kappa_{e A_r\Delta\phi}$ estimates coincide with large errors in q estimates. This observation supports the suggestion of McCallum et al., (2012), that unreasonable variations in $\kappa_{e A_r\Delta\phi}$ coincide with erroneous flux estimates, with the largest $\kappa_{e A_r\Delta\phi}$ occurring when the flow rate changes rapidly. The range of plausible κ_e for any given site can be reduced in the VFLUX 2 post processing options to assist with decisions concerning the reliability of flux estimates.

Fig. 4 shows the results from a gradual change in q_{meas} that occurs over the duration of a temperature cycle (from cycle 8 to 9). The layout of Fig. 4 is identical to Fig. 3, with the exception of the extent of the time axis (x axis). The time axis spans from cycle 5 to 16, to highlight the influence of the gradual change in q_{meas} .

The results in Fig. 4b highlight challenges in the use of $\Delta\phi$. It is known that q_{meas} remains positive (i.e. downwelling) throughout the entire column experiment, although in a field setting it is likely that the flow direction may be unknown. If the convention suggested by Soto-López et al. (2011) is used, where A_r is used to determine flow direction, and $\Delta\phi$ to determine q , a sudden change in $q_{\Delta\phi}$ would occur after cycle 11 as q_{A_r} estimates become negative (suggesting upwelling). Fig. 4b also shows that the q_{A_r} method produces a more reliable range of fluxes during the gradual change in q_{meas} (cycles 8 to 12).

As with the results shown in Fig. 3, the large errors from the $q_{A_r\Delta\phi}$ and $q_{\Delta\phi}$ methods during the gradual change experiment coincide with $\kappa_{eA_r\Delta\phi}$ values that are outside of the practical κ_e range; this indicates $\Delta\phi$ information is compromised during the hydraulic shift, introducing error into the $q_{A_r\Delta\phi}$ method. As $q_{A_r\Delta\phi}$ gradually approaches q_{meas} after the change in q_{meas} , the $\kappa_{eA_r\Delta\phi}$ time series approaches the calculated κ_e value calculated from the measured thermal parameter values from Lautz (2012).

Monte Carlo error analysis for all methods is based on assumed uncertainty distributions for select sediment and thermal parameters in the heat transport equation (n , λ_0 , C_s , and C_w ; Table 2). The apparent uncertainty of $q_{A_r\Delta\phi}$ is lower than q_{A_r} or $q_{\Delta\phi}$ for nearly all q estimates during the column experiments, and this is particularly apparent at lower q (e.g. Figure 3b and 4b). The lower uncertainty of $q_{A_r\Delta\phi}$ is expected, given that *a priori* knowledge of λ_0 is not required for the flux calculations and therefore fewer variables are considered uncertain in the Monte Carlo analysis for $q_{A_r\Delta\phi}$. The uncertainty of $q_{A_r\Delta\phi}$ is notably reduced at lower q

(cycles 5-7 versus cycles 12-14, Figure 4b); this is due to the smaller temperature oscillations (ΔT) that occur in the sediment-water matrix under low flow conditions. When ΔT is small, model sensitivity to n , C_s , and C_w (the thermal parameters considered uncertain in the combined methods) is minimal and uncertainty of $q_{A_r\Delta\phi}$ values due to thermal parameter uncertainty is minimized. While the A_r and $\Delta\phi$ methods are still sensitive to uncertainty in λ_0 at low flux, λ_0 is not an input for the combined methods and therefore not a source of uncertainty. In contrast, when ΔT is large at higher flux rates, uncertainty of the heat capacity of the sediment-water matrix (derived from n , C_s , and C_w) becomes the primary source of error and uncertainties of the A_r , $\Delta\phi$, and combined methods are essentially identical.

Although the Monte Carlo analysis suggests $q_{A_r\Delta\phi}$ has a lower degree of uncertainty at low q_{meas} , we can clearly see from Fig. 4 that the $q_{A_r\Delta\phi}$ values are substantially different from q_{meas} , with greater model residuals than for q_{A_r} . Although uncertainty of q estimates due to thermal parameter uncertainty has been reduced using the combined methods, uncertainty derived from the shifting phase lag through time under transient flow conditions causes error in $q_{A_r\Delta\phi}$. For this reason, caution should be used when interpreting flux uncertainty using only thermal parameter uncertainty analysis. The combined methods reduce the flux estimate uncertainty associated with uncertainty in the thermal parameters when ΔT is small. But, that apparent certainty in flux estimates could be misleading, given the other potential sources of uncertainty in the analysis.

3.2. Influence of non-sinusoidal temperature signal

Fig. 5 shows the results from an experiment with a relatively constant q_{meas} , but an irregular heating of the source water.

In Fig. 5a, cycles 19 and 24 show the greatest change in the temperature signal. The effect of the temperature signal at cycle 19 can be clearly observed in the $q_{\Delta\phi}$, q_{A_r} , and $q_{A_r\Delta\phi}$ estimates, where all methods produce lower q estimates during this period. In the ‘weather’ experiment shown in Fig. 5b, the mean q_{meas} was 1.77 m d^{-1} . The greatest errors (difference between q_{meas} and mean from the Monte Carlo simulation) at cycle 19 were 0.03, 0.13 and 0.40 m d^{-1} for q_{A_r} , $q_{A_r\Delta\phi}$, and $q_{\Delta\phi}$ respectively. The mean errors, averaged across all time steps follows the same pattern ($\bar{q}_{A_r} < \bar{q}_{A_r\Delta\phi} < \bar{q}_{\Delta\phi}$), with mean errors of 0.07, 0.14 and 0.28 m d^{-1} . The $q_{\Delta\phi}$ method also has the widest uncertainty bands throughout the experiment, with a typical range of 1.03, compared to 0.48 for the q_{A_r} method, and 0.59 m d^{-1} for the $q_{A_r\Delta\phi}$ method. In Fig. 5c there is a change in $\kappa_{eA_r\Delta\phi}$ at cycle 19, although this would be difficult to distinguish from the cyclic variations in $\kappa_{eA_r\Delta\phi}$ that occur in Figs. 3c and 4c.

The findings from the reanalysis of the Lautz (2012) sand column experiments of non-steady flows (Figs. 3 and 4, Table 3), and non-sinusoidal temperature signals (Fig. 4) highlight that non-steady flow rates have the greatest influence of the validity of q estimates from the use of A_r , $\Delta\phi$ or $A_r\Delta\phi$. While q estimates are influenced by variations in the temperature signal, it has been previously highlighted that this effect could be reduced by calculating average q during periods of relatively stable flows Lautz (2010).

3.3 Influence of multi-dimensional flow in heterogeneous media

The results presented in Fig. 6 are from the HGS model run for a longitudinal section along the length (x) of the model at $y = 1.0$ m. Output at $y = 2.0$ and 3.0 m are included in the online materials.

Fig. 6 shows time averaged values from `vfluxmc.m` (with 200 realizations) q estimates (6a, c, e, g, i), and κ_e estimates (6b, d, f, h, j). For the κ_e panels in Fig. 6, the κ_e calculated from the HGS input parameters is shown in black (κ_{eHGS}), the $\kappa_{eA_r\Delta\phi}$ from the McCallum et al. (2012) or Luce et al. (2013) method is shown in purple. As Eqns. 8 and 10 are functions of A_r and $\Delta\phi$, and not C , C_w or λ_0 , these solutions provide a unique time series of $\kappa_{eA_r\Delta\phi}$, rather than a distribution as is the case for q_{A_r} , $q_{\Delta\phi}$, or $q_{A_r\Delta\phi}$. In all cases in Fig. 6, the sign of $q_{\Delta\phi}$ is determined from the sign of q_{A_r} .

For the $x = 0.5$ m q estimates shown in Fig. 6a, all methods follow the general pattern of q_{HGS} . The range of $q_{A_r\Delta\phi}$ and $q_{\Delta\phi}$ are essentially identical, with over-estimates of downwelling occurring for the upper 0.2 m of the profile for q_{A_r} . At $x = 2.0$ m (Fig. 6b, where vertical fluxes are more moderate compared to the $x = 0.5$ m profile), the q_{A_r} and $q_{A_r\Delta\phi}$ estimates are similar for the entire profile, with a wider distribution of flux estimates of $q_{\Delta\phi}$. The benefits of the $A_r\Delta\phi$ methods are highlighted for the $x = 5.0$ m data (Fig. 6e), where the uncertainty in estimates are essentially zero. For the upwelling profiles ($x = 8.0$ and 9.5 m, 6g, 6i) q estimates are either not possible, or highly variable below $z = 0.4$ m. This shows a limitation of diurnal temperature time series methods, as flux magnitude and direction will affect the damping of the surface signal with depth, and if the diurnal signal amplitude is less than

twice sensor precision it cannot be reliably measured. For an extensive investigation into the limitations of diurnal signal methods for upwelling cases, the reader is referred to Briggs et al. (2014). Here q estimates are only shown to $z = 0.45$ m below which the diurnal signal is not present; errors clearly increase toward this limit of detection. In the upwelling profiles (Figs. 6g, 6i), the combined $A_r\Delta\phi$ methods perform well in the upper 0.35 m. For example, for the $q_{A_r\Delta\phi}$ method, the errors from the mean from the Monte Carlo method at $z = 0.15$ in the bottom row of Fig. 6 ($x = 9.5, y = 1.0$) was -0.11 m d^{-1} where q_{HGS} was -0.74 m d^{-1} .

The $\kappa_{eA_r\Delta\phi}$ estimates in Fig. 6 (b, d, f, h, j) show that spatial variations in κ_e estimates can also be indicative of erroneous q estimates from the all methods. In the $x = 0.5$ m data (Fig. 6b), the sudden increase in $\kappa_{eA_r\Delta\phi}$ at $z = 0.25$ m occurs at the same depth as the greatest overestimate of $q_{A_r\Delta\phi}$. For the upwelling profiles ($x = 8.0$ and 9.5 m, i.e. Figs. 6h and 6i) the estimates of $\kappa_{eA_r\Delta\phi}$ that either occur outside of the practical κ_e limits, or where $\kappa_{eA_r\Delta\phi}$ oscillates from one depth to the next also coincide with erroneous flux estimates.

The κ_e estimates in Fig. 6 show that spatial variations in $\kappa_{eA_r\Delta\phi}$ can also be useful in determining where q estimates are reliable. However, sudden changes in $\kappa_{eA_r\Delta\phi}$ also occur when rapid changes in q occur. For example, the change in q_{HGS} that occurs between $z = 0.27$ and $z = 0.35$ m is 1.69 m d^{-1} .

4. Discussion and recommendations

From the investigation into time-varying fluxes in Section 3.1, the use of the $\Delta\phi$ method led to the greatest variability in q estimates. The use of $\Delta\phi$ also offers other challenges, including the fact that a method to determine flow direction is also required. The results from the gradual q change experiment (Fig. 4) highlight difficulties in selecting the sign for flow from the A_r method, and the magnitude from the $\Delta\phi$ method. For the time-varying q conditions tested here, the A_r method performed better than the $A_r\Delta\phi$ method during periods of (and after) non-steady flow, due to apparent adverse influence of the $\Delta\phi$ data on $A_r\Delta\phi$ results.

Investigations into the influence of a non-sinusoidal temperature boundary condition (Section 3.2) showed that the influence of a non-sinusoidal temperature source did not lead to significant variations (in time) in q estimates across all of the time series methods for the temperature signals used in this study, supporting the findings of Lautz (2012). Of the three methods considered here, the $\Delta\phi$ method led to the widest distribution of q estimates, with comparable flux estimates from the A_r and $A_r\Delta\phi$ methods. Analyses of the Lautz (2012) sand column data demonstrates that large variations in $\kappa_{eA_r\Delta\phi}$ with time can suggest that q estimates are likely to be inaccurate. In the analyses presented here, large variations in $\kappa_{eA_r\Delta\phi}$ occurred in cases where q was both under- and over-estimated.

In the investigations into the influence of streambed heterogeneity on q estimates (Section 3.3), again, the $\Delta\phi$ method performed poorly. Irvine et al. (2015) showed that when there is no uncertainty in thermal parameters, that the use of the A_r method typically produced q estimates with lower errors than the $A_r\Delta\phi$ method. The results presented here, where there was uncertainty in thermal parameters, show that the $A_r\Delta\phi$ method typically produces q

estimates with smaller errors than the A_r or $\Delta\phi$ methods when temperature variations over a time cycle are small (i.e. at low flux rates, like period B, Fig. 3), and this is due to the removal of uncertainty in λ_0 , which is not required as an input for the $A_r\Delta\phi$ methods. The results from Fig. 6 show that large spatial variability in $\kappa_{eA_r\Delta\phi}$ estimates also coincides with unreliable flux estimates, or that there is insufficient strength in the diurnal signal for reliable q estimates.

Analytical temperature time series-based methods of determining fluid exchange between surface water and groundwater are typically applied where one or more of the underlying assumptions of the methods are not met. Generally, the performance of the $\Delta\phi$ method was poor (relative to the A_r or $A_r\Delta\phi$ methods) under all conditions tested. When fluxes are variable with time, the A_r method produces more accurate results. When flows are stable in time, and near zero, estimates of $\kappa_{eA_r\Delta\phi}$ are typically very accurate in the cases investigated here. This point could be utilized, whereby estimated $\kappa_{eA_r\Delta\phi}$ values from stable periods in association with measured n and estimated solid-grain heat capacity (C_s) could be used to constrain λ_0 , which in turn would improve flux estimates from the A_r method. This process will require the estimation of other parameters; however, the measurement of some thermal parameters can assist this process. Of the parameters required as input in VFLUX (1 and 2) (C_w , n , C_s , β , and λ_0), n is likely the simplest to measure. For example a core sample can be weighed before and after the sample is oven dried. The properties of water (C_w , ρ_w , and λ_w) vary on a very small range for the temperature variations observed in streambeds. In zones of strong upwelling (not investigated here), where phase lags are relatively short and difficult to characterize accurately with most current sensing technology (Briggs et al. 2014), methods that utilize both A_r and $\Delta\phi$ information (e.g. $A_r\Delta\phi$, $\kappa_{eA_r\Delta\phi}$) may be particularly prone to error.

597

598 Additional uses of $\kappa_{eA_r\Delta\phi}$ output include analysis based on a “plausible” range of κ_e , strong
599 temporal deviations from this plausible range provides an objective metric to indicate time
600 periods when q estimates may be unreliable. The `vflux_post.m` program allows for the results
601 from a VFLUX run to be visualized. In VFLUX 2, plots are produced which show time series
602 of κ_e from the McCallum et al. (2012), and Luce et al. (2013) methods in comparison to a
603 large range of plausible κ_e values across all expected streambed sediment types. The user has
604 the option to enter a new range of n , c_s , ρ_s and λ_s to reduce the plausible range of κ_e values for
605 a particular site.

606

607 VFLUX 2 also includes the equation to assess streambed thickness (Δz) above a lower,
608 buried sensor, which can be used to assess scour assuming a thermally-mixed water column as
609 presented by Luce et al. (2013), and Tonina et al. (2014). While these approaches were
610 originally designed to determine streambed scour and deposition, further opportunities are yet
611 to be explored. VFLUX 2 calculates Δz using an approach outlined in Luce et al. (2013, their
612 Eqn. 57) using the user input C_w , C_s and λ_θ to determine κ_e . It is expected that oscillations in
613 Δz will occur when large changes in q occur (as is the case with $\kappa_{eA_r\Delta\phi}$) which may also
614 provide a means to determine when q estimates are unreliable. However, a systematic offset
615 between the predicted and known Δz may also provide the opportunity to refine thermal
616 property inputs for VFLUX 2 to improve both q_{A_r} and $q_{A_r\Delta\phi}$, as both require estimates of C_s
617 to convert v_t to q .

618

619 **Conclusions and implications for field applications**

We present an analysis of the influence of non-steady flows, non-diurnal temperature signals, and non-vertical flow in heterogeneous media on estimates of vertical water flux from temperature time series methods included in the new VLFUX 2 program. Even with their inherent complexities, the sand column and numerical model investigations presented here represent simplified cases relative to field applications. It should be highlighted that all field sites are ‘non-ideal’ in that any of the abovementioned non-ideal conditions violate the assumptions of the analytical solutions discussed in this paper. However, we have highlighted that the use of the combined $A_r\Delta\phi$ thermal diffusivity output can assist in the identification of time periods where flux estimates may be unreliable.

There are many methods to determine groundwater-surface water exchange using temperature time series. This study clearly demonstrates limitations of methods based on $\Delta\phi$ alone, and with alternative methods available, we do not generally suggest its use. During transient flow conditions, the use of the A_r method is suggested. The $A_r\Delta\phi$ methods perform best during steady flow conditions, and when q is near zero in low-flux environments. The ability to define low fluxes (e.g. $<0.1 \text{ m d}^{-1}$) more precisely than with the individual methods may benefit problems such as contaminant transport where even low flux rates are important to water quality. The VFLUX 2 package integrates the combined $A_r\Delta\phi$ methods into a familiar format without added user input, allowing for facile adoption by the surface water-groundwater interaction community.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. EAR-0901480, the U.S. Geological Survey (USGS) National Research Program, and the Groundwater Resources Program and Toxic Substances Hydrology Program. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government. The paper benefitted from the feedback from two anonymous reviewers, and the thoughtful review of an earlier version of the manuscript by Randy Hunt.

References

- Anderson, M.P., 2005. Heat as a ground water tracer. *Ground water*, 43(6): 951-968.
- Birkel, C., Soulsby, C., Irvine, D.J., Malcolm, I., Lautz, L.K., Tetzlaff, D., 2015. Heat-based hyporheic flux calculations in heterogeneous salmon spawning gravels. *Aquat.Sci.* doi: 10.1007/s00027-015-0417-4
- Bravo, H.R., Jiang, F., Hunt, R.J., 2002. Using groundwater temperature data to constrain parameter estimation in a groundwater flow model of a wetland system. *Water Resour. Res.*, 38(8).
- 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 Resour. Res.*, 48.
- Briggs, M.A., Lautz, L.K., Hare, D.K., Gonzalez-Pinzon, R., 2013. Relating hyporheic fluxes, residence times, and redox-sensitive biogeochemical processes upstream of beaver dams. *Freshw. Sci.*, 32(2): 622-641.

664 Briggs, M.A., Lautz, L.K., Buckley, S.F., Lane, J.W., 2014. Practical limitations on the use of
665 diurnal temperature signals to quantify groundwater upwelling. *J Hydrol*, 519: 1739–
666 1751.

667 Conant, B., 2004. Delineating and quantifying ground water discharge zones using streambed
668 temperatures. *Ground water*, 42(2): 243-257.

669 Cuthbert, M.O., Mackay, R. 2013. Impacts of nonuniform flow on estimates of vertical
670 streambed flux. *Water Resour. Res.*, 49, doi:10.1029/2011WR011587.

671 Deutsch, C.V., Journel, A.G., 1998. *GSLIB Geostatistical Software Library and User's*
672 *Guide*, 2nd ed. New York: Oxford University Press.

673 Fitts, C.R., 2013. *Groundwater Science*. Elsevier, New York.

674 Goto, S., Yamano, M., Kinoshita, M., 2005. Thermal response of sediment with vertical fluid
675 flow to periodic temperature variation at the surface. *J. Geophys. Res. B Solid Earth*
676 110, 1–11.

677 Gordon, R. P., Lautz, L. K., Briggs, M. A., McKenzie, J. M., 2012. Automated calculation of
678 vertical pore–water flux from field temperature time series using the VFLUX method
679 and computer program, *J. Hydrol.*, 420, 142–158.

680 Gordon, R.P., Lautz, L.K., Daniluk, T.L., 2013. Spatial patterns of hyporheic exchange and
681 biogeochemical cycling around cross-vane restoration structures: Implications for
682 stream restoration design. *Water Resour. Res.*, 49(4): 2040-2055.

683 Hatch, C.E., Fisher, A.T., Revenaugh, J.S., Constantz, J., Ruehl, C., 2006. Quantifying
684 surface water–groundwater interactions using time series analysis of streambed thermal
685 records: method development. *Water Resour. Res.* 42 (10), W10410.

686 Hatch, C.E., Fisher, A.T., Ruehl, C.R., Stemler, G., 2010. Spatial and temporal variations in
687 streambed hydraulic conductivity quantified with time–series thermal methods. *J*
688 *Hydrol*, 389(3–4): 276–288.

689 Hopmans, J.W., Simunek, J., Bristow, K.L., 2002. Indirect estimation of soil thermal
690 properties and water flux using heat pulse probe measurements: Geometry and
691 dispersion effects. *Water Resour. Res.*, 38(1).

692 Hunt, R.J., Krabbenhoft, D.P., Anderson, M.P., 1996. Groundwater inflow measurements in
693 wetland systems. *Water Resour. Res.*, 32(3): 495-507.

694 Irvine, D.J., Cranswick, R.H., Simmons, C.T., Shanafield, M.A., Lautz, L.K., 2015. The
695 effect of streambed heterogeneity on groundwater– surface water exchange fluxes
696 inferred from temperature time series. *Water Resour. Res.* doi:
697 10.1002/2014WR015769.

698 Irvine, D.J., Lautz, L.K., 2015. High resolution mapping of hyporheic fluxes using streambed
699 temperatures: Recommendations and limitations. *J. Hydrol.*, 524, 137–146, doi:
700 10.1016/j.jhydrol.2015.02.030.

701 Keery, J., Binley, A., Crook, N., Smith, J. W. N., 2007. Temporal and spatial variability of
702 groundwater–surface water fluxes: Development and application of an analytical
703 method using temperature time series, *J. Hydrol.*, 336(1–2), 1–16.

704 Koch, F., Voytek, E.B., Day-Lewis, F.D., Healy, R., Briggs, M.A., Lane, J.W., Werkema, D.
705 (2015), 1DTempPro V.2: New Features for Parameter Estimation, Heterogeneity, and
706 Time-Varying Exchange, Groundwater, doi:10.1111/gwat.12369.

707 Krabbenhoft, D.P., Babiarz, C.L., 1992. The role of groundwater transport in aquatic mercury
708 cycling. *Water Resour. Res.*, 28(12): 3119-3128.

709 Lapham, W.W., 1989. Use of temperature profiles beneath streams to determine rates of
710 vertical ground-water flow and vertical hydraulic conductivity, U.S. Geological Survey
711 Water-Supply Paper 2337, Denver.

712 Lautz, L.K., 2010. Impacts of nonideal field conditions on vertical water velocity estimates
 713 from streambed temperature time series. *Water Resour. Res.*, 46,
 714 doi:10.1029/2009WR007917.

715 Lautz, L.K., 2012. Observing temporal patterns of vertical flux through streambed sediments
 716 using time-series analysis of temperature records, *J. Hydrol.*, 464, 199–215.

717 Lautz, L.K., Ribaud, R.E., 2012. Scaling up point-in-space heat tracing of seepage flux
 718 using bed temperatures as a quantitative proxy. *Hydrogeol. J.*, 20(7): 1223-1238.

719 Luce, C. H., Tonina, D., Gariglio, F., Applebee, R., 2013. Solutions for the diurnally forced
 720 advection–diffusion equation to estimate bulk fluid velocity and diffusivity in
 721 streambeds from temperature time series, *Water Resour. Res.*, 49, 488–506,
 722 doi:10.1029/2012WR012380.

723 McCallum, A.M., Andersen, M.S., Rau, G.C., Acworth, R.I., 2012. A 1–D analytical method
 724 for estimating surface water—groundwater interactions and effective thermal
 725 diffusivity using temperature time series, *Water Resour. Res.*, 48.

726 McCallum, A.M., Andersen, M.S., Rau, G.C., Larsen, J.R., Acworth, R.I., 2014. River–
 727 aquifer interactions in a semiarid environment investigated using point and reach
 728 measurements. *Water Resour. Res.*, 50(4): 2815–2829, doi: 10.1029/2012WR012007.

729 Rau, G.C., Andersen, M.S., McCallum, A.M., Acworth, R.I., 2010. Analytical methods that
 730 use natural heat as a tracer to quantify surface water–groundwater exchange, evaluated
 731 using field temperature records. *Hydrogeol. J.*, 18(5): 1093–1110, doi:
 732 10.1029/2011WR011038.

733 Rau, G.C., Andersen, M.S., Acworth, R.I., 2012. Experimental investigation of the thermal
 734 dispersivity term and its significance in the heat transport equation for flow in
 735 sediments. *Water Resour. Res.*, 48.

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

740 Roshan, H., Rau, G.C., Andersen, M.S., Acworth, I.R., 2012. Use of heat as a tracer to
741 quantify vertical streambed flow in a two-dimensional flow field. *Water Resour. Res.*
742 48, doi:10.1029/2012WR011918.

743 Shanafield, M., Hatch, C.H., Pohl, G. 2011. Uncertainty in thermal time series analysis
744 estimates of streambed water flux. *Water Resour. Res.* 47,
745 doi:10.1029/2010WR009574.

746 Soto-López, C.D., Meixner, T., Ferre, T.P.A., 2011. Effects of measurement resolution on the
747 analysis of temperature time series for stream-aquifer flux estimation. *Water Resour.*
748 *Res.*, 47, doi: 10.1029/2011WR010834.

749 Stallman, R.W., 1965. Steady one-dimensional fluid flow in a semi-infinite porous medium
750 with sinusoidal surface temperature, *J. Geophys.Res.* 70 (12), 2821–2827.

751 Stonestrom, D.A., Constantz, J., 2003. Heat as a tool for studying the movement of ground
752 water near streams. *U.S. Geological Survey Circular* 1260.

753 Swanson, T.E., Cardenas, M.B., 2011, Ex-Stream: A MATLAB program for calculating fluid
754 flux through sediment–water interfaces based on steady and transient temperature
755 profiles. *Comput. Geosci.* 37, 1664–1669.

756 Therrien, R., McLaren, R.G., Sudicky, E.A., Panday, S.M., 2006. *HydroGeoSphere*.
757 Groundwater Simulations Group, University of Waterloo, Waterloo, Canada.

- Tonina, D., Luce, C., Gariglio, F., 2014. Quantifying streambed deposition and scour from stream and hyporheic water temperature time series. *Water Resour. Res.*, 50, 287–191, doi: 10.1002/2013WR014567.
- Vogt, T., Schneider, P., Hahn–Woernle, L., Cirpka, O.A., 2010. Estimation of seepage rates in a losing stream by means of fiber–optic high–resolution vertical temperature profiling. *J. Hydrol*, 380(1–2): 154–164.
- Wagner, W., Cooper, J.R., Dittmann, A., Kijima, J., Kretzschmar, H.J., Kruse, A., Mares, R., Oguchi, K., Sato, H., Stocker, I., Sifner, O., Takaishi Y., Tanishita I., Trubenbach J., Willkommen T., 2000. The IAPWS industrial formulation 1997 for the thermodynamic properties of water and steam. *J. Eng. Gas Turbines Power*. 122: 150-182, doi: 10.1115/1.483186.
- Waples, D.W., Waples, J.S., 2004. A review and evaluation of specific heat capacities of rocks, minerals and subsurface fluids. Part 1: Minerals and nonporous rocks. *Natural Resources Research*, 13(2): 97–122.
- Young, P.C., Pedregal, D.J., Tych, W., 1999. Dynamic harmonic regression. *J. Forecast.* 18 (6), 369–394.
- Young, P.C, Taylor, C.J., Tych, W., Pegregal, D.J., McKenna, P.G., 2010. The Captain toolbox. Centre for Research on Environmental Systems and Statistics, Lancaster University, UL. <<http://www.es.lancs.ac.uk/cres/captain>>.

783 **Table and figure captions**

784

785 **Table 1:** Parameters in calculation of limits on κ_e values

786

787 **Table 2:** Thermal properties of sand column experiment and numerical model (From Lautz

788 (2012))

789

790 **Table 3:** q estimates and errors from ‘step experiment’ in Fig. 3

791

792

793 **Figure 1:** Configuration of the Lautz (2012) sand column experimental set up (adapted from

794 Lautz (2012)).

795

796 **Figure 2:** Showing a) model dimensions, boundary conditions and locations of observation

797 point profiles for considered simulations (flow paths for a homogeneous simulation shown),

798 and b) considered heterogeneous $\ln(K)$ field.

799

800 **Figure 3:** Reanalysis of the Lautz (2012) ‘step change’ experiment. a) shows temperatures at

801 0.05 and 0.10 m, b) shows the observed flux (q_{meas}) through the sand column (black) with the

802 calculated fluxes from the use of the A_r (green), $\Delta\phi$ (blue) combined $A_r\Delta\phi$ methods (purple),

803 c) shows the calculated κ_e ($\text{m}^2 \text{d}^{-1}$) from the thermal properties from the Lautz (2012)

804 experiment in black, the $\kappa_{eA_r\Delta\phi}$ (purple), and the plausible range of κ_e (gray).

805

Figure 4: Reanalysis of the Lautz (2012) ‘gradual change’ experiment. a) shows temperatures at 0.05 and 0.10 m, b) shows the observed flux (q_{meas}) through the sand column (black) with the calculated fluxes from the use of the A_r (green), $\Delta\phi$ (blue) combined $A_r\Delta\phi$ methods (purple), c) shows the calculated κ_e ($m^2 d^{-1}$) from the thermal properties from the Lautz (2012) experiment in black, the $\kappa_{eA_r\Delta\phi}$ (purple), and the plausible range of κ_e (gray).

Figure 5: Reanalysis of the Lautz (2012) ‘weather experiment’. a) shows temperatures at 0.05 and 0.10 m, b) shows the observed flux through the sand column (black) with the calculated fluxes from the use of the amplitude equations (green), phase equations (blue) and paired use of amplitude and phase (purple), c) shows the calculated κ_e ($m^2 d^{-1}$) from Lautz (2012) in black, the κ_e calculated from the paired use of amplitude and phase (purple), and the plausible range of κ_e (gray).

Figure 6: Analysis of HydroGeoSphere model output at $y = 1.0$ m. Left column shows the vertical component of flux (q_z) from HydroGeoSphere (black), ± 95 % confidence limits for q_{A_r} (green), ± 95 % confidence limits for $q_{\Delta\phi}$ (blue), and ± 95 % confidence limits for $q_{A_r\Delta\phi}$ (purple). Right column shows κ_e from HydroGeoSphere (black), $\kappa_{eA_r\Delta\phi}$ (purple), and the practical limits of κ_e (gray). From top to bottom, data are shown for $x = 0.5$ (a, b), 2.0 (c, d), 5.0 (e, f), 8.0 (g, h) and 9.5 m (i, j).

Table 1: Parameters in calculation of limits on κ_e values

Parameter, <i>symbol</i> (unit)	Parameter range	Source, comment
Porosity, n (–)	0.2– 0.5	Fitts (2013)
Specific heat capacity of solid, c_s (J kg ^{–1} °C ^{–1})	731– 1078	Waples and Waples (2004), range for quartzite
Specific heat capacity of water, c_w (J kg ^{–1} °C ^{–1})	4178– 4216	Calculated for 1 – 40 °C at 1 atm using Excel plugin (Wagner et al. 2000)
Thermal conductivity of solid, λ_s , (W m ^{–1} °C ^{–1})	2.18– 8.39	Hopmans et al. (2002), Rau et al. (2012), Stonestrom and Constantz (2003)
Thermal conductivity of water, λ_w , (W m ^{–1} °C ^{–1})	0.56– 0.63	Calculated for 1 – 40 °C at 1 atm using Excel plugin (Wagner et al. 2000)
Density of solid, ρ_s (kg m ^{–3})	2625– 2680	Stonestrom and Constantz (2003), Waples and Waples (2004)
Density of water, ρ_w (kg m ^{–3})	992– 1000	Calculated for 1 – 40 °C at 1 atm using Excel plugin (Wagner et al. 2000)

Table 2: Thermal properties of sand column experiment and numerical model (From Lautz (2012))

Parameter (<i>symbol</i>)	Value	Standard deviation
Porosity (n)	0.30	0.05
Dispersivity (β)	0.0 m	0.0 m
Baseline thermal conductivity (λ_0)	$3.4 \text{ W m}^{-1} \text{ }^{\circ}\text{C}^{-1}$	$0.6 \text{ W m}^{-1} \text{ }^{\circ}\text{C}^{-1}$
Volumetric heat capacity of solid (C_s)	$3.6 \times 10^6 \text{ J m}^{-3} \text{ }^{\circ}\text{C}^{-1}$	$0.4 \times 10^6 \text{ J m}^{-3} \text{ }^{\circ}\text{C}^{-1}$
Volumetric heat capacity of water (C_w)	$4.2 \times 10^6 \text{ J m}^{-3} \text{ }^{\circ}\text{C}^{-1}$	$0.2 \times 10^6 \text{ J m}^{-3} \text{ }^{\circ}\text{C}^{-1}$

Table 3: q estimates and errors from ‘step experiment’ in Fig. 3

Time Period (cycles)	Period name	$\bar{q}_{meas}$	Estimates (m d ⁻¹)			Errors (m d ⁻¹)		
			$\bar{q}_{A_r}$	$\bar{q}_{\Delta\phi}$	$\bar{q}_{A_r\Delta\phi}$	$\bar{q}_{A_r}$	$\bar{q}_{\Delta\phi}$	$\bar{q}_{A_r\Delta\phi}$
4.5 – 12.7	A	2.93	2.93	3.52	3.35	0.00	0.56	0.43
12.7 – 21.5	B	1.56	1.61	1.93	1.67	0.06	0.38	0.12
21.5 – 26.0	C	3.67	3.11	4.44	4.06	-0.56	0.77	0.39

figure 1

figure 2

Note: Not to scale

figure 3

figure 4

figure 5

figure 6