

Contents lists available at [SciVerse ScienceDirect](#)

Journal of Hydrology

journal homepage: [www.elsevier.com/locate/jhydrol](http://www.elsevier.com/locate/jhydrol)

## Automated calculation of vertical pore-water flux from field temperature time series using the VFLUX method and computer program

Ryan P. Gordon<sup>a,\*</sup>, Laura K. Lautz<sup>a</sup>, Martin A. Briggs<sup>a</sup>, Jeffrey M. McKenzie<sup>b</sup>

<sup>a</sup>Syracuse University, Department of Earth Sciences, 204 Heroy Geology Laboratory, Syracuse, NY 13244, USA

<sup>b</sup>McGill University, Earth and Planetary Sciences, 3450 University Street, Montreal, Quebec, Canada H3A 2A7

### ARTICLE INFO

#### Article history:

Received 25 May 2011

Received in revised form 8 November 2011

Accepted 28 November 2011

Available online xxxxx

This manuscript was handled by Philippe Baveye, Editor-in-Chief, with the assistance of Renduo Zhang, Associate Editor

#### Keywords:

Heat tracing

Groundwater–surface water interaction

Hyporheic

MATLAB

Dynamic Harmonic Regression

Distributed temperature sensing

### SUMMARY

Heat is a useful tracer for quantifying groundwater–surface water interaction, but analyzing large amounts of raw thermal data has many challenges. We present a computer program named VFLUX, written in the MATLAB computing language, for processing raw temperature time series and calculating vertical water flux in shallow sub-surface-water systems. The step-by-step workflow synthesizes several recent advancements in signal processing, and adds new techniques for calculating flux rates with large numbers of temperature records from high-resolution sensor profiles. The program includes functions for quantitatively evaluating the ideal spacing between sensor pairs, and for performing error and sensitivity analyses for the heat transport model due to thermal parameter uncertainty. The program synchronizes and resamples temperature data from multiple sensors in a vertical profile, isolates the diurnal signal from each time series and extracts its amplitude and phase angle information using Dynamic Harmonic Regression (DHR), and calculates vertical water flux rates between multiple sensor pairs using heat transport models. Flux rates are calculated every 1-to-2 h using four similar analytical methods. One or more “sliding analysis windows” can be used to automatically identify any number of variably spaced sensor pairs for flux calculations, which is necessary when a single vertical profile contains many sensors, such as in a high-resolution fiber-optic distributed temperature sensing (DTS) profile. We demonstrate the new method by processing two field temperature time series datasets collected using discrete temperature sensors and a high-resolution DTS profile. The analyses of field data show vertical flux rates significantly decreasing with depth at high-spatial resolution as the sensor profiles penetrate shallow, curved hyporheic flow paths, patterns which may have been obscured without the unique analytical abilities of VFLUX.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

While heat has been used for many years as a tracer of groundwater–surface water interaction (Anderson, 2005), several recent developments have improved the process of estimating vertical water flux to or from surface water using temperature time series, including the introduction of one-dimensional analytical solutions to the heat transport equation (Hatch et al., 2006; Keery et al., 2007), improved methods for signal processing of raw temperature time series (Keery et al., 2007; Young et al., 1999), and the deployment of many sensors in high-resolution profiles (Briggs et al., accepted for publication; Vogt et al., 2010). These advancements have made the estimation of pore-water flux easier, more accurate, and more useful, but their combined benefits have not yet been compiled into a step-by-step workflow that can be easily auto-

ated for processing large amounts of thermal data or performing complicated sensitivity analyses.

In streams, heat is a naturally occurring, non-reactive tracer that is easy to measure and model, making it a practical tool for studying water fluxes through the streambed (Constantz, 2008). Numerical models of fluid and heat flow (e.g. Healy and Ronan, 1996; Vogel et al., 2010) have often been calibrated to fit temperature patterns at specific boundaries (e.g. Niswonger and Prudic, 2003; Ronan et al., 1998; Vogel et al., 2011). Simplified analytical models have also been developed for idealized boundary conditions, such as the steady state model of Bredehoeft and Papadopoulos (1965) and the transient model of Stallman (1965). Steady state models were originally employed in the geothermal zone, but have since been used to quantify exchange with surface water (e.g. Schmidt et al., 2007). Lapham (1989) used vertical temperature profiles beneath streams to quantify the rate of vertical water flow using a numerical approximation of the one-dimensional, analytical heat transport equation of Stallman (1965). Analytical extensions to the Stallman model have since been developed (Hatch

\* Corresponding author. Tel.: +1 315 443 2672 (O).

E-mail address: [rpgordon@syr.edu](mailto:rpgordon@syr.edu) (R.P. Gordon).

et al., 2006; Keery et al., 2007) that solve for the one-dimensional flux of fluid between two vertically-spaced temperature sensors using the phase and amplitude changes of transient thermal signals.

Analytical modeling of transient heat transport takes advantage of the daily fluctuation in temperature that takes place in stream water. As streams heat and cool during the diurnal cycle, a quasi-sinusoidal temperature signal with a 24-h period propagates into the streambed. The depth, speed, and strength of the propagation of the thermal front depends on the heat capacity and conductivity of water and sediment, as well as the volume of fluid moving vertically through the sediment (Stonestrom and Constantz, 2003). By measuring the attenuation of amplitude and the increase in phase angle of the temperature signal with depth, it is possible to quantify the magnitude and direction of the vertical component of water flux (Hatch et al., 2006; Keery et al., 2007).

Temperature is easy and inexpensive to measure in streams and the streambed, and the analytical models of Hatch et al. (2006) and Keery et al. (2007) are relatively simple. However, processing raw temperature time series and extracting the information necessary to compute flux, namely amplitude and phase, is not trivial. Fluctuations in stream temperature are caused not only by daily variation in solar incidence, but by seasonality, weather patterns, precipitation and snowmelt, stream shading, long-wave radiation, changes in latent heat, upstream watershed characteristics, and even wind (Caissie, 2006). Furthermore, temperature sensors have different noise characteristics, depending on the technology used. Real-world temperature time series therefore contain a variety of signals and noise, and filtering techniques are required to isolate the signal of interest, in this case the 24-h, diurnal oscillation (Hatch et al., 2006; Keery et al., 2007). After filtering, objective procedures are needed to extract signal amplitudes and phase angles to compute flux rates.

A variety of methods have been used by previous researchers to filter and analyze temperature time series to calculate fluid flux rates. Hatch et al. (2006) isolated the diurnal signal using a cosine taper band-pass filter and then selected daily temperature maxima and minima (using a semi-automated computer program) to calculate amplitude attenuation and time lag with depth. Fanelli and Lautz (2008) similarly identified daily extrema manually, but did not filter the raw temperature data before applying the Hatch et al. (2006) method, producing quantifiable errors (Lautz, 2010). Keery et al. (2007) used Dynamic Harmonic Regression (DHR) (Young et al., 1999) both to isolate the diurnal signal and to extract amplitude and phase information. Vogt et al. (2010) similarly used the amplitude and phase output from a DHR analysis of temperature time series to solve for flux using the analytical solution of Stallman (1965). Swanson and Cardenas (2010) avoided filtering altogether by fitting a stationary sine function to individual day-long periods of temperature records, and using the amplitude and phase of the sine wave in the analytical equations of Hatch et al. (2006) and Keery et al. (2007).

Recently, scientists have used heat modeling methods in research on such diverse topics as streambed chlorobenzene concentrations (Schmidt et al., 2011), wetland phosphorus mobility (Maassen and Balla, 2010), aquifer recharge from perennial pools (Rau et al., 2010), potential culvert removal (Anderson et al., 2010), hyporheic exchange around geomorphic features (Crispell and Endreny, 2009; Fanelli and Lautz, 2008; Lautz et al., 2010), streambed infiltration (Vogt et al., 2010), groundwater discharge at stream cross-sections (Jensen and Engesgaard, 2011), and contaminated groundwater contribution to streams (Lautz and Ribaud, submitted for publication). However, these studies have all used different methods for collecting, filtering, and analyzing temperature time series and estimating vertical flux, some with distinct flaws. Each researcher has had to devise individual solu-

tions to—or simply ignore—a host of challenges created by field temperature records, and they have been forced to develop their own computational workflows to deal with the volume of thermal data collected. Furthermore, researchers are increasingly collecting very large datasets from many sensors in high-resolution temperature profiles using fiber-optic distributed temperature sensing (DTS) methods (e.g. Briggs et al., accepted for publication; Vogt et al., 2010). This proliferation in data gives researchers greater flexibility when selecting the ideal spacing between temperature sensors (Hatch et al., 2006), and allows them to detect changes in vertical flux with depth to greater precision, but also creates a computational problem in which the number of potential flux calculations increases quadratically with the number of sensors.

We believe that heat transport modeling would be used more often by a wider range of researchers if an automated, step-by-step program were freely available to easily and consistently calculate vertical fluid flux from raw temperature time series, without resorting to shortcuts in signal processing. Such an automated workflow in the form of a computer program would make processing very large amounts of data from many temperature sensors more practical and flexible. The wide availability and use of such a program would support a more consistent application of time series analysis methods among different studies by hydrologists, ecologists, geochemists, engineers, and others. By automating the calculation of vertical fluid flux, the program may also help avoid computational error.

In this paper, we present a computer program that automates the entire process of calculating vertical flux rates from raw temperature time series in the shallow beds of streams or other surface water bodies. The method builds upon previous work (Hatch et al., 2006; Keery et al., 2007; Vogt et al., 2010) and adds new techniques for processing large numbers of temperature records from high-resolution sensor profiles with greater spatial and temporal resolution. We automate the new method using a computer program named VFLUX (Vertical Fluid [Heat] Transport Solver). The program formats, synchronizes, and resamples temperature data from multiple sensors in a vertical temperature profile (TP), then isolates the diurnal signal from each time series using DHR, extracts amplitude and phase information, and calculates vertical flux rates between multiple sensor pairs using a “sliding analysis window” and the analytical models of Hatch et al. (2006) and Keery et al. (2007). Flux rates are calculated every 1-to-2 h, depending on the sampling rate of the original data, or at every sample interval if the original data was sampled more coarsely than every 2 h. The program is written in the MATLAB computing language, and is designed to integrate with other command-based MATLAB functions or scripts, allowing it to be easily modified and incorporated into custom workflows. To illustrate this point, we demonstrate two MATLAB functions that run VFLUX iteratively in order to perform analyses of error due to thermal parameter uncertainty, as well as sensitivity of the model to each thermal parameter. The VFLUX program will be useful to other researchers who need a flexible and robust method for automatically calculating vertical flux rates from profiles of temperature sensors, including complex error analysis, and especially for practitioners of high-resolution DTS sensors. We demonstrate our method and the functions of VFLUX by processing two field temperature time series datasets collected using different sensor technologies. These examples illustrate the challenges associated with modeling real-world time series, and how they can be overcome using the VFLUX method and program.

At the time of publication, only one automated method for processing raw temperature time series and calculating vertical flux has been published as a computer program, named Ex-Stream (Swanson and Cardenas, 2011). VFLUX differs from the program of Swanson and Cardenas in several ways. Ex-Stream is operated

through a graphical user interface, while VFLUX is a command-line tool executed using well-documented commands, which allows it to be called by custom scripts and easily incorporated into other MATLAB programs. VFLUX employs DHR for extracting the diurnal signal from a time series and estimates changing flux at hourly time scales, while Ex-Stream operates by fitting a static sine function to the raw data day-by-day (Swanson and Cardenas, 2011), which may not fit well to parts of the time series that deviate significantly from a perfect sinusoid, such as during weather pattern shifts, and only allows flux calculations at daily intervals. Ex-Stream also uses two steady-state solutions (Bredenhoeft and Papadopoulos, 1965; Schmidt et al., 2007), in addition to the analytical solutions of Hatch et al. (2006) and Keery et al. (2007). Finally, VFLUX is designed for calculating flux between many variably-spaced sensor pairs in a single profile of temperature sensors, which is useful for detecting changing flux with depth and for identifying the ideal sensor spacing for a particular experiment, and is necessary when employing many sensors, such as in a high-resolution DTS profile. Ex-Stream only calculates flux between one pair of sensors in a profile at a time.

To our knowledge, VFLUX is the first published computer program for estimating water flux that automates a complete approach to data pre-processing of raw temperature time series, including multiple sensor synchronization, resampling, and filtering of a single tracer frequency with a robust method like DHR. It is also the first method that uses a “sliding window” to provide flux results at many depths with high spatial resolution. The program includes new tools for evaluating the sensitivity of the heat transport model to sensor spacing, which allows the user to intelligently determine the ideal sensor spacing for a particular application. The VFLUX package also includes two separate functions that perform analyses of model sensitivity to thermal parameters and estimates of error due to parameter uncertainty. Automated error analysis and sensitivity programs such as these have not yet been published, and many recently published papers do not include error or uncertainty estimates for reported flux values (e.g. Schmidt et al., 2011; Vogt et al., 2010). The analysis of field temperature data in this paper is also the first high-resolution presentation, to our knowledge, of vertical flux rates that decrease with depth due to penetration through curved hyporheic flow cells.

## 2. Methods

### 2.1. Field data collection

A set of thermal records was collected in the bed of Ninemile Creek, a fourth-order stream (contributing area to site: 177 km<sup>2</sup>) in Marcellus, New York between August 27 and September 18, 2009. The streambed at the site was composed of cobbly, sandy gravel with some silt. Thermal parameters of the saturated sediment were estimated through field observation and the guidelines provided by Lapham (1989), and are summarized in Table 1. Seven iButton Thermochron band-gap temperature sensors and loggers (Maxim Integrated Products, Inc., Sunnyvale, CA) were embedded in a steel rod to make a vertical TP (Fig. 1A). The steel TP rod was driven into the streambed at the head of a riffle so that one sensor was 0.05 m above the streambed interface, and the others were positioned at 0.05, 0.10, 0.15, 0.20, 0.25 and 0.30 m below the streambed interface. The sensors have a resolution of 0.0625 °C, a manufacturer-reported accuracy of ±0.5 °C and a manufacturer-reported thermal response time of 130 s, much shorter than our sampling interval. In a laboratory test of 80 iButtons, including those used in the present study, all sensors were precise within a range of 0.12 °C in an ice bath and at room temperature. The sensors were programmed to record temperature every

10 min, for a total of 3070 measurements at each sensor. All seven sensors were synchronized prior to deployment, so that measurements were taken on the same time schedule. There were no breaks in sampling, so the measurements were evenly spaced in time throughout.

The second set of thermal records was collected in the bed just upstream of a small beaver dam in Cherry Creek, a second-order stream (contributing area to site: 31.1 km<sup>2</sup>) in Lander, Wyoming between July 11 and August 11, 2010 (see Briggs et al., accepted for publication, for further details). The streambed was composed of interbedded organic-rich silt, sand, and gravels. Thermal parameters of the saturated sediment were again estimated through field observation and the relationships presented by Lapham (1989), and are summarized in Table 1. Temperature measurements were collected using a fiber-optic distributed temperature sensing (DTS) method, in which the temperature-dependent backscatter of a laser pulse is used to continuously estimate temperature along a glass fiber (Dakin et al., 1985; Selker et al., 2006). The fiber used at Cherry Creek was housed in stainless-steel cladding and wrapped in a tight coil around a one-meter long, threaded PVC rod (Fig. 1B). An Agilent Distributed Temperature Sensor (model N4386A, Agilent Technologies, Inc., Santa Clara, CA) was used to measure temperature every meter along the cable length, yielding a temperature measurement every 0.014 m along the length of the vertical TP rod. The rod was driven vertically into the streambed and positioned such that temperature was measured at 57 locations between 0.014 and 0.78 m depth below the streambed interface. The DTS instrument in this experimental configuration had a precision of ±0.2 °C, estimated using the manufacturer’s software, and was calibrated using mixed thermal baths monitored with iButton Thermochrons (±0.5 °C accuracy, 0.0625 °C resolution, 130 s thermal response time). The DTS system recorded temperature every 20 min, for a total of 2182 measurements at each of the 57 depths along the TP. Due to the continuous nature of DTS, measurements of temperature at different depths were all synchronized in time; however, because the battery of the DTS unit needed to be changed daily, some sample times were missed. There were 30 gaps in the data that were less than 81 min in length, and 3 gaps due to equipment malfunctions that were between 203 and 524 min in length. The missing samples were linearly interpolated from the original data and inserted into the time series, so that the time series processed with VFLUX was evenly spaced in time.

### 2.2. The VFLUX program

The temperature time series from Ninemile Creek and Cherry Creek were processed with the VFLUX program to calculate vertical water flux rates at multiple depths in the streambeds of the study sites. VFLUX is distributed as a MATLAB toolbox, a set of functions written in the MATLAB computing language that are designed to run in the MATLAB environment. The program was developed in MATLAB 7.10.0 (R2010a) (The MathWorks, Inc., Natick, MA). The VFLUX functions are executed by typing commands at the MATLAB command prompt, and can also be called by external functions or integrated into custom scripts. The functions are distributed as open-source code (in MATLAB M-files) that is free to use, easily modified, and well-commented. The VFLUX program files are included in Supplementary material available with the online version of this article, and the most up-to-date version may be downloaded from the following web site: [http://hydrology.syr.edu/lautz\\_group/vflux.html](http://hydrology.syr.edu/lautz_group/vflux.html). Full documentation is provided along with the program, which explains in detail the program’s commands, inputs, outputs, options, and switches.

Given a set of temperature time series from a single vertical sensor profile and the depth of each sensor, the VFLUX program will perform the following six major steps (see Fig. 2): (1) format

**Table 1**

Values of the physical parameters used to calculate flux rates at Ninemile Creek, NY and Cherry Creek, WY. Values listed in table are also the base or mean values used in the sensitivity and Monte Carlo analyses for Ninemile Creek. The standard deviation used in the Monte Carlo analysis is shown in parentheses. Thermal properties were estimated based on guidelines in Lapham (1989) and values in Fetter (2001) and CRC (2011).

Property	Symbol	Units	Value	
			Ninemile Creek	Cherry Creek
Total porosity	$n$	Dimensionless	0.20 (0.04) <sup>a</sup>	0.35
Baseline thermal conductivity	$\lambda_o$	$\text{J s}^{-1} \text{m}^{-1} \text{°C}^{-1}$	2.26 (0.31) <sup>b</sup>	1.30
Thermal dispersivity	$\beta$	m	0.001 (0) <sup>c</sup>	0.001
Volumetric heat capacity of sediment	$C_s$	$\text{J m}^{-3} \text{°C}^{-1}$	$2.09 \times 10^6$ ( $3.1 \times 10^4$ ) <sup>d</sup>	$2.09 \times 10^6$
Volumetric heat capacity of water	$C_w$	$\text{J m}^{-3} \text{°C}^{-1}$	$4.18 \times 10^6$ ( $2.1 \times 10^4$ ) <sup>e</sup>	$4.18 \times 10^6$

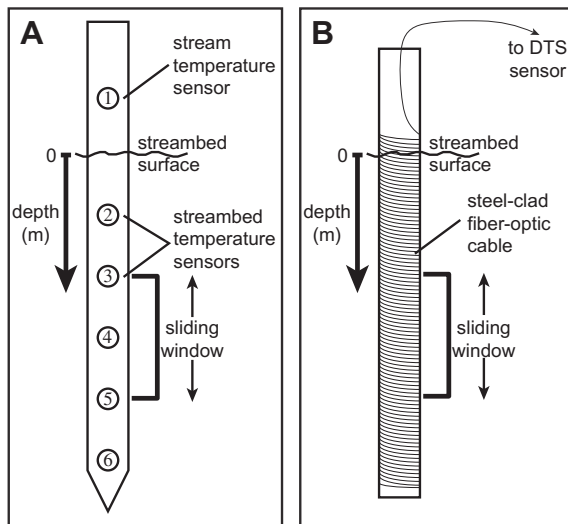
<sup>a</sup> Low and high  $n$  values for sensitivity analysis are 0.18 and 0.34, respectively.

<sup>b</sup> Low and high  $\lambda_o$  values for sensitivity analysis are 1.21 and  $2.47 \text{ J s}^{-1} \text{m}^{-1} \text{°C}^{-1}$ , respectively.

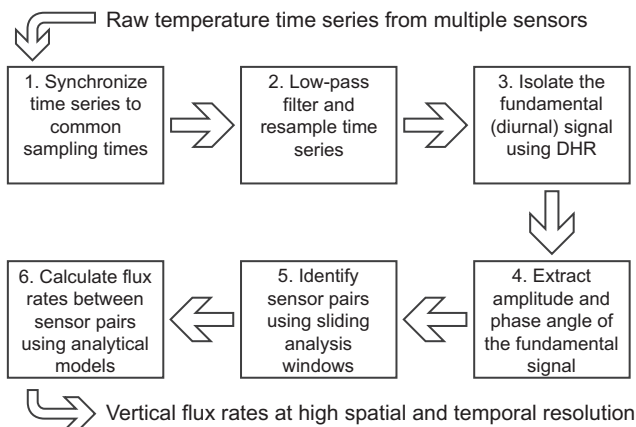
<sup>c</sup> Low and high  $\beta$  values for sensitivity analysis are 0 and 0.1 m, respectively.

<sup>d</sup> Low and high  $C_s$  values for sensitivity analysis are  $2.03 \times 10^6$  and  $2.15 \times 10^6 \text{ J m}^{-3} \text{°C}^{-1}$ , respectively.

<sup>e</sup> Low and high  $C_w$  values for sensitivity analysis are  $4.14 \times 10^6$  and  $4.23 \times 10^6 \text{ J m}^{-3} \text{°C}^{-1}$ , respectively.



**Fig. 1.** (A) A diagram of a typical vertical temperature profile (TP), showing five iButton Thermochron temperature sensors installed in the streambed and one in the water column, which gives the temperature of the water at the streambed interface. The sensors are not necessarily evenly-spaced. The sliding window (in this case, two sensor-spacings long) identifies sensor pairs between which flux will be calculated. (B) A diagram of a high-resolution DTS temperature profile, showing a coil of fiber-optic cable wrapped around a threaded PVC rod. Using a distributed temperature sensing device, the temperature of short segments (0.024 m) of the rod length can be measured, which are equivalent to temperature sensors. The sliding window again identifies sensor pairs for flux calculations.



**Fig. 2.** A flowchart of the six major steps in the VFLUX method.

and synchronize all the time series to a single vector of sampling times; (2) low-pass filter and resample the time series; (3) isolate the fundamental signal (the signal of interest, typically diurnal) using DHR; (4) extract amplitude and phase information for the fundamental signal using DHR; (5) identify pairs of sensors based on one or more “sliding analysis windows”; and (6) calculate vertical water flux rates between the identified sensor pairs. Additional programs included with the VFLUX package perform sensitivity analyses on the thermal input parameters, and calculate confidence intervals for flux estimates based on thermal parameter uncertainty.

### 2.2.1. Format and synchronize time series

Time series data for each temperature sensor are input as a vector (a one-dimensional array) of evenly-spaced sample times, in days, and a corresponding vector of observed temperatures, in degrees C. If all the sensors were sampled at the same time schedule (i.e., the time vectors for all sensors have the same start time, end time, and sampling rate), then VFLUX simply formats the time series into a MATLAB structure array to be processed by the remainder of the program. However, if different sensors in the temperature profile were sampled at different time schedules (i.e., they have different start or end times, and/or different sampling rates), then VFLUX synchronizes all the time series by reducing them to the “lowest common denominator”; that is, it trims all the input series to the shortest time range common to all and, if necessary, linearly interpolates all the input series at the same sampling rate. This creates a single time vector that is common to all the temperature data. The interpolation method can easily be changed to a spline interpolation or another method.

### 2.2.2. Low-pass filter and resample time series

The second step in the VFLUX program is to reduce the sampling rate to approximately 12–24 samples per fundamental cycle, if the original sampling rate was greater. This step improves the filtering results because VFLUX uses the standard frequency-domain optimization method within the Dynamic Harmonic Regression (DHR) model. This optimization method is sensitive to oversampling, and functions best when the sampling rate is not much higher than the frequency of the oscillation of interest (Young et al., 1999; Włodzimierz Tych, personal communication, March 3, 2011). If the time series are oversampled, then the signal of interest becomes compressed in the frequency domain, and the optimization method is ineffective at identifying the model hyperparameters. Although it is beneficial in this case to resample at a lower rate before applying DHR, oversampling in the collection of raw data is still desirable, because it allows for the identification and reduction of noise by low-pass filtering during the resample

process, and it avoids the greater problem of undersampling. If the raw data is undersampled during collection (fewer than approximately 12 samples per cycle), then important information about the signal may be irrecoverably lost (Box et al., 1994).

VFLUX first calculates an appropriate integer factor by which to reduce the sampling rate (called *rfactor*). The program then decimates the time vector by this rate, keeping every *rfactor*-th sample time and discarding the rest. Alternately, *rfactor* can be specified by the user when calling the VFLUX program. VFLUX then resamples each temperature vector using an anti-aliasing, lowpass FIR (finite impulse response) filter designed with a Kaiser window and decimates the filtered signal by *rfactor*, discarding samples that are not needed. VFLUX reduces the edge effects of the filtering process by reflecting and mirroring the temperature vector at both ends as padding, and then discarding this padding after running the resample command.

The low-pass filter reduces high-frequency noise that is inherent in the natural temperature signal or is an artifact of the sensor technology. The anti-aliasing properties of the filter prevent noise from reflecting around the Nyquist frequency and disrupting the signal identification in the next step.

### 2.2.3. Isolate the fundamental tracer signal

The third step in the VFLUX program is to isolate a tracer signal of a single fundamental frequency (typically the diurnal signal). This is performed using DHR as implemented in the Captain Toolbox (Young et al., 2010), a set of MATLAB functions developed at Lancaster University. DHR is a method for non-stationary time series analysis that is particularly useful for extracting harmonic signals from dynamic environmental systems (Young et al., 1999). DHR was first used in hydrologic heat modeling by Keery et al. (2007). The driving temperature oscillation in a stream system changes in time due to weather and seasonality, and vertical water flux rates may also change in time; therefore, the amplitude and phase of the diurnal signal in the streambed are also time-varying, and a non-stationary approach to signal extraction is necessary. DHR, a simplification of the Unobserved Component model, has the following form:

$$y_t = T_t + C_t + e_t \quad (1)$$

where  $y_t$  is the observed time series,  $T_t$  is a trend or zero-frequency component,  $C_t$  is a cyclical component, and  $e_t$  is an irregular, white-noise component (Young et al., 1999). The cyclical term is modeled as a sum of the fundamental signal and its associated harmonics:

$$C_t = \sum_{i=1}^N [a_{i,t} \cos(\omega_i t) + b_{i,t} \sin(\omega_i t)] \quad (2)$$

where  $a_{i,t}$  and  $b_{i,t}$  are stochastic time-varying parameters (TVPs) and  $\omega_1, \omega_2, \dots, \omega_N$  are the fundamental frequency ( $\omega_1$ ) and its harmonics ( $\omega_i = i \omega_1$ ) up to the Nyquist frequency ( $\omega_N$ ). The trend component  $T_t$  can also be considered as a zero-frequency term ( $\omega_0 = 0$ ) incorporated into the cyclical term sum. This DHR model can be thought of as a non-stationary extension of the discrete Fourier transform, where the amplitude and phase of each time series component themselves change with time. Identification of the TVPs is achieved in a stochastic state space formulation using two-step Kalman filtering and fixed-interval smoothing (Young et al., 1999).

VFLUX calls the DHR function for each time series to separate the fundamental cyclical signal from the trend, noise, and harmonics. In practice, only the harmonic components that are present in the original temperature data need to be identified, typically  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$ . Therefore, by default VFLUX attempts to identify a trend, the fundamental signal ( $\omega_1$ ), and the first and second harmonics ( $\omega_2$  and  $\omega_3$ ) using an auto-regression (AR) frequency spectrum created with the Captain Toolbox. VFLUX displays a

diagnostic plot of the AR spectrum for each time series, allowing the user to evaluate whether the appropriate number of harmonics was identified. The specific harmonics, and the method that VFLUX uses to identify them, can be modified by the user. VFLUX then fits the DHR model to the AR spectrum, by optimization to a non-linear least-squares objective function using the Captain Toolbox (Young et al., 1999). The AR spectrum and the model fit are plotted and displayed to the user. Finally, VFLUX filters and isolates the trend, fundamental signal, identified harmonics, and noise components. These components are then plotted and displayed to the user.

### 2.2.4. Extract amplitude and phase information

The fourth step in the VFLUX program is performed during the DHR analysis above. The amplitude and phase of any harmonic component at any discrete time can be calculated by the equations:

$$A_{i,t} = \sqrt{a_{i,t}^2 + b_{i,t}^2} \quad (3)$$

$$\phi_{i,t} = \tan^{-1} \left( \frac{a_{i,t}}{b_{i,t}} \right) \quad (4)$$

where  $A_{i,t}$  is the amplitude and  $\phi_{i,t}$  is the phase angle for the component with frequency  $\omega_i$  at time  $t$  (Vogt et al., 2010). The Captain Toolbox function 'dhr' supplies the amplitude and phase angle of the fundamental frequency at each time step in the resampled temperature time series, producing an amplitude time series and a phase time series for each sensor. From Eq. (4), the phase angle reported by DHR must be between  $-\pi/2$  and  $\pi/2$ ; it is therefore possible that the phase angle of a time series reported by DHR may jump from  $\pi/2$  to  $-\pi/2$  or from  $-\pi/2$  to  $\pi/2$  during the duration of the time series. VFLUX converts the phase angle into a total phase offset for each sample of each time series by adding  $\pi$  whenever the phase time series jumps from  $\pi/2$  to  $-\pi/2$ , and subtracting  $\pi$  whenever the phase time series jumps from  $-\pi/2$  to  $\pi/2$ .

### 2.2.5. Identify sensor pairs

The analytical methods of Hatch et al. (2006) and Keery et al. (2007) (the Hatch and Keery methods) calculate water flux between a single pair of vertically-spaced temperature sensors. If more than two sensors are deployed in the same profile, multiple combinations of sensors can be used to calculate vertical flux. The multiple sensor pairs may have different midpoint depths, different separation distances, and/or they may overlap in space (for an example, see Jensen and Engesgaard, 2011). When a TP contains many sensors, such as in a high-resolution DTS profile, the number of possible combinations of sensors can become very large (equal to  $N(N-1)/2$  in a profile of  $N$  sensors). The main benefit of using DTS technology in a vertical temperature profile is the high spatial resolution and therefore the large number of effective sensors available, which allow for many possible sensor spacings at different depths. A large number of possible sensor pairs with a range of separation distances can be very useful to help identify the ideal sensor spacing for a particular flux rate and set of field conditions (Hatch et al., 2006). It is also important to have control over the depth of the sensor pairs in order to detect changes in the vertical component of hyporheic flux, which may vary significantly with depth (Buffington and Tonina, 2009).

The large number of potential pairs in a high-resolution DTS profile leads to a very large number of calculations. VFLUX is designed to automatically identify any number of sensor pairs through the use of one or more "sliding analysis windows". The window(s) are specified by the user in units of "sensor-spacings" when calling the VFLUX program. The actual distance between sensors is not used for identifying pairs, so sensors do not need to be evenly-spaced. For example, a window of 2 sensor-spacings is depicted in Fig. 1A. As this analysis window slides from the top of the

TP to the bottom, it identifies all the sensor pairs that are separated by 2 spacings. With a total of 6 sensors in the profile, this window would identify sensors 1 and 3, 2 and 4, 3 and 5, and 4 and 6 as pairs. Multiple windows can be defined in a single VFLUX run; if we included a window of 1 sensor-spacing in the above example, then VFLUX would also identify sensors 1 and 2, 2 and 3, 3 and 4, 4 and 5, and 5 and 6 as pairs. The maximum number of possible pairs in this example is 15, which would all be identified by defining windows of 1–5 sensor-spacings.

When processing the data from Ninemile Creek with VFLUX, every possible window was used, i.e., 1–6 sensor spacings, which correspond to sensor separation distances of 0.05, 0.10, 0.15, 0.20, 0.25 and 0.30 m. This leads to 21 total sensor pairs. In the case of Cherry Creek, the high-resolution DTS profile was processed with windows of 8, 9, 10, 11, and 12 sensor spacings, which correspond to sensor separation distances of 0.110, 0.124, 0.138, 0.152, and 0.166 m, for a total of 235 sensor pairs. VFLUX performed flux calculations at each 2-h time step for every pair of sensors that was separated by any of these distance values. A total of 4809 flux calculations were made with the data from Ninemile Creek, and 85,540 flux calculations with the data from Cherry Creek.

### 2.2.6. Calculate vertical flux

The sixth and final step in the VFLUX program is to calculate vertical water flux between the identified sensor pairs. The flux value calculated for each pair is assigned to a point equidistant between the sensors, referred to as the center-of-pair depth. The Hatch and Keery methods are both analytical solutions to the one-dimensional heat transport equation:

$$\frac{\delta T}{\delta t} = \kappa_e \frac{\delta^2 T}{\delta z^2} - q \frac{C_w}{C} \frac{\delta T}{\delta z} \quad (5)$$

where  $T$  is temperature ( $^{\circ}\text{C}$ ),  $t$  is time (s),  $\kappa_e$  is the effective thermal diffusivity of the saturated sediment ( $\text{m}^2 \text{s}^{-1}$ ),  $z$  is depth (m),  $q$  is fluid flux ( $\text{m s}^{-1}$ ),  $C$  is the volumetric heat capacity of the saturated sediment ( $\text{J m}^{-3} \text{ }^{\circ}\text{C}^{-1}$ ), and  $C_w$  is the volumetric heat capacity of the water (Goto et al., 2005; Stallman, 1965).  $C$  is calculated as the mean of  $C_w$  and  $C_s$ , the volumetric heat capacity of the sediment grains, weighted by total porosity. Both Hatch et al. (2006) and Keery et al. (2007) developed methods to solve for  $q$  in Eq. (5) by measuring the attenuation of the amplitude of a quasi-sinusoidal temperature signal as it propagates vertically through the streambed, or by measuring the speed at which it propagates.

The analytical solution provided by Hatch et al. (2006) solves for the vertical water flux between two sensors as a function of either amplitude or phase differences between the sensors' temperature signals:

$$q = \frac{C}{C_w} \left( \frac{2\kappa_e}{\Delta z} \ln A_r + \sqrt{\frac{\alpha + v^2}{2}} \right) \quad (6)$$

$$|q| = \frac{C}{C_w} \sqrt{\alpha - 2 \left( \frac{4\pi\Delta t\kappa_e}{P\Delta z} \right)^2} \quad (7)$$

where  $q$  is vertical fluid flux in the downward direction ( $\text{m s}^{-1}$ ),  $A_r$  is the ratio of amplitudes (a measure of amplitude attenuation) between the lower sensor and the upper sensor (dimensionless),  $\Delta z$  is the distance between the two sensors in the streambed (m),  $v$  is the velocity of the thermal front ( $\text{m s}^{-1}$ ),  $\Delta t$  is the time lag (a measure of the speed of signal propagation) between the two temperature signals (s),  $P$  is the period of the temperature signal (s), and  $\alpha$  is defined by:

$$\alpha = \sqrt{v^4 + \left( \frac{8\pi\kappa_e}{P} \right)^2} \quad (8)$$

Effective thermal diffusivity,  $\kappa_e$ , which incorporates conductivity and dispersivity, is given by:

$$\kappa_e = \left( \frac{\lambda_0}{C} \right) + \beta |v_f| \quad (9)$$

where  $\lambda_0$  is the baseline thermal conductivity (i.e., thermal conductivity without the dispersive effect of fluid flow) of the saturated sediment ( $\text{J s}^{-1} \text{m}^{-1} \text{ }^{\circ}\text{C}^{-1}$ ),  $\beta$  is thermal dispersivity (m), and  $v_f$  is the linear particle velocity ( $\text{m s}^{-1}$ ) (Hatch et al., 2006).

Note that flux,  $q$ , is a true volumetric flux rate (volume per area per time), while linear particle velocity,  $v_f$ , is a true velocity (distance per time), and both are distinct from the thermal front velocity,  $v$ , despite all having the same units. The relationships between these terms are:

$$v_f = \frac{q}{n_e} \quad (10)$$

$$q = \frac{C}{C_w} v \quad (11)$$

where  $n_e$  is the effective porosity (dimensionless) (Hatch et al., 2006).

The analytical solution to the Stallman (1965) equation provided by Keery et al. (2007) is similar to the Hatch method, but does not include thermal dispersivity. Vertical water flux is again calculated as a function of amplitude or phase differences (Keery et al., 2007):

$$\left( \frac{H^3 \ln A_r}{4\Delta z} \right) q^3 - \left( \frac{5H^2 \ln^2 A_r}{4\Delta z^2} \right) q^2 + \left( \frac{2H \ln^3 A_r}{\Delta z^3} \right) q + \left( \frac{\pi C}{\lambda_0 P} \right)^2 - \frac{\ln^4 A_r}{\Delta z^4} = 0 \quad (12)$$

$$|q| = \sqrt{\left( \frac{C\Delta z}{\Delta t C_w} \right)^2 - \left( \frac{4\pi\Delta t\lambda_0}{P\Delta z C_w} \right)^2} \quad (13)$$

where  $H = C/\lambda_0$ . Note that in both Eqs. (7) and (13), the phase methods only give the magnitude of flux, and not its direction.

In order to apply the Hatch and Keery methods, VFLUX first must calculate the amplitude ratio and time lag at each time step (i.e., at each sample) in the resampled and filtered fundamental signal. At each sample (time  $t$ ), the time lag,  $\Delta t$ , is defined by:

$$\Delta t = \frac{P}{2\pi} (\phi_{z+\Delta z,t+\Delta t} - \phi_{z,t}) \quad (14)$$

where  $\phi_{z,t}$  is the phase angle of the sensor at depth  $z$  and at time  $t$  (see Eq. (4)), and  $\phi_{z+\Delta z,t+\Delta t}$  is the phase angle of the sensor deeper by  $\Delta z$  at time  $t + \Delta t$  (Vogt et al., 2010). However, because  $\Delta t$  must be known to determine the phase angle at  $t + \Delta t$ , and because the phase angle typically changes slowly in time on a daily scale, we assume it is generally safe to calculate the time lag using the estimate:

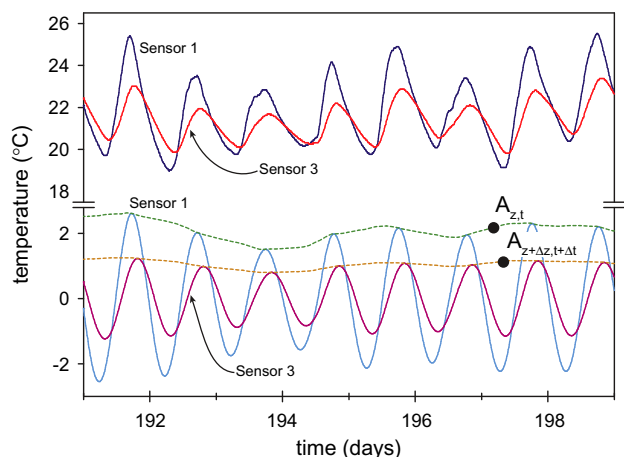
$$\Delta t \approx \frac{P}{2\pi} (\phi_{z+\Delta z,t} - \phi_{z,t}) \quad (15)$$

which is the formulation used in VFLUX. The program then calculates the amplitude ratio,  $A_r$ , by:

$$A_r = \frac{A_{z+\Delta z,t+\Delta t}}{A_{z,t}} \quad (16)$$

where  $A_{z,t}$  is the amplitude of the sensor at depth  $z$  and at time  $t$  (see Eq. (3)), and  $A_{z+\Delta z,t+\Delta t}$  is the amplitude of the sensor deeper by  $\Delta z$  at time  $t + \Delta t$  (Keery et al., 2007). The amplitude ratio and time lag are illustrated in Fig. 3.

VFLUX calculates vertical water flux,  $q$ , for each sample in the resampled time series using the Hatch amplitude and phase methods (Eqs. (6) and (7)) and the Keery amplitude and phase methods (Eqs. (12) and (13)). The sediment and thermal properties total porosity, dispersivity, baseline thermal conductivity, and



**Fig. 3.** Two discrete time series from sensors 1 and 3 in Fig. 1A. The raw temperature time series are shown in the upper part of the plot. In the lower part are shown the diurnal frequency components of the two time series, extracted using Dynamic Harmonic Regression (DHR). The dashed lines are the amplitudes of the diurnal components, extracted using DHR. The pair of points illustrates the time lag and amplitude attenuation between equivalent parts of the temperature signal as it progresses through different depths.

volumetric heat capacity of the sediment and water must be input when running VFLUX, and the other parameters in Eqs. (6), (7), (12), and (13) are calculated by the program.

Because of the relationship in Eq. (11), the thermal front velocity is found on both sides of Eqs. (6) and (7) and they must be solved iteratively. VFLUX solves Eqs. (6) and (7) by subtracting  $q$  from both sides of the equation and then finding the value of  $q$  at which the right side of the equation equals zero. Eq. (12) must be solved by finding the roots of the third-order polynomial on the left side. At least one root must be real, and the other two may be real or complex, although only a result with one real root represents a true physical flux rate (Keery et al., 2007). VFLUX solves the third-order polynomial and only accepts a single real root as the value for  $q$ . If all three roots are real, then VFLUX warns the user and automatically assigns a null value (Keery et al., 2007). Eq. (13) is solved for  $|q|$  explicitly as written. If any value for  $A_r$ ,  $\Delta t$ , or  $q$  cannot be calculated for any reason during a VFLUX run, then a null value is assigned, a warning is displayed to the user and the reason is described. The most common reasons that VFLUX fails to calculate flux are that  $A_r$  is greater than or equal to one or  $\Delta t$  is less than or equal to zero (both usually caused by non-ideal sensor spacing), or an attempt to find roots failed to converge on a real value.

There is an optional switch in the VFLUX program that can be set to also exclude flux values (write a null value) if the calculated flux estimate is outside a specified range of optimal sensitivity. This sensitivity check is only available for the Hatch amplitude method, and must be enabled in the VFLUX code. As described in Results and Discussion below, Hatch et al. (2006) identified a range of flux values for which the analytical heat transport model is most sensitive. This range depends on the sensor spacing  $\Delta z$  and the specific thermal properties, and is based on the slope of the amplitude ratio versus flux curve, or the derivative  $dA_r/dq$  (Hatch et al., 2006). If this slope is too low, which occurs at very high and low  $A_r$  values, then small errors in the amplitude ratio can cause large errors in flux. If this option is enabled, VFLUX will calculate  $dA_r/dq$  and assign a null value if the derivative is below a certain value,  $0.001 \text{ d m}^{-1}$  by default (after Hatch et al., 2006). Ultimately, violations of this sensitivity range are caused by sensors that are spaced too close together (the amplitude ratio is too close to 1) or too far apart (the amplitude ratio is too close to 0), therefore, this low-derivative check acts like an additional constraint on the ideal sensor spacing.

When the VFLUX program runs to completion, it creates four data matrices containing the results of the flux calculations: one matrix each for the Hatch amplitude method, Hatch phase method, Keery amplitude method, and Keery phase method, plus a vector containing the time at which each flux calculation was made. VFLUX also creates metadata matrices that describe the sensor pairs used for each set of flux calculations, and the thermal parameters used in the heat transport models. At the end of a successful VFLUX run, the program can optionally enter a results visualization routine. If selected by the user, the visualization routine displays several plots, containing raw and filtered time series, amplitude and phase information, and final flux results through time for all four methods. Default MATLAB zooming, panning, and editing tools can be used to customize the plots. The visualization routine also can display bar charts of successful flux calculations for different sensor pairs and window sizes, to help in the analysis of ideal sensor spacing, as discussed below. These plots show the percent of time steps for which flux was successfully calculated for each sensor depth and each window size. This visualization illustrates which sensor pair combinations are within the ideal sensitivity range of the models. Other combinations of results can be displayed after exiting VFLUX by plotting flux through time using the 'plot' commands in MATLAB.

It is recommended that the first 2 or 3 days of flux data from the beginning and end of the data collection periods be discarded, due to the edge effects of digital filtering (Hatch et al., 2006; Keery et al., 2007), although that choice is left up to the user and is not implemented in VFLUX. In the case of the field data from Ninemile Creek and Cherry Creek used in this study, the first and last 2 days of flux results were discarded.

### 2.2.7. Sensitivity and uncertainty estimation programs

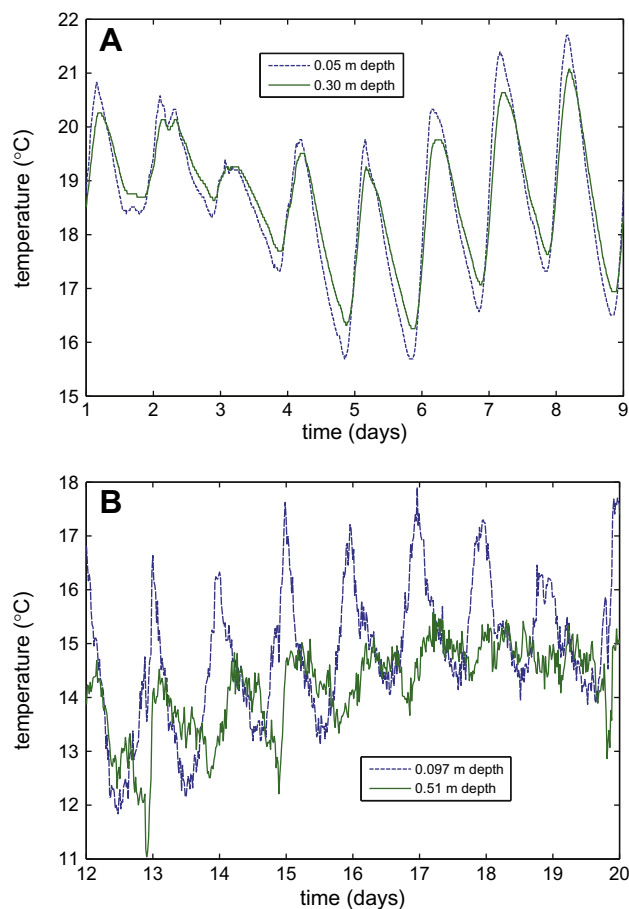
The VFLUX package contains two additional programs, "vfluxsens", for performing sensitivity analyses, and "vfluxmc", for performing Monte Carlo error estimations. These programs both run the VFLUX method iteratively in order to produce different estimates of flux from different thermal parameters. The sensitivity analysis program is run using similar inputs to the standard VFLUX, but instead of single thermal and sediment input parameters, the user inputs a low and high value for each parameter in addition to the estimated or measured value ("base" value). The program first runs VFLUX only for the base values, then subsequently runs VFLUX twice for each parameter, using the input high and low values, while maintaining all other parameters at the base values. At the end, for each sensor pair, the program displays a series of plots, each showing maximum and minimum values of flux through time for each of the varied parameters. The sensitivity analysis was performed for the data collected at Ninemile Creek, with sediment and thermal parameter ranges as shown in Table 1. The high and low values in Table 1 represent the full range of typical physical values for coarse-grained sediment and water between 0 and 25 °C from Lapham (1989), Fetter (2001), and CRC (2011).

The Monte Carlo error estimation program can be used to create statistical confidence intervals around flux estimates through time. Such confidence intervals are needed to determine if flux estimates are significantly different from zero, or to compare differences in flux between different sites or even different depths. Many recent publications that report flux values calculated with analytical heat transport methods do not include confidence intervals (e.g. Crispell and Endreny, 2009; Engelhardt et al., 2011; Schmidt et al., 2011; Vogt et al., 2010), making it difficult to determine if flux estimates are significant. The Monte Carlo estimation program has similar inputs to the standard VFLUX and sensitivity analysis programs, but takes as inputs mean and standard deviation values for each thermal and sediment parameter, which can be based on estimations or actual laboratory or field measurements. The program

performs 1000 realizations (by default). For each realization, the program selects a set of random parameters from normal distributions with mean and standard deviation, as input. All but two of the parameters are uncorrelated, so that the random values selected in each realization are unrelated. Thermal conductivity and porosity are inversely correlated around their respective means, following guidelines in Lapham (1989), which show that thermal conductivity varies with dry bulk density. Porosity also affects the calculated value of  $C$ , the volumetric heat capacity of the sediment–water matrix. In each of the realizations, the program runs the standard VFLUX analysis with the randomly generated parameter values. Flux values are calculated for each sensor pair and each point in time during each of the 1000 realizations, and then standard deviation values are calculated through all 1000 realizations. Finally, approximate 95-percent confidence intervals are created around the mean flux values by adding and subtracting two standard deviations from the mean flux values. This Monte Carlo analysis was performed for the data collected at Ninemile Creek, with parameter means and standard deviations as shown in Table 1.

### 3. Results and discussion

Two examples of raw temperature time series from Ninemile Creek in New York and Cherry Creek in Wyoming are plotted in



**Fig. 4.** Examples of raw temperature time series data from (A) Ninemile Creek at 0.05 m and 0.30 m depth and (B) Cherry Creek at 0.097 m and 0.51 m depth. Note in both plots the amplitude attenuation and time lag between the shallower and deeper sensors, the asymmetric saw-toothed pattern of daily increases and decreases in temperature, the changing trend through time, and the amount of noise present in each time series.

Fig. 4. A diurnal temperature signal with a 24-h period of oscillation is obvious at 0.05 and 0.30 m depths at Ninemile Creek (Fig. 4A) and at 0.097 m depth at Cherry Creek, and it can also be made out by eye despite the noise at 0.51 m depth in Cherry Creek (Fig. 4B). In both cases, the deeper time series has smaller amplitude and is shifted forward in time relative to the shallower time series, as we would expect from the models of Hatch et al. (2006) and Keery et al. (2007). In both raw data sets, the time series exhibit non-sinusoidal features such as noise, asymmetry, and changes in trend from day-to-day (i.e., the daily mean temperature changes from one day to the next). The DTS data particularly appear to contain significant high-frequency noise, due to the relatively low precision ( $\pm 0.2$  °C) of the DTS system used in this specific configuration. Both datasets have enough noise, natural or sensor-derived, to occasionally create local maxima and minima that do not correspond to the daily maximum and minimum temperatures. In both stream systems, temperature rises faster in the morning and falls more slowly in the evening, leading to asymmetric or saw-toothed patterns. This periodic asymmetry gives significant amplitude to the first few harmonics of the fundamental diurnal signal, which will be seen during the DHR analysis below. In both data sets there are also variations in the temperature trend (which can be thought of as a moving average daily temperature), which are caused by changing weather patterns and differing amounts of solar radiation. All of these deviations from an ideal sinusoid are common in natural temperature time series and can cause error in flux calculations (Lautz, 2010).

Both Hatch et al. (2006) and Keery et al. (2007) note that the amplitude ratio method and the phase lag method differ in their sensitivity to flux depending on sensor spacing, signal frequency, and flux rate, with the amplitude ratio being more sensitive at lower rates on the order of  $1 \times 10^{-5}$  m s<sup>-1</sup> (Hatch et al., 2006), which are more typical of general stream values found in the literature (e.g. Fanelli and Lautz, 2008; Hatch et al., 2006; Keery et al., 2007; Lautz et al., 2010). Furthermore, Lautz (2010) demonstrated that the amplitude ratio method is more reliable in non-ideal conditions, which likely exist at most field sites, and Hatch et al. (2006) notes that filtering can create spurious phase shifts. Most researchers have focused on the amplitude ratio method in their research (e.g. Fanelli and Lautz, 2008; Keery et al., 2007; Schmidt et al., 2011; Vogt et al., 2010), and we do the same here by concentrating this discussion on the amplitude ratio results.

A number of researchers have investigated potential errors in water flux estimates using heat modeling methods, including the effects of error in temperature measurement, thermal diffusivity, and sensor spacing (Shanafield et al., 2011), sediment heterogeneity (Ferguson and Bense, 2011; Schornberg et al., 2010), and non-vertical flow (Lautz, 2010; Rau et al., 2010). These non-ideal conditions are certainly present to some degree at both field sites, particularly non-vertical flow (see Section 3.6, below), but the potential uncertainty is not discussed in detail. However, we do conduct a sensitivity and uncertainty analysis for errors in sediment and thermal properties, which demonstrate how VFLUX can be a useful tool for investigating uncertainty and error in modeling or field studies such as Lautz (2010) or Shanafield et al. (2011).

#### 3.1. Sensor synchronization

In both data sets, from Ninemile Creek and Cherry Creek, the temperature sensors in each TP were synchronized by nature of the experimental setup, so synchronization of the time series using VFLUX was unnecessary. The first step performed by VFLUX was therefore simple; the program copied the time series into a MATLAB structure to be processed by the remainder of the program. In other potential cases, however, especially those that use discrete

sensors and data loggers like iButtons, the sensors may not collect measurements in sync. In these cases, synchronization is an important step in processing for two reasons. First, it is important that all time series be sampled at the same rate so that the discrete-time DHR analysis can be performed consistently on all the time series in a TP. Different discrete sampling rates have different bandwidth capacities, and therefore may contain different frequency spectra (Box et al., 1994), making comparison between DHR results more difficult. Second, DHR as implemented in the Captain Toolbox is blind to the absolute time of each point in a time series, and calculates phase angles based on the position of the first data point in each time series (Young et al., 2010). If the first data point is shifted in absolute time from one time series to another, the Captain Toolbox will not recognize this and the phase shift,  $\Delta t$ , in Eq. (15) will be calculated incorrectly. Finally, it is obviously necessary that the time series from all sensors be complete over the same time period, so that comparisons between the signals can be made for every day and every possible pair of sensors.

### 3.2. Oversampling and high-frequency noise

The resample routine is an important part of the VFLUX process because it reduces oversampled data and removes high-frequency noise. Oversampling reduces the effectiveness of DHR optimization and filtering, as described above in Methods. The raw data from Ninemile Creek was sampled every 10 min, for a sampling rate of 144 samples per day, which is much higher than the rate of approximately 12-to-24 samples per day that is ideal for describing a perfect diurnal signal. However, oversampling in the original data collection stage has no disadvantage as long as the data are resampled prior to a DHR analysis, and oversampling has the advantage of better describing the diurnal signal in relation to high-frequency noise, which is then removed by the low-pass filter. When analyzing the Ninemile Creek time series, VFLUX chose an integer resample factor (*rfactor*) of 12, so that the reduced sampling rate was 12 samples per day. Fig. 5A shows an example of data from Ninemile Creek (0.05 m depth) at its original sampling rate and after being resampled and low-pass filtered by VFLUX. Because the original data collected at Ninemile Creek was not overly noisy, the resampled time series appears to be very similar to the original series in form, but simply described by fewer data points. The time series from Cherry Creek, however, include a lot of high frequency noise (Fig. 5B). Temperature at Cherry Creek was sampled every 20 min, for a sampling rate of 72 samples per day. The value of *rfactor* was therefore calculated to be 6, and the reduced sampling rate was again 12 samples per day. Comparing a subset of the resampled data with the raw data in Fig. 5B shows that the low-pass FIR filter has reduced the high-frequency noise and smoothed the signal, in addition to lowering the sampling rate.

### 3.3. Non-sinusoidal raw temperature signals

Departure from an ideal sinusoidal temperature signal creates several challenges when calculating flux rates from real-world temperature time series using the Hatch and Keery methods, especially if relying on daily extrema to calculate amplitude attenuation and time lag (e.g., Fanelli and Lautz, 2008; Hatch et al., 2006; Lautz, 2010) or if fitting a static sine function to the raw time series (e.g., Swanson and Cardenas, 2010). First, the analytical solutions to the one-dimensional heat transport equation are based on the assumption that the temperature oscillations at the surface and at depth are sinusoidal waves of a single frequency (Hatch et al., 2006; Keery et al., 2007; Stallman, 1965). However, real-world temperature time series contain many harmonic and trend components, as previously discussed, requiring the frequency of interest to be isolated by filtering. Second, the presence of local maxima

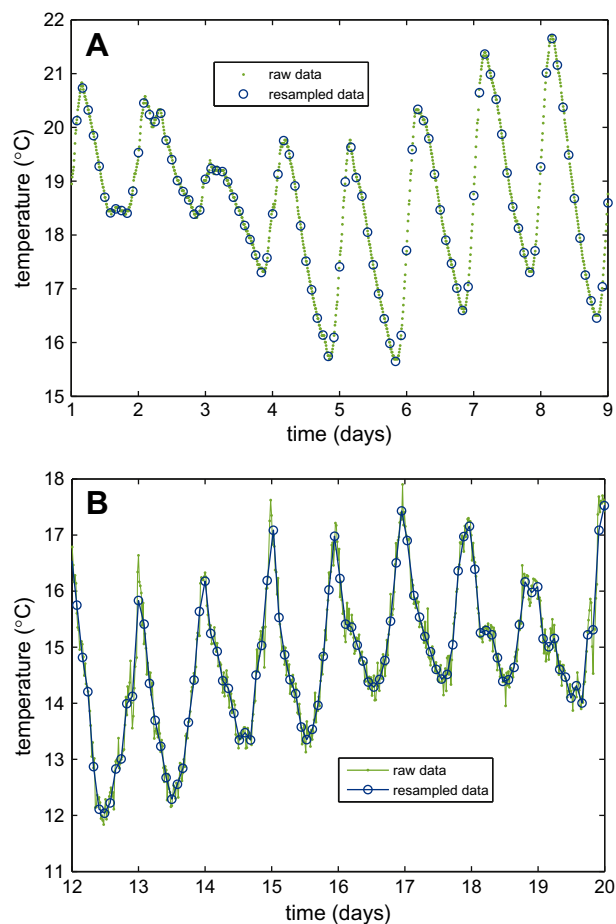


Fig. 5. Examples of temperature time series data before and after being low-pass filtered and resampled. (A) From Ninemile Creek at 0.05 m depth. (B) From Cherry Creek at 0.097 m depth. The low-pass filter has visibly reduced the high-frequency noise in the Cherry Creek data.

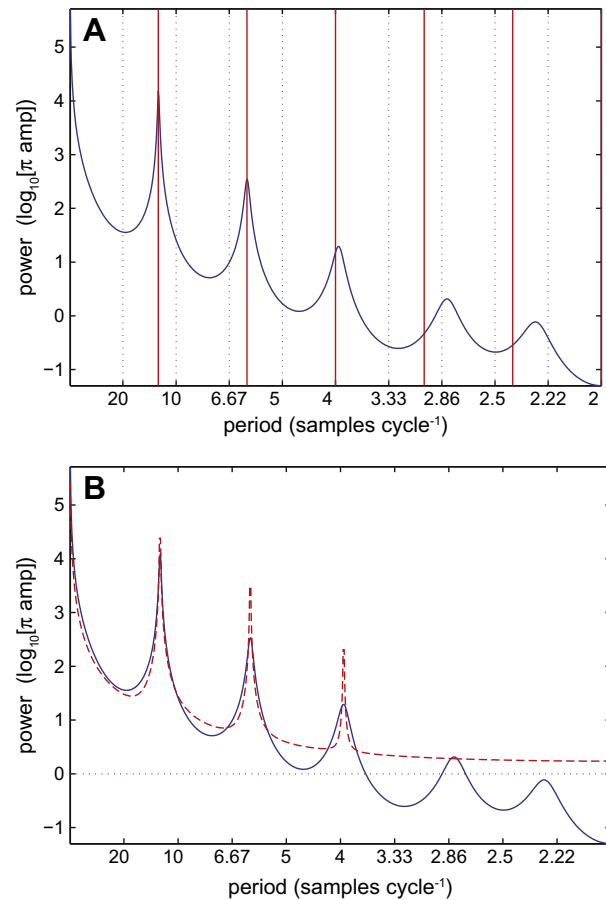
and minima (such as in the second and third days of Fig. 4A) make the detection of daily extrema more challenging with an automated peak-detection program, although this can certainly be overcome with good programming. Third, the saw-toothed pattern found in many natural temperature series (including those from Ninemile and Cherry Creeks) mean that maxima and minima do not occur 12 h apart; therefore, a different phase angle would be inferred from the maximum time than from the minimum time. Furthermore, the time lag ( $\Delta t$ ) between a shallow and deep sensor is not the same when computed with the maximum and the minimum of an irregular signal, because the slower temperature change before each minimum allows the streambed to respond faster, with a shorter time lag (Lautz, 2010). Fourth, when the temperature trend changes through time, estimates of the signal amplitude (made by halving the difference between daily extrema) can be very inaccurate, and can change dramatically from one day to the next despite a constant flux rate. Using Fig. 4A as an example, the temperature difference between the maximum at 3.1 days and the following minimum is higher than twice the true diurnal signal amplitude because the trend is decreasing quickly over that time period. Alternately, if one were to use the difference between the minimum at 2.9 days and the maximum at 3.1 days, the estimated amplitude would be too low, for the same reason. This example also shows that calculating the amplitude using a maximum and the following minimum can often give a different amplitude estimate for the same day than using a minimum and the following maximum. Abrupt changes in trend, such as those that

occur with frontal weather patterns, can entirely mask the presence of a daily maximum or minimum, making it impossible on that day to calculate flux using extrema. When the raw time series deviates substantially from a pure sinusoid, it may also be very difficult to fit the raw data with a static sine function (Swanson and Cardenas, 2011), leading to substantial errors in the determination of amplitude and phase using that method, as well (for example, a static sinusoid fits poorly to any of the 24-h periods shown in Fig. 4A between days 2 and 5).

These challenges illustrate how an accurate filtering mechanism is important for accurate time series analysis, and how an objective, consistent method of extracting amplitude and phase is needed to calculate flux rates. DHR is a robust method of filtering that can identify harmonic components in a complex time series with greater accuracy than a band-pass filter (Keery et al., 2007). DHR can successfully identify the diurnal signal and separate it from trend and harmonics even when the trend is changing quickly, so flux can be calculated on days when it might otherwise be impossible. DHR is also fully automated and can function objectively and consistently on a wide range of datasets without the need for human selection of extrema or other temperature patterns (Keery et al., 2007). VFLUX has been programmed to make the application of DHR straightforward for the user for this particular application.

VFLUX processed the time series from each Ninemile Creek sensor using the same DHR settings. An auto-regression (AR) spectrum of data from the 0.05 m sensor is shown in Fig. 6A, showing the peaks in power for various components of the temperature record. A similar figure is produced by VFLUX for each time series using the 'arspec' function from the Captain Toolbox and is shown to the user during program execution. Note the strong peaks in power at the fundamental frequency (the diurnal oscillation at a period of 12 samples per cycle) and its first two harmonics (with periods of 6 and 4 samples per cycle). Two lesser peaks can be seen in the spectrum near the third and fourth harmonics (3 and 2.4 samples per cycle), but their power is very low and their identification by periodicity is not clear. The half-peak in the spectrum at the zero frequency (infinite period) to the left of the fundamental peak represents the trend component in Eq. (1), or a zero-frequency component in Eq. (2). The strong harmonic response at 2 and 3 times the fundamental frequency is due to the asymmetric but periodic saw-tooth temperature patterns, as any periodic function that deviates from a pure sinusoid can be decomposed into a sum of sinusoidal harmonics at multiples of the fundamental frequency (for details of basic Fourier theory, see, e.g., Boas, 1983). Fig. 6B shows the fit of the DHR model (Eqs. (1) and (2)) to the time series AR spectrum, optimized by VFLUX using the DHR optimization function 'dhropt' from the Captain Toolbox. The fit of the model to the data appears very good along the trend and the fundamental and first harmonic peaks, and good for the second harmonic, as well. The optimized value of the non-linear (logarithmic) least squares objective function is 171.5. DHR has successfully differentiated the fundamental diurnal frequency from its trend and asymmetric components. The results of the DHR analysis for the Ninemile Creek 0.05 m sensor are shown in Fig. 7. The DHR model matches the actual data well, and has extracted the diurnal signal even in difficult sections with local extrema and changing trend, such as between 1 and 5 days.

The AR spectrum and the DHR model fit to data from 0.097 m depth in Cherry Creek are shown in Fig. 8A. The trend and diurnal signal are strong, with a weaker first harmonic and a very weak second harmonic. The optimized DHR model fits the data well over the trend and diurnal peak, but not as well over the two harmonics. The non-linear least squares objective function is 166.4. A similar figure for data from 0.51 m depth in Cherry Creek is shown in Fig. 8B. The diurnal signal is much weaker at this depth compared

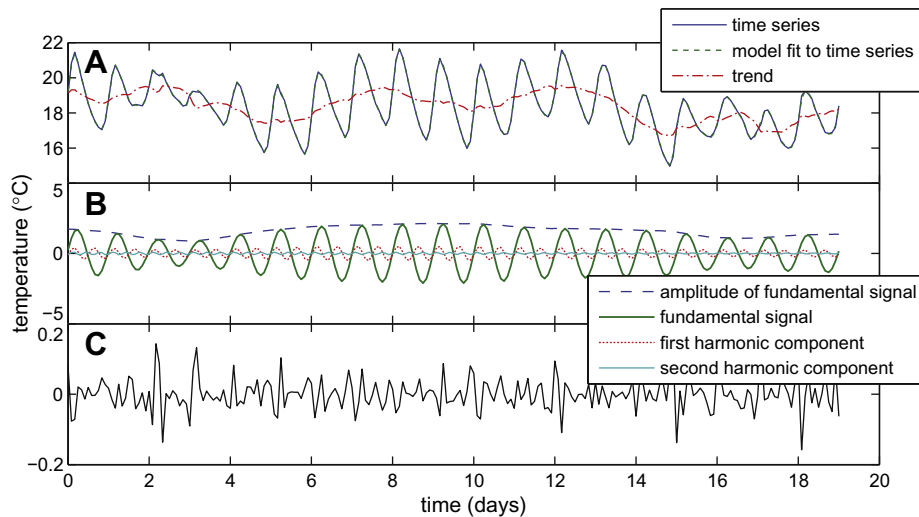


**Fig. 6.** (A) The AR(12) spectrum (auto-regression spectrum of order 12) of the 0.05 m time series from Ninemile Creek, showing peaks near the fundamental frequency (12 samples per cycle) and the first and second harmonics (6 and 4 samples per cycle). The slope of the spectrum up and left towards the zero frequency (infinite period) is the trend component of the DHR model. The vertical lines are indices showing the expected periods of the fundamental signal and its harmonics (12, 6, 4, 3 and 2.4 samples per cycle). (B) The optimized DHR model for the AR spectrum in (A) (dashed line), showing a good model fit to the data (solid line) along the trend, fundamental, and first two harmonic peaks.

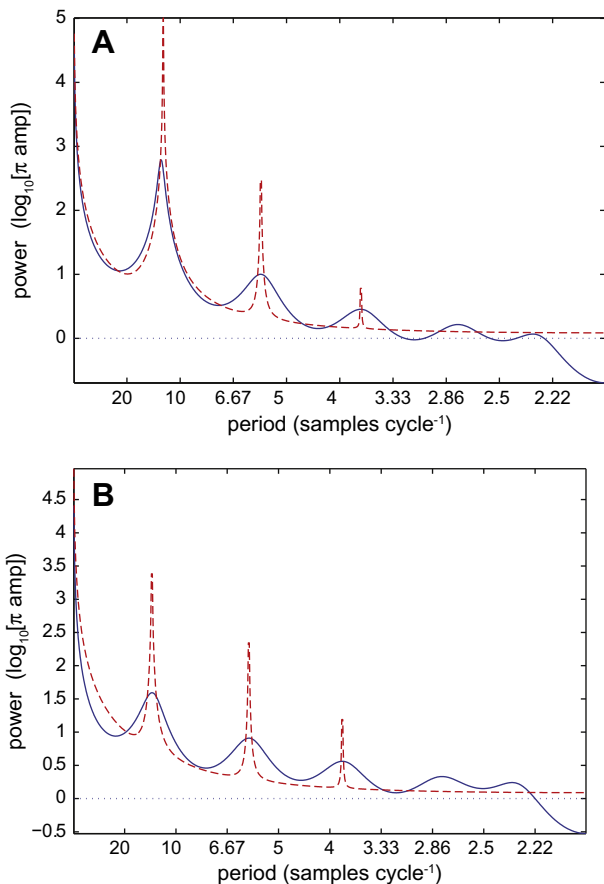
to the sensor noise, and the DHR optimization is not as effective at modeling the trend or fundamental signal. As a result, the fitted spectrum in Fig. 8B does not fit the actual spectrum well over the trend or diurnal frequency peaks. The value of the objective function, at 355.4, is more than twice the value for the fit to the time series at 0.097 m. At all sensors below 0.51 m depth at Cherry Creek, the mean amplitude of the diurnal signal is less than 0.2 °C, the precision of the DTS equipment. As a result, the diurnal signal cannot be easily filtered from the noise. We suggest in general that signal components with amplitudes below the sensor precision should be used with caution. The filtered components from 0.51 m depth are shown in Fig. 9.

### 3.4. Sample-by-sample flux calculation

After filtering the temperature time series with DHR, it would certainly be possible to identify daily maxima and minima and thereby calculate daily amplitudes and phase angles; however, another significant advantage of using DHR is that it allows VFLUX to calculate amplitude and phase at each sub-daily time step in the resampled data (by Eqs. (3) and (4)), rather than estimating them daily. There are several benefits to calculating flux at the sub-daily time scale.



**Fig. 7.** Results of DHR filtering of the 0.05 m time series from Ninemile Creek. (A) The original resampled time series, the DHR model fit to the time series in the time domain, and the DHR-identified trend. (B) The fundamental (diurnal) frequency and its amplitude, and the first two harmonics. (C) The residual, or the difference between the original data and the model fit.



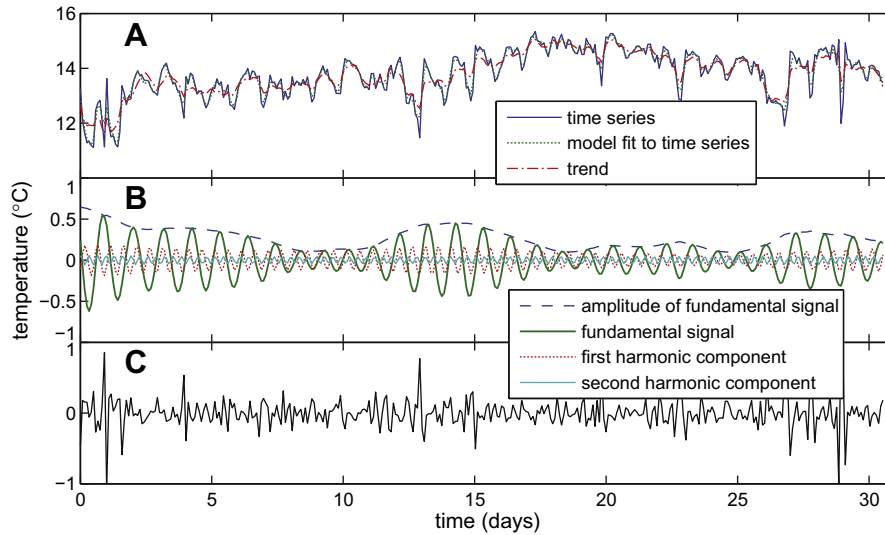
**Fig. 8.** (A) The AR(12) spectrum of the 0.097 m time series from Cherry Creek (solid line), with the optimized DHR model fit (dashed line). (B) The AR(12) spectrum of the 0.51 m time series from Cherry Creek (solid line), with the optimized DHR model fit (dashed line).

We have already discussed how DHR avoids the subjective human identification of extrema, but sample-by-sample calculations also have the benefit of being a more precise-in-time estimate of the character of a time series. Note how in Figs. 7 and 9,

the amplitude of the filtered diurnal signal changes with time. We know from the DHR model (Eqs. (1) and (2)) that the amplitude of a single-frequency component is time-varying. When an estimate of amplitude is calculated as half of the difference between a daily maximum and minimum, the amplitude is effectively averaged over the time period between the extrema; however, with DHR, the amplitude and phase of each frequency component is determined at each sample, which represents a period of time only as long as the sampling period (2-to-4 h in VFLUX). By employing DHR, VFLUX can calculate flux rates between a sensor pair from the amplitude attenuation and time lag at any sample in the time series, and it thereby avoids relying heavily on the time and temperature of only two measurements per day, i.e., the daily extrema. The high temporal resolution of sample-by-sample flux calculations is most useful during days with unusual weather patterns, when it may provide flux results for fractions of the day, while the daily extrema method may fail to give any flux measurement at all.

Sub-daily flux calculations might be needed in systems where flux rates change quickly, such as in tidal estuaries, dam-controlled reservoirs or rivers, or in flashy streams during discrete precipitation events. However, it should be noted that, to our knowledge, no research has been performed on the applicability of the Hatch or Keery analytical methods to systems where flux rates change significantly at sub-daily scales. Keery et al. (2007) did note that the Stallman equation is based on the assumption that flux does not change at scales shorter than the diurnal period, and “it is therefore neither necessary nor appropriate to calculate vertical flux for every time step.” However, VFLUX would be an ideal tool in any field or modeling investigation of the Hatch or Keery method in dynamic-flux environments.

One clear value of making many independent estimates of a daily flux rate is that the mean or median of such estimates, which have inherent noise, may best estimate the value in question. We recommend that estimates of physical vertical water flux at a field site be made by finding the central tendency of several days of flux calculations (as in Lautz, 2010). Shown in Fig. 10 are the amplitude and phase angle of a pair of sensors from Ninemile Creek through time, and the flux rates that were calculated by VFLUX at each time step. The slight differences between flux calculated with the Keery and Hatch methods in Fig. 10C are due to the differing treatment of



**Fig. 9.** Results of DHR filtering of the 0.51 m time series from Cherry Creek. The fundamental component and its amplitude are identified, but the model fit to the raw data is not ideal, with some large spikes in the residual. Note that the average amplitude of the diurnal signal is approximately 0.27 °C, only slightly greater than the  $\pm 0.2$  °C precision of the DTS equipment. (A) The original resampled time series, the DHR model fit to the time series in the time domain, and the DHR-identified trend. (B) The fundamental (diurnal) frequency and its amplitude, and the first two harmonics. (C) The residual, or the difference between the original data and the model fit.

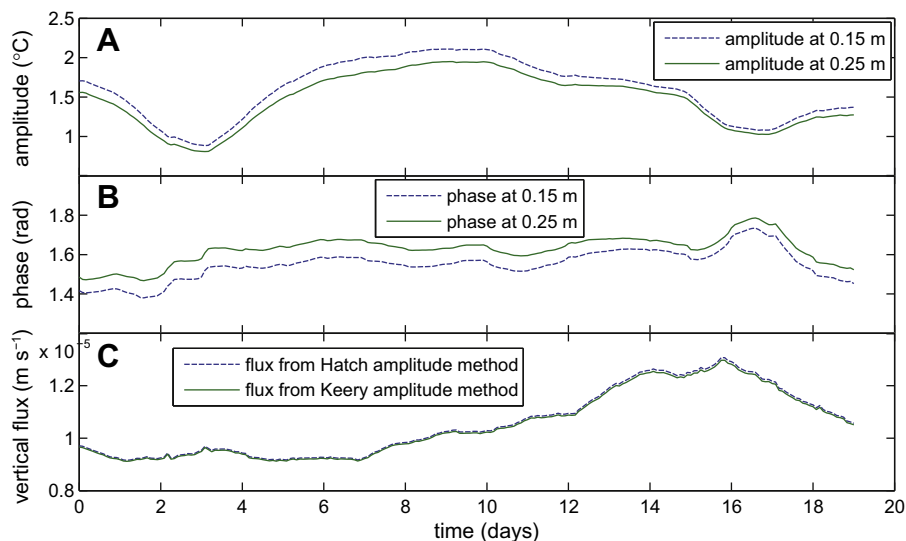
thermal dispersion by the two models (see the discussion in Keery et al., 2007).

### 3.5. Sensor pair identification and ideal sensor spacing

The Hatch and Keery methods both require concurrent information from two vertically-spaced temperature sensors to make a single determination of vertical water flux. Choosing the best spacing between the sensors ( $\Delta z$ ) requires balancing several considerations. The flux calculated for a single sensor pair is an estimate of the volume of water per area moving vertically over the distance  $\Delta z$ , or the average vertical flux over  $\Delta z$ . VFLUX assigns the flux estimate to a point equidistant between the sensors, referred to as the center-of-pair depth; however, as  $\Delta z$  increases, so does the averaging distance represented by that single point. A long averaging distance might not be appropriate if the vertical component of flux changes along the distance  $\Delta z$ . If the purpose of an experiment is

to observe changes in vertical flux with depth, or if it is critical that a flux estimate represent a certain depth with precision, then the goal should be to have the smallest sensor spacing (window) possible.

On the other hand, the amplitude attenuation and time lag between two sensors must be large enough to be resolved by the resolution and response time of the sensors (Hatch et al., 2006). If the sensors are too close together, then the amplitude ratio can effectively approach one (or even surpass one due to measurement error), and the time lag can likewise approach zero. If the sensors are too far apart, then the tracer signal may be damped out below the sensor resolution at the lower depth, and the amplitude ratio can be effectively zero. If sensor spacing is too large, vertical resolution is also lost. Furthermore, the analytical heat transport models are most sensitive within defined ranges of  $A_r$  and  $\Delta t$  that vary depending on the magnitude and direction of flux and the thermal properties of the streambed sediment (Hatch et al., 2006; Keery



**Fig. 10.** (A) A plot of the amplitude of the diurnal signal through time from sensors at Ninemile Creek at 0.15 and 0.25 m depth. (B) A plot of the phase angle of the diurnal signal from the same two sensors. (C) Flux calculated using the Hatch and Keery amplitude methods for the pair of sensors in (A) at a center-of-pair depth of 0.20 m.

et al., 2007). For example, the Hatch amplitude method loses sensitivity as  $A_r$  approaches one or zero, near which small changes in the amplitude ratio can lead to large changes in the calculated flux due to a very low  $dA_r/dq$ . Selection of the best arrangement of sensors in a TP is therefore a particular problem, because the flux rates may not be known prior to equipment installation, or the flux rates may change with time or depth. Therefore, for the most flexible analysis Hatch et al. (2006) recommend the deployment of many sensors, so as to have the greatest number of pairs with the widest range of  $\Delta z$  values possible. However, the large number of potential pairs (1596 in the case of Cherry Creek) leads to a very large number of calculations.

As described in Methods, VFLUX has several features that allow the user to determine an ideal sensor spacing range. For any time step where the amplitude ratio is 1 or 0 (the sensor spacing is too small or too large), the program will not calculate either of the amplitude methods, and a null value is written. Likewise, for any time step where the time lag is 0 or negative, the program will not calculate either of the phase methods, and a null value is written. The application of the Hatch amplitude method in the VFLUX code also includes an optional sensitivity check, which will write a null flux value if the derivative  $dA_r/dq$  is below a certain value ( $0.001 \text{ d m}^{-1}$ , by default), which indicates very poor sensitivity of the model due to non-ideal sensor spacing. The selection of an ideal window size for a specific application therefore involves achieving a sufficient percentage of non-null flux values while still maintaining a small-enough window to supply sufficient vertical resolution in space.

VFLUX displays bar charts of the percent of non-null flux calculations for different sensor pairs and sensor spacings (Fig. 11). For each window size, a separate bar chart shows the percent of successful calculations made at each center-of-pair depth, and the number of bars and their spacing indicates the resolution in the vertical dimension (i.e., spatial resolution is higher with smaller window sizes and a greater number of closely spaced depths). In relatively high-flux environments, such as the examples from Cherry and Ninemile Creeks, the ideal sensor spacing is larger than for low-flux environments; therefore, at small windows sizes, a large number of flux calculations are outside the model sensitivity range. However, because vertical flux changes with depth at both sites, we want high spatial resolution to resolve these changes, and larger window sizes may obscure or blur transitions in flux with depth.

At Ninemile Creek, amplitude ratios were generally high (ranging from 0.716 to 1.01 for all pairs and time steps), indicating that water flux was strong and in the downward direction. Because there are only 21 possible pairs in a TP of seven sensors, every possible window was used in VFLUX, i.e., 1, 2, 3, 4, 5, and 6 sensor spacings, which correspond to  $\Delta z$  values of 0.05, 0.10, 0.15, 0.20, 0.25 and 0.30 m. At a  $\Delta z$  of 0.05 m, 17% of time steps had an  $A_r \geq 1$  and the mean  $A_r$  was 0.965; at a  $\Delta z$  of 0.10 m, 9% of time steps had a  $A_r \geq 1$  and the mean  $A_r$  was 0.944; by a  $\Delta z$  of 0.20 m, there were no time steps at which flux could not be calculated and the mean  $A_r$  was 0.903. The bar charts in Fig. 11 break down these statistics for each window by depth. With a window of 1 sensor spacing ( $\Delta z$  of 0.05 m), Fig. 11 shows that fewer than 35% of the flux calculations were within the model sensitivity range at the shallowest center-of-pair depth (0.025 m), where vertical flux was highest. A substantial percentage of calculations were also null at 0.075 m depth. With a window of 2 ( $\Delta z$  of 0.10 m), the shallowest depth of 0.05 m has an approximately 60% success rate, and with a window of 3 ( $\Delta z$  of 0.15 m), the shallowest depth is now up to 90%. We made the subjective decision that a  $\Delta z$  of 0.10 m best balanced spatial resolution and model sensitivity, because more than half of flux values were calculated within the model sensitivity range at even the shallowest depth (with the highest

vertical flux rate), and five different center-of-pair depths gave enough vertical resolution to describe changing flux patterns with depth.

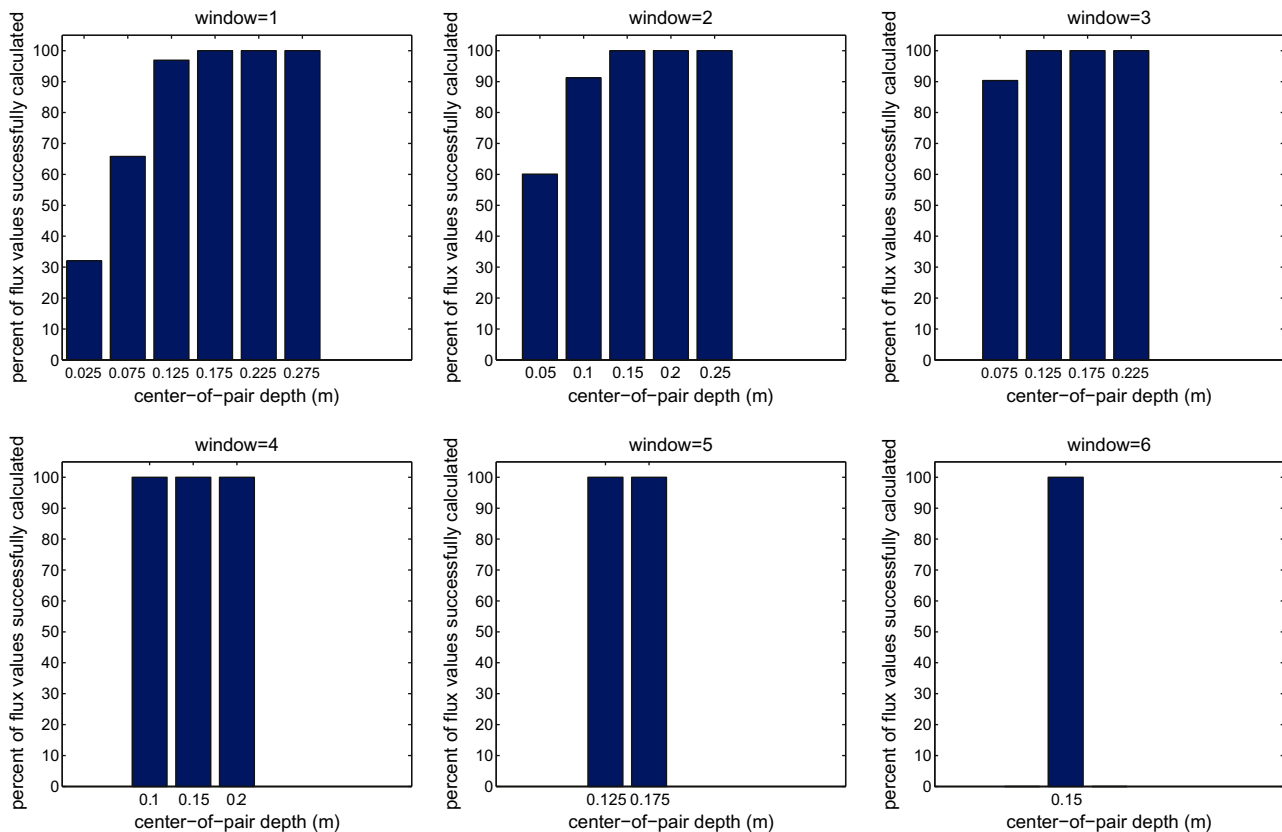
As stated in Methods, the Cherry Creek DTS data were processed with windows of 8, 9, 10, 11, and 12 sensor spacings, which correspond to  $\Delta z$  values of 0.110, 0.124, 0.138, 0.152, and 0.166 m. These spacings were chosen because the amplitude ratios observed for this rod were generally high and lag times were short, meaning that the water flux at the TP location was strongly downwelling, and large sensor spacings were required to detect a change in amplitude. The pairs separated by the smaller  $\Delta z$  values of 0.110 and 0.124 m had many time steps at which flux could not be computed because the amplitude ratio was close to or greater than one. The values above one are due to slight errors in measurement or digital filtering. At  $\Delta z$  of 0.110 m, 21% of time steps for pairs with a center-of-pair depth less than 0.20 m had an  $A_r \geq 1$ , and the mean  $A_r$  among these time steps was 0.963; at a  $\Delta z$  of 0.138 m, 18% of time steps had an  $A_r \geq 1$  and the mean  $A_r$  was 0.954; and at a  $\Delta z$  of 0.166 m, only 12% of time steps shallower than 0.20 m had an  $A_r \geq 1$  and the mean  $A_r$  was 0.943. We made the subjective decision that a  $\Delta z$  of 0.138 m gave the best balance between vertical spatial resolution and model sensitivity for the entire TP, although it would also be possible to use a combination of multiple  $\Delta z$  values at different times and/or depths to maximize the sensitivity of the results in both space and flux rate (see Briggs et al., accepted for publication).

These examples show how the deployment of many sensors in a TP, coupled with the computing flexibility of a program like VFLUX, can help identify the best sensor spacing for a particular experiment and field site. In the cases of Cherry Creek and Ninemile Creek, high amplitude ratios due to high downward flux rates made the choice of sensor spacing even more important.

### 3.6. Flux results and a comparison of methods

The results of flux calculations at Ninemile and Cherry Creeks show that vertical flux rates at both sites are indeed strongly downwelling (approximately  $5 \times 10^{-5} \text{ m s}^{-1}$  and  $2 \times 10^{-5} \text{ m s}^{-1}$ , respectively, at the shallowest depths). A box plot of 2-h flux estimates over the entire collection period (excluding the first and last 48 h) versus the center-of-pair depth from Ninemile Creek is shown in Fig. 12A. A similar plot from Cherry Creek is shown in Fig. 12B. These two figures only include flux estimates calculated with the Hatch amplitude ratio method (Eq. (6)). As can be seen in the box plots, the distribution of flux estimates for each sensor pair is not normally distributed but has a positive skew with large positive outliers. The outliers are a result of decreasing sensitivity of the amplitude ratio model as the amplitude ratios become very close to 1, and could have been removed by employing the Hatch amplitude method derivative check (see the discussion above). In Fig. 12, we have chosen to keep all outliers in order to illustrate the range of potential results, and we therefore use the median to represent the central tendency of a population of flux estimates.

At both sites, the vertical flux rates at the shallowest depths are the largest (Fig. 12), with a median rate over time of  $5.1 \times 10^{-5}$  ( $+1.4 \times 10^{-5}$  or  $-7.2 \times 10^{-6} \text{ m s}^{-1}$ ) at 0.025 m depth at Ninemile Creek and  $1.7 \times 10^{-5}$  ( $+5.5 \times 10^{-6}$  or  $-1.9 \times 10^{-6} \text{ m s}^{-1}$ ) at 0.097 m depth at Cherry Creek (the plus-or-minus errors span the interquartile range). As the depth increases away from the streambed interface, the flux rate decreases towards zero. We interpret these spatial patterns to be due to relatively shallow, curved hyporheic flow paths that are not purely vertical but have a horizontal component that increases with increasing depth (also see Briggs et al., accepted for publication). The profile at Ninemile Creek was installed at the head of a riffle in a pool-riffle sequence, where we would expect the initiation of a shallow, curved



**Fig. 11.** An example of the sensor spacing diagnostic charts produced by the visualization routine of VFLUX. This set of charts was produced using the Hatch amplitude method results from the Ninemile Creek data. Each bar chart for a specific window size (sensor spacing) shows the percent of flux values through time that were within the sensitivity range of the model at each center-of-pair depth.

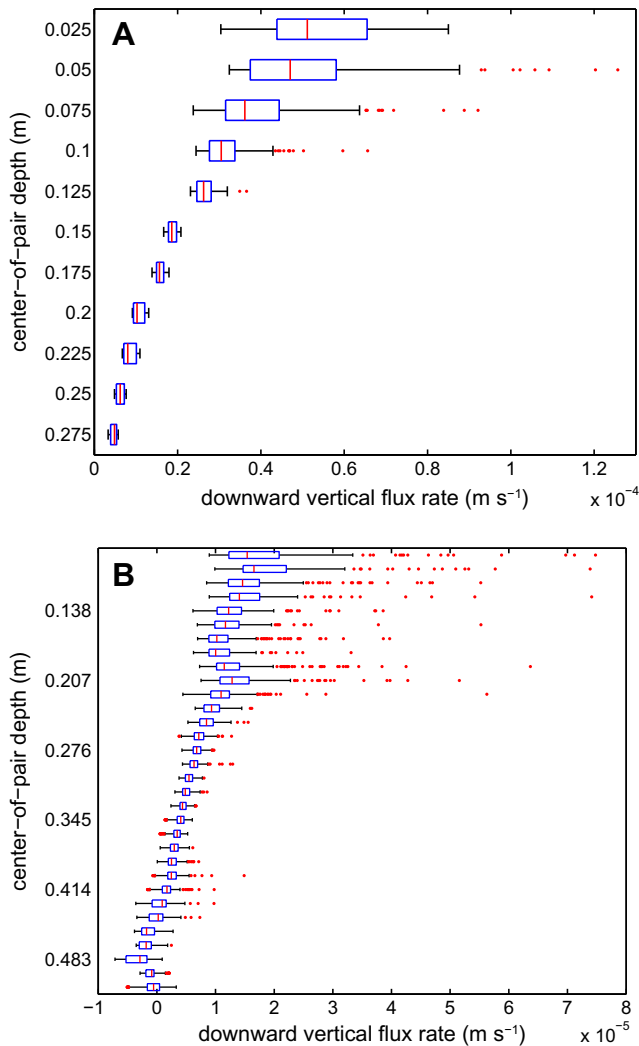
hyporheic flow cell. The profile at Cherry Creek was installed just upstream of a beaver dam, which creates a step in hydraulic head and likely initiates similar shallow hyporheic flow that becomes horizontal with depth. The flux rates near the streambed interface are therefore likely to be the best estimate of exchange between the hyporheic zone and the stream itself. Compared to previous methods in which the sensor spacing increased with increasing depth (Vogt et al., 2010), the use of a sliding window in the VFLUX method makes it possible to identify with greater precision changes in flux with depth, because it is easy to use the smallest sensor spacing—with the greatest spatial precision—that is still sensitive to the heat-transport model in use (Briggs et al., accepted for publication).

Fig. 13 shows calculated vertical flux rates through time for a single pair of sensors at 0.15 and 0.25 m depth (center-of-pair depth is 0.20 m) at Ninemile Creek, calculated with the Hatch amplitude method. Also plotted are two sets of flux rates that were calculated on a daily basis by selecting the maximum and minimum temperature in each 24-h period (similar to the methods of Fanelli and Lautz, 2008; Hatch et al., 2006; Lautz, 2010). The daily flux calculations were made on both the raw, unfiltered time series and on the diurnal signal after filtering it with DHR. Although the true flux values in the field are not known independently from our heat transport estimates, it is still useful to compare the estimates made from both the sample-by-sample and daily methods using both filtered and unfiltered thermal data, since all of these techniques have been used in the published literature (e.g. Fanelli and Lautz, 2008; Hatch et al., 2006; Keery et al., 2007). The daily flux calculations for the filtered time series agree fairly well with the results from VFLUX, except in the last 5 days, where they are slightly lower than the flux rates from VFLUX. The median flux

value over the whole collection period is not significantly different between the two methods (Mann–Whitney rank test,  $p = 0.718$ ). However, the daily flux calculations made with the unfiltered time series do not agree well with the results from VFLUX in Fig. 13. The unfiltered flux estimates are more variable in time than both of the filtered flux estimates, and the median flux value is significantly different between the two methods at the 5-percent significance level ( $p = 0.029$ ).

### 3.7. Sensitivity and uncertainty analysis for thermal properties

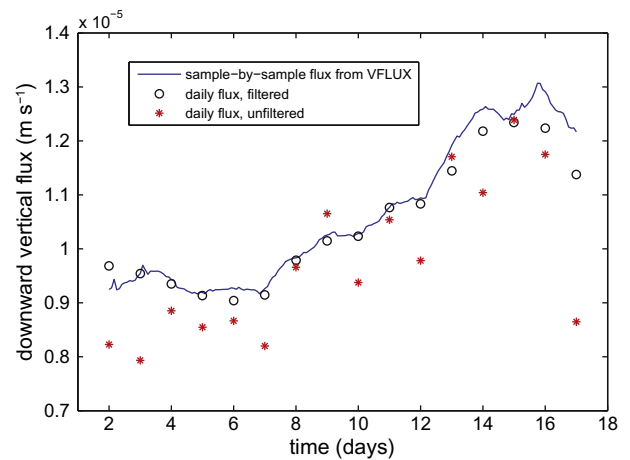
We employed the VFLUX programs “vfluxsens” and “vfluxmc” to conduct basic sensitivity analyses on sediment and thermal properties and to estimate confidence intervals for flux estimates through time for the data from Ninemile Creek. The results of the sensitivity analysis for a  $\Delta z$  of 0.1 m at a center-of-pair depth of 0.2 m are presented in Table 2. At this magnitude of downward flux ( $1.03 \times 10^{-5} \text{ m s}^{-1}$  calculated with the base parameters) estimated flux values are most sensitive to thermal conductivity, closely followed by porosity, which determines the total heat capacity of the saturated matrix. Flux is least sensitive to the heat capacity of sediment and water, because these values vary little in nature (Lapham, 1989). Although only results from one pair of sensors are presented, these general patterns hold true for all of the flux results from Ninemile Creek. As the magnitude of downward flux increases, flux becomes more sensitive to porosity and thermal conductivity. A similar sensitivity analysis was performed for Cherry Creek (not shown), where flux magnitudes were smaller than at Ninemile. At flux rates above approximately  $5 \times 10^{-6} \text{ m s}^{-1}$ , the Cherry Creek analysis shows a similar pattern to the Ninemile analysis, with increasing model sensitivity to porosity and thermal



**Fig. 12.** Box plots of 2-h flux estimates over the entire collection period (excluding the first and last 48 h), showing the decrease in vertical flux with depth at both study locations. The depth value of each box is the center-of-pair depth. The boxes represent the interquartile range and the vertical lines the median value. The whiskers extend to the most extreme data value that is beyond the boxes by less than 1.5 times the interquartile range, and any outliers beyond the whiskers are plotted individually as dots. (A) At Ninemile Creek, only showing sensor pairs where  $\Delta z$  is 0.05 or 0.10 m. (B) At Cherry Creek, only showing sensor pairs with a  $\Delta z$  of 0.138 m and with a center-of-pair depth of 0.51 m or less.

conductivity with increasing downward flux. However, at flux rates below  $5 \times 10^{-6} \text{ m s}^{-1}$ , the model is more sensitive to thermal conductivity with decreasing flux rates, suggesting that model sensitivity to conductivity is at a minimum near  $5 \times 10^{-6} \text{ m s}^{-1}$  (for the specific conditions at Cherry Creek). At Ninemile Creek, flux magnitudes were not small enough to observe this effect.

One question that arises when comparing different flux estimates from one site to another, or even from one depth to another, is whether differences in flux estimates are due to truly different magnitudes of water flux, or simply to different sediment thermal properties. For example, in order to say with greater certainty that vertical flux does indeed decrease with depth at Ninemile Creek (Fig. 12A), we can test whether the apparent differences in flux could be due to changes in sediment texture alone. Select results from the Monte Carlo analysis of data from Ninemile Creek are shown in Fig. 14. Flux estimates are plotted through time for two sensor pairs with center-of-pair depths at 0.15 and 0.20 m ( $\Delta z$  of 0.10 m), with upper and lower uncertainty limits ( $\pm 2$  standard

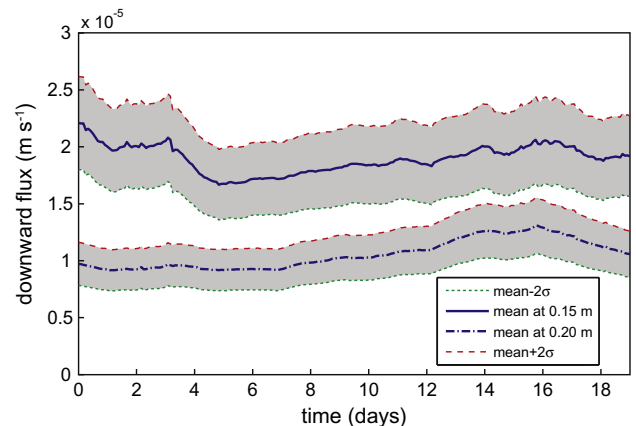


**Fig. 13.** Vertical flux rates through time calculated for the pair of sensors at 0.15 and 0.25 m depth at Ninemile Creek. Flux rates are calculated with the VFLUX program, and also using the daily maxima and minima of the raw temperature time series and the time series after filtering with DHR.

**Table 2**

Results of the sensitivity analysis for Ninemile Creek data ( $\Delta z$  of 0.1 m, center-of-pair depth of 0.2 m).

Property varied	Median flux value through time ( $\text{m s}^{-1}$ )	
	With low parameter value	With high parameter value
(None, using base values from Table 1)	$1.03 \times 10^{-5}$	
Total porosity	$1.01 \times 10^{-5}$	$1.11 \times 10^{-5}$
Baseline thermal conductivity	$0.86 \times 10^{-5}$	$1.05 \times 10^{-5}$
Thermal dispersivity	$1.02 \times 10^{-5}$	$1.07 \times 10^{-5}$
Volumetric heat capacity of sediment	$1.02 \times 10^{-5}$	$1.04 \times 10^{-5}$
Volumetric heat capacity of water	$1.034 \times 10^{-5}$	$1.018 \times 10^{-5}$



**Fig. 14.** Results from the Monte Carlo analysis of flux uncertainty for 2 sensor pairs with center-of-pair depths at 0.15 and 0.20 m ( $\Delta z$  of 0.10 m). The upper and lower uncertainty limits represent the mean flux estimate  $\pm 2$  standard deviations.

deviations). The plot shows that uncertainty is greater at greater downward flux rates (at the shallower depth), but there is no overlap between the confidence intervals from the two depths. Results therefore suggest that the different flux rates measured at 0.15 and 0.20 m are statistically different. Error or uncertainty analyses such as the one presented here are an important component of physical

flux estimation, and confidence intervals should generally be given along with reported flux values. The VFLUX software package and the Monte Carlo analysis program demonstrated here make the calculation of confidence intervals automatic and straightforward.

#### 4. Conclusion

Time series of streambed temperatures have been successfully used by many researchers to estimate vertical water flux rates in shallow streambeds (e.g. Fanelli and Lautz, 2008; Hatch et al., 2006; Keery et al., 2007; Vogt et al., 2010), and these studies have each made advances in heat-transport modeling, signal processing, and experimental deployment. We have introduced a step-by-step workflow for analyzing temperature time series and calculating vertical water flux, which incorporates the advancements made in previous work and adds automation and new techniques for processing large amounts of data from high-resolution sensor profiles. The method has been written into a computer program called VFLUX, which automates the entire process of calculating vertical flux rates from raw temperature time series. VFLUX is the first published computer program for calculating water flux from temperature time series to completely automate data pre-processing, including robust filtering with Dynamic Harmonic Regression (DHR), and the first method to use a sliding window to identify changing flux rates with depth at high spatial resolution. VFLUX also includes functions for data and result visualization, new tools for evaluating model sensitivity and ideal sensor spacing, and functions for sensitivity analysis and confidence interval calculation.

We have shown using two field examples that each step of the method is important for calculating accurate flux rates from real-world temperature data, which are typically characterized by several irregular signals that vary in time with different amounts of noise. The automated method processes time series consistently for every sensor in a profile, and it can calculate flux between every possible sensor pair separated by every possible sensor spacing, which is particularly useful when the ideal sensor spacing for a specific field site is unknown. An automated program such as VFLUX is of great assistance when deploying a large number of sensors, as in a high-resolution DTS profile, due to the very large number of potential calculations. We have shown how VFLUX can help the practitioner quantitatively evaluate the ideal spacing between sensor pairs, and how the program can be used in situations where many repeated calculations are desired, such as studies of model sensitivity to thermal parameters and Monte Carlo error estimation.

We have also shown how this method, which provides high spatial and temporal resolution, is both more accurate and reliable than previous methods at capturing changing flux rates through time and changing vertical flux with depth. Our analysis of field data from Ninemile Creek, NY, and Cherry Creek, WY demonstrate how the VFLUX method can be used in situations where shallow hyporheic flow paths change direction from vertical near the surface to horizontal with depth, such as in shallow, curved hyporheic flow paths. For example, high-resolution DTS data from Cherry Creek processed with this method show vertical flux decreasing from approximately  $1.7 \times 10^{-5} \text{ m s}^{-1}$  very near the sediment–water interface to near zero by approximately 0.5 m depth. A Monte Carlo uncertainty analysis at Ninemile Creek shows that the observed decrease in flux with depth cannot be due to changing sediment thermal properties alone. Unlike other methods used in the literature, VFLUX also supplies flux results at high temporal resolution, and therefore could also be used to investigate flux rates that change on hourly time scales in estuary environments or dam-controlled water bodies. We believe that the release of this method and the availability of the VFLUX program will make heat

transport modeling a more approachable and widely-used technique for calculating vertical water flux rates using temperature.

#### Acknowledgements

Field data were collected in collaboration with Timothy Daniluk and Danielle Hare. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0750965, and is partially based upon work supported by the National Science Foundation under Grant Nos. EAR-0911612 and EAR-0901480 and the Canadian Foundation for Innovation.

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhydrol.2011.11.053.

#### References

- Anderson, M.P., 2005. Heat as a ground water tracer. *Ground Water* 43 (6), 951–968.
- Anderson, W.P., Anderson, J.L., Thaxton, C.S., Babyak, C.M., 2010. Changes in stream temperatures in response to restoration of groundwater discharge and solar heating in a culverted, urban stream. *J. Hydrol.* 393 (3–4), 309–320.
- Boas, M.L., 1983. *Mathematical Methods in the Physical Sciences*. Wiley, New York, 793 pp.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, NJ, 598 pp.
- Bredehoeft, J.D., Papadopoulos, I.S., 1965. Rates of vertical groundwater movement estimated from the Earth's thermal profile. *Water Resour. Res.* 1 (2), 325–328.
- Briggs, M.A., Lautz, L.K., Gordon, R.P., McKenzie, J.M., Hare, D.K., accepted for publication. High resolution fiber-optic distributed temperature sensing of hyporheic flux patterns in varied space and time around beaver dams. *Water Resour. Res.*
- Buffington, J.M., Tonina, D., 2009. Hyporheic exchange in mountain rivers II: effects of channel morphology on mechanics, scales, and rates of exchange. *Geogr. Compass* 3 (3), 1038–1062.
- Caissie, D., 2006. The thermal regime of rivers: a review. *Freshw. Biol.* 51 (8), 1389–1406.
- Chemical Rubber Company (CRC), 2011. *CRC Handbook of Chemistry and Physics* (Online). CRC Press, Boca Raton, FL.
- Constantz, J., 2008. Heat as a tracer to determine streambed water exchanges. *Water Resour. Res.* 44, W00D10.
- Crispell, J.K., Endreny, T.A., 2009. Hyporheic exchange flow around constructed in-channel structures and implications for restoration design. *Hydrol. Process.* 23 (8), 1158–1168.
- Dakin, J.P., Pratt, D.J., Bibby, G.W., Ross, J.N., 1985. Distributed optical fiber Raman temperature sensor using a semiconductor light-source and detector. *Electron. Lett.* 21 (13), 569–570.
- Engelhardt, I. et al., 2011. Comparison of tracer methods to quantify hydrodynamic exchange within the hyporheic zone. *J. Hydrol.* 400 (1–2), 255–266.
- 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.
- Ferguson, G., Bense, V., 2011. Uncertainty in 1D heat-flow analysis to estimate groundwater discharge to a stream. *Ground Water* 49 (3), 336–347.
- Fetter, C.W., 2001. *Applied Hydrogeology*. Prentice Hall, Upper Saddle River, NJ, p. xvii, 598.
- Goto, S., Yamano, M., Kinoshita, M., 2005. Thermal response of sediment with vertical fluid flow to periodic temperature variation at the surface. *J. Geophys. Res.-Solid Earth* 110 (B1), B01106.
- 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 Resour. Res.* 42 (10), W10410.
- Healy, R.W., Ronan, A.D., 1996. Documentation of computer program VS2DH for simulation of energy transport in variably saturated porous media—Modification of the US Geological Survey's computer program VS2DT, US Geological Survey Water-Resources Investigations Report 96-4230, pp. 36.
- Jensen, J.K., Engesgaard, P., 2011. Nonuniform groundwater discharge across a streambed: heat as a tracer. *Vadose Zone J.* 10 (1), 98–109.
- 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. *J. Hydrol.* 336 (1–2), 1–16.
- Lapham, W.W., 1989. Use of temperature profiles beneath streams to determine rates of vertical ground-water flow and vertical hydraulic conductivity. *US Geol. Surv. Water Supply Pap.*, 2337.
- Lautz, L.K., 2010. Impacts of nonideal field conditions on vertical water velocity estimates from streambed temperature time series. *Water Resour. Res.* 46, W01509.

- Lautz, L.K., Kranes, N.T., Siegel, D.I., 2010. Heat tracing of heterogeneous hyporheic exchange adjacent to in-stream geomorphic features. *Hydrol. Process.* 24 (21), 3074–3086.
- Lautz, L.K., Ribaud, R.E., submitted for publication. Scaling up point-in-space heat tracing of seepage flux using bed temperatures as a quantitative proxy. *Hydrogeol. J.*
- Maassen, S., Balla, D., 2010. Impact of hydrodynamics (ex- and infiltration) on the microbially controlled phosphorus mobility in running water sediments of a cultivated northeast German wetland. *Ecol. Eng.* 36 (9), 1146–1155.
- Niswonger, R., Prudic, D.E., 2003. Modeling heat as a tracer to estimate streambed seepage and hydraulic conductivity. In: Constantz, J., Stonestrom, D.A. (Eds.), *Heat as a Tool for Studying the Movement of Ground Water Near Streams*, US Geological Survey Circular 1260, pp. 81–89.
- 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. *Hydrogeol. J.* 18 (5), 1093–1110.
- Ronan, A.D., Prudic, D.E., Thodal, C.E., Constantz, J., 1998. Field study and simulation of diurnal temperature effects on infiltration and variably saturated flow beneath an ephemeral stream. *Water Resour. Res.* 34 (9), 2137–2153.
- Schmidt, C., Conant, B., Bayer-Raich, M., Schirmer, M., 2007. Evaluation and field-scale application of an analytical method to quantify groundwater discharge using mapped streambed temperatures. *J. Hydrol.* 347 (3–4), 292–307.
- Schmidt, C., Martienssen, M., Kalbus, E., 2011. Influence of water flux and redox conditions on chlorobenzene concentrations in a contaminated streambed. *Hydrol. Process.* 25 (2), 234–245.
- Schornberg, C., Schmidt, C., Kalbus, E., Fleckenstein, J.H., 2010. Simulating the effects of geologic heterogeneity and transient boundary conditions on streambed temperatures – implications for temperature-based water flux calculations. *Adv. Water Resour.* 33 (11), 1309–1319.
- Selker, J.S. et al., 2006. Distributed fiber-optic temperature sensing for hydrologic systems. *Water Resour. Res.* 42 (12), W12202.
- Shanfield, M., Hatch, C., Pohl, G., 2011. Uncertainty in thermal time series analysis estimates of streambed water flux. *Water Resour. Res.* 47.
- Stallman, R.W., 1965. Steady one-dimensional fluid flow in a semi-infinite porous medium with sinusoidal surface temperature. *J. Geophys. Res.* 70 (12), 2821–2827.
- Stonestrom, D.A., Constantz, J., 2003. Heat as a tool for studying the movement of ground water near streams. *US Geol. Surv. Circ.* 1260, 96.
- Swanson, T.E., Cardenas, M.B., 2010. Diel heat transport within the hyporheic zone of a pool-riffle-pool sequence of a losing stream and evaluation of models for fluid flux estimation using heat. *Limnol. Oceanogr.* 55 (4), 1741–1754.
- 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. *Comput. Geosci.* 37 (10), 1664–1669.
- Vogel, T., Brezina, J., Dohnal, M., Dusek, J., 2010. Physical and numerical coupling in dual-continuum modeling of preferential flow. *Vadose Zone J.* 9 (2), 260–267.
- Vogel, T., Dohnal, M., Votrubova, J., 2011. Modeling heat fluxes in macroporous soil under sparse young forest of temperate humid climate. *J. Hydrol.* 402 (3–4), 367–376.
- 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.
- Young, P.C., Pegregal, 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, UK. Internet. <<http://www.es.lancs.ac.uk/cres/captain>>.