We would like to thank both Dr. Cirpka and the anonymous second reviewer for their useful comments, which we respond to below. We have carefully examined and incorporated their concerns, which has substantially improved this manuscript.

Response to comments by Dr. Olaf A. Cirpka:

General analytic solution:

> *Response*: In this comment, an alternative analytical solution is derived by Dr. Cirpka. Indeed there are many analytical solutions to the numerical convection-conduction equation, and each incorporates a different set of assumptions. In this study, we have chosen two specific analytical solutions- the solutions by Bredehoeft and Papadapulos (1965) and the two recent, similar adaptations of the Stallman solution (1965). These solutions have been increasingly used in recent heat as a tracer literature in a variety of conditions. Although these methods are commonly employed to analyze field data, their relevancy to the specific field conditions is often not discussed, and sometimes they are used even when the field data violates several of the basic assumptions (for example, Lautz and Ribaudo, 2012). Therefore, we believe it is a timely, relevant, and interesting topic for the HESS readership to consider the limitations of these two analytical solutions, and quantify how much error is expected by the use of them under these conditions.

Response: In light of this comment, we have added a sentence to the end of the introduction further stressing the relevance of the particular analytical solutions we have chosen, "Although these are not the only, or perhaps even the best analytical solutions for these conditions, they are the most commonly used, and therefore their performance is of interest as an example of the potential limitations of one-dimensional, analytical solutions for using heat as a tracer of streambed flux."

Main Comments:

1. The authors consider cases in which steady-state and diurnal contributions are overlain. It does not make sense to use analytical solutions for either only steadystate or only sinusoidal boundary conditions in such cases. This becomes evident in the numerical example. The reported miss-fits indicate either systematic bias (in case of the spectral approach of HK) or temporal fluctuations (in case of the steady-state approach of BP). Of course, one has to separate the time-fluctuating and steady-state contributions when using either of the two approaches. In the field example, the authors implicitly remove

the steady-state contribution when applying the amplitude-dampening approach. However, they don't try to make use of the combined, existing analytical solution.

Response: Although it is advocated later in this review that the numerical example should be omitted, this comment shows that it is indeed very clear and illustrates the intended point. These are the two analytical solutions that are commonly used in the literature and are therefore of most relevance in the field for this example. This point is now clearly stated in the introduction, and the existence of other solutions is also discussed.

2. The authors make wrong statements about underlying assumptions in the analytical expressions. A prominent example is that there would be a difference whether conductive flux is predominantly upwards or downwards. That is simply wrong. It is important whether convection is upwards or downwards; and it is important whether a mean temperature gradient exists. But the analytical expressions don't mind whether groundwater or stream water is cooler in average.

Response: We don't state anywhere in the manuscript that the analytical expression "mind" whether groundwater or stream water is cooler. We *do* state that the assumption of no vertical temperature gradient in the Stallman solution will be violated if the groundwater temperature is warmer than the surface water. This is unavoidable. This gradient is somewhat removed by filtering the data; only the desired frequencies remain in the filtered data, and therefore we also examine whether the filtered data performs better under these conditions.

2. The key problem in the field application is temporal variability of flow. The head differences fluctuate on time scales of days. The estimated travel time of temperature is also in the range of one day. Thus, any analysis assuming quasi steadystate flow must be flawed. This is the true difficulty of the appliction, but the problem is not posed that way. I highly recommend that the authors change their virtual toy problem to something similar: A test case in which flow is transient, in which "data" are attempted to be fitted with a solution based on steady-state flow.

Response: Field data that were relatively constant were chosen for this study. There was a variation in flow of less than five percent over the study period, which is barely "transient" and was unlikely to be a major source of error. Additional sources of error, including specific measured parameters and literature values, is discussed in detail.

3. A core difficulty in using either only the steady-state contribution, or only the amplitude dampening, or only the phase shift is that the temperature propagation behavior

depends on two coefficients: vf and Ke. The effective thermal diffusivity Ke depends on hydromechanic dispersion, porosity, and the mineral composition of the streambed. These properties are not too well constrained. In particular the method based on steady-state heat transport yields exclusively the thermal Peclet number vf z=Ke rather than vf (and thus qz) itself. A combined analysis of all three components yields, at least in theory, a handle on both vf and Ke.

Response: This is correct, and one reason why we have presented the theoretical example. In the theoretical example, we know the exact values of both v and Ke, therefore we can best illustrate errors in the estimation of v that stem from model assumptions that are violated.

Specific Comments

1. throughout the text: Replace "diel" with "diurnal".

Response: Diel has been left in the text, because we are referring to a pattern that is on a 24-hour cycle. Diel: Occurring on a 24-hour cycle, as opposed to diurnal (day) or nocturnal (night) occurrences. Source: McGraw-Hill Dictionary of Scientific and Technical Terms, 6th edition (2002), published by The McGraw-Hill Companies, Inc.

2. throughout the text: There is no "flat" temperature, there is a constant one.

Response: “Flat” has been changed to “constant” throughout the text as suggested.

3. throughout the text: Use consistent notation including consistent choices when a velocity is positive or negative.

Response: The sign of the velocity in the Bredehoeft and Papadapulos solution has been reversed for consistency. A positive velocity now indicates upward flux throughout the text.

4. page 4306, lines 3-4: I doubt that there are so few studies in which temperature is used as a natural tracer under upwelling conditions (Conant, 2004; Schmidt et al., 2006, 2007).

Response: This is true, there are several studies in which temperature is used as a tracer under upwelling conditions, and your suggestions are now cited in the text. There is now further discussion of alternative methods, which nicely incorporates Conant (2004). However, for all of the cited studies (Conant, 2004; Schmidt et al., 2006, 2007), the surface water was warmer than the groundwater and therefore conduction was in the downward direction despite convection in the upward direction. Our study focuses on the particular case of upwards convection and conduction.

4. page 4306, line 9: add "mean" before "temperature gradient"

Response: “Mean” added as suggested.

5. page 4306, line 15: An estimation within one order of magnitude is not good at all.

Response: “Within one order of magnitude” has been replaced by the actual values of error.

7. page 4306, lines 17-18: I don’t think that that statement that one should consider the physical processes to be measured before designing a measurement set-up belongs into the abstract of any scientific paper. This is self speaking for any scientist.

Response: This sentence has been removed.

8. page 4306, line 21: "using" rather than "both"

Response: This sentence has been modified to read, “Heat as a tracer has been used in many numerical and analytical studies of streambed water flux…”

9. page 4307, line 21: "small inverse gradients", what do the authors mean? Small absolute values of the gradient, which would make sense to me, or large gradients? The same phrase appears later on again, and remains confusing.

Response: In both cases this phrase has been changed to “for inverse temperature gradients near the surface.”

10. page 4307, line 22: Whether groundwater temperatures are higher or lower than stream temperatures is absolutely irrelevant for the analysis. The question at hand is whether there is a mean gradient at all, positive or negative.

Response: We disagree. In the case where the conduction is downward, we would expect the diel temperature signal to be moving into the subsurface driven by this force, and how far that temperature signal penetrates is dependent on the ratio of conduction to convection (ie the thermal Peclet number). In the case where both conduction and convection are upward, there is no process forcing the diel temperature signal into the subsurface. Therefore when a diel signal is observed at shallow depths under these circumstances, it is important to investigate how various models will match that signal and what error is incurred by these conditions.

11. page 4307, line 23: replace "less" by "smaller"

Response: Replaced as suggested.

12. page 4307, line 26: add "uniform" before "vertical infiltration velocity"

Response: Added as suggested.

13. page 4308, line 1: "no mean temperature gradient" rather then just "temperature Gradient

Response: “Mean” added as suggested.

14. page 4308, line 4: "larger" rather than "greater"

Response: Replaced as suggested.

15. page 4308, line 7: Again, whether mean conduction is upwards or downwards is absolutely irrelevant.

Response: see response to comment 10.

16. Sections 2.2 & 2.3: I highly recommend presenting the combined solution right away. Also flipping the sign of vz does not contribute to clarity. There is no need to stick exactly to the nomenclature of the cited articles.

Response: vz has been made positive in the equation as suggested.

17. page 4310, line 4: I have no idea what "anisothermal flow" means. Do you mean "temperature independent flow"?

Response: A thermodynamic process in which the temperature of the system remains constant and thermal equilibrium is maintained is termed an isothermal process. Anisothermal means that it does not have a single, fixed temperature. This is one of the assumptions specifically described in Bredehoeft and Papadapulos (1965) and in consequent studies that employ the Bredehoeft and Papadapulos solution; therefore, we believe it is appropriate here.

> 18. page 4310, line 10: Another unclear small inverse gradient

Response: this phrase has been changed to, ”for inverse temperature gradients near the surface.”

> 19. page 4310, line 18: "When dispersivity is neglected": Sorry, the quoted analytical solution does not neglect dispersion/conduction at all.

Response: The Keery et al. (2007) *formulation* of this solution does neglect dispersivity (page 6 of that paper in the text following equation 12) and discusses exactly why they do neglect the term (to make the equation computationally easier). This is indeed the main difference between the Keery et al. (2007) and Hatch et al. (2006) formulations. To make this clearer, this sentence has been changed to, “When dispersivity is neglected, the Hatch et al. (2006) and Keery et al. (2006) formulations for solving the Stallman equation yield identical results.”

> 20. page 4310, following Eq. (4): Ke and vf should be introduced earlier on, where the convection-conduction equation is discussed for the first time.

Response: Ke and vf are not in equation 4 as formulated. Because qz is used across all equations, we believe it is less confusing to introduce qz in equation 4 (instead of vf, which is only used in the HK solution). Similarly, the specific heat and density are used in the Bredehoeft solution, and Ke itself is computed from the thermal conductivity divided by the product of the specific heat and density.

> 21. page 4311, lines 4-8: The authors did not understand that the steady-state contribution and the sinusoidal contribution are additive. This results from the underlying pde to be linear and the orthogonality of different frequencies in Fourier analysis.

Response: In this sentence we are describing a relevant assumption of the Stallman model, which is that there is no temperature gradient with depth. On p.2822 of Stallman (1965), below equation (3), this assumption is stated: “Furthermore, it is assumed tacitly that temperature fluctuations at all depths will be in equilibrium with the sinusoidal fluctuations of constant amplitude generated at the land surface.” The examination of error caused by violation of this assumption is also a main focus of Lautz (2010). This is important to iterate because there *is* a temperature gradient in the field data- unless the groundwater and surface temperatures are the same, there will always be a temperature gradient if conduction and convection are both upwards- we are examining the error caused by the violation of this assumption.

> 22. Section 2.4: I don't understand the purpose of the virtual test case at all. Steady state is steady state, and time periodic is time periodic. You must not confuse those two. This is so obvious that there is no need to perform a numerical test.

Response: The theoretical test case allows for the comparison of the models under specific conditions, with known input values of Ke and v, and therefore illustrates differences caused by specific, stated assumptions in the analytical models.

> 23. Page 4314, lines 17-18: This statement is utterly wrong. Eq. (3) does not depend on the sign of the temperature gradient. It depends on steady-state flow and uniform coefficients.

Response: We are not saying that Eq. (3) depends on the temperature gradient, only that the heat and fluid flow must be in the same direction for the equation to be valid. This is stated on page 4310, lines 5- 7 and again on page 4314, lines 17-18. Perhaps this was unclear in the second instance, therefore these lines on page 4314 have been clarified to read: "In this case, fluid flow between the surface and 0.040 m was upward while heat conduction was downward, and the assumption that heat and fluid flow are in the same direction was violated."

> 24. Section 3.2.1: You got me lost here at some point. Please make first clear what you are doing. Obviously, a time-varying heat-flux boundary condition is fitted to the data. However, there are head measurements in the stream and at depth. Thus, rather than relying on those heads, which should determine the volumetric flux, the specific discharge is tweeked in such a way that some temperature breakthrough curves fit. That implies inconsistencies with the head measurements. I am not convinced that this "calibrated" model is the right reference.

Response: This is the conventional method of using a numerical model of heat and water transport to determine streambed (water) flux (ie Niswonger and Prudic, 2003; Constantz, 2008; Anderson, 2005). The head measurements provide the water transport boundary conditions, the surface and groundwater temperatures provide the heat transport boundary conditions, and then q is solved for to match the temperature traces at intermediate depths. This process is quite clearly explained in the text.

25. Section 3.2.2: Taking the solution relying on steady-state flow worked OK when considering some data tripples, but not with all of them. Again, this is inconsistent. The analytical solution is for the entire profile and should thus be fitted to all data points. If that leads to bad results, the expression is not valid. However, you could still work with the combined steady-state/sinusoidal expression (and maybe fit Ke while you are calibrating the model).
Let's assume that the good agreement with the steady-state solution is real: Does this imply quasi steady-state behavior? If the time scales of heat transfer are smaller than the time scales of the velocity fluctuations, the latter would not be particularly surprising.

Response: It is true that the solution should be applied to the entire profile. However, often the "entire profile" only refers to very shallow data. In this analysis we are addressing, "How would it change our vertical flux estimates if our 'entire profile' was only for over a shallower vertical distance?" We have endeavored to select data in which, as explained in the methods, the flow varies only minimally over the study period. We did not try to fit Ke because the purpose was to use the same parameter values in each of the models and compare the difference in model outputs between the individual models.

26. Section 3.2.3: I don't quite see why the authors restrict themselves to analyzing the amplitude dampening of the dirunal signal. The phase shift does provide additional information, and of course one can use the combination of the steadystate and diurnal contributions. The statement on page 4317, line 13-15, that the majority of studies based on analyzing the periodic signal makes use of the amplitude only may be doubted. We have always taken the time shift as primary measurement, because it is the most intuitive quantity (Vogt et al., 2010a,b, 2012). However, if possible, one should always take both the amplitude and dampening.

Response: This is true, it would be useful to estimate the flux using both amplitude and time shift. However, as stated, in this case the time shift did not yield any result for the majority of timesteps because there was not a consistent shift in peak temperature at depth. In total we give four reasons for not using the time shift on page 4317 lines 10-15.

In response to the comment on the Vogt et al. papers, Vogt et al. (2010) use a model based on constant vertical seepage rate, regardless of whether non-uniform vertical flow is present and the velocity is integrated over the entire depth, independent of local velocity variations. Obviously, improvement in vertical velocity resolution is warranted and in fact Vogt et al.

(2010) specifically state the importance of the vertical variation of the horizontal groundwater flow velocities estimated through their vertical profiling. However, Vogt et al. (2010) do provide a useful suggestion of using high vertical resolution temperature measurements when possible. As the authors state, this would help to avoid or identify discrepancies such as we found when estimating vertical fluxes using only sparse measurements in the vertical direction. A reference to these findings of Vogt et al. (2010) has been added to the discussion section.

> 27. Page 4320, line 7. I don't quite see why a sensitivity analysis should be outside the scope of the current study. The authors could easily drop the virtual test case.

Response: The goal of the study was to identify differences in estimated flux between models, and under specific conditions illustrated in the field data. A sensitivity analysis of the unknown parameters, which were kept consistent between models, would not help to elucidate this goal.

Anonymous reviewer two:

Comments: 1) “Can we use the BP and HK analytical solutions when both streambed convection and conduction are upward?” First, the criteria used to answer this question are not clear. For example, the authors state that “flux estimates over the entire vertical streambed column can be within one order of magnitude of the numerical model under some conditions.” Do the authors consider “within an order of magnitude” to be acceptable? And more importantly, what are the “some conditions” under which the estimates were acceptable? Both the criteria used to evaluate the analytical solutions, and the conditions under which the solutions were acceptable, should be made more clear and more quantitative.

On a related note, it would seem that there would be some threshold upward temperature gradient above which the analytical solutions would not be satisfactory. However, I am not sure that testing only two temperature gradients (0 and 7 deg C) as the authors have done can adequately address this issue. Why did the authors not repeat these calculations for a range of temperature gradients? I would recommend such an exercise before publication.

2) “How important are the sampling depths under these conditions?” In the conclusions, the authors state that “temperatures at intermediate depths in the sediment column provided necessary information on heat transport and water flux through the sediments.” But it is not clear what is meant by “intermediate” – does this simply mean that more than one temperature sensor is needed? If so this doesn’t tell us anything about the depths that should be required. The authors also mention the well-known “need for consideration of sensor spacing and selection in determining flux through streambed sediments,” but it is not clear what new information or insight this study provides with respect to this consideration. For example, how deep do the sensors need to be? The answer to this important question is not clear.

3) “How do results from each of these analytical models compare to the numerical estimates under these conditions?” While this is the most straightforward of the three questions, it seems to me to be a restatement of the first question. I would recommend combination of the two.

Response: The stated objectives have been combined and simplified to:

- How well do these common analytical approaches perform when the convective and conductive fluxes are upward?
- How important is the length of the vertical profile of observed temperatures under these conditions?

The discussion has been rewritten and now specifically addresses these two points. First, the ability of the analytical solutions to match the numerical estimate is addressed for the scenarios that use the entire vertical profile. This is quantified in terms of error. Then, the variability in flux estimates produced by the shorter profiles is discussed. It is also mentioned that deeper profiles might better capture the driving forces for water transport if the object is to capture the flux of water between the surface and the groundwater. Further, a discussion of the limits of the Stallman equation (as formulated by Hatch et al., 2006 or Keery et al., 2006) is included. No absolute limits are presented, because the ability of the equation to correctly estimate fluxes is dependent on the magnitude of conduction and convection and the thermal properties (ie how far the diel signal from the surface propagates into the subsurface). Because we are not trying to find these absolute limits, we do not think it would improve the manuscript to add further temperature gradients to the theoretical example. The theoretical example is simply there to elucidate the effects of a thermal gradient on predicted temperatures from the analytical solutions. The patterns, not the absolute error between the numerical and analytical solutions, is the important message of the theoretical example.
The abstract has also been edited for clarity and the phrase “within an order of magnitude” has been removed. Instead, the range of error is given.

Minor points:

P4312, L3: define RMSE when first used

Response: Now written out on line 22.

P4314 L10-11: “Surface water discharge during this period was 10.60+- 0.38 m3 s-1” Please cite the source of this value.

Response: This is from Shope (2009) and is now cited.

P4316 L14-15: “Mean differences in fluxes between Hydrus and BP3 was 0.01 deg C and even less for BP2.” This is confusing – are you comparing fluxes or temperatures?

Response: This refers to the differences in temperatures at each timestep and has been corrected to read, “A comparison of BP3 simulated temperatures and observed data at the 0.210 m deep sensor for each timestep yields an overall mean difference of only 0.01°C for the study period, and even less for BP2.”

P4320 L5: Does it have to be “hyporheic” flow? Wouldn’t non-vertical discharging groundwater cause the same problem?

Response: Yes, nonvertical, discharging groundwater would also cause error. There was generally more error when only the shallower probes were used, which is why we considered hyporheic flow a potential source of error. This sentence has been modified to read, "nonvertical or hyporheic flow".

> P4320, L28 to P4321, L12: This summary of previous work doesn't seem appropriate for the Conclusions section. I would recommend moving to the introduction.

Response: Moved as suggested.

References cited:

Anderson, M.: Heat as a groundwater tracer, Ground Water, 43(6) 951–968, 2005.

Bredehoeft, J. D., and Papadopulos, I. S.: Rates of vertical groundwater movement estimated from the earth's thermal profile, Water Resources Research, 1, 325-328, 1965.

Conant Jr, B.: Delineating and quantifying ground water discharge zones using streambed temperatures, Ground Water, 42(2), 243–257, 2004.

Constantz, J.: Heat as a tracer to determine streambed water exchanges, Water Resources Research, 44, W00D10, doi: 10.1029/2008WR006996, 2008.

Hatch, C. E., Fisher, A. T., Revenaugh, J. S., Constantz, J., and Ruehl, C.: Quantifying surface water–groundwater interactions using time series analysis of streambed thermal records: Method development, Water Resources Research, 42, W10410, doi:10410.11029/12005WR004787, 2006.

Keery, J., Binley, A., Crook, N., and Smith, J. W. N.: Temporal and spatial variability of groundwater-surface water fluxes: Development and application of an analytical method using temperature time series, J. Hydrol., 336, 1-16, 2007.

Lautz, L. K.: Impacts of nonideal field conditions on vertical water velocity estimates from streambed temperature time series, Water Resources Research, 46, 10.1029/2009wr007917, 2010.

Lautz, L.K., and Ribaudo, R.E.: Scaling up point-in-space heat tracing of seepage flux using bed temperatures as a quantitative proxy, Hydrogeology Journal, DOI 10.1007/s10040-012-0870-2, 2012.

Niswonger, R., and Prudic, D.E.: Modeling heat as a tracer to estimate streambed seepage and hydraulic conductivity. In: Constantz, J., Stronestrom, D.A. (Eds.), Heat as a tool for studying the movement of ground water near streams, United States Geological Survey Circular 1260, pp. 81-89, 2003.

Schmidt, C., Bayer-Raich, M., and Schirmer, M.: Characterization of spatial heterogeneity of groundwater-stream water interactions using multiple depth streambed temperature measurements at the reach scale, Hydrology and Earth System Sciences, 10, 849-859, 2006.

Schmidt, C., Conant, B., Bayer-Raich, M., and Schirmer, M.: Evaluation and fieldscale application of an analytical method to quantify groundwater discharge using mapped streambed temperatures, J. Hydrol., 347(3-4), 292–307, doi:10.1016/j.jhydrol.2007.08.022.

Shope, C. L.: Water and heat fluxes in a channel bar compared with adjacent streambed and streambank hydrogeology, Program of Hydrologic Sciences, University of Nevada, Reno, NV, 226 pp., 2009.

Stallman, R. W.: Steady one-dimensional fluid flow in a semi-infinite porous medium with sinusoidal surface temperature, Journal of Geophysical Research, 70, 2821-2827, 1965.

Vogt, T., Schneider, P., Hahn-Woernle, L., and Cirpka, O.A.: Estimation of seepage rates in a losing stream by means of fiber-optic high-resolution vertical temperature profiling, J. Hydrol., 280, 154-164, 2010.