We would like to thank the reviewer for the constructive and useful comments. We have posted our replies below the respective comments.

General comments

1. I think that the methodology section “Tomographical inversion procedure” should be better articulated. More specifically, there is no apparent link between “travel time inversion” and “early diagnostics”. For example, it is not clear to me why the step 3 of early diagnostic is needed if we invert for first arrivals. I think that an additional subsection stating the inverse problem in terms of equations, clearly showing the data and model parameters used (K or log K), the forward operator, and how the inverse problem is solved is necessary, especially for a paper proposing a new inversion methodology.

We agree with the reviewer and we have improved the section “inversion”. The details are discussed in response to the specific comments below.

2. Throughout the paper, the level of misfit (RMS, error weighted data misfit, chisquare, for example) between observed and calculated data is never given (scatter plots would be interesting, too). I think this is an important point for several reasons: 1) the comparison between different inversions is only valid if they have a similar level of misfit 2) the use of the staggered grid approach probably lead to a level of misfit which is higher than one would obtain with a finer grid, because the staggered grid is not able to reproduce fine scale structures 3) there is no insurance that the mean of the inversion on the different staggered grid actually explains the data (as the mean of a posterior distribution is not necessarily a sample of this posterior), therefore a forward simulation on the final results (fine grid) should be performed to assess the RMS and 4) This is a commonly used value to assess the quality of an inversion (standard in geophysical inversion).

For the validation section a new scatter plot was added (see reply to comment #20). No significant difference in the RMS misfit of the travel time inversions was detected. The staggered grid method has already been proven by several studies to be ideal to solve similar problems such as hydraulic tomography or tracer tomography (Brauchler et al., 2003, 2013a, 2013b; Hu et al., 2015; Jiménez et al., 2013). Using finer grids instead of staggering would require regularization terms and cannot guarantee a reconstruction of the real heterogeneity without prior information. Otherwise, the limited number of source-receiver combinations would result in an ill-posed problem. In the validation case, the investigation of the RMS error has been done on the fine grid result. In the last section, the quality assessment is focused on the usefulness of the result instead of the quantitative inversion performance. Hence, the aim was to produce a practical application window. However, we agree with the reviewer that including additional data and using finer grids would be an interesting aspect to investigate in the future.

3. The methodology makes some assumptions. For some of them, no reference is given and no demonstration is made in the paper, so that nothing really supports the statement. I would like those assumptions to be more thoroughly discussed, for example the limitations arising from them, how they could be overcome (use of geophysical data, joint inversion of hydraulic and tracer tomography, joint

inversion of salt and thermal tracers, regularized inversion, etc.). Some of these issues as well as some other comments are described in my specific comments.

The main goal of this study was to provide a standalone procedure which uses the tomographical thermal tracer data. Although strong assumptions have been made during the inversion (advective transport only, neglect of density and viscosity effects, homogeneous distribution of thermal rock parameters and uniform hydraulic gradient), the proposed method was capable of providing good reconstructions. Demonstration of the robustness of the inversion procedure is a core message of our work. We agree that the result quality could be increased by involving additional data in the inversion, but we also would like to point out that satisfactory results can be achieved using only the thermal tracer tomography data.

Specific comments

1. P3L40-41. It might be worth mentioning geophysical technique here in addition to DTS. See for example Hermans et al. (2014).

Hermans, T., Nguyen F., Robert, T. and Revil, A. 2014. Geophysical methods for monitoring temperature changes in shallow low enthalpy geothermal systems. Energies, 7, 5083-5118.

Added according to the comment.

2. P8.L151-153. See my comment 3 above. For example, would it be possible to invert jointly for porosity and hydraulic conductivity? Even if smaller, the effect of porosity might not be negligible as it appears as squared (effect of the retardation coefficient), what is the effect on the hydraulic conductivity distribution? Also, the gradient (does fix means constant?) will have an influence near the injection well, where the obtained solution is sometimes relatively erroneous. The gradient could be computed at the end of each iteration as input for the next one. Please discuss.

With the limited amount of data available due to the fact that repetitions of thermal tracer tests are cost and time demanding, solving the problem for a second parameter would be difficult. Experience from hydraulic tomography studies shows that inverting on porosity is possible by involving additional information besides travel time (e.g. pressure attenuation) or by applying a model calibration step after the tomographic inversion (Brauchler et al., 2007; Hu et al., 2011; Jiménez et al., 2013).

If we consider a possible porosity range between 0.2-0.4, the squared range spans only 0.04-0.16, which is still much less than the expected differences in *K*. With respect to the gradient, the authors agree that the distribution of hydraulic gradient can have an effect on the result quality, and that it could be corrected with iterative updates, but our sensitivity analysis on injection rates (section 4.4) showed that under realistic circumstances, this distortion effect is negligible.

3. P9L171-172. Is the peak breakthrough time already refer to the derivative of temperature? Precise to avoid confusion with maximum temperature that would be obtained with a Dirac or time-limited injection. What do you mean by real breakthrough time? Maybe show the equation of T’, too.

The peak breakthrough time refers to the derivative of temperature. Revised as:

"Mitigation thus can be done by using an earlier characteristic time of the thermal front instead of the (peak of the first derivative) breakthrough time, thus using the fastest component of the heat transport - advection."

4. P10L188. Do you have a reference that shows the non-sensitivity to those parameters?

These parameters are either have higher orders or multiplied by velocities at higher orders. By neglecting the terms with higher orders of velocities, all of these parameters are cancelled out.

5. P11L208-210. Why is this third step needed? Do you invert for the peak time or the early time? Is it this corrected peak time which is calculated through equation 2? Please make the link between the two sections.

New sentence added: *"Step 3 allows the travel time to be related with the transport process, and to return a real and scaled K value instead of just information about the heterogeneity contrasts."*

6. P12L219-220. Do you mean hydraulic conductivity instead of velocity? I guess that velocity refer to the analogy to the traditional use of the SIRT algorithm in seismic and GPR inversion (equation 1), but I think it would be more consistent to use hydraulic conductivity everywhere (as in the figure for example), considering that other parameters (porosity and gradient) are not taken into account. How do you update the values (reference)?

The SIRT algorithm solves for mean tracer velocity, and then this velocity is transformed to K with a scaling step where constant porosity, gradient and heat capacity are assumed. A detailed description of the SIRT algorithm can be found in the appendix of (Hu et al., 2011).

7. P12L221-224. Give a reference on inversion problems such as Menke (1989) or Aster et al. (2005)

Menke, W. 1989. Geophysical data analysis: discrete inverse theory. Academic press, Inc. San Diego, revised Edition.

Aster, R., Borchers, B., and Thurber, C. 2005. Parameter estimation and inverse problems. Elsevier Academic Press, Amsterdam.

References are added to the revised manuscript.

8. P12L226-228. The use of a limited number of parameters still limit each solution in the way that it is not possible to describe small scale heterogeneity, what would be possible on a finer grid with appropriate prior information and regularization (for example using non-smooth operator). See my comment on the misfit, too.

The choice of using staggered grid methods was to provide solution stability with the limited amount of source-receiver combinations. See also answer to comment 14.

9. P13L245. The tomographic matrix? None of your equation mentions this. Hence, the necessity to describe the equations of the inverse problem. I guess, that by the tomographic matrix, you mean the matrix mapping the model into the data d=f(m), i.e. the forward operator.

The ij element of the tomographic matrix is the length of the i-th trajectory path in the j-th pixel of the tomogram (Jiménez et al., 2013).

10. P13L262-263. Confusing, please rephrase. I guess that "similar hydraulic properties" refers to the properties inside a given hydrofacies, but it could be understood as "the different hydrofacies have similar properties".

Revised accordingly.

11. P14L280-L283. Those values correspond to the analog. What values were used for inversion (constant value for the whole aquifer for porosity, thermal conductivity and heat capacity). Beta does not appear in the equations.

Beta was not used for the inversion. Arithmetic mean values were used for the thermal parameters and the head difference between the two wells was used to calculate the head gradient.

Modified line 295:

"The constant head values are specified to impose an average hydraulic gradient according to Table 2, but for the inversion, the measured cross-well head difference was used."

12. P15L305-307. Considering buoyancy and viscosity effects?

Clarification is added to line 306: "Note that independent of the dimensions of the reconstructed sections, the full 3-D analog model was always used to simulate the thermal tracer propagation and resulting travel times, considering buoyancy and viscosity effects."

13. P16L332. The influence of the injection rate on the gradient i was neglected during inversion (i based on the boundary condition only)? What is the influence? Please precise.

These effects are discussed in chapter 4.3 dealing with the sensitivity analysis of injection rates.

14. P18L384. Globally, the results on a staggered grid really look like the results that would be obtain on a finer grid with smoothing regularization. What is the benefit in terms of time of using the staggered grid approach? Such approach is probably not able to produce sharp contrasts as observed in the aquifer analog due to the high resolution grid, whereas a high resolution grid with an appropriately chosen regularization (e.g., MGS as developed in geophysics) can potentially lead to better results. Also, the staggered approach probably tends to average the results, so that high and low K are under-represented in the final solution. Also, there is no certainty that the mean of the 16 tomograms actually fits sufficiently the data.

Experience from earlier studies shows that the staggered grid method is an ideal choice for discretization of hydrogeological problems (Brauchler et al., 2003, 2013a, 2013b; Hu et al., 2015; Jiménez et al., 2013).

15. P19L387-399. The null-space energy represents zones not covered by flow paths. The origin can be either the source-receiver positions or the presence of low K zones. In figures 5b, d, f, the null-space zone is always larger in the upper part where low K are observable. Therefore, a low-K value may be represented by a high null-space energy, but still be relatively reliable. A sensitivity approach (using the Jacobian for example) could potentially revealed this by showing that those parameters actually influence the data.

In this study we focused on tomogram areas in which preferential flow paths are reconstructed (see line 30 and line 611). K-values of the preferential flow trajectories are of major interest in many applications such as solute transport prediction or remediation. Investigation of low-K zones without transport trajectories would be a very interesting aspect to examine, but it is beyond the scope of this work.

The following text is added to line 477:

"Although the 2-D nor the 3-D inversion were capable of reconstructing the top low-K zone, the distribution of the reconstructed transport trajectories can be used to identify these locations. Although revealing more information about these zones is beyond the scope of this work, it would be an interesting aspect to examine in the future."

16. P20L413. In P1, the highly conductive zone is continuous, but not between any used injection-receiver pair. The method thus misses how this layer becomes thinner near the measuring well.

The two receivers are inside this layer on the reconstruction because they are both very close to the high-K zone and affected by it. The thickness of the layer between the two wells is too small to properly reconstruct its changes with the used number of sources and receivers.

17. P20L414-417. Are the K values of inversion mean values? How are there calculated (arithmetic, geometric mean)? Which cells were considered (cells corresponding to the true zone)? Not straightforward as the values of K are continuous. I find relatively surprising that the inversion is not able to retrieve the homogeneous conductivity value of the bottom layer as it is fully constrained by 4 injection-receiver pairs.

Mean values were taken from the reconstructed zones. It is true that the bottom zone contains 4 injection-receiver pairs, but the transport trajectories close to the top are distorted by the higher K zone above.

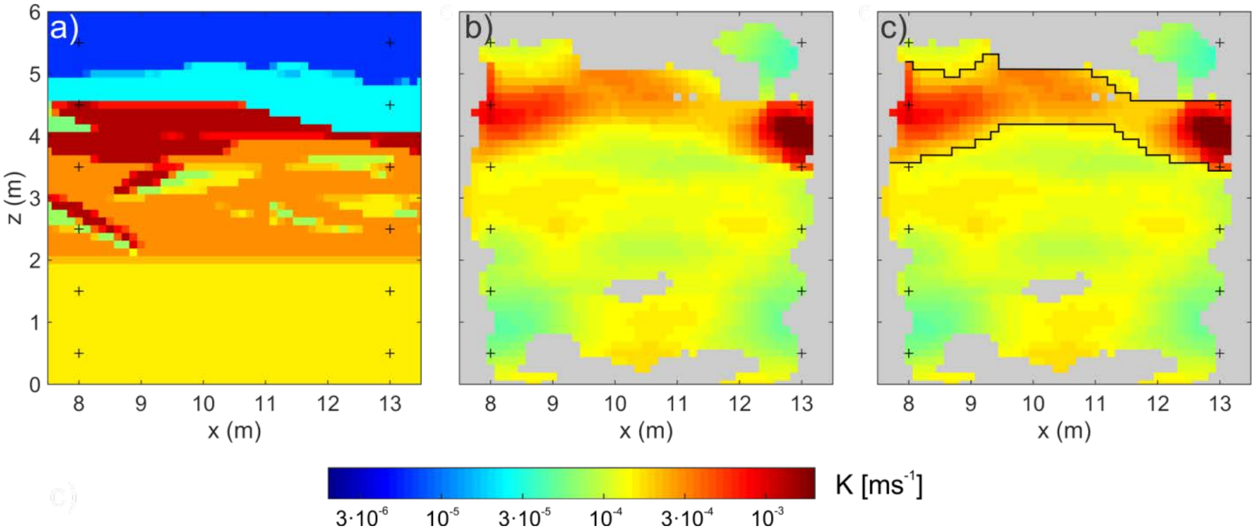
18. P20L422-428. The effect of taking early times is not shown in the paper, since the method is not compared with a classic inversion of BTC. I would thus mitigate this statement.

The more exact time detection of ETD (see line 204) was used as a benefit during the inversion:

“The identification of the peak time through the derivative T' *is challenging due to the flatness of the curve at the maximum value of the peak. However, using the early time diagnostics (step 1), only the value of the peak must be known for Eq. (8).”*

19. P20L429-439. A suggestion for the figure would be to show the boundary between facies on the inverted tomograms.

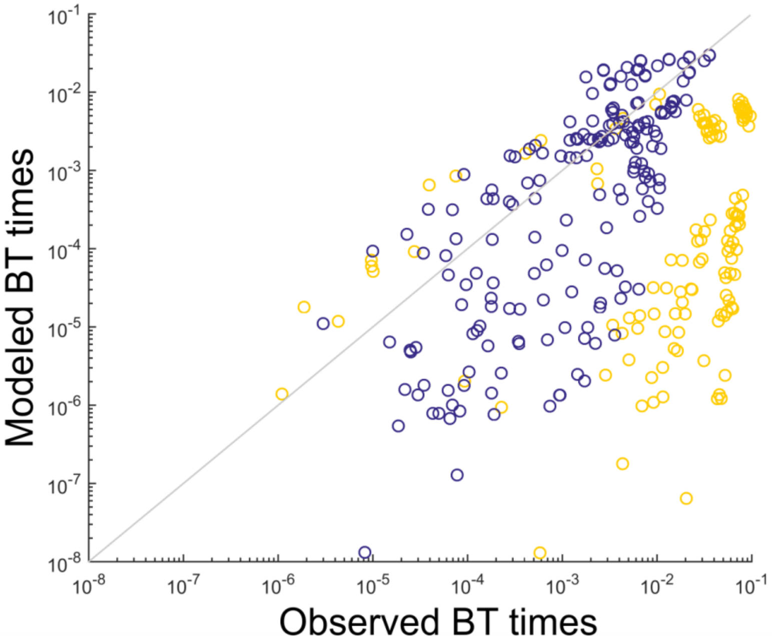
We were also considering highlighting the facies boundaries, but chose not to, in order to allow the reader to inspect the results from an unbiased perspective. One example with boundaries is shown below:



Results from profile 1: a) analog profile, b) reconstructed tomogram, c) tomogram with highlighting of the high-K facies.

20. P23L489. Consider also a scatter plot (true vs. estimated) for a better visualization.

The following scatter plot is added to the revised manuscript as figure 7b:



The figure shows the relation between the modeled and the observed breakthrough times (BT) on a logarithmic scatter plot. The yellow points mark the known outliers, such as observations of the unreconstructed low-K zones (also presented on the histogram at the original figure 7).

21. P24L516. It is never explicitly said that buoyancy and viscosity effect were considered in the simulation of the experiment on the 3D analog model.

See our answer to your comment #12.

22. P25L541. Why approximately constant and not constant? What deviations from constant are there?

There are no deviations; the word "approximately" is removed.

23. P25L538-546. To me, the results of figure 9f is not that bad. It is closer to the truth near the measuring well, detecting the low hydraulic conductivity zone in the upper part. In contrast to figure 9e, it does not image the preferential flow paths into the zone of low hydraulic conductivity. It is true that it is not satisfactory near the injection well. To me, the most coherent solution might be figure 9d, showing that some distortions could already happen in figure 9e.

The assumption of homogeneous head gradient distribution is weaker if we increase the injection rate. It is important to find a compromise, with an injection rate where the temperature change can be measured but the head distribution is still close to uniform (Note that this assumption is not valid if there is heterogeneity, but this effect is small and ignored). These distortions have less effect at the observation side, and thus the results remain good.

24. P29L622. Viscosity has a direct effect on hydraulic conductivity, so it should have an effect on the inversion based on advection, too.

Viscosity should have an effect on the result quality at large temperature differences. On figure 8, this is probably the reason for the differences in the resulting K values with different temperatures.

Technical comments

1. P17L343. Specify here how many pairs are left (instead of line 375).

Revised accordingly.

2. P20L420. Replace vales by values.

Corrected.

3. P23L482. Do you mean figure 6b?

No, this sentence is related with the validation wells, displayed in light grey in Figure 4b.

4. P24L516. Do you mean Figure 8?

Corrected.

References

Brauchler, R., Liedl, R. and Dietrich, P.: A travel time based hydraulic tomographic approach, Water Resour. Res., 39(12), n/a–n/a, doi:10.1029/2003WR002262, 2003.

Brauchler, R., Cheng, J.-T., Dietrich, P., Everett, M., Johnson, B., Liedl, R. and Sauter, M.: An inversion strategy for hydraulic tomography: Coupling travel time and amplitude inversion, J. Hydrol., 345(3-4), 184–198, doi:10.1016/j.jhydrol.2007.08.011, 2007.

Brauchler, R., Böhm, G., Leven, C., Dietrich, P. and Sauter, M.: A laboratory study of tracer tomography, Hydrogeol. J., 21(6), 1265–1274, doi:10.1007/s10040-013-1006-z, 2013a.

Brauchler, R., Hu, R., Hu, L., Jiménez, S., Bayer, P., Dietrich, P. and Ptak, T.: Rapid field application of hydraulic tomography for resolving aquifer heterogeneity in unconsolidated sediments, Water Resour. Res., 49(4), 2013–2024, doi:10.1002/wrcr.20181, 2013b.

Hu, L., Bayer, P., Alt-Epping, P., Tatomir, A., Sauter, M. and Brauchler, R.: Time-lapse pressure tomography for characterizing CO2 plume evolution in a deep saline aquifer, Int. J. Greenh. Gas Control, 39, 91–106, doi:10.1016/j.ijggc.2015.04.013, 2015.

Hu, R., Brauchler, R., Herold, M. and Bayer, P.: Hydraulic tomography analog outcrop study: Combining travel time and steady shape inversion, J. Hydrol., 409(1-2), 350–362, doi:10.1016/j.jhydrol.2011.08.031, 2011.

Jiménez, S., Brauchler, R. and Bayer, P.: A new sequential procedure for hydraulic tomographic inversion, Adv. Water Resour., 62(PA), 59–70, doi:10.1016/j.advwatres.2013.10.002, 2013.