Interactive comment on "Contaminant source localization via Bayesian global optimization" by Guillaume Pirot et al.

Guillaume Pirot et al.

guillaume.pirot@unil.ch

Received and published: 20 September 2017

The original comments of referee 1 are in black and our answers to these comments are in blue.

General comments

1, In the introduction section, you have introduced some classification for groundwater pollution source identification. In terms of the Bayesian approach you used, you also need to introduce the state of art of contaminant source localization based on Bayesian approaches (eg., Cupola et al. [2015], Zeng et al. [2012] , Raziyeh Farmani et al.[2009]), and detail the difference and merit of the approach you used by comparing with those Bayesian-based approaches.

The term 'Bayesian' is typically used when a method involves Baye's rule. And, as the referee points out, many Bayesian approaches have been used throughout groundwater sciences and notably for contaminant source localization. However the term "Bayesian Optimization" is very specific and does not refer to the general combination of Bayesian methods and optimization. It refers to the special case where the objective function itself is seen as random and endowed with a prior distribution, updated along evaluations (See Shahriari et al. 2016 for a review). Kriging can be referred to as Bayesian inference. Approaches classified in both optimization or probabilistic methods may then be qualified as Bayesian. This is why we think it would not particularly help to compare our method specifically to other Bayesian approaches out of the scope of Bayesian optimization. We will update the introduction to clarify the meaning of the term 'Bayesian' and to explain why a comparison with other Bayesian approaches would not make much sense.

In Butera et al. [2013] (who developed the SRSI approach, a "stochastic procedure which finds the source location and the release history by means of a Baseyian geostatistical approach" as stated by Cupola et al. [2015]), kriging is used to infer release functions that are combined with specific zonal transfer functions. Flow and transport numerical simulations need to be computed for all candidate locations of the transfer functions. In our approach, computations are not limited to isolated locations, and it is possible to explore solution at any location of the search grid (which is limited by the resolution of the flow and transport numerical model).

The approach proposed by Zeng et al. [2012] uses a Bayesian formulation of the inverse problem to estimate the probability density function (pdf) of the location and release time of a contaminant. Optimization approaches, such as ours, provide optimal solutions but it do not allow to estimate the pdf.

The paper by Raziyeh Farmani et al.[2009] does not seem relevant for comparison, as it does not deal with contaminant source identification, but with water management.

2, Why do you choose source A (89, -36) and B (100, 10) as reference contaminant sources? Besides, it is better to show source A and B in Figure 1.

We chose two different contaminant source reference locations to test the influence of the source location versus the geology first on the misfit objective function and second on the ability of the proposed approach to deal with more or less complex objective functions. This is an argument we will add in the related paragraph in Section 2 (Synthetic test cases).

3,On page 10, can you explain why there is large departure between reference source location and minimum of the objective function location.

This is because the resolution of the search grid is not the same as the grid resolution of the flow and transport numerical model. It is explained in lines 13 to 15 of page 10 and this choice is motivated between line 23 of page 5 and line 3 of page 6.

Specific comments:

1, On page 1, it is better to briefly explain how analytical solution and regression approaches works, just as you have done to the other three categories.

Certainly, we could precise: 'in which a set of equations can be solved analytically or whose parameters can be estimated by least-square regression'.

2, Please be careful to use "To the best of our knowledge" on page 2 and 3.

This is what we use on page 2 lines 27 and 32, we will correct this on page 3 line 12.

3, It is not appropriate to set longitudinal dispersivity as 1 m when your resolution of the aquifer is 1m*1m.

We do not agree for two reasons. First, we are mainly modeling the spreading of the contaminant due to an explicit description of geological heterogeneity at a small scale. Therefore, the longitudinal dispersivity is taken at the smallest possible value with our mesh size. Second, the same parameters are used for all the simulations. The only

C3

unknown is the position of the contaminant source. Therefore, the comparison between the various numerical simulations is fair. To conclude, we could use a larger dispersity but it would spread the solute in an exaggerate manner and mask the effect of the heterogeneity. The problem would be much less interesting.

4, Please show both source A and B in Figure 3 and Figure 1.

This is something we can add if the referees think it can improve the understanding of the experimental setup. In that case, for consistency between the figures, it might make sense to add the search zone as well, and to add both search zone and reference locations on Figure 2 as well.

5, The dots in Figure 5 are with three different colors, and you just explained the meaning of blue and white dots. Please explain the left one.

Indeed, a part of the legend is missing. The following legend (Fig. 1 of this answer) will be added to the figure.

References used in the answer:

Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. "Taking the human out of the loop: A review of bayesian optimization." *Proceedings of the IEEE* 104, no. 1 (2016): 148-175.

Cupola, Fausto, Maria Giovanna Tanda, and Andrea Zanini. "Laboratory sandbox validation of pollutant source location methods." *Stochastic environmental research and risk assessment* 29, no. 1 (2015): 169-182.

Zeng, Lingzao, Liangsheng Shi, Dongxiao Zhang, and Laosheng Wu. "A sparse grid based Bayesian method for contaminant source identification." *Advances in water resources* 37 (2012): 1-9.

Farmani, Raziyeh, Hans Jørgen Henriksen, and Dragan Savic. "An evolutionary Bayesian belief network methodology for optimum management of groundwater con-

C4

tamination." *Environmental Modelling & Software* 24, no. 3 (2009): 303-310.

Butera, Ilaria, Maria Giovanna Tanda, and Andrea Zanini. "Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach." *Stochastic Environmental Research and Risk Assessment* 27, no. 5 (2013): 1269-1280.

C5

○ 9 points initial design　　　● 41 first iterations　　　● 50 last iterations

✕ contaminant reference location　　　□ objective function minimum

Fig. 1. missing legend