Dear Editor and authors,

This work can be interesting. It aims to localize the contaminant source of (non-reactive) solute in a synthetic confined aquifer through a Bayesian optimization approach, with the support of 25 measurements locations. The measurements (with zero measurement error) are taken from the 'true' case where the model inputs are assumed to be completely known. To what I have studied, authors wanted to use four scenarios (coping with different nonlinearity level of objective functions) to explore the capability of the Bayesian optimizations in localizing contaminant sources. However, in my opinion, using the objective functions stem from those four scenarios to convince the reader that these objective functions can be referred as a benchmark is weak. I notice that the authors' response to General Comments (3) of referee #2 do make sense in some extent. However, the associated nonlinearity level in these objective functions can hardly be previously classified (or say ranked) which weaken attractiveness of this work. To improve the quality of this work, I would like to suggest the authors to further perform scenarios considering measurement error in several different magnitudes and/or various measurement network having various number of measurement locations. My points to provide this suggestion are majorly attributed to that (*i*) measurement error and number of measurement locations are two very important factors in a realistic problem; (*ii*) the objective function stemmed from localizing contaminant sources is indeed a function both measurement error and number of measurement locations. Generally, the higher measurement error the higher nonlinearity level of objective functions; the less number of measurement the higher nonlinearity level of objective functions. Then, authors can explore robustness of Bayesian optimization approach in several classified nonlinearity levels, which can improve the quality of this work and increase its attractiveness of being a good reference in localizing contaminant sources.
I further give the following specific comments line by line.

Line 1 to line 2 on page 1: Please keep the consistency between terminology "transmissivity" and "hydraulic conductivity" throughout the main text.

Line 5 to line 6 on page 1: Why the objective function you proposed can be used as a benchmark? Beside they own multiple local minmia, are there other special reasons? There are many studies contributed to localize contaminant sources in heterogeneity medium. It can be better if authors can explain this in Abstract.

Line 9 to line 10 on page 4: Why did authors use 100 replications? Please explain this.

Figure 1: There are only three transmissivity values (i.e. 1E-1, 1E-3 and 1E-5) in Fig. 1, right? If so, I would like to suggest the authors to replace the color bar with a legend in color box to avoid confusion.

The caption of Figure 2: Where are the boundary conditions? Please check Figure 2.

About section 4, Can you further descript the implementation of the optimization procedure in a Flow Chart? It would be better to show computational procedure in a Figure than solely in a lot of words.

Line 20 to line 21 on page 9: Please write the preset maximum number of iteration here.