We are grateful for the constructive comments of the two reviewers. They helped us to significantly improve the quality of the paper, and we tried to modify our manuscript and reply in detail to every single comment when carrying out major revisions of the manuscript for submission to HESS.

Referee #1

The referee raised in total 10 specific comments. In the following we will comment on each of the points. For convenience, we cite the referee comment first (italic).

1. Comment:

During reading the paper it did not become clear to me why in this study a...
light-on instead of a light transmission method was used. Only at the end, in the Conclusions, it is mentioned that the light transmission method would be promising and that it was not chosen here as the construction of the chamber did not allow for it. I would expect that the light-through method would lead to more reliable results. The problem of 3D effects (which is not mentioned here, was it tested?) would be less severe, a uniform illumination could be achieved more easily, the flare effects would be less severe. Also, very good literature is available about light-through measurements to determine fluid contents (for example by the group of R. Glass and R. Detwiler, J. Selker and coworkers, among others).&#8221;

We fully agree with the referee on the comment on methodology comparison, especially the suggested comparison of light transmission and light reflection methods. According to the referee the problem of 3D effects would be less severe using the transmission method. We theoretically agree with this comment, however in our experiments 3D effects were not observed. The problem of 3D effects was visually tested by comparing the position of the salt-dye front at a distinct time step at the front window and at the back window of the tank. The inlet and outlet openings are placed in the center of the side edges, and the tank is filled with homogeneous porous media, which prevents 3D effects. Further, the referee argues that a uniform lightning could be achieved more easily applying the light through method. We agree, but since we derive the parameters of the intensity vs. concentration curve for each observation point separately a uniform illumination is not mandatory for our image analysis approach. The only prerequisite is that there are no fluctuations in lightning from one picture to the other. To summarize this it is essential to have a temporally uniform illumination, whereas the spatially non-uniform illumination is considered by the image analysis approach. We fully agree with the referee on the reduction of flare effects using the transmission method. However, the light transmission method is beyond the scope of this work. In the revised version of the manuscript we added a discussion of light on vs. light through methodology in the light of published literature.
2. Comment:

The image analysis discussed in these papers (R. Glass and R. Detwiler, J. Selker and coworkers) could be extended to analyze solute concentration in one phase. There probably exist already studies about this. I guess the light-on method has also advantages, for example, it could be used for non-transparent filling material. A comparison of the methods and a discussion of the advantages of the method used here would be helpful as a motivation for the use of the light-on method.

Since most of the relevant experimental work in fluid mechanics is done using the light on method (Schincariol et al., 1993, Swartz and Schwartz, 1998, Wildenschild and Jensen, 1999, Simmons et al., 2002, Rahman et al., 2005, McNeil et al., 2006, Goswami and Clement, 2007) we consider our paper as contribution to assess the applicability, reliability and the limitations of the methodology. Therefore, the focus of the paper should be on the light on method and its inherent limitations. However, we agree with the referee that a more detailed discussion of the advantages vs. disadvantages of the method would improve the paper. This is added in the revised manuscript.

3. Comment:

The correction of fluctuations in brightness (Section 3.2) is not so convincing. It is demonstrated here only for the color cards (Figure 3), which are placed close to the reference card. It would be interesting to see a demonstration of the applicability of the correction for the concentration measurement at an observation point, which is not placed in the center and where the illumination difference to the reference card is high. As written above, Rahman et al., 2005, use a similar optical method to determine the solute concentration in a lab experiment. They use a more complex procedure to compensate for the fluctuations of the light source than described here in Section 3.2. The method used here would correspond to their equation (13) with gamma = 1 and alpha = 0. It would be useful to have a comment on that.

Rahman et al. (2005) used the \(\gamma\)-calibration model to correct the color repre-
presented images:

\[
\begin{align*}
\text{with } a, b, \gamma \text{ as correction parameters, assumed spatially uniform, } I \text{ is the measured intensity. The values of the parameters are determined by fitting the measured intensities of color cards to the ideal model. Schincariol et al. (1993) or McNeil et al. (2006) among others convert the measured intensities to optical density in order to account for fluctuations in brightness from one image to the other. Optical density } D \text{ is non-linearly related to intensity } I \text{ by:}
\end{align*}
\]

where \(a\) is simply a constant of proportionality. Schincariol et al. (1993) and McNeil et al. (2006) argued that the standardization of images to optical density is necessary because variations in lightning, exposure and film development result in non-uniform image intensity between successive images. However, modern digital cameras collect linear intensity values and store those measurements linearly on the chip. The Nikon D70, used in our experiments, provides two types of data of one image: 1. The linear raw data stored as .nef 2. Automatically, non-linearly adjusted images as .jpg RAW data (which Nikon call NEF data, an acronym for Nikon Electronic File) is the output from each of the original red, green and blue sensitive pixels of the image sensor, after being read out by the array electronics and passing through an analogue to digital converter. Now one of two things can be done with the RAW data. It can be stored on the memory card, or it can be further processed to yield a JPEG image. If the data is stored as a JPEG file, it goes through the Bayer interpolation, is modified by in camera set parameters such as white balance, saturation, sharpness, contrast etc, is subject to JPEG compression and then stored. The advantage of saving JPEG data is that the file size is smaller and the file can be directly read by many programs or even sent directly to a printer. The disadvantage is that there is a quality loss, the amount of loss depending on how much compression is used and the data are not linear any more. The more complex correction method used in Rahman et al. (2005) or the standardization of images to optical density is only necessary if: 1. Analogue images are taken and the film has to be developed and scanned to convert the image
to digital data. Non-linearly adjusted images (.jpg) are used in our case we used the linear raw data. Therefore, the intensity correction based on reference white intensity (I\text{ref}) is sufficient:

Further, a more complex non-linear correction is contra productive, because it brings artificial non-linearity into the data. Nowadays, optical densities are only used for image processing in terms of visual improvement of colors or brightness (personal communication Dr. Rosenthaler, Visual Media Lab, Uni Basel). We analyzed the impact and the functionality of our brightness correction method at several points along the image and the method delivers comparable results at all points. Therefore, we assume that it can be used for the entire image. We added this discussion in the new manuscript.

4. Comment:

"The applicability of the fluctuations of the brightness (Section 3.2) would also be more convincing if a mass balance for a test experiment would be shown. The input and output of concentration are known, therefore the total mass of salt water in the flume is known. It would be interesting to see if this mass is recovered with the optical method used here and if possible mass errors are in the range of the error due to the lens flare effect."

The impact of lens flare should be more pronounced if only a small portion of the domain is filled with the dark solution. Whereas, the effect should be reduced if the tank is filled by the dark solution. In order to analyses this we conducted a fifth experiment, E5. The dye-salt solution enters the domain over one inlet at the margin of the tank with a constant well-measured flow rate. The concentration is 100 g/l of NaCl. Each 30 sec an image is taken and we analyzed the images after 10 and 45 minutes. Figure 10 in the revised manuscript shows the spatial distribution of the saltwater (red). The 10 min image delivers an underestimation of the total mass of -4.2 % compared to the mass entering the domain. This is within the range of expected lens flare errors. After 45 min the bright region of the tank is significantly reduced and the mass error
accounts for 1.6 %. The method overestimates the total mass. Since the porosity is used to calculate the mass possible small errors in this parameter could cause the overestimation. However, this would affect the mass after 10 min in the same way. Due to this analysis the calibration parameters determined for each pixel are reliable and possible errors can be explained with the impact of lens flare.

5. Comment:

I am not very convinced of one of the major findings in the paper that the resolution of the observation point is crucial for the precision of the intensity measurement (Abstract, Conclusions and Section 3.3). The resolution areas analyzed here correspond roughly to one grain size up to an area of 10 x 10 grain sizes, which is roughly an REV of the porosity. The fluctuations of the intensities become insignificant only after averaging over the REV, which is probably not so surprising and could be transferred to other experimental setups or fillings.

The statistical analysis clearly demonstrates the decline of precision with increasing resolution. Therefore, we consider the conclusion that the resolution is crucial for the precision of the measurement as valid. There is no or at least little discussion on this in most of the studies found in the peer literature. Schincariol et al. (1993) and McNeil et al. (2006) suggested a general 3x3 pixel median smoothing. In our case 3x3 pixels are not sufficient. Based on the statistical analysis the adequate resolution is a trade off between the precision of the measurement (lower resolution) and the ability to derive detailed information from the images (higher resolution). We agree with the referee that the right resolution corresponds to the REV. However, this has to be determined for the specific grain size used in the experiment. Thus, the statistical analysis is necessary and the resolution of the observation point needs to be determined specifically for each experiment.

6. Comment:

The mentioning of the different experiments (E1 to E4) in Section 2 without
further explanation is confusing. The same is true for Section 3.4. It would be helpful if it would be explained in Section 2 what experiments were carried out.

This is improved in the new version of the manuscript. We introduced the experiments in section 2 as recommended by the referee.

7. Comment:

In Rahman et al., 2005, the images had to be corrected for rotation or translation movement. Was this not necessary in these experiments?

Our camera was fixed and not removed or touched during the experiments (see old manuscript pp. 4181, l. 13-15): For image processing it is important that images, taken at different times, match on a pixel by pixel basis. Therefore, a computer program (Nikon camera controlPRO) controlled the camera, which was not touched or removed during the entire experiment including the calibration procedure. Thus, neither translation nor rotation was necessary.

8. Comment:

I do not agree with the conclusion that both optical and electrical resistivity methods yield the same concentrations. Unfortunately in Figure 15 only the upper and lower limit of the optical curve is given. But if I understand the error estimation correctly, the measurement curve would be closer to the lower bound. So there is quite a difference between the concentration measured with electrical resistivity and the optical method in P2. Also in P3 the mass is different. It could be argued that the electrical resistivity curve is broader due to the larger sampling volume. However, if this was the only effect, the mass underneath the curves should be the same. This seems not to be the case. Which method is more reliable? Could the difference be due to the flare effect although the mask was applied?

From supplementary experiments (see section 3.5) we found that the observation holes around the observation point might influence the measurement and flare effects are
possible although the mask is applied. Therefore, we present the concentrations derived from the image analysis method in boundaries and do not show the measured concentrations itself. The upper boundary contains the error related to flare effects and it is close to the concentrations measured with the resistivity method. As the resistivity measurement cell (RMC) concentrations are within the boundaries of the optical method at point P2 we conclude that both methods yield comparable concentrations within the precision constraints of the optical method. From the experimental setup we know that the concentration should be 100 g/l at P2 (this is also confirmed by the numerical simulation of the experiment). The resistivity method reproduces the 100 g/l. Therefore, we consider the electrical method as more precise. The drawback is that the measurement radius is not known and therefore it is not useful to compare these measurements with numerical simulations in order to benchmark numerical codes. We do not understand the argument that at point P3 the mass underneath the curves should be equal. Since the measurement areas significantly differ between both approaches we would not expect the mass to be equal underneath the curves. For the revision of the paper we rephrased chapter 5 in order to make it clearer.

9. Comment:

Why is the median chosen to average the intensities in Section 3.3 and Figure 4? The change of the averaged value in Figure 4B is confusing. If the mean was used instead of the median, one would expect in Figure 4B to find always the same mean with decreasing error bars.

The median was used because it is not as much influenced by extremes as the mean. It was used for all analyses. We rephrased on pp. 4183, l.4 averaged values to medians.

10. Comment:

It would be useful to have the calibration and standard deviations in Figure 5 also for the other measurement points in order to have an impression how big the variability is. This could even be included in one figure.
This is included in the new version.
The technical comments are considered in the revised version of the manuscript.