"Assessing the effect of flood restoration on surface-subsurface interactions in Rohrschollen Island 'Upper Rhine River – France) using integrated hydrological modeling and thermal infrared imaging", by B. Jeannot et al.

Reply to the comments raised by Rev #2
The discussion reported hereafter remind us with the comments of the Reviewer typed in straight font as our answers appear below in italic. In the reviewing process of HESS, a reply should be sent before being allowed to propose a revised manuscript. This is why some of our answers might appear as declarations of intend to which we would then try to stick in writing the revised version of the manuscript.

Jeannot et al. in the manuscript "Assessing the effect of flood restoration on surface subsurface interactions in Rohrschollen Island (Upper Rhine River – France) using integrated hydrological modelling and thermal infrared imaging" evaluate the efficiency of restoration actions adopted in Rohrschollen Island, in the Upper Rhine River specifically in terms of surface-subsurface flow exchange in the hyporheic zone. The surface subsurface interaction is quantified with a fully-distributed hydrological model, the Normally Integrated Model (NIM). The exfiltration areas are of great interest for this study, as they represent the opportunity for hyporheic exchange enhancement. Using an innovative approach, the output of the validated model in terms of exfiltration is compared with the information derived from thermal infrared imaging.

- *We thank the Reviewer for considering our approach innovative and for emphasizing the great interest associated with fine-scale modeling of surface-subsurface water exchanges in the context of hydrosystem restorations*

General comments
I appreciated reading this manuscript, I found it very instructive and I think it is a valuable contribution to HESS. Authors address a key issue for ecological restoration and suggest an innovative framework to evaluate the impact of anthropogenic activities. I have some suggestions I would like to see addressed in a revised manuscript. My first observation concerns the model setup. I might have misunderstood, but the surface-subsurface flow interaction model seems to be not considering some input/output terms related to the surface processes, e.g., the precipitation and evapotranspiration. If they are somehow included in the model parameters this should be pointed out in the manuscript and, if not, this hypothesis should be clearly stated and justified.

- *We thank the Reviewer for his (her) positive and encouraging appraisal. Regarding the boundary conditions at the surface of the Island, it is right that we neglected rainfall infiltrations and evapotranspiration (ET) processes. As we mentioned in another reply (to comments of Rev #1), the main reason to this simplification comes from the volume of water injected from the river to the Island, even under continuous base flow injections of 2 m^3s^{-1}. This flow rate mainly infiltrates in the subsurface through the bed of the injection channel and the BGW stream network. This base injection of 2 m^3s^{-1} feeds the Island for a total amount of 6.3×10^7 m^3s^{-1} in a year, which would be equivalent to 15 m annual rainfall infiltration over the whole Island. Even if only 20% of the injected volume reached the subsurface, this would still correspond to 3 m rainfall… For the short-terms events considered in this study and the forced flow conditions due to intense river water injections, one can consider that rainfall and ET are negligible. We agree that this rapid evaluation of injected volumes is not given in the manuscript. It could* *help the reader to understand why rainfall and ET are neglected. We also agree that modeling long-term behavior, in the case of recessed river water feeding, should include hydro-meteorological forcing.*

Another point related to the model setup concerns the time step and the time horizon of the simulation. These aspects are not discussed at all in the presentation of the model setup. Being this a study of the effects of some management policies, the reader expects a long-term problem setup. Moreover, the time step and the time horizon might influence the results and this should be discussed in the manuscript. In other words, the short time horizon implies calibrating the model on one event only, opening a debate about the robustness of the calibration, although the performance in validation might be convincing.

- *It must be acknowledged that the topic of this study is limited to short-term responses of the hydrosystem and the impact of artificial flash flooding on stimulated flow interactions between surface and subsurface compartments of the system. Evaluating quantitatively these impacts imposes modeling exercises with highly resolved models in time and space. In the present case dealing with a flat riverine island, high resolution in space allows for accurate delineation of wetlands and ponds that trap water in the surface compartment, and for accurate locations of infiltration versus exfiltration areas. The short time step employed goes with the capability of grasping highly transient behaviors including inversion of flow according to the contrasted dynamics between surface and subsurface compartments of the hydrosystem. That kind of simulation cannot be performed over large periods of time and is not designed for that. It is also noteworthy that Rohrschollen Island is a landscape recently restored (as many other examples in the World) which is still rapidly evolving due to the impacts of restoration works on the geometry of the system, its geomorphology, its land cover, etc. Another way to say that is to point out that the system cannot be simulated for long-term predictions on the basis of its current state. This notwithstanding, short-term simulations (as proposed in this study) and for various scenarios inheriting from information on the system over various past periods may serve so establish some probabilities (or statistics) on the responses of the system. For example, we show with our short-term simulations that flash flooding the Island with a peak injection of 80* m^3s^{-1} *fosters exfiltration over at least the next 45 days. Duplicating that kind of calculations for different periods in a year could allow to define the number and the dates of peak injections for maintaining a prescribed exfiltration water volume over a year. This will not address the robustness of the model for long-term predictions, but at least this gives some meaning and interest to short-term simulations when long-term forecasts and evaluations are requested. Perhaps slightly trimming the Introduction at a few places (as suggested by another reviewer) and adding at the end of the Introduction a short paragraph written in the sense evoked above could help to show that modeling accurately short events is not necessarily in complete opposition with long-term considerations on the modeled system.*

My second observation concerns the sensitivity of the results of the hydrological model to the calibrated parameters, such as the hydraulic conductivity. The authors claim that the exfiltration areas are the result of rapid infiltration which produces an important increase in groundwater level. However, the exfiltration areas do not always coincide with the temperature anomalies observed in the infrared imaging. The authors point out possible causes of the mismatch but do not talk about the sensitivity of the model to the calibrated parameters. One could claim that the observed results might be due to an overestimation of the hydraulic conductivity in certain specific areas of the catchment. The discussion of this issue is required in the revised version of the manuscript.

- *Rev#2 is right when he suggests that there exist multiple reasons which could explain discrepancies between simulated exfiltration areas and temperature anomalies. We also agree that the model was not subjected to a complete sensitivity analysis which would assume, for robust results with non-linear processes, that the analytical differentiation of the state variables with respect to model parameters are calculated (the blunt method of perturbations is usually flawed or imprecise in these very transient and non-linear problems). Regarding over- or underestimations of hydraulic conductivity values, groundwater heads monitored over the Island are not very helpful to better condition conductivity, as they are strongly impacted by boundary conditions and water levels in streams…Nevertheless, the macroscopic hydraulic diffusion (the ratio of conductivity to specific storage) is correctly fitted as shown by the good match of observed heads both in time and amplitude. The point is that thermal anomalies are visible at a scale on the order of less than 10 m, which is also the scale of local heterogeneity of clay, sand, gravel, and pebble deposits in alluvial systems. A numerical model handling local heterogeneity at that scale should employ a mesh of 1-2 m sized elementary cells. In view of the available data, the mesh could not be assigned with a parameter value per cell, except in a stochastic framework generating various guesses on the parameter distributions and seeking the best ones. This is not the aim of this study which explores the possibility of simulating very transient behaviors of the hydrosystem and high contrasts of characteristic times associated with the flow processes. As is the case in many actual case studies, the model inherits from the few information available and is expected to render results at the large scale that roughly match up local data. Pinpoint accuracy is not expected. An additional paragraph could be added to better emphasize that :1- the measurement support of thermal anomalies is not that of the model , 2- flawed flow parameters could also explain discrepancies between simulations and observations, but at the scale of the Island, and given the hydraulic information available on the system, it is hard to obtain better results.*

Specific comments

L84-93: It would be interesting to have some examples of how increasing exchanges in the hyporheic zone contributes to the restoration projects.
L109: Wrong citation: "Fattichi" change into Fatichi.
L115: change computer into computational.
L124-126: Refer citations to the specific effect which is taken into account (water table dynamics, flood frequency, ecosystem services..).
L138-142: Better to express research objectives, possibly related to the discussion paragraphs: (1) model performance, (2) comparison between model results and TIR images and (3) comparison of different management options in terms of input quantity (and frequency, to add).

- *The five above comments could easily be accounted for in the revised manuscript by correcting typos, giving a few additional references, and rephrasing a few paragraphs.*

L296-299: The results from field experiments helped predefining the ranges of variations of some parameters of the model, but still some uncertainty exist on their calibration. It is worth commenting on the uncertainty coming from the calibration of the parameters of the model here and/or in the discussion section. Moreover, some parameters that I assume were calibrated (e.g., the Manning coefficients) are not mentioned in Section 2.2.2. Maybe a table with the starting range of variation of all the calibrated parameters and the calibrated value could be useful.

- *We agree with Rev#2 that a table and a few comments on it could be helpful.*

L277-305: Here the simulation time step and time horizon should be discussed, because the reader does not know which is the reference time scale, also because restoration processes are usually associated with long time scales.

- *As told above when answering to a general comment on short-terms versus long-terms expectations regarding restored systems, we could add our view on the topic at the end of the Introduction. Having said that, a short reminder in due place (e.g., at the very beginning of Section 2.2.2) on the objective of the modeling exercise, the use of a highly resolved model run over short simulation periods, and its eventual usefulness for long-term predictions, could clarify the reading.*

L315: "After a first simulation employing the initial parametrization (defined in Section 2.2.2)" see comments below concerning the explanation of the model parametrization. It should be more exhaustive.

- *We agree that the paragraph is not fully clear and could be rewritten by stating that the calibration was performed by trial and error on the two types of sensitive model parameters (found by prior exploratory calculations) that are the hydraulic conductivity and the exchange coefficient. For the sake of simplicity, these parameters are spatially distributed as uniform values over subareas of the domain (zonation). The initial delineation of these subareas is consistent with the spatial distribution of sedimentary facies in Fig.2. However, when calibrating the model, a few zones have been added, especially along clogged portions of the BGW, to match local hydraulic head data (the only reliable hydraulic information available). The other parameters were prescribed at common values reported in the literature for similar environments. As preliminary calculations showed the weak sensitivity of these parameters, they were not calibrated…*

L320-322: "Only the hydraulic conductivity and the exchange coefficient between surface and subsurface were slightly adjusted while trying to preserve the initial spatial zonation" Not clear statement. Which is the initial spatial zonation? How was it defined? Does it mean hydraulic conductivity and the exchange coefficient are the only manually-calibrated parameters? What about the other parameters of the model? How were they fixed?

- *See our reply to the previous comment. We think that a better rewriting of the calibration procedure would shed light on the questions raised by Rev#2.*

L341-360: The discussion is consistent, but Figure 5 and Figure 6 are switched, so Figure 5 refers to the validation and Figure 6 to the calibration. This Section lacks of comments on the impact of the simulation horizon. The calibration on a single event does impact the results of calibration and validation. Please, comment on this.

- *Regarding Figs 5 and 6, this is an error when uploading the Figs for submission. Regarding the calibration and validation on two single events, we emphasized above that the aim of our study was not building a model for robust long-term predictions. The focus is here put on the feasibility of simulations highly resolved in time and space, probably the best way to decipher how flash floods impact at a fine scale over time and space the hydraulic behavior of the system. This focus could be reminded here. One could also remind that for systems under restoration processes and that still evolve in* *their geometry, environment, and external forcing conditions, applying a current model (designed and calibrated on current information) over long-term horizons would miss that the system changed within the forecasting period.*

L354 and L357: It would be interesting to see the values of the three components of the KGE and their variation from calibration to validation.

- *This request was also pointed out by another reviewer and the three components could be given in the revised version.*

L371-377: Among this factors also the sensitivity to the model parameters should be pointed out. Somebody could claim that the observed dynamic might be due to an overestimation of the hydraulic conductivity implying higher infiltration and consequently exfiltration on a much larger area than the one where thermal anomalies are observed. Moreover, it is difficult to quantify the uncertainty related to the airborne TIR images, which were collected in a single survey. Maybe some comments on this uncertainty might help the reader to evaluate the results robustness.

- *As discussed earlier when dealing with the second main observation raised by Rev#2, model sensitivity to parameters is not straightforward to calculate in the present context. Having said that, discrepancies between simulated exfiltration areas and TIR images concern tiny spots that could be related to very local contrasts in hydraulic conductivity values. These small-scale contrasts cannot be mimicked in a model at the scale of the Island. Even though TIR images were carefully processed, one cannot overlook the fact that they represent a snapshot that could change within hours. We touch here the limitation of facing models with data that are not built (collected) on the same elementary support over time and space. In that case, we think that the robustness of the results is not in the fact that the model accurately represents data over a single scenario, but in the fact that the model roughly represents data over multiple different scenarios (events). Unfortunately, we only have at reach a single set of TIR. A comment in that sense could be added in the revised manuscript.*

L406-408: It could be helpful adding in Fig.10 also the injected flow reported in Fig.3.

- *This would probably overload the Fig, with a third scale for injected flow rates. But we could easily mark with bars along the time scale when the main injection peaks occur.*

L424: ".noting that the new channel was excavated in highly conductive sedimentary formations" this information comes from field experiments or from the parameters of the calibrated model?

- *This assertion comes from observations when excavating the channel mainly in pebble bars. This observation could be added.*

L464: "..but maintains increased areas of exfiltration over extensive periods" the definition "extensive periods" would have a more precise meaning if the problem in terms of time horizons was discussed while defining the boundaries of the problem. The observation is suitable also for "long periods" at L467. Section 3.4 is very interesting because it tests two different mitigation strategies in terms of input rate and volume, but under the title "Suggestions for management practices" some more information is expected, for example in terms of exchange frequency required over one year in order to observe ecological enhancement. Adding some information in this direction completes also the conclusions of Section 3.3, where the authors state "When forced injections enhance the development of wetlands and maintain high rates of exfiltration over long periods, from the mere hydrological standpoint, restoration works are successful", but how often does it happen? How often should it happen in order to enhance the ecological status of the environment?

- *We agree with that, the notion of "extensive periods" or 'long-periods" being here defined with reference to the complete period over which a flash flood impacts the hydrosystem (a short-term event, in other words). We think that after clearly stating the motivation our study in the Introduction, and after reminding what is targeted when presenting the model settings, any reader should clearly grasp that the study is focused on short-events. In the same way as we also discussed earlier in this reply, it appears interesting to show how simulations of short-events as samples of the long-term behavior of the system may serve to partly foresee this behavior. We could had in Section 3.4 our view on the "statistics" (see above) provided by duplicating short-term simulations with the aim of guiding practices in relation with the long-term management of the system (as the example proposed above defining the number of flood events needed to maintain a prescribed volume of exfiltration in a year).*

Figure 10: Add the pattern of the inflow. Right y-axes change m3/s into m3/s. If possible, it would be nice to have enhanced image quality of Figure 5, 6, 8, 9 and 11.

- *See above, bars along the time scale would provide the dates of the peak injections without overloading Fig. 10. Regarding Figs in general, the quality of the originals is far better than the low-resolution pdf format proposed by the Journal for reviews. The original quality seems to be enough for further eventual electronic publication.*