Tuesday, July 30, 2013

Dear Markus,

On behalf of my co-authors, I would like to thank you and the three anonymous reviewers for your many thoughtful comments. You have raised a number of important issues that I hope we can address satisfactorily in a revised version of the paper.

It is clear from the reviews that we did not do a sufficiently good job of describing the purpose and intended audience for PERSiST. We hope that providing a little background can help to explain why this paper has been written.

**Background**

First and foremost, the paper was written because PERSiST is being used. Research groups in Sweden, Finland, Norway and the UK are using PERSiST for water quality and climate change studies. As these studies are written up and submitted for publication, they will need to cite a model description paper. One of the benefits of the EGU open access model is that discussion papers are published and possible to cite though the authors hope that, with some revision, the article will also be acceptable for publication in HESS.

With regard to environmental modeling, we reject the “one model fits all” hypothesis. This has two consequences. First, we believe there is a need for flexible modeling frameworks such as SUPERFLEX (Fenicia et al. 2011) and PERSiST which are able to represent different perceptual models of catchment hydrology. Second, we believe that no single piece of rainfall runoff modeling software is suitable for all modeling tasks.

While we and our colleagues study catchments in a wide range of climatic conditions, we are unaware of a single model that is suitable for runoff simulation in Mediterranean, wet temperate and boreal conditions. We have found models suitable for runoff simulation in boreal and temperate conditions but have encountered challenges when simulating flow and nitrogen fluxes in Mediterranean catchments (Bernal et al. 2004). PERSiST has been designed as an easy to use, flexible modeling framework suitable for performing catchment-scale hydrological simulations in support of biogeochemical research based on long-term monitoring data for a range of climate types found in Europe.

We work with research groups and managers who are interested in the effects of changing climate and land management on the water quality of catchments in a range of climatic conditions. Typically, we rely on government monitoring agencies for the data used in our simulations. In practice, this means the data we have to hand are limited to daily measurements of flow, weekly or less frequent measurements of water chemistry and meteorological time series that are ideally but not always collected within the catchment of interest. One overarching research goal for us and our collaborators is how to best use long-term monitoring agency data for the evaluation of possible consequences of climate and land use change. To that end, PERSiST has been designed for use with a minimal amount of stream flow and meteorological data.

One important design goal for PERSIST is that the model must be easy to use. This meant that a lot of our effort has been focused on developing an intuitive Windows-based user interface. It has been our experience that other well-used and well-respected rainfall runoff models are not used as often as they might be because of poor user interfaces.

The decision to use a specific piece of software is conditioned by both the problem domain and the social environment of the modeler. Different software has different strengths. The appropriate tool for simulating groundwater flows in a well-instrumented research catchment may well be different from that for simulating snow dynamics in an alpine environment. In our experience, having access to someone who has some skill with a particular model makes it a lot more likely that model will be used. Put another way, it has been our experience that it is very difficult to learn to use any model properly without some sort of hands-on guidance and training. Thus, the most appropriate tool from a scientific standpoint may not be used because no one is available who has the necessary skills to use it, or train others in its use. PERSiST is being used by multiple research groups in Fennoscandia and the UK not because it is scientifically head and shoulders above other models but because it is easy to use, support is available, and is an appropriate tool for the task of generating catchment-scale hydrological simulations.

You and the reviewers were quite right to criticize the sensitivity analysis presented as being ill-posed. We believe that many tasks in environmental modeling are ill-posed inverse problems. O'Sullivan (1986) identifies an inverse problem as one where there the modeler is "… interested in recovering a whole function given a finite number of noisy measurements on functionals". To re-state O'Sullivan, we are interested in recovering stream flow and catchment moisture status from uncertain and incomplete measurements of stream flow, air temperature and precipitation. Runoff prediction is ill-posed insofar as there are no analytic solutions to the governing equations, unique solutions do not exist and the solutions can be excessively sensitive to small perturbations in the initial conditions of the data. Thus, we accept the concern that we have addressed an ill posed problem. In light of reviewer comments, we accept that we could have addressed a slightly less ill-posed question but we do not believe it is realistic to hope for well-posed solutions to most hydrological problems.

From our own experience and the published literature, it is quite clear that no single rainfall runoff model is ideal for use in all catchments. While Fenicia et al. (2011) have expressed concerns about allowing modeling objectives to control model structure, we do not believe these concerns are warranted. We are in agreement with Beven (2000) about the uniqueness of place. Philosophically, we are committed to a form of instrumentalism which states that the true values of parameters such as catchment moisture content cannot be measured, but reasonable approximations can be estimated. Furthermore, we hold that calibration is an important tool for the identification of effective parameter values and that effective parameters are characterized by a range of more or less plausible values. There is no justification for unique effective parameter values but model applications based on single parameter sets can still have merit. We are probably not as familiar with the PUB literature as we should be; much of our work has focused on the difficulties of working with gauged catchments and some of the challenges in extrapolating properties across apparently similar catchments (i.e. Oni et al. 2013).

**Model Description**

The goals of this first PERSiST paper are to (i) document the model, (ii) describe a simple Monte Carlo analysis tool and (iii) present a first application of the model and sensitivity tool. Follow-on papers from us and our collaborators use PERSiST as part of model chains for biogeochemical and climate change simulations. We suspect there is also an opportunity to explore the suitability of different perceptual representations of flow routing within catchments on runoff.

It is clear that we needed to do a better job of model description. Specifically, we should have stated that the model calculates rainfall from runoff by moving water between compartments in an arbitrary order. PERSiST operates sequentially. The reach network is represented as a directed tree graph, with the root at the catchment outlet. Stream flow is estimated in terminal (furthest upstream) reaches first. The graph is then traversed and stream flow estimated in all other reaches based on stream flow from upstream reaches and inputs from the local sub catchment. Within each sub catchment, runoff is estimates sequentially for each land cover type. For each day of simulation: (i) rainfall, snow accumulation and snowmelt are estimated based on measured precipitation; (ii) rainfall and snowmelt are routed through the uppermost buckets in a land cover type; (iii) ET from the uppermost bucket is estimated; (iv) outflow from the bucket is estimated. Steps (i)-(iv) are repeated for each bucket in the land cover type. We agree with Fenicia et al. (2011) that this approach is lacking in physical realism as water is routed sequentially, not simultaneously between possible stores.

We agree with the comments about the model application and sensitivity analysis. Specifically, we recognize the need for a different set of uncertainty analyses. Following the advice of reviewer 3, we are in the process of running three sensitivity analyses where we generate ensembles of plausible parameter sets based on calibration to (i) the bottom 10% of flows in each reach, (ii) the top 10% of flows in each reach and (iii) all measured flows. The details of the sensitivity analysis and preliminary results are described later.

**PERSiST and INCA**

It is clear that we did not do a good enough job explaining why PERSiST exists and how it differs from other models. The intended audiences for PERSiST are scientists and catchment managers who use the INCA family of models (Whitehead et al. 1998, Wade et al. 2002) to assess the potential effects of climate and land-management change on surface water quality. The INCA family of models has been widely used in Europe and North America to simulate inorganic nitrogen, phosphorus, sediment, organic carbon, mercury, etc. in soils and surface waters and to assess the potential effects of climate and land management on surface water quality.

PERSiST was written primarily (but not exclusively) to address challenges with the use of the INCA (Whitehead et al. 1998, Wade et al. 2002) family of models. There are two main challenges when using INCA. First, it relies on external inputs from a rainfall-runoff model. Second, the perceptual model of catchment hydrology in INCA is not well suited to catchments outside the temperate and boreal ecoregions (Bernal et al. 2004, Medici et al. 2008, 2010). We did not call the model "INCA-Hydrology" as we hoped to have a tool that was able to represent a range of perceptual models, and not just the one incorporated in INCA.

INCA relies on external time series of hydrologically effective rainfall (the fraction of precipitation which contributes to runoff) and soil moisture deficits (the difference between the current depth of water and the water holding capacity). In the past, these time series have been obtained from rainfall-runoff models including HBV, WSFS, MORECS and IHACRES. There are a number of conceptual and practical problems with the use of any of these models to generate time series inputs for INCA. First, the conceptual representation of water stores differs between INCA and the aforementioned models. While it is possible to obtain credible time series of hydrologically effective rainfall and soil moisture deficits using any of the current generation of lumped rainfall runoff models, it is not entirely satisfactory to use one perceptual models of the runoff generation process for hydrological estimation and another for water chemistry simulations.

Sequential calibration of water quality models in which hydrology is calibrated first and then biogeochemical parameters are adjusted has been shown to give less than ideal water quality simulations (McIntyre et al. 2005). This problem has also been noted by Tominaga et al. (WRR, under review), amongst others. A linked model based on INCA-Nitrogen and PERSiST is currently being tested for its ability to simulate water and inorganic nitrogen fluxes in temperate and Mediterranean catchments (Erlandsson et al. in prep.). Another linked model for dissolved and particulate phosphorus transport is also under development. Linked hydrological and biogeochemical models will facilitate the simultaneous calibration of water quantity and quality time series, and permit multi-objective calibration as the hydrological calibration can be constrained by the credibility of the water quality calibration.

**PERSiST and Other Models**

Current use rainfall runoff models may not be well suited to water quality modeling as they may fail to successfully simulate biogeochemically important summer low flow events. We have struggled with this problem in model applications in Canada and the UK. To the best of our knowledge, PERSiST is unique insofar as it simulates these hydrologically unimportant but biogeochemically vital low flow events.

Inundation, or the flooding of riparian areas, can be an important mechanism for redistribution of suspended material and potentially solutes. One of the design goals of PERSiST was to permit the simulation of inundation events. A detailed presentation of the inundation routines will be deferred to a future paper in which we present a simulation of catchment-scale sediment transport.

PERSiST has also been designed to simulate infiltration, or the movement of water from a stream to a surrounding riparian area. Infiltration appears to be an important control on surface runoff, especially in Mediterranean catchments.

At a practical level, PERSiST has been written to provide a greater control over hydrologic simulations than has previously been available to groups using the INCA model. Several groups in Sweden, Norway, Finland and the UK are already using PERSiST to generate time series of hydrologically effective rainfall and soil moisture deficits for INCA simulations. This is in part due to the ease of use of PERSiST, and the difficulties that may be encountered when trying to obtain hydrological data from national operational hydrology data centres.

We believe that PERSiST is easier to use and more flexible than HBV, IHACRES, MORECS or WSFS. This is not to suggest that PERSiST is better than any of these models, but it does fill a need for the groups we work with, the broader hydrological and water quality modeling community.

There are a large number of semi-distributed stream flow and solute transport models, many of which are widely used. With the exception of the SUPERFLEX (Fenicia et al. 2011) framework, we are not aware of any other modeling frameworks possessing the desired degree of flexibility in representing different perceptual models of the runoff generation process. However, Hrachowitz et al. (2013) suggest there are other modeling frameworks we need to explore.

We have had some exposure to the HYPE family of models (Arheimer et al. 2010, Lindström et al. 2010). In our opinion, HYPE is excellent for regional predictions and use in modeling systems where it is necessary to predict flows from a large number of catchments. PERSiST has been designed primarily for single catchment simulations. As such, it has a more friendly user interface than HYPE. HYPE also uses a fixed number of buckets for routing flow through a catchment.

PERSiST shares some characteristics of HBV-Light (Seibert and Vis. 2012) insofar as it has a graphical user interface which facilitates immediate feedback about the effect of parameter changes on simulated streamflow. Unlike HBV-Light, PERSiST is able to simulate stream flows at multiple points within a river network.

Because of its internal “steepest descent” flow routing algorithm, TAC-D (Uhlenbrook et al. 2004, Uhlenbrook and Sieber 2005) is probably more suited to mountainous catchments or other regions of high relief. We suspect it would be hard to obtain adequate results using TAC-D in the very flat catchments modeled by van der Velde et al. 2012. It would be informative to test PERSiST in catchments with very high and very low relief. We suspect that PERSiST might have difficulty in credibly simulating snowmelt from catchments with large elevational gradients or with flow simulation in extremely flat catchments with significant amounts of artificial drainage. We have data at our disposal to test PERSiST model performance in very flat and very steep catchments. These model applications might form the basis of future papers.

To the best of our knowledge, Hellebrand et al. (2011) have very different modeling goals than us. While they are interested in regionalization and prediction in ungauged basins, we only envision PERSiST being used in catchments where there are 2+ years of monitoring data.

We agree with you and the reviewers that solute time series can be used to constrain model parameterizations. Thank you for bringing Shaw et al. (2008) to our attention. They perform a similar study to one of our colleagues, who also used chloride as a semi-conservative tracer (Jin et al. 2011). We must caution against the uncritical use of chloride as a tracer. Chloride is not conservative (Svensson et al. 2012). This may not be problematic when there are significant chloride inputs, but can lead to inappropriate conclusions when chloride is in short supply.

Prior to receiving these reviews, we were not aware of the pesticide modeling conducted by Bertuzzo et al. (2012). There are a number of insights in their 2012 paper and earlier publications which we hope to incorporate into our pesticide modeling efforts.

**Reviewer 1**

1. From the abstract and introduction, the expectation is that the Author will address the problem of modeling simultaneously runoff and solute transport, presenting and applying an integrated model. However, from the presentation and application I only see a rainfall-runoff model, and no solute transport simulations are presented.

> Yes, this is true. We apologize for not having been sufficiently clear about the purpose of the model. PERSiST has been designed as a runoff simulator that can be used for solute transport simulations. The version of the model presented here only simulates runoff.

2. It is not clear to me what the real novelty of the paper is. The model is clearly different from others, but there are many models already available, and it is not clear why this particular model would be better than others, and in what respects.

3. In the Abstract, the Authors appear to suggest that their model would be an advancement with respect to the Superflex framework. As far as my understanding of Superflex goes, this framework would not be in contrast to distributed applications. In addition, I don't see in what the flexibility of Persist would consist.

As we have stated earlier, we do not believe that there is a single "one size fits all" model. None of us have used the SUPERFLEX model, but it is clear from our reading that SUPERFLEX is an important tool for detailed understanding of well-studied research catchments. We do not mean to suggest that our model is an advance over the SUPERFLEX framework. Our understanding is that SUPERFLEX and PERSiST have fundamentally different purposes. Our reading of the literature suggests that SUPERFLEX is primarily used in fundamental hydrological research. PERSiST has been developed as an alternative to operational hydrological models and for use with long-term monitoring agency data. Like SUPERFLEX, the flexibility in PERSiST consists of giving the modeler the ability to use different perceptual models of the runoff generation process ultimately needed for solute and particulate transport modelling.

4. Reducing hydrological modeling to the description of the relation of P=R-E is a bit reductive. Particularly since P is not equal to R-E, especially at the time scales the Authors are considering.

OK, we can remove that. However, we would argue that all successful modeling, whether in hydrology or other fields, is dependent on successful abstraction and reductive approaches. Furthermore, we would suggest that over the multi-year simulation period presented here, P is approximately equal to R-E as the long term change in catchment storage is approximately zero.

5. There is no mention about the numerical methods used to solve model equations, but from Equation 3 it is clear that the Authors are using the fixed time explicit Euler method. The Authors should be aware that this approach is highly inaccurate.

No, we are not using a fixed time Euler method and yes, we are aware that the fixed time explicit Euler method can be highly inaccurate. We are solving fluxes in a particular order using discrete equations. We are fully aware of the concerns about the using a discrete approach in which equations are solved in an arbitrary order (Fenicia et al. 2011).

We suspect that attempts to simultaneously solve all model equations as a system of ODEs could prove computationally challenging. In the model application presented here, we divide the Thames into 8 reaches. Each reach may have up to three landscape types and each landcover type contains three buckets. Each reach also contains a river segment. As each reach may receive different precipitation inputs, it is necessary to solve 8*3*3+8=80 equations.

We must caution that a perception of an appropriate ODE solver as a panacea for numerical problems can lead to difficulties of interpretation and communication. Most rainfall runoff models, ours included, represent snowmelt and accumulation. Because the model switches between snow accumulation and melt depending on air temperature, the function to be integrated becomes discontinuous. This can lead to challenges in the numerical solution (Hairer et al. 2009 pp. 196-200) which do not appear to be adequately appreciated in the hydrological modeling community.

Similarly, the question of how to calibrate models based on simultaneous solution of sets of ODEs must be approached with care. There is no guarantee that flow at an arbitrary time will be exactly the same as mean daily flow (for a graphic example of this, see Baggaley et al., 2009). Ideally, calibration would not be based on a single point in time arbitrarily sampled from ODE output but would aggregate outputs so as to be compatible with the time step of the flow observations used for model calibration.

6. The MCMC is basically a standard Metropolis Hastings algorithm, used to sample the Nash and Sutcliffe objective function surface. I do not understand the claims done in the discussion, in the paragraph starting with the sentence "There are a number of different schools of thought about Monte Carlo analysis. . ." . I don't think the Author are inventing or proposing anything new here.

We agree; we are basically using a standard Metropolis Hastings (MH) algorithm. As such, we are not presenting anything new. What we suggest is new is that we have used an ensemble of MH chains to identify plausible parameter sets. We suggest that our approach has more in common with the Horberger/Spear/Young identification of behavioural parameter sets than with the literature on MCMC methods in hydrology. The novelty in our method is that we use an ensemble of Metropolis Hastings algorithms to generate a set of plausible parameter sets. Note that "plausible" is not the same as "optimal". While we believe there is a single, optimal parameter set with which the goodness of fit between observed and modeled data is maximal, we do not see any point in devoting a lot of effort to a search for this. Beven (2006) inter alia is very clear regarding the epistemological shortcomings of searching for a single "optimal" parameter set.

To address the last sentence of the preceding reviewer comment, we agree with Beven et al. (2008) that "At the current time it is difficult to make a reasoned choice between methods of uncertainty estimation for real applications because of a lack of understanding of the real information content of data in conditioning models". We see strengths and weaknesses to all existing tools, including our own, for uncertainty or sensitivity analysis. We merely wanted to recognize the existence of the debate in the literature.

7. Line 13 of page 8652. The upper and lower limit of the objective function should be plus infinite and zero. Please correct.

Actually no, we maximized the objective function (p 8653, l10-11). This function was the sum of one minus the Nash Sutcliffe (NS) statistic for each of the of the eight streamflow time series ($\Sigma$(1-NS)). As the NS statistic ranges from a minimum of minus infinity to a maximum of one, one minus NS will have a lower value of minus infinity and a maximum of zero.

8. The Authors write that the Persist framework has been designed to be as simple as possible but no simpler. There is no evidence supporting this statement.

This is a little difficult to respond to as we need to present negative evidence. Using an ad-hoc approach as opposed to solving a series of ODEs (using an appropriate solver) is one example of design simplicity. We are aware that this may lead to tradeoffs, and would be happy to discuss this further with the reviewer. We recognize that transit times between buckets may be important (see Fenicia et al. 2011). We realize this level of complexity is appropriate for the phenomena simulated by SUPERFLEX but do not have sufficient reason to include it in PERSiST.

9. The Authors are trying to calibrate simultaneously 108 parameters with a single objective function. I think the calibration problem may be ill-posed, and alternative strategies should be investigated.

We agree that the calibration strategy is ill posed and are investigating alternate calibration strategies but feel we must clarify some points. First of all, we are not using a single objective function. In the submitted version, we used 8 individual NS statistics. The acceptance test in the MCMC is based on the sum of these 8 individual values. We would suggest that 8 objective functions are being used to make one decision about whether to accept or reject a candidate parameter set.

The modeling exercise presented here can be thought of as an inverse problem in which inferences are made about a phenomenon from partial or incomplete information (O'Sullivan 1986). We would suggest that most inverse problems in environmental modeling are not well posed. We agree with O'Sullivan (1986) who states "In an ill posed inverse problem, a classical least squares … solution may not be uniquely defined" and with Beven (2006), who has done so much to raise awareness in the hydrological community about the problems of equifinality.

**Reviewer 2 General comments**:

1) The paper claims that PERSiST would be very flexible and able to simulate solute transport. However, the model presented is a pure semi-distributed rainfall- runoff model. There are other semi-distributed models, e.g. SWAT, that probably have similar flexibility and they are additionally designed to simulate solute transport. Therefore, the innovation of this paper remains unclear.

   As we have stated earlier, we do not believe that there is a single best "one size fits all" model. SWAT has clearly proven its usefulness in simulating water, solute and sediment transport in a wide range of catchments. All of us have a background with the INCA family of models (Whitehead et a. 1998; Wade et al. 2002) and very limited familiarity with SWAT. To the best of our knowledge, SWAT lacks many of the features of PERSiST. Specifically, we do not believe that SWAT gives a modeler any flexibility in the perceptual model through the assembly and connection of an arbitrary number of buckets to simulate land phase hydrology. Our understanding is that SWAT hydrology is based on runoff curves. While it was not extensively discussed in the model application presented here, PERSiST is able to simulate inundation when a river overflows its banks, and infiltration from a river to riparian areas. There is some evidence that SWAT has this ability, also.

2) Bucket type of models, as applied for PERSiST, have limited capability to simulate transit times of water, which in turn is important for rate-limited, reactive transport. Characteristic time constants as presented in Figure 8b are likely much too short. This is suggested from studies by Howden et al. on nitrate transport in the Thames River basin.

   We agree with Howden et al (2005) (and with our earlier work) that groundwater in the Thames catchment is characterized by very long transit times, potentially on the order of 100 years or more.

Following Soulsby et al.  (2009), we define transit time is a measure of the time elapsed between a water molecule entering and leaving a catchment, and mean transit time (MTT) as the total storage divided by the flux of water, Thus, at steady state MTT will be equal to the depth of water in a bucket (L) divided by runoff from the bucket (L/T)

The characteristic time constant in PERSiST does not immediately provide insight into transit times, but instead describes the behavior of the falling limb of the hydrograph. In PERSiST, the transit time distribution is determined from characteristic time constants, storage volumes and the rate at which water enters the system.  Storage volume is related to the depth of water in a bucket. In PERSiST, the water in a bucket can be split into freely draining and retained fractions depending on water depth. It is assumed that the all water in a bucket is well mixed, thus water molecules in both the freely draining and retained fractions will eventually leave the bucket. The characteristic time constant only applies to water in the freely draining fraction.

With PERSiST, it is possible to simulate arbitrarily long MTT by using large values for the depth of retained water. For example, if 500 mm water /year enters and leaves a bucket, a MTT of 100 years can be simulated if the retained depth in PERSiST is set to 50 m.

**Reviewer 3 General comments:**

The title and the abstract imply that the paper presents a flexible and directly coupled flow-solute model. However, the reader only discovers quite a bit later that the paper essentially deals with a flexible rainfall-runoff model framework ultimately developed as a delivery model to biogeochemical models such as INCA. Flexible model frameworks are an approach forward to testing hypotheses about catchment functioning and the dominant hydrological processes governing the hydrograph, but it is not clear what the differences, advantages, disadvantages and novelty of the model framework presented are compared to other recent flexible model frameworks such as FUSE, FLEX, SUPERFLEX and most recently DYNAMIT (which incorporated solute transport). I therefore suggest that it should be clearly stated throughout the manuscript that a flexible rainfall-runoff model framework is presented using a more thorough discussion in the light of the wider literature.

Yes, we agree with these concerns. We hope the earlier parts of this letter satisfactorily explain the manner in which we could address them.

Unfortunately, we have been unable to find published literature on DYNAMIT. Any references or suggestions would be greatly appreciated.

- The model calibration strategy seems flawed, which needs at least clarification and/or new analysis. First, the model is calibrated using only the Nash-Sutcliffe performance measure (see specific comments below). I strongly suggest considering using multi-objective calibration especially if peak and low flows are of interest for posterior solute transport modelling. Second, the total number of calibrated parameters is rather large (108) which necessarily results in parameters that cannot be uniquely identified during calibration. Third, the flexible model structure selection procedure is not clear to me. I suggest to clearly show which are the potential model structures, which structure was ultimately selected during calibration and how was this best performing (or most parsimonious?) model structure selected.

Thank you for pointing out the flaws in the model calibration strategy. We are currently engaged in a new round of model calibrations which we hope address some of the shortcomings of that presented in the submitted version of the paper.

We have re-thought our calibration strategy. First of all, we have reduced the number of land cover types from twelve to three. This resulted in a reduction from 108 to 45 parameters that were allowed to vary during calibration. We realize that 45 parameters are still a lot, and will not lead to a unique solution. There are 7 parameters for each land cover type (time constants for the quick, soil water and groundwater buckets; Degree Day ET, ET threshold temperature and, ET drought modifier and soil water drought runoff fraction), giving a total of 21 terrestrial parameters. There are 3 parameters for each reach/subcatchments representative of the “a” and “b” parameters in the flow velocity equation and a subcatchments specific rainfall multiplier for a total of 24 parameters.

- There are some inconsistencies in terms of the paper structure. For example, in the method section there is no mentioning of the sensitivity analysis, but a table (6) and three figures (8a to c) are dedicated to show results from a sensitivity analysis. Further to this, the result section is very brief and could be more elaborate or merged with the discussion section altogether.

Yes, we need to re-structure the paper. The methods section should include a description of the PERSiST model and the procedure used for the sensitivity analysis. Furthermore, it has become clear that we will need to expand the site description.

**Reviewer 3 Specific comments**:

Page 8637, Line 1-3: I think only Hrachowitz et al simulated water and solute transport and with good results.

We need to re-phrase this section. You are quite correct to point out that there is a considerable literature showing successful simulation of water and solute transport with a plethora of models in a wide range of catchments.

Page 8639, Line 9: Is the only difference to SUPERFLEX a semi-distributed catchment representation?

No, there are a number of differences between PERSiST and SUPERFLEX, however, the possibility of semi-distributed catchment representations may not be one of them. The SUPERFLEX literature and the comments of reviewer 1 seem to imply that SUPERFLEX can be used for representing catchments in a semi-distributed manner.

From our reading of the SUPERFLEX literature, it appears that there are a number of differences between the two models. SUPERFLEX uses a more sophisticated flow routing algorithm incorporating numerous routing functions. It appears that considerable effort has been expended in identifying and testing appropriate ODE solvers in SUPERFLEX.

Page 8647, Line 18-20: Are those fractions fixed? In this case based on what criteria?

No, these fractions are not fixed but they, and the number of buckets to use when transforming precipitation to runoff must be specified by the model user based on his or her perceptual model of the catchment. Published estimates of the baseflow index can provide guidance as to appropriate fractions but the values could be better constrained based on measured differences in water quality between buckets. Simultaneously varying the fractions and characteristic time constants can lead to significant equifinality. Routing fractions, characteristic time constants and retained water depths can all be better constrained with the aid of water chemistry data. Elrandsson et al. (in prep) are exploring the use of a linked version of INCA-N and PERSiST to simulate inorganic nitrogen dynamics in the Thames basin and are showing the importance of soil solution chemistry measurements in constraining hydrological parameters.

Page 8652, Line 15: The NSE performance measure is biased towards peak flows, but low flows as you state on page 8640, line 2 are important for biogeochemical processes. I invite the authors to consider using a multi-objective calibration strategy evaluating more aspects of the hydrograph other than peak flows.

Thank you, this is a good idea. We are in the process of running three sets of calibrations using the top 10%, bottom 10% and all observed flows using a simplified representation of land cover types and a reduced number of variable parameters.

Page 8652, Line 23: Was really a total of 108 parameters simultaneously calibrated? I strongly suggest revision of this ill-posed calibration strategy.

Yes, 108 parameters were calibrated simultaneously. Please see above for a discussion of the revised calibration strategy.

Page 8653, Line 2-5: Please clarify this sentence.

As noted by one of the reviewers, a simple Metropolis Hastings algorithm was used in the Monte Carlo analysis. We chose not to use a Gibbs sampler as it seemed this would introduce unnecessary complexity into the analysis.

Page 8658, Line 26-29: These sentences seem identical to page 8638.

Yes, they will be removed or reworded if we receive editorial permission to submit a revised version.

Figure 6a and 7a: I suggest representing the maximum and minimum simulations as bands to improve visibility

Yes, we will do this.

**Editor Comments**

(1) It is not clear what the actual working hypothesis of this paper is or what the authors are trying to convey. And how does solute transport play a role here as suggested in the title?

From reading your and the other reviewers' comments, we realize this is a shortcoming in our original submission. We hope the discussion of the model purpose at the start of this document adequately clarifies this point.

(2) In how far is PERSiST new, novel, innovative, different to or better than any other conceptual model approach? In principal, every conceptual model can be (and many actually were) applied in a (semi-) distributed way (e.g. Uhlenbrook et al., 2004; Lindstroem et al., 2010; Hellebrand et al., 2011).

We would suggest that PERSiST is an incremental improvement to existing models. We did not mean to suggest that other conceptual models have not or cannot be used in a semi-distributed manner. If we were aware of another model which met our criteria, we would have used it instead of embarking on the long, arduous process of developing, testing and publishing a new model.

We believe PERSiST is unique insofar as it the only model we are aware of which provides all of the following: (i) a flexible model structure, (ii) an easy to use graphical interface, (iii) the ability to simulate inundation and infiltration, (iv) ability to calibrate to multiple flow time series

(3) As also pointed out by the reviewers, neither proper credit is given to existing integrated rainfall-runoff/solute transport applications nor are these discussed in context with the presented method. It is important not only to refer to earlier models but also to highlight and discuss in detail the differences (advantages/disadvantages) of these previous models to PERSiST. In particular, the statement on

P.8637, l.2-3 ("[...] they have not necessarily been well suited to simulating solute transport") is surprising as many previous models demonstrated good ability in reproducing well streamflow AND solute transport. Therefore, the manuscript needs to be put much more in context of earlier work and I would encourage the authors to consider at least the following references for detailed discussion:

Uhlenbrook and Sieber, 2005; Dunn et al., 2007; Shaw et al., 2008; Fenicia et al.,2010; Birkel et al., 2011; Arheimer et al., 2012; McMillan et al., 2012; Van der Velde et al., 2012; Hrachowitz et al., 2013a; Bertuzzo et al., 2013

We did not mean to suggest that all previous models have been poorly suited to simulation of streamflow and solute transport. We are familiar with some of the work of the Aberdeen group (Dunn et al. 2007; Birkel et al. 2011. McMillan et al. 2012). The relationships between PERSiST and the models of Arheimer et al. 2012, Bertuzzo et al. 2013 and Van der Velde et al. 2012 are discussed earlier. We realize very clearly the need to do a better job of putting our work in context, and giving appropriate credit to existing models.

(4) In the light of considerable progress over the past decade and as pointed out by the reviewers, the calibration strategy is too simplistic to adequately constrain a model with 108(!!) parameters (see for example the review given by Hrachowitz et al., 2013).

Yes, the calibration strategy was too simplistic, and thanks for bringing Hrachowitz et al. (2013) to our attention. There are a number of points raised in that paper that are very relevant for our paper. First of all, we hope our model is compatible with the sentiments expressed in section 3.2.1 of your paper. We do not propose PERSiST as a universal solution to the rainfall runoff modeling problem. We do believe, however, that there is an unfilled niche in the rainfall runoff model ecosystem and that PERSiST is a useful tool for addressing some rainfall runoff modeling problems, specifically those having to do with the simulation of water quality time series derived from long term monitoring data.

We believe the calibration strategy proposed by reviewer 3 is closer to that espoused in Hrachowitz et al. (2013). The uncertainty analysis is incomplete (typically, it takes 2-3 weeks to perform all the model runs) but some preliminary conclusions can be drawn. First, the reduced parameter set analysis performs approximately as well as the original 108 parameter calibration in overall fitting of modeled and observed flows. Second, objective functions based on squared distances appeared to be more amenable to improvement using the Monte Carlo algorithm presented here than objective functions based on absolute deviations. Third, the PERSiST model appeared to perform better when simulating the top 10% of flows than for the bottom 10%. We would argue that this apparent discrepancy in model fit is due to differences in the adequacy of the perceptual model used here for simulating high and low flows. At high flows, it appears that the perceptual model is adequate, and that flow in the Thames is largely controlled by the manner in which rainfall is routed through the surrounding catchment and upstream reaches. At low flows, this relationship breaks down. We believe this is an important insight into the adequacy of our perceptual model. At high flows, our perceptual model which states that flow can be simulated as a function of precipitation routed through three different landcover types may well be appropriate. At low flows, it appears that other factors are more important than just routing of rainfall through the catchment. Human actions and complex geology may have a significant effect during periods of low flow. There are 45 locks on the Thames which are used for flow regulation during periods of low flow, and significant abstractions for drinking water. We must be clear that we do not believe this finding invalidates our approach but instead shows the value of calibration for gaining process understanding in hydrology.

Using a certain calibration algorithm, you will surely enough find some mathematically feasible posterior distribution of feasible parameters. However, the parameterization will be subject to significant equifinalty. How do you make sure that your mathematically feasible parameterizations are also *realistic* parameterizations (cf. Beven, 2006; Gupta et al., 2008, Andreassian et al., 2012; Gharari et al., 2013). In line with the reviewers I would therefore also suggest to review your calibration strategy and consider more calibration objectives and/or criteria to constrain your model, thereby increasing the confidence in your models to represent reality in the best possible way.

We must disagree that a credible calibration algorithm will always identify a mathematically feasible posterior distribution of feasible parameters. Our experience suggests that a model with significant structural flaws cannot be calibrated successfully (i.e. NS >0), no matter what range of parameter values is used. We would furthermore suggest that it is not possible to calibrate a rainfall runoff model when the rainfall time series does not match runoff. We have performed an experiment to test this hypothesis in which a modeler was assigned four catchments to simulate. For three of the catchments, the precipitation time series was obtained from the same location as the flow measurements. For the fourth catchment, a different precipitation time series was used. The modeler was able to obtain mathematically feasible posterior distributions for the three instances where rainfall and runoff were observed at the same site but not at the fourth site where rainfall was obtained at a different location.

We agree that the use of 12 land cover types was not the best possible way of representing the reality of the Thames catchment for flow simulations. Such a complex land cover representation may be appropriate for solute (nitrogen or phosphorus) simulations but is clearly not justified for flow simulations. Preliminary results from a sensitivity analysis using only three land cover types representative of the different underlying geology (chalk, gravel and sedimentary bedrock) suggest that similar NS statistics can be obtained with the simplified land cover type representation as were obtained in the simulations presented in the submitted version of the paper.

Andreassian et al. (2012) was an enjoyable paper to read, but it appears they still hold out hope for mathematically optimal parameter sets. As we have explained earlier, we are not convinced of the value of searching for a mathematical optimum which minimizes the distance between uncertain measured values and model outputs. If we can pursue their gold mining analogy, we would like to share a few thoughts with you. One of us used to have a number of friends who made their living from prospecting for gold. Panning for gold is an important tool for prospectors. While it looks easy, knowing where and how to pan for gold takes a lot of skill, and could perhaps be seen as a metaphor for expert judgment applied to manual calibrations. There is a very easy test to distinguish fool's from real gold. Fool's gold (or iron pyrite) has a cubic crystal structure whereas gold is amorphous. If you hold the rock up to a bright light and rotate it, fool's gold will appear to change shape because of the changing reflection from its cubic crystal structure, whereas gold will not do so. I suppose this could be a metaphor for the need for hydrological knowledge when evaluating model parameterizations. Panning for gold and searching through piles of rock for visible gold may not be an optimal way to get rich. Gold is valuable enough that rocks with no visible evidence of gold can still be worth mining. Finding these rocks is dependent not on identifying one single optimal location, but on the careful study of a range of geological evidence to identify plausible locations that may contain gold. We would like to suggest that our strategy of identifying ensembles of plausible parameter sets is akin to this latter method. We may not have gold nuggets, but have gained additional insight into catchment function.

I would encourage the authors to use the opportunity of the rapid delivery of reviewer comments to actually engage in direct discussion with the reviewers during the remaining Open Discussion Period on HESSD over the issues they raised, potentially clarifying many of the earlier concerns. In any case detailed responses to the reviewer comments need to be submitted before the end of the Discussion Period. Note that a revised version of the manuscript with detailed description of changes should NOT be uploaded at this point, but only after the Editor Decision is available after the end of the Discussion Period.

We look forward to on-going discussions with you and the other reviewers. It is clear that this and any subsequent rounds of discussions will give us the opportunity to improve our paper. Due to the large number of changes requested by the reviewers and our schedules, we would like to request an extension of the Open Discussion Period to the end of September. If this is feasible, it will give us sufficient time to adequately incorporate all your insights into a revised manuscript. Once again, we would like to thank you and the three anonymous referees for the effort you have invested in reviewing our manuscript.

Sincerely,

Martyn Futter

References

Andreassian, V., Le Moine, N., Perrin, C., Ramos, M.-H., Oudin, L., Mathevet, T., Lerat, J., and Berthet, L.: All that glitters is not gold: the case of calibrating hydrological models, Hydrol. Process., 26, 2206–2210, oi:10.1002/hyp.9264, 2012.

Arheimer, B., Dahne, J., Donnelly, C., Lindstrom, G., and Stromqvist, J.: Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input data quality and scale, Hydrol. Res., 43, 315–329, 2012.

Baggaley, N. J., S. J. Langan, M. N. Futter, J. M. Potts, and S. M. Dunn. "Long-term trends in hydro-climatology of a major Scottish mountain river."*Science of the Total Environment* 407, no. 16 (2009): 4633-4641.

Bernal, S., A. Butturini, J. L. Riera, E. Vázquez, and F. Sabater. "Calibration of the INCA model in a Mediterranean forested catchment: the effect of hydrological inter-annual variability in an intermittent stream." *Hydrology and Earth System Sciences Discussions* 8, no. 4 (2004): 729-741.

Bertuzzo, E., Thomet, M., Botter, G., and Rinaldo, A.: Catchment scale herbicides transport: theory and application, Adv. Water Res., 52, 232–242, doi:10.1016/j.advwatres.2012.11.007, 2013.

Beven, K.J. "Uniqueness of place and process representations in hydrological modelling." *Hydrology and Earth System Sciences Discussions* 4, no. 2 (2000): 203-213.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.

Beven, .K.J., P.J. Smith, and J.E. Freer. "So just why would a modeller choose to be incoherent?." *Journal of hydrology* 354, no. 1 (2008): 15-32.

Birkel, C., Tetzlaff, D., Dunn, S. M., and Soulsby, C.: Using lumped conceptual rainfall-runoff models to simulate daily isotope variability with fractionation in a nested mesoscale catchment, Adv. Water Resour., 34, 383–394, 2011.

Dunn, S.M.,McDonnell, J.J., and Vache, K.B., 2007. Factors influencing the residence time catchment waters: a virtual experiment approach. Water Resources Research, 43, W06408.

Fenicia, F., D. Kavetski, and H.G. Savenije. "Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development." *Water Resources Research* 47, no. 11 (2011): W11510.

Hairer, E., S.P. Nørsett, and G. Wanner, Solving Ordinary Differential Equations I: Nonstiff Problems, 2nd ed. (Springer, Berlin, 2009).

Hellebrand, H., et al., 2011. A process proof test for model concepts: modelling the meso-scale. Physics and Chemistry of the Earth, 36, 42–53.

Hrachowitz, M., Savenije, H.H.G., Bloschl, G., McDonnell, J.J., Sivapalan, M.,Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D.,Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., and Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB) A review. Hydrological Sciences Journal, 58 (6), 1–58, doi: 10.1080/02626667.2013.803183, 2013.

Howden, N. J. K., T. P. Burt, F. Worrall, M. J. Whelan, and M. Bieroza. "Nitrate concentrations and fluxes in the River Thames over 140 years (1868–2008): are increases irreversible?." *Hydrological Processes* 24, no. 18 (2010): 2657-2662.

Jin, Li, P.G. Whitehead, D.I. Siegel, and S. Findlay. "Salting our landscape: An integrated catchment model using readily accessible data to assess emerging road salt contamination to streams. " *Environmental Pollution*159, no. 5 (2011): 1257-1265.

Lindstrom, G., et al., 2010. Development and test of the HYPE (Hydrological Predictions for the Environment) model water quality model for different spatial scales. Hydrology Research, 41, 295–319.

McGuire, K.J., and J.J. McDonnell. "A review and evaluation of catchment transit time modeling." *Journal of Hydrology* 330, no. 3 (2006): 543-563.

McMillan, H. K., Tetzlaff, D., Clark, M., and Soulsby, C.: Do time variable tracers aid the evaluation of hydrological model structure? A multimodel approach, Water Resour. Res., 48, W05501, doi:10.1029/2011WR011688, 2012.

Medici, C. S. Bernal, A. Butturini, F. Sabater, M. Martin, A.J. Wade, F. Francés "Modelling the inorganic nitrogen behaviour in a small Mediterranean forested catchment, Fuirosos (Catalonia)" Hydrol. Earth Syst. Sci., 14 (2010), pp. 1–15

McIntyre, N., B. Jackson, A. J. Wade, D. Butterfield, and H. S. Wheater. "Sensitivity analysis of a catchment-scale nitrogen model." *Journal of Hydrology*315, no. 1 (2005): 71-92.

O'Sullivan, F. "A statistical perspective on ill-posed inverse problems."*Statistical science* (1986): 502-518.

Oni, S.K., M.N. Futter, L.A. Molot, and P.J. Dillon. "Adjacent catchments with similar patterns of land use and climate have markedly different dissolved organic carbon concentration and runoff dynamics."*Hydrological Processes* (2013).

Seibert, J., and M. J. P. Vis. "Teaching hydrological modeling with a user-friendly catchment-runoff-model software package." *Hydrology and Earth System Sciences* 16, no. 9 (2012): 3315-3325.

Shaw, S., Harpold, A. A., Taylor, J. C., andWalter, M. T.: Investigating a high resolution, stream chloride time series from the Biscuit Brook catchment, Catskills, NY, J. Hydrol.,348, 245–256, 2008.

Soulsby, C., D. Tetzlaff, and M. Hrachowitz. "Tracers and transit times: windows for viewing catchment scale storage?." *Hydrological Processes* 23, no. 24 (2009): 3503-3507.

Svensson, T., G.M. Lovett, and G.E. Likens. "Is chloride a conservative ion in forest ecosystems?." *Biogeochemistry* 107, no. 1-3 (2012): 125-134.

Uhlenbrook, S., Roser, S., and Tilch, N., 2004. Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. Journal of Hydrology, 291, 278–296.

Uhlenbrook, S. and Sieber, A.: On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model, Environ. Model. Softw., 20,19–32, 2005.

Van der Velde, Y., Torfs, P. J. J. F., van der Zee, S. E. A. T. M., and Uijlenhoet, R.: Quantifying catchment-scale mixing and its effects on time-varying travel time distributions, Water Resour. Res., 48, W06536, doi:10.1029/2011WR011310, 2012. Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 8635, 2013.

Wade, A. J., P. Durand, V. Beaujouan, W. W. Wessel, K. J. Raat, P. G. Whitehead, D. Butterfield, K. Rankinen, and A. Lepisto. "A nitrogen model for European catchments: INCA, new model structure and equations." *Hydrology and Earth System Sciences* 6, no. 3 (2002): 559-582.

Whitehead, P. G., E. J. Wilson, and D. Butterfield. "A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I—model structure and process equations." *Science of the total environment* 210 (1998): 547-558.’