Author Response to Editor Decision: Reconsider after major revisions (further review by editor and referees) (18 Feb 2018) by Harrie-Jan Hendricks Franssen

Comments to the Author:

Dear Dr Davison,

Your manuscript “Parameter-state ensemble data assimilation using Approximate Bayesian Computing for short-term hydrological prediction” has been subjected now to review by three reviewers. Two of them recommended major revision and one of them minor revision. I think the paper can be published after major revision including additional review. However, given the reviewer comments, rejection of the manuscript is still likely if the concerns of the reviewers are not resolved.

The main points to be handled are:

1. Clarification of the methodology at several points, as indicated by the reviewers (especially reviewer #3). Advantages compared to other methods should be clarified. This means that the discussion of the methodology should be placed in a broader context.

Author response: The methodology has been clarified at several points. Broader context has been provided in the introduction (p2, lines 6 – 13), section 2.1 of the methodology (p3, lines 12 – 25), and throughout section 2.7 explaining the ensemble selection methodologies (p6 – 9, and the additions of Figures 1, 2, 3 and 5).

2. The use of a 3-day windows only to select parameters should be justified. This is typically not enough to characterize both base flow and peak flows. The selected example is therefore unfortunate.

Author response: The analysis has been expanded to also examine 10, 20, 30 and 40 day windows.

3. A better justification of the selected ten parameters is needed. This could for example be done by providing additional sensitivity analysis.

Author response: The analysis has been expanded to include a selection of 5, 10, 20, 30, 40 and 50 parameters based on the lowest RMSE values.

In your answer to the main points and detailed comments, please indicate how comments have been handled exactly, indicating also whether text has been deleted and what the position of newly included text blocks is. I am looking forward to the new version of the paper.

Author response: The changes are too numerous to include here and a document with changes tracked has also been uploaded.

hess-2017-482: Authors' responses to Reviewer Comments for "Parameter-state ensemble data assimilation using Approximate Bayesian Computing for short-term hydrological prediction"

Responses to RC1

Thank you for the positive "overall recommendation." In answer to your major and minor concerns:

Major Concerns

1) I had to go back to the methods sections several times to interpret what methods were covered by "P-SEDA" and how the method was applied. Algorithm 1 is extremely helpful, but I would suggest a second algorithm or a flow chart that is specific to the P-SEDA method applied, and perhaps specific to the 3-day moving window application presented in the manuscript. The manuscript should be more explicit about how P-SEDA is different from Algorithm 1.

Author Response: A second algorithm is added in section 2.7 to illustrate how the P-SEDA method is applied and how this is different than Algorithm 1. New Figures (1, 2, 3 and 5) have also been added to illustrate the algorithm and its implementation.

2) I cannot tell from the description in the methods section if the filters are applied sequentially or if they always draw from the same set of 10,000 simulations. This is stated more clearly in the conclusions. Because the majority of particle filters are applied sequentially, it should be clear early on that this is not the case in this paper.

Author Response: Some additional context has been provided in the introduction, section 2.7 and conclusions that describe the approach as a hybrid of the particle filter and variational DA methods.

3) I would not expect three days of streamflow to be enough to determine reasonable streamflow parameters. The parameters that produce good baseflow are rarely the same parameters that produce good flood peaks. Please provide more justification for testing this method

Author Response: This is an excellent point. The analysis has been expanded to include longer filter periods.

Minor concerns

1) Tables 4 5, lines 10-25, p. 13, and Figure 8, Conclusions: terminology suddenly changes. "Projection methods" were never defined; previously the four P-SEDA methods were all referred to as "filters". It is unclear which is the "3-day filter" and which are projections. I assume they correspond to the previously defined filters as follows, but I am not certain:
a. 3-day filter = minimized uncertainty filter
b. 3-day projection = preceding streamflow filter
c. bulk projection = bulk calibration filter
d. 3-day projection with constrained parameters = parameter and preceding streamflow filter.

Author Response: The terminology has been changed to explain the ensemble selection methodologies more clearly. The minimized uncertainty filter is now called the “optimal hind-cast,” the bulk projection has been removed, and the parameter and preceding streamflow filter has been renamed the “hindsight parameter constraint and preceding 3-day streamflow filter.”

2) Line 21, p. 6: Instead of saying “ensemble data assimilation filters”, say “P-SEDA” filters. Otherwise, it is not clear that P-SEDA encompasses all 4 filter approaches.

Author Response: The change in text has been made as suggested.

3) Lines 30-31, p. 13: I assume that you do not use the bulk calibration filter here because it performed poorly, but it is probably worth stating that.

Author Response: Yes. This is correct. However, in this latest version of the paper, the bulk calibration filter has been removed as an example.

4) Line 32, p. 17: The reason you chose to focus on 2014 should be stated in methods.

Author Response: The last two paragraphs of section 4.3 will be moved to the end of section 2.8 in the methods.

5) How do the authors propose to implement the parameter and preceding streamflow filter? Operationally, one would not extract parameters for all days during a year’s precipitation events before setting the parameter range for this filter. Would it be based on the previous year’s filter or would the parameter prior distribution be updated based on the days leading up to the current date? How would that impact the result? This is touched on in the conclusions, and it is probably beyond the scope of this study, but it would be helpful if this limitation were mentioned in the methods section.

Author Response: Yes. The implementation of this filter in an operational setting is not feasible. It was mainly included to show that a small subset of the 10,000 LHS runs could be used more effectively than the full 10,000 parameter sets. The final sentence of section 2.7.3 is our response to this comment: “This ensemble represents an approach that cannot be used in a forecasting context, but does represent a proxy for other parameter-constraining methods that are explored in the discussion.”

6) Line 8, p. 19: Shouldn’t it be the minimized uncertainty filter that shows the model is capable of simulating streamflow for any 3-day period.

Author Response: Yes. This has been corrected in the text.

Typos, grammar, etc.

1) Line 23, p. 9: “eror” -> “error”

Author response: This will be corrected.

2) Line 25, p. 13: I’m pretty sure you mean 4c to 4e and 7.

Author response: This has been corrected.

3) Lines 20-21, p. 15: ": : :including the state of basin storage in the assessment of equifinality clearly shows that the parameter-state sets are not equal." I'm not sure what is meant by "assessment of equifinality." Also, please refer either to a figure, a table or a citation that supports this.

Author response: Thank you for pointing this out. This sentence is not clear and the paragraph will be adjusted as follows:

"The issue of widely-varying simulated basin storage (Figure 4c) also highlights the issue of equifinality, which is defined here as the idea that many different model simulations can produce acceptable results (Beven, 1993). The model is able to find many parameter-state sets that fit the streamflow for short periods of time. If only streamflow observations are available, the selected simulations are equifinal. However, including the state of basin storage clearly shows that the parameter-state sets are not equal. If soil moisture observations are also available and used, then these simulations are not equifinal and the selected simulations can be further constrained."

4) Figure 1: Color of lakes and rivers in legend should match their color in Fig. 1c.

Author Response: The figure has been changed accordingly.

5) Figure 2: The toolbar at the top should be removed.

Author Response: The figure has been changed accordingly.

6) Figure 5: Plot storage on the same scale in a and b. I'm guessing the solid black line is precipitation and the dots are storage simulated by the best 10 parameter sets. A legend would help.

Author response: The figure has been changed accordingly.

7) Table 4: "assessement" -> "assessment"

Author response: This has been corrected.

hess-2017-482: Authors' responses to Reviewer Comments for "Parameter-state ensemble data assimilation using Approximate Bayesian Computing for short-term hydrological prediction"

Responses to RC2

Major Concerns

1. The authors state that a drawback of traditional filtering methods that update the model state only or update state and parameters simultaneously provide state and parameter values that are not consistent (described in the Introduction, and re-iterated in the Discussion). But this "inconsistency" is the exact nature of filtering-based data assimilation as opposed to variational approaches and smoothing methods. The objective of filtering is to provide the best estimate of the state (or state and parameters) at a certain time, taking the underlying model predictive and observational uncertainties into account. In the approach presented in the paper, only model parameter uncertainty is accounted for, and the results clearly show that this is not sufficient to produce skilful predictions from the filtering. These results are not surprising, since the filter does not explicitly account for one of the major error sources in hydrological modelling, the error in the precipitation forcing. Data assimilation methods that update the model state directly account for this error.

Author response: The concern raised by the reviewer points to two sentences of relatively minor significance to the paper. These two sentences will be removed. Additional context has been provided to describe the approach as a hybrid of particle filter and variational DA. We disagree that the approach only accounts for model parameter uncertainty. The P-SEDA approach also accounts for changes to the model states by selecting joint parameter-state sets that vary both states and parameters for each selected parameter set (see Figure 7c for example).

2. The filtering approach developed uses a 3-day window to select the top 10 best parameter sets. The use of a 3-day window is not justified, and it seems questionable whether such a short window is sufficient considering the different time scales of runoff responses, ranging from slowly varying baseflow to fast responding overland flow contribution. The optimal window size will depend on flow regime.

Author Response: This is an excellent point. The analysis has been expanded to include longer filter periods.

3. A latin hypercube sampling approach is applied for generating the population of parameter sets from which the top 10 parameter sets are selected in the filtering approach. The authors discuss the limitation of the LHS approach. I wonder why this limitation has not been addressed in the work. The results of the bulk calibration filter that corresponds to a classical calibration-validation approach clearly show the limitation of the LHS approach.

Author response: We would have to re-do the entire study from scratch to address the limitation in this study. Given the limitation of LHS, however, it is reassuring that the approach works as well as it does with the longer filter periods. This limitation will be addressed in future work.

Detailed comments

1. Page 2, line 19-21. Not clear exactly what you mean by this statement (see General comments above).

Author response: This sentence is relatively insignificant for the paper and will be removed.

2. Page 5, line 17-18. Explain "CLASS tile" and "GRU".

Author response: This sentence is relatively insignificant for the paper and will be removed.

3. Page 6, line 1-2. How were the parameters and parameter intervals chosen for the LHS sampling? Based on a preliminary sensitivity analysis?

Author response: An earlier version of the paper included more details about the simple study conducted to select the parameters. A small amount of detail will be reinstated by adding the following sentence at the end of the paragraph on page 6, line 2. "The parameters that were perturbed were based on the lead author's experience with the model. Parameter intervals were set based on the ranges found in sources identified under the source column of Table 3. In the case of user specified parameters, these were set by the lead author."

4. Page 6, line 11. Abbreviation "H-EPS" not defined.

Author response: Thank you for catching this. H-EPS on Page 6, line 11 will be replaced with "Hydrological-Ensemble Prediction System (H-EPS)"

5. Page 6, line 19-20. How did you justify that the choice of the 10 best parameter sets is optimal?

Author response: The choice of the 10 best parameters was arbitrary. We have expanded the analysis to include 5, 10, 20, 30, 40 and 50 parameter sets. The method did not show much sensitivity to the number of parameter sets.

6. Page 7, line 9. Abbreviation "CaPA" not defined.

Author response: CaPA is defined on page 5, line 25. No changes to the text will be made.

7. Page 10, line 8-13. A long explanation. Rephrase.

Author response: We will rephrase the entire paragraph as follows:

"Note that these periods do not necessarily correspond to the rising-limb and recession periods of the hydrograph since the river does not always respond strongly to the precipitation for the time period of study in this basin. As a result, for lack of better terminology, these periods are hereafter referred to as "rain-influenced" and "rain-free". It would be more correct to say "periods during and immediately after the rainfall within the 3-day period" and "otherwise rain-free," but this terminology would be cumbersome throughout the remainder of the paper. Furthermore, it is also important to note that the terms “rain-influenced” and “rain-free” only refer to a time period rather than the discharge of the river. The time periods that these terms refer to are the stretch of time under consideration in the analysis.”

8. Page 10, line 20-25. Include a paragraph where you introduce the test period and test events.

Author response: The existing (short) paragraph will be altered as follows:

“Recall that MESH is run in a continuous simulation mode for the period of June 2002 to November 2014, with a more detailed analysis of the ensemble selection methodologies from June 1 to October 31, 2014. Within this time period, there are five significant precipitation events. The beginning and ending of the precipitation events are considered as follows:”

9. Page 10, line 26-29. Description of the reference forecast is out of place here. Move to the previous section where it is already introduced (page 9, line 5-6).

Author response: The following sentences will be removed from Page 10, line 26-29

“The skill is calculated with an unskilled reference forecast, which in this study is taken to be the measured streamflow at 00 UTC and 12 UTC each day as the forecast for the next 72 hours. This reference forecast is a persistence forecast, which assumes the streamflow is persistent for the forecast period.”

and inserted in the first paragraph of page 9 as follows:

“First, a qualitative analysis is undertaken to take advantage of the human brain’s ability to synthesize information. The results are then quantitatively verified using the Ensemble Verification System (EVS, Brown et al., 2010). To examine the quality of the ensemble mean when compared with the corresponding observation, the mean error (ME) is calculated. Then the quality of the ensemble distribution is calculated using rank histograms. Finally, the skill relative to using the current streamflow as the forecast is calculated using the mean Continuous Ranked Probability Skill Score (CRPSS). The unskilled reference forecast in this study is taken to be the measured streamflow at 00 UTC and 12 UTC each day as the forecast for the next 72 hours. This reference forecast is a persistence forecast, which assumes the streamflow is persistent for the forecast period.”

10. Page 11, line 27-29. Not clear how the water storage value is calculated. Is it a state variable in the model? Or is it assessed using the water balance calculations described in the discussion?

Author response: This storage is a state variable in the model. The text on Page 11, line 28 will be changed by replacing the words “… water storage values for each… ” with “water storage state variables for each…”

11. Page 13, line 4-6. Not clear how the 91 parameter sets are chosen. And how can this approach be applied in an operational setting?

Author response: The 91 parameter sets are chosen by confining the values of the normalized parameters based on the author's interpretation of Figure 6. The text will be adjusted as follows:

"Based on a subjective visual analysis of these box-plots, the 10,000 parameter sets are reduced to 91 parameter sets by confining the values of the normalized parameters as follows…"

A new Figure (Figure 5) has also been added to illustrate the approach.

This approach cannot be applied in an operational setting and simply provides some assurance that the method has the possibility of being useful, as discussed in section 4.2.

12. Page 13, line 11 and line 21. Use "reference forecast" instead of "unskilled forecast".

Author response: "reference forecast" will be used instead of "unskilled forecast."

13. Page 15, line 28-29. An example of using SMOS for DA in a hydrological model can be found in Ridler et al. (2014).

Author response: Thank you for this reference. It will be added to the list of references in the paper.

14. Page 16, line 8-10. The use of LHS is identified as one of the key limitations of the approach developed. So why wasn't this issue further investigated (see General comments above)?

Author response: Please see response to the General comment above.

15. Page 18, line 6-7. Why is it an advantage that parameters and state variables are consistent (see General comments above)?

Author response: Please see response to the General comment above.

16. Page 18, line 12. Abbreviation "H-LSS" not defined.

Author response: H-LSS is defined on page 2, line 3.

17. Tables 2-3. Very detailed information, and difficult to understand without knowledge of the model applied. I suggest to move this to Supplementary material together with a brief description of the model applied.

Author response: There is a very brief description of the model in section 2.3. Tables 2-3 can be moved to supplementary material if required.

18. Table 5, caption. Delete “low-skill”.

Author response: “low-skill” will be deleted in the caption of Table 5.

hess-2017-482: Authors' responses to Reviewer Comments for "Parameter-state ensemble data assimilation using Approximate Bayesian Computing for short-term hydrological prediction"

Responses to RC3

1. Section 2.1: The first few sentences describe briefly how the P-SEDA filter works. States and parameters are drawn from some multivariate initial distribution - and then analyzed for use in a projection period. How is this analysis done? I think a Figure may really help to communicate to readers how the P-SEDA method is implemented. The two sentences, "The analysis is completed and the process repeated for the next appropriate time-step in the continuous simulations" and "In this manner, both the parameters and states are drawn from the entire M simulations for the projection period". Not clear to me.

Author response: Additional context describing the filter as a hybrid of the particle filter and variational data assimilation, and a figure will be added in section 2.1. A second algorithm will also be added in section 2.7. We trust that the additional context, this new figure, and the second algorithm added to section 2.7 clarifies how the P-SEDA method is implemented.

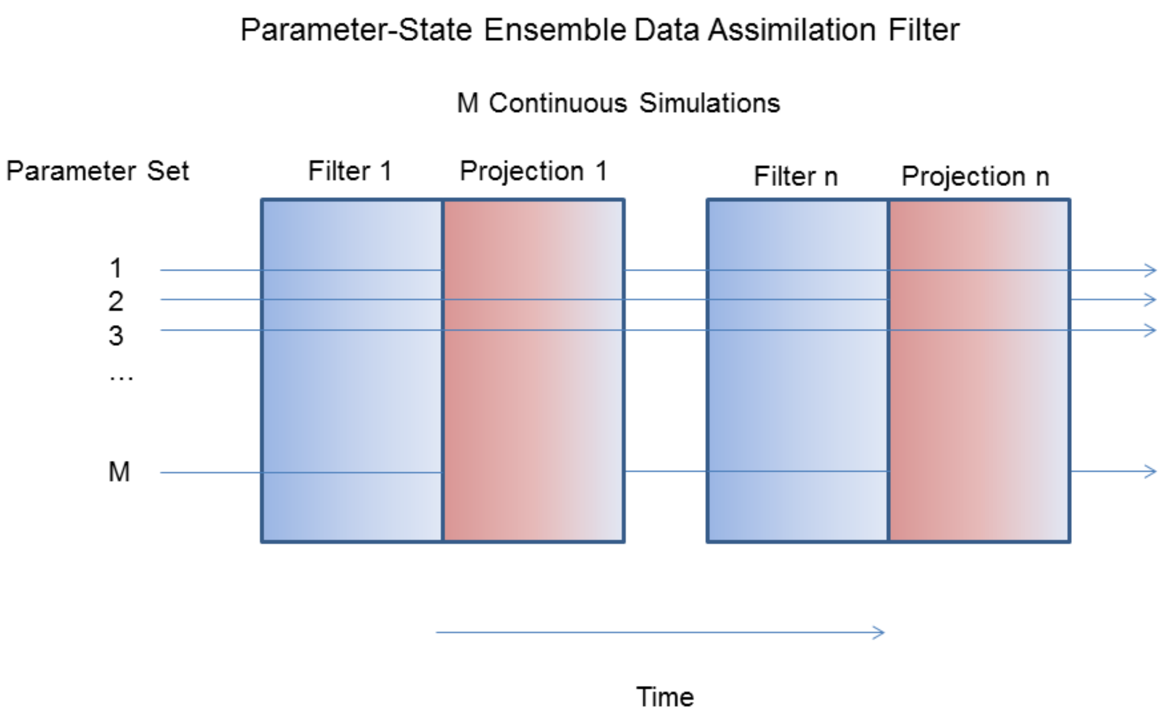

Figure 1: Schematic of the P-SEDA filter. M simulations are run continuously from a model, of which the filter chooses a number from which to analyze a projection. The process is then repeated for subsequent filter periods, noting that the M simulations run continuously through the previous projection periods even though they are not all selected for the previous projection period analysis.

2. Line 10: so you are talking here about the normalized weights. A particle filter uses three different weights: incremental weight (for current datum only), unnormalized weight (normalized

weight prior to datum x incremental weight -> summarizes weight of entire trajectory) and normalized weights -> normalization of unnormalized weights before moving on to the next datum.

Author response: It is not clear that this comment requires a response. Any additional thoughts that the reviewer can provide to guide the authors would be appreciated.

3. Line 12: Resampling is the crux to an efficient implementation of the particle filter. Otherwise, many trajectories will receive a negligible weight and the PF does not approximate closely the target PDF.

Author response: The description of the method has been re-worked as a hybrid of the particle filter and variational DA. This will hopefully address the reviewer's comment. If not, any additional thoughts would be appreciated.

4. Line 14 - 16 "The approach presented here is the same, but without resampling and always returning to the original particles as updated by the model and assigning a weight of zero or one to each particle based on the filter (i.e. using a rectangular filter)." is unclear to me. This goes back to my earlier comment. From what is presented, I do not understand how the authors implement such approach. Thus, no resampling is done? How do you return to the original particles. As with comment 1 above, can you give a detailed example, in text or in Figure that explains how this works. For example, at a time, t, we have the state forecast and associated parameter values + an incoming observation. What does the filter do then? How does it return to the original particles? How are the weights assigned? How is resampling avoided? etc.

Author response: We always return to the original particles by performing a continuous simulation of all of the particles. We hope that the addition of Figures 1, 2, 3, 5, Algorithm 2, and the description of the approach as a hybrid DA method makes the process more clear.

5. Algorithm 1: How is s(y_i) computed? And how do you find the theta's from the k_m nearest neighbors of s0?

Author response: In the pure form of ABC, s(y_i) is simply a statistical property, such as mean or standard deviation, of the simulation. This is compared to the same statistical property of the observation (s0). So if the mean were the statistical property being compared between the simulation and the observation, then the k_m nearest neighbors of s0 would be the k simulations that have the mean that is closest to the observations. However, in our approach, we replace a comparison of statistical properties with a cost function (Root Mean Squared Error).

6. I do not understand where ABC comes in. Is this in the selection of s0? and the s_y's? And how is the likelihood function formulated? This is done by simulation, yet, I miss the details necessary to understand and comprehend what has exactly been done.

Author response: All of algorithm 1 represents the ABC algorithm. The likelihood function is approximated by the model. The additional text and figures added to the paper should make it easier to understand and comprehend what has exactly been done.

7. Latin Hypercube sampling is argued as being highly inefficient. That is true if you want to approximate a target PDF, nevertheless, if you just want to sample the parameter space, then this may be one of the best methods you can use.

Author response: Operational models are generally concerned with predictive ability and thus are more concerned with approximating a target PDF rather than sampling the parameter space. As such, future work in this area should consider methods other than LHS. No changes are planned to address this comment.

8. In Section 2.7 the authors describe how they construct the ensemble. None of the four approaches listed are described in detail. Hence, I do not understand what is being done. "minimized uncertainty filter". Need a detailed explanation, step by step how we go about initial states and parameters to a minimized uncertainty filter. Same holds for the other three listed methods. Without this the results in this paper will not be understood, nor are impossible to be reconstructed by the reader.

Author response: The following Figures will be added to Section 2.7, which we expect will make the details of the various approaches more clear.

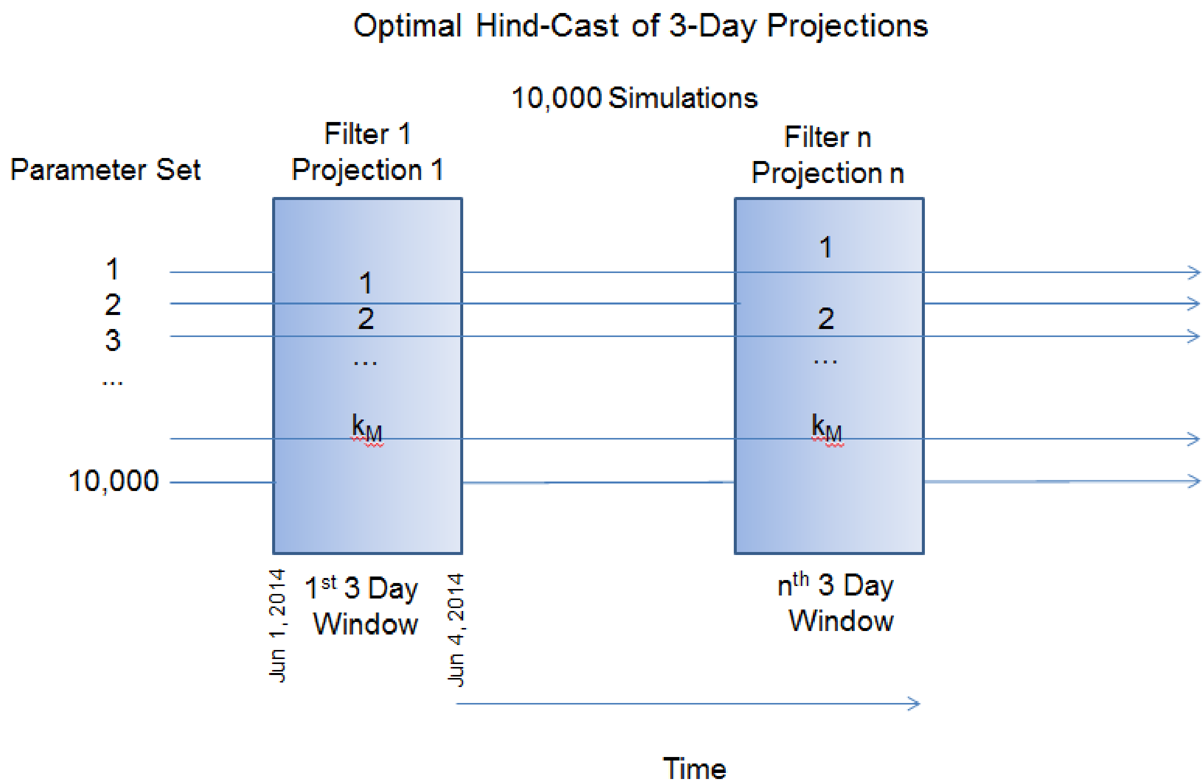

Figure 2: Schematic of the P-SEDA filter for the optimal hind-cast of 3-day projections used in this study. 10,000 simulations are run continuously through the MESH model, of which the filter chooses a number (k_M) for the hind-cast analysis. The process is then repeated for subsequent filter periods.

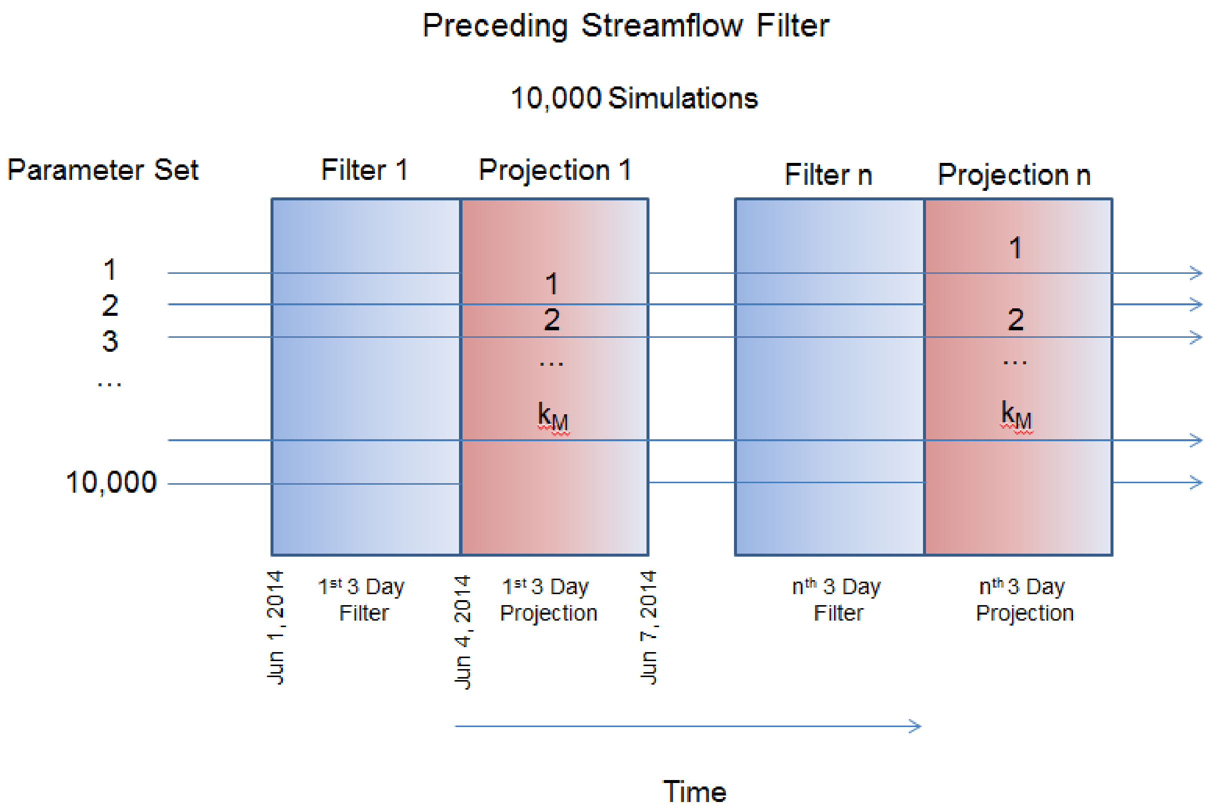

Figure 3: Schematic of the P-SEDA filter for the preceding streamflow filter. 10,000 simulations are run continuously through the MESH model, of which the filter chooses a number (k_M) from which to analyze a projection. The process is then repeated for subsequent filter periods, noting that the M simulations run continuously through the previous projection periods even though they are not all selected for the previous projection period analysis.

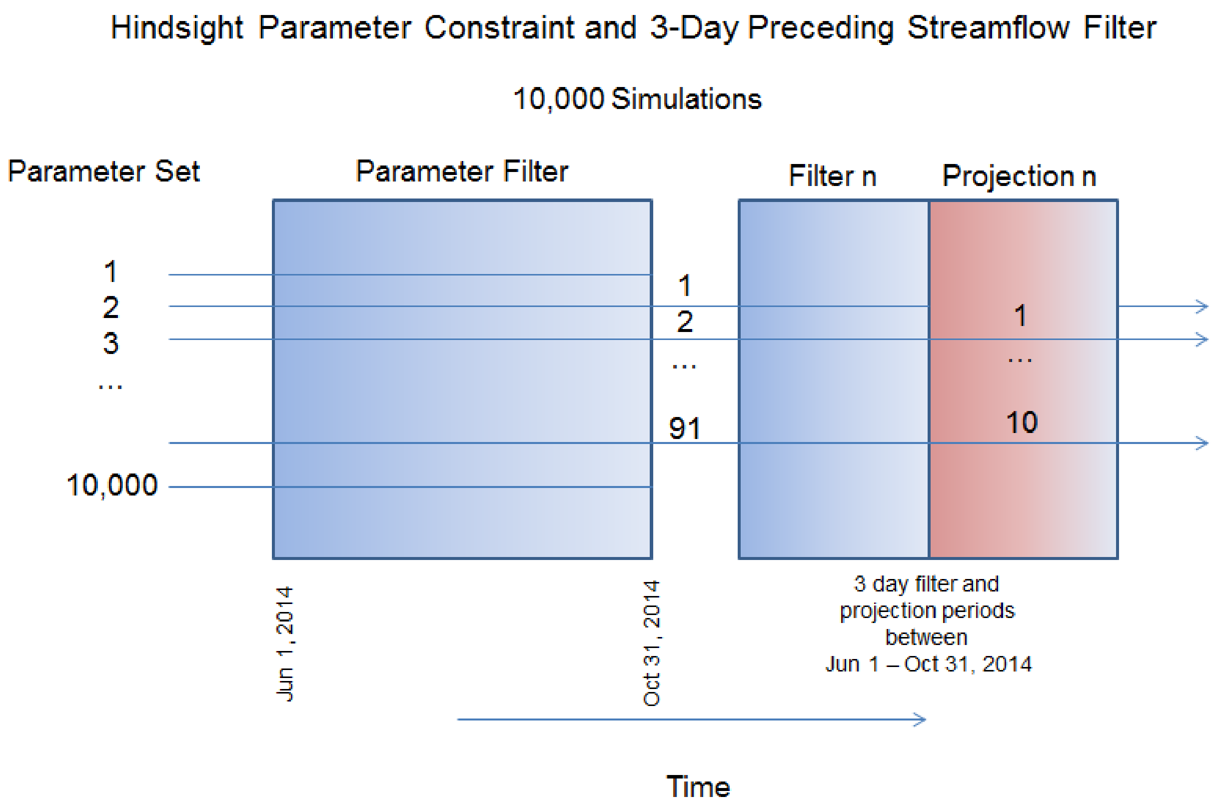

Figure 5: Schematic of the P-SEDA filter for the hindsight parameter constraint and preceding 3-day streamflow filter used in this study. 10,000 simulations are run continuously through the MESH model. The original 10,000 parameter sets are reduced to 91 parameter sets based on a hindsight analysis of parameters that are shown to be important during precipitation events between June 1 and October 31, 2014. These remaining 91 parameter sets are then selected for analysis in the preceding streamflow filter as if M = 91 in Algorithm 2.

Parameter-state ensemble data assimilation using Approximate Bayesian Computing for short-term hydrological prediction

Bruce Davison[1], Vincent Fortin[2], Alain Pietroniro[1], Man K. Yau[3], and Robert Leconte[4]

[1]Environment and Climate Change Canada, Saskatoon, Saskatchewan, Canada.
[2]Environment and Climate Change Canada, Montreal, Quebec, Canada.
[3]McGill University, Montreal, Quebec, Canada.
[4]Université de Sherbrooke, Sherbrooke, Quebec, Canada.

Correspondence to: Bruce Davison (bruce.davison@canada.ca)

Abstract. The main sources of uncertainty in hydrological modelling can be summarized as structural errors, parameter errors, and data errors. Operational modellers are generally more concerned with predictive ability than model errors, and Data Assimilation (DA) methods are commonly employed to merge models with observations to improve predictive ability. This paper presents an example of Approximate Bayesian Computing (ABC), or a ~~simplified~~ hybrid of a (simplified) Particle Filter (PF) and variational DA, to simultaneously assimilate model states and parameters, calling the method Parameter-State Ensemble DA (P-SEDA). The case study is from June to October, 2014 for a small (1 324 km^2) watershed just north of Lake Superior in Ontario, Canada using the Canadian semi-distributed hydrologic land-surface scheme MESH. The study examines how well the approach works given various levels of certainty in the data; beginning with certainty in the streamflow and precipitation, followed by uncertainty in the streamflow and certainty in the precipitation, and finally uncertainty in both the streamflow and precipitation. The approach is found to work in this case when streamflow and precipitation is fairly certain, while being more challenging to implement in a forecasting scenario where future streamflow and precipitation is much less certain. The main challenge is determined to be related to parametric uncertainty and ideas for overcoming this challenge are discussed.

1 Introduction

A fundamental problem making good streamflow predictions in process-based models rests with the various sources of uncertainty in modelling the flow. These sources of uncertainty have been described in a number of papers (e.g. Beck, 1987; Krzysztofowicz, 2001; Vrugt et al., 2005; Liu and Gupta, 2007; Velázquez et al., 2009). In particular, Liu and Gupta (2007) consider a general framework of seven model components which include the system boundary (B), inputs (u), initial states (x_0), parameters (Θ), structure (M), states (x), and outputs (y). Five of these model components (B, u, x_0, Θ, and M) must be predefined and their uncertainties cascade to x and y. Since the inputs, initial states, and observations used to verify the model outputs can often be considered as data errors, and the system boundary can be considered a source of structural uncertainty, the main sources of errors in hydrologic modelling can be summarized as structural errors, parameter errors, and data errors. Operational hydrological models are generally more concerned with predictive ability than correctness of the model structure (Gupta et al., 2008, p 3804). As such, parameter and data errors are often the focus for operational hydrological predictions.

Within this context of managing parameter and data uncertainty, it is the purpose of this paper to propose a novel approach to short-term hydrological prediction in a relatively small, data-sparse watershed (1,324 km^2). The approach involves using a hydrologic land-surface-scheme (H-LSS) and simultaneous estimation of parameters and state variables through Data Assimilation (DA) using Approximate Bayesian Computation (ABC, Biau et al., 2015), or a hybrid of a simplified version of the Particle Filter (PF, Arulampalam et al., 2002) and vatiational DA (Asch et al., 2016).

DA is one way to improve hydrological predictions by merging models with observations, and ~~can~~ DA methods can categorized in a number of different ways (Liu and Gupta, 2007; Rakovec et al., 2015; Sun et al., 2016; Asch et al., 2016). One way of categorizing DA methods is by variational or sequential (statistical) methods. Variational approaches minimize the differences between observations and model output over a period of time, while sequential approaches assimilate observations as they are obtained. Another way of categorizing DA methods is by their time dependence. Usually, smoothing problems attempt to make predictions for the past, filtering problems attempt make predictions for the present, and forecasting problems attempt to make predictions for the future.

DA can include methods that help resolve problems related to estimating states, assessing parameters and identifying the appropriate model structure (Liu and Gupta, 2007). Most applications of DA focus on merging state variables in a model with corresponding observations, while a few methods combine state and parameter estimation to improve predictions (e.g. Vrugt et al., 2005; Moradkhani et al., 2005a, b; Drécourt et al., 2006; Labarre et al., 2006; Qin et al., 2009; Nie et al., 2011; Xie and Zhang, 2013; Bi et al., 2014).

The strategies for combined state and parameter assimilation generally fall into three main categories (Liu and Gupta, 2007). One strategy, such as that used by Vrugt et al. (2005), is to use standard techniques to simultaneously optimize parameters and assimilate states. In this strategy, an outer loop is used to optimize parameter sets while an inner loop is used to assimilate the state variables for each calibration parameter set at each time-step. Another strategy is to use dual filters (e.g. the dual Ensemble Kalman Filter or dual Particle Filter) to update states and parameters independently (e.g. Moradkhani et al., 2005a, b; Qin et al., 2009). In these cases, the parameters and states are continuously updated as new observations become available. The third strategy is most often called "state augmentation" and uses regular data assimilation methods where parameters are considered state variables and added to the state vector (e.g. Drécourt et al., 2006; Nie et al., 2011; Bi et al., 2014). A single filter is then used to update the parameters and states simultaneously. ~~One drawback of traditional DA (of states only) and of the aforementioned parameter and state DA methods, however, is that the resulting parameters and states are not necessarily compatible with one-another.~~

In this paper, a new and very simple method of simultaneous state and parameter DA ~~, that ensures parameter and state compatibility,~~ is presented for short-term hydrological ensemble prediction (up to 3 days). We call this DA-approach the Parameter-State Ensemble Data Assimilation (P-SEDA) filter and make use of ABC, which is also a ~~simplified PF~~hybrid approach of a (simplified) PF and variational DA. The approach is described with the intention of making clear how to implement the filter with a wide variety of models in data-rich or data-sparse watersheds, and examined here using a parameter-intensive hydrologic land-surface scheme in a data sparse watershed. The case studies include model structural and parameter errors, which are inevitable regardless of the model being used or the basin being modelled, to evaluate the ability of the ~~filter to work under such conditions~~P-SEDA approach.

The intial case studies are hindcasting exercises that reduce data and model structural uncertainty as much as possible, followed by a more realistic forecasting example (albeit in hindcasting mode) that incorporates data input uncertainty using Environment and Climate Change Canada's (ECCC's) meteorological Regional Ensemble Prediction System (REPS) to drive the model.

2 Methodology

2.1 The Parameter-State Ensemble Data Assimilation (P-SEDA) Filter

The P-SEDA filter works in the following manner, as illustrated in Figure 1. First, a number of parameter sets (M) are pre-defined to be used for continuous simulation with a model. Filtering criteria are used to determine which of the parameter sets and their associated state variables will be used to generate an ensemble of streamflows for analysis in a projection period. The analysis is completed for the projection period and the process repeated for the next appropriate time-step in the continuous simulations. Note that the M model simulations continue through the projection period in the continuous simulations. The filter simply chooses which of the M continuous simulations to select for the projection analysis. In this manner, both the parameters and states are drawn from the entire M simulations for the projection period. ~~This is very similar to traditional particle filtering methods.~~

Compared to the traditional DA approaches described in the introduction, the P-SEDA approach can be considered a hybrid approach to DA (Asch et al., 2016). It is a hybrid in the sense that it has characteristics of both a particle filter and variational DA. A traditional particle filter has the following four steps: 1) generate an initial set of particles (or parameter sets) and run the model for a short time (e.g. one timestep) to produce model output for the variable of interest, 2) assign a weight between zero and one to each particle such that higher weights are given to parameter sets that produce model outputs more closely matching the observed variable, 3) resample the parameter space with respect to the weights (i.e. produce a new set of particles with parameters that are closer to the parameter sets that produced higher weights), and 4) propogate the new particles using the model, thus repeating the cycle. The approach presented here is the same, but without resampling and always returning to the original particles as updated by the model and assigning a weight of zero or one to each particle based on the filter~~(i.e.using a rectangular filter).~~. The similarity to variational DA comes from the fact that a cost function is minimized based on a moving window of time from the past to the present.This is consistent with the statement in Liu and Gupta (2007) that 'theoretically, variational DA methods can be used for filtering problems if a new smoothing problem is defined sequentially at each observation time point.' This is exactly what is being proposed for the P-SEDA approach, but with a moving window of time for observational input to the DA process rather than an ever-expanding window of time.

For a single filter-projection period, this approach is also described by ABC. As described by Vrugt and Sadegh (2013), for scenarios focused on parameter uncertainty (as is the case in this paper) the posterior parameter distribution $p(\theta|y)$ given the streamflow y is estimated using Bayes theorem:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

where $p(\theta)$ is the prior distribution of parameters, $p(y)$ is the normalization constant, and $p(y|\theta) \equiv L(\theta|y)$ is the likelihood function. In situations where the likelihood function cannot be computed (again, as is the case in this paper) the likelihood is approximated using a model.

Biau et al. (2015, Algorithm 2) provides a widely-used algorithm for the approach, shown below as Algorithm 1. The symbols not yet described in Algorithm 1 include a statistic representing the observations, s_0, and the statistic representing the simulations $s(y_i)$.

Algorithm 1 Pseudo-code of a generic ABC algorithm

Require: A positive integer M and an integer k_M between 1 and M.
for i = 1 to M **do**
generate θ_i from the prior $p(\theta)$;
generate y_i from the approximate likelihood $p(y|\theta)$
end for
return The θ_i's such that $s(y_i)$ is among the ~~k$_M$~~ k_M-nearest neighbors of s_0.

In the context of the P-SEDA filter for hydrological prediction, θ_i is the i^{th} parameter set. The prior $p(\theta)$ represents the infinite possible parameter sets based on ranges selected by the user and the likelihood is approximated by the model used to generate the y_i simulated streamflow values (giving rise to the "Approximate" in ABC). In Biau et al. (2015), only the k_M-nearest neighbors between the statistic representing the observations, s_0, and simulations, $s(y_i)$, for the filter period are kept for analysis. The case study presented below alters the selection criteria slightly by using a distance function (Root Mean Squared Error) to determine the discrepancy between the observations and the simulations, rather than independent statistical properties (such as mean and standard deviation) of the two.

In the P-SEDA approach, we can disregard the notion of finding a distribution of parameter sets that fits the entire streamflow record of interest. Instead, we look for a set of plausible parameter sets, locate a certain number of these that generate the "best" results for the filter period under consideration, and then evaluate how well these parameters and states perform for the projection period. The process is then repeated to find new parameter sets and states for consideration in successive projection periods. There are a number of ways in which filter and projection periods can be formulated. ~~Four~~ A number of such formulations are described in section 2.7, which should clarify the generic process described in this paragraph. ~~As already mentioned, this approach can be considered as a simplified PF or ABC.~~

2.2 Case Study Basin Description

The study watershed is 1,324 km^2 and is drained by the Little Pic River near Coldwell, Ontario, Canada, just north of Lake Superior. The streamflow gauge (02BA003) is between the communities of Terrace Bay and Marathon and has been operated by the Water Survey of Canada from 1972 to the present. There are no precipitation measurements in the basin, but the surrounding region's annual precipitation ranges from 654 to 879 mm, with the mean summer rainfall ranging from 231 to 298 mm (Crins et al., 2009). The mean annual temperature ranges from -1.7 to 2.1 °C (Mackey et al., 1996). The sub-surface sits on Precambrian Shield with significant amounts of volcanic rock, greenstone, siltstone and shale (Sutcliffe, 1991). The dominant landcover in the basin is mixed forest, followed by coniferous forest, water, sparse forest and deciduous forest (Crins et al., 2009). The streamflow regime is characterized by frozen conditions through the winter months (November to April), but has been known to produce a spring freshet as early as March. Summer and autumn peaks can be on the same order of magnitude as the spring freshet, but are more often smaller. The peak flow is usually in May and the highest daily discharge recorded is 269 m^3/s on June 30, 2008.

2.3 The Semi-Distributed Hydrologic Land-Surface Scheme

The model used to simulate the streamflow is the semi-distributed hydrological land-surface scheme MESH (Pietroniro et al., 2007), configured with the Canadian Land-Surface Scheme (CLASS, Verseghy, 1991; Verseghy et al., 1993), the hydrologic routing from WATFLOOD (Kouwen et al., 2002), and additional hydrological processes to better simulate surface and sub-surface lateral flow across the landscape to the river (Soulis et al., 2000, 2011).

The basin geophysical characteristics needed for MESH include a digital elevation model (DEM), landcover classification, and soil information. The DEM comes from the Canadian Digital Elevation Data (CDED) at a scale of 1:50,000 and based on the NAD83 horizontal reference datum (Natural Resources Canada, 2015). The landcover classification comes from the LCC2000-V product originating from classified Landsat 5 and Landsat 7 satellite images and the soils information comes from the ecodistricts classification of the national ecological framework for Canada (Agriculture and Agri-Food Canada, 2015). The basin fits within ecodistrict 389 - Long Lake.

Table 1 shows the estimated percentages of each landcover present in the basin as defined by the LCC2000-V product. Based on this classification, the two dominant landcovers are coniferous and broadleaf forest, which are often mixed. Without knowing more specific information about the landcover, the mixed forests are assumed to be fifty percent coniferous and fifty percent broadleaf, resulting in an estimate of forty-eight percent coniferous and thirty-nine percent broadleaf. These values are then arbitrarily rounded up to fifty percent coniferous and forty percent broadleaf in the model representation of landcover. The remaining ten percent of landcover inevitably includes parametric uncertainty due to the model's inability to properly represent the eight percent of the basin that is covered by small lakes.

Figure 6 illustrates a) the location of the basin, b) ecodistrict boundary and model grid, c) river network and gauge location, and d) landcover. Sub-grid variability of each grid is handled via the CLASS tile with each grid being represented by a single ecodistrict GRU.

2.4 Forcing Data

The meteorological inputs for MESH include incoming shortwave radiation, incoming longwave radiation, precipitation, temperature, barometric pressure, specific humidity and wind speed. The timestep of the model is set to 30 minutes. For the first two case studies minimizing input data uncertainty, most of these meteorological inputs were derived from ECCC's Global Environmental Multi-scale (GEM) Numerical Weather Prediction (NWP) model (Côté et al., 1998a), stitching together the 6–17 hour UTC forecasts from twice-daily runs beginning in January, 2002 and linearly interpolating between hours to obtain half-hourly values. Precipitation is obtained from the Canadian Precipitation Analysis (CaPA, Mahfouf et al., 2007; Lespinas et al., 2015), which is an assimilation of ground-based observations and GEM precipitation forecasts. For the third case study including forcing input data uncertainty, an ensemble of meteorological inputs is obtained from ECCC's Meteorological Regional Ensemble Prediction System (REPS). The REPS provides 72 hour forecasts twice daily.

2.5 Parameter Selection

H-LSS's contain many parameters and there is a large body of scientific literature examining various techniques for effectively estimating parameters (for a brief review, see Matott et al., 2009). The method that was used in this study is Latin Hypercube Sampling (LHS, McKay et al., 1979). Twenty-eight parameters were perturbed based on the results of a simple study (not shown) comparing the perturbation of 6, 15 and 28 parameters. Table 2 shows the parameter values that were fixed during the simulations while Table 3 shows the ranges for parameters that were perturbed. The parameters that were perturbed were based on the lead author's experience with the model. Parameter intervals were set based on the ranges found in sources identified under the source column of Table 3. In the case of user specified parameters, these were set by the lead author.

It is worth noting up-front that this approach to parameter perturbation is very inefficient. Sampling via LHS is a variation of uniform random sampling that is traditionally used in the generalized likelihood uncertainty estimation (GLUE) methodology (Beven and Binley, 1992). Tolson and Shoemaker (2008) provide a very thorough explanation of the limitations of LHS and other methods of combatting the inefficiency of the tradtional GLUE uniform random sampling. The purpose of this study, however, is to examine the P-SEDA methodology. Implications of the parameter sampling methodology are examined in the discussion after the results are presented.

2.6 Projection Periods for Short-term Hydrological Prediction

This paper is focused on short-term hydrological ensemble prediction (up to 3 days), with an interest in using the ECCC meteorological REPS to force a more comprehensive Hydrological-Ensemble Prediction System (H-EPS). As such, projection periods are defined as the three-day windows of time from the beginning of each ECCC-REPS run at 0 UTC and 12 UTC. The red and pink bars in Figure 4 illustrate tweleve projection periods beginning on 0 UTC, July 20 to 12 UTC, July 25, 2014. The remainder of Figure 4 is described in section 2.7.2.

2.7 Ensemble Selection Methodologies

The total population of model runs is generated by setting M to ~~10,000~~ $10,000$ in Algorithm 1 and using LHS to generate the 10,000 parameter sets from a uniform prior distribution of 28 parameters based on the ranges shown in Table 3. To approximate the likelihood, MESH is run with each of the ~~10,000~~ $10,000$ parameter sets in a continuous simulation mode to generate streamflow values (y) for the period of June 2002 to November 2014.

The Data Assimilation is performed by filtering the total population of ~~10,000~~ $10,000$ model runs to generate an ensemble of the ~~10~~ "best" model runs ~~for each~~ to be used in the subsequent projection period. ~~The following four ensemble selection methodologies, or ensemble data assimilation filters,~~ Algorithm 2 represents the specific implementation of Algorithm 1 used in this paper.

Algorithm 2 Pseudo-code of the ABC algorithm implementation in this paper

---

Require: A positive integer $10,000$ (M) and an integer (k_M) between 1 and M (In this study, a sensitivity analysis of k_M is performed by setting $k_M = 5, 10, 20, 30, 40,$ and 50).

for i = 1 to $10,000$ **do**

generate parameter set i (θ_i) from all possible parameter sets ($p(\theta)$) using LHS;

generate streamflow i (y_i) using the model (MESH)

end for

return The parameter sets with the lowest k_M RMSE values between observed and simulated streamflow values from the $10,000$ model runs for the filter period. (This is the filter.)

The following P-SEDA configurations are examined in this study and will be described shortly:

1. ~~Minimized uncertainty filter~~ Optimal hind-cast of 3-day projections
2. ~~Bulk calibration filter~~
3. Preceding streamflow filter
4. ~~Parameter and preceding streamflow filter~~ Hindsight parameter constraint and preceding 3-day streamflow filter
5. Hydrologic-Ensemble Prediction System (H-EPS)

An initial evaluation of the P-SEDA filters requires some sources of uncertainty to be minimized. In particular, streamflow observations and precipitation uncertainty are considered; and questions around the model's ability to manage snow processes are simply avoided.

The quality of a streamflow observation is commonly known to be affected by ice, the occurrence of which is noted in the Water Survey of Canada database of streamflow. For the Little Pic River watershed, ice on the river can occur as early as November and as late as April. In addition, prior to the spring freshet, the streamflow contains almost no information that could assist in predicting future streamflow. The future flow primarily depends on other land-surface characteristics such as snow water equivalent and frozen ground. Because the immediate interest is in evaluating the P-SEDA filters when snow or ice on the ground is not present, the analysis is only performed from June 1 to October 31 for ~~selected years~~2014, with some qualitative analysis beginning on May 1, 2014.

Each of the four ~~filter~~ P-SEDA configurations is described below.

2.7.1 ~~Minimized Uncertainty Filter~~Optimal Hind-cast of 3-day Projections

Since there are no precipitation observations in the basin, but there are some precipitation gauges nearby which are used in the generation of CaPA, the filters are initially tested with reduced precipitation uncertainty by forcing the model with CaPA.

In addition, streamflow uncertainty is minimized by filtering with known streamflow ~~for the time periods of interest~~in a hind-casting excercise. In other words, ~~in this configuration, the filter period corresponds with the projection period and we use the observed streamflow to filter the parameter-state combinations to generate the ensemble. The resulting minimized uncertainty filter is a hind-casting excercise~~ the resulting ensemble is used to determine if the parameter selection methodology (LHS) allows the model to produce streamflow values that match observations ~~.~~

2.7.2 ~~Bulk Calibration Filter~~

~~The bulk calibration filter represents a more traditional calibration and validation exercise in which the calibration period is the filter period and the validation period is the projection period. In our example, the top~~ given advanced knowledge of precipitation and streamflow. This process is illustrated in Figure 2. In this case, 10~~daily RMSE values from the 10~~,000 ~~LHS simulations for Jun 1, 2002 to June 1, 2009 are filtered to see how they perform for the projection periods from June 1 to October 31, 2014.~~ simulations are run continuously through the MESH model, of which the filter chooses a number (k_M) for the hind-cast analysis. The process is then repeated for subsequent filter periods. Note that in this hind-casting exercise, the filter periods correspond to the projection periods.

2.7.2 Preceding Streamflow Filter

Figure 2 illustrates the preceding streamflow filter. In this study, 10,000 simulations are run continuously through the MESH model, of which the filter chooses a number (k_M) from which to analyze for a projection period. The process is then repeated for subsequent filter periods, noting that the M simulations run continuously through the previous projection periods even though they are not all selected for the previous projection period analysis.

To give more detail to the sequencing within the filter-projection cycle, Figure 4 illustrates the ~~third filter considered~~ preceding streamflow filter considered for a 3-day filter period (other filter period lengths are examined in the results). In this Figure, twelve filter periods are shown in orange and green for July 17 to July 25, 2014. The first filter period is represented by the orange bar near the top of the Figure beside the ABC1 label. This filter period runs from 0 UTC on July 17 to 0 UTC on July 20, 2014. During this three-day period, the "best" parameter sets are selected based on how the model simulates the observed streamflow. The simulations for these top performing parameter sets are then extended for three more days, which is considered to be the projection period.

This process is then repeated 12 hours later. The second filter period is represented by the orange bar illustrated just below the first filter period, and labeled ABC2 in the Figure. This filter period runs from 12 UTC on July 17 to 12 UTC on July 20, 2014. During this three-day period, the "best" parameter sets are selected based on how the model simulates the observed streamflow. Although there is considerable overlap between the first and second filter periods, the second filter period begins and ends 12 hours after the first filter period, producing a new ensemble of "best" parameter sets. The simulations for these new parameter sets are then extended for three days, which is considered to be a new projection period that also begins and ends 12 hours after the previous projection period.

The process is then continually repeated every 12 hours as shown by the remainder of the bars shown in the rows labeled ABC3 to ABC12. Each instance of the filter and projection periods represents a single application of ABC, which is why each row is labeled as such. We call this filter the preceding streamflow filter because the projection periods shown in red and pink occur immediately after the filter periods.

One important consideration, that becomes relevant in the analysis, is the six hours of precipitation that occurred on July 22 from 14 UTC to 20 UTC. This is illustrated by the small blue bar at the top and near the middle of Figure 4. Some of the projection periods "see" this precipitation event (illustrated by the green bars) and some do not (illustrated by the orange bars). As will be explained more fully in section 2.8.2, the analysis is split according to the sub-periods of the filter and projection periods that a) occur during and just after the precipitation event (the light green and pink bars); and b) that occur when it is otherwise rain-free (orange, dark green and red bars).

2.7.3 ~~Parameter and Preceding Streamflow Filter~~

~~The fourth filter~~ To properly examine the effectiveness of the P-SEDA approach, a sensitivity analysis is performed for the length of the filter period as well as for k_M in Algorithms 1 and 2. The length of the filter period is tested for 3, 10, 20, 30 and 40 days while k_M is tested for values of 5, 10, 20, 30, 40 and 50.

2.7.3 Hindsight Parameter Constraint and Preceding 3-Day Streamflow Filter

The third ensemble considered is very similar to the preceding streamflow filter with a filter period of 3 days. However, to constrain the parameter space further than determined by ~~the~~ LHS, the population of parameter sets is reduced by selecting a sub-set from the initial population of 10,000 parameter sets. The sub-set is selected by confining the parameter values based on model simulations that respond well during precipitation events in 2014. ~~Details~~ Figure 5 illustrates this filter and more details are provided in section 3.3 of the results. This ensemble represents an approach that cannot be used in a forecasting context, but does represent a proxy for other parameter-constraining methods that are explored in the discussion.

2.8 ~~H-EPS~~

2.7.1 Hydrologic-Ensemble Prediction System (H-EPS)

Of course it is not often known for sure if precipitation will occur in the future, and certainly not the amount of precipitation that will occur. As a result, ECCC's Meteorological REPS is also used with the June 1 to October 31, 2014 data in a hindcasting mode to examine how the PSEDA approach can be used in a true forecasting context. At 00 UTC and 12 UTC every 12 hours from May 1 to October 31, 2014, the top 10 parameter-state pairs from the filter period and different projection periods are run for 3-days using the forcing data from the 20 members of the REPS, for a total of 200 H-EPS members. This analysis is performed for ~~all filters except for the bulk calibration filter.~~ a hind-sight parameter constraint and preceding 3-day streamflow filter and k_M value of 10 for illustrative purposes.

The only other use of ECCC's REPS as a part of an H-EPS is found in Abaza et al. (2013), in which the Canadian operational meteorological Global Ensemble Prediction System (GEPS) was compared with the REPS and the deterministic 15 km GEM NWP forcing of the province of Quebec's operational streamflow forecasting system. The study found that both the GEPS and REPS outperformed a deterministic run for eight watersheds ranging in size from 355 to 5820 km^2. The REPS was also found to be superior to the GEPS in terms of its ability to predict forecast uncertainty.

One issue highlighted in the conclusion of Abaza et al. (2013) is that the REPS was found to produce unusually high precipitation spikes. This issue of excessive precipitation was, in many cases, determined to be caused by the physics perturbation scheme that was used to generate the ensemble (Erfani et al., 2014) and was fixed in the version of the REPS that was officially released on December 4, 2013. The update to the REPS is one of the main reasons for focusing on 2014 as a period of interest.

2.8 Verification

Demargne et al. (2010) differentiates diagnostic verification for evaluating the performance of a system from real-time verification for helping end-users make decisions about the future. The verification performed here is done for the first of these objectives; evaluating the performance of a system.

2.8.1 Verification of the Ensemble Selection Methodologies

First, a qualitative analysis is undertaken to take advantage of the human brain's ability to synthesize information. The results are then quantitatively verified using the Ensemble Verification System (EVS, Brown et al., 2010). To examine the quality of the ensemble mean when compared with the corresponding observation, the mean error (ME) is calculated. Then the quality of the ensemble distribution is calculated using rank histograms. Finally, the skill relative to using the current streamflow as the forecast is calculated using the mean Continuous Ranked Probability Skill Score ($\overline{\text{CRPSS}}$). The reference forecast in this study is taken to be the measured streamflow at 00 UTC and 12 UTC each day as the forecast for the next 72 hours. This reference forecast is a persistence forecast, which assumes the streamflow is persistent for the forecast period.

The ME measures the average difference between a set of forecasts and corresponding observations. In this case, it measures the average difference between the mean average of the ensemble forecast ($\overline{Y}$) and the observation (x) as follows:

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}(\overline{Y}_i - x_i)$$

The ME may be positive, zero, or negative. A positive value represents an ensemble mean that is positively biased while a negative error represents an ensemble mean that is negatively biased. A value of zero represents an absence of bias in the ensemble mean.

The rank histogram measures the reliability of an ensemble forecasting system. It involves counting the fraction of observations that fall between any two ranked ensemble members in the forecast distribution. For an ensemble forecast containing

m ensemble members ranked in ascending order, there are $m-1$ spaces between any two ranked ensemble members and two spaces at the ends (above and below the ensemble forecast range) for a total of $m+1$ spaces (s_1, ... , s_{m+1}). The corresponding observation h for each ensemble forecast will fall within one of the spaces.

$$h_i = \frac{1}{n}\sum_{j=1}^{n} 1\{x_j \in s_{ij}\}$$

where h_i is the fraction in the i^{th} bin, x_j is the j^{th} observed value, s_{ij} is the i^{th} gap associated with the j^{th} forecast, and $1\{\cdot\}$ is a step function that gives a value of 1 if the condition is met and 0 otherwise.

The mean Continuous Ranked Probability Skill Score ($\overline{\text{CRPSS}}$) measures the performance of one forecasting system compared to another forecasting system in terms of the mean Continuous Ranked Probability Score ($\overline{\text{CRPS}}$). The $\overline{\text{CRPS}}$ measures the average square eror of a probability forecast across all possible event thresholds. The $\overline{\text{CRPSS}}$ comprises a ratio of the $\overline{\text{CRPS}}$ for the forecasting system to be evaluated $\overline{\text{CRPS}}_{\text{EVAL}}$, and the $\overline{\text{CRPS}}$ for a reference forecasting system, $\overline{\text{CRPS}}_{\text{REF}}$.

$$\overline{\text{CRPSS}} = \frac{\overline{\text{CRPS}}_{\text{REF}} - \overline{\text{CRPS}}_{\text{EVAL}}}{\overline{\text{CRPS}}_{\text{REF}}}$$

As a measure of the average square error in probability, values for the $\overline{\text{CRPS}}$ approaching zero are better. As a result, values for the $\overline{\text{CRPSS}}$ closer to one are better as this illustrates that $\overline{\text{CRPS}}_{\text{EVAL}} < \overline{\text{CRPS}}_{\text{REF}}$.

2.8.2 Splitting the Analysis Based on Rainfall

The qualitative analysis shown in the following Results section illustrates that there is a significant difference in the abilities of the filters to effectively project streamflow when it is raining and when it is not raining. As a result, the quantitative analysis is split into two parts: 1) for periods in which it is raining and just afterwards (for the remainder of each respective filter or projection period), and 2) for periods in which it is otherwise not raining.

Note that these periods do not necessarily correspond to the rising-limb and recession periods of the hydrograph since the river does not always respond strongly to the precipitation for the time period of study in this basin. As a result, for lack of better terminology, these periods are hereafter referred to as “rain-influenced” and “rain-free~~.~~”. It would be more correct to say “periods during and immediately after the rainfall within the 3-day period” and “otherwise rain-free,” but ~~that would be too cumbersome in~~ this terminology would be cumbersome throughout the remainder of the ~~text, so we ask for your indulgence in the potentially confusing use of the terms “rain-influenced” and “rain-free.”~~ paper. Furthermore, it is also important to note that the terms “rain-influenced” and “rain-free” only refer to a time period rather than the discharge of the river. The time ~~period~~ periods that these terms refer to are the ~~portions of the 3-day time period~~ stretch of time under consideration in the analysis.

Recall the description for Figure 4 in Section 2.7.2. for illustrating the difference between “rainfall” and “rain-free”. The filter periods are rain-free for the July 22 rain event in ABC1 through to ABC12 as shown by the orange and dark green bars, while the filter periods are rain-influenced in ABC7 through to ABC12 as shown in the light green bars. Similarly, the projection periods are rain-influenced in ABC1 to ABC6 as shown by the pink bars and rain-free in ABC1 to ABC12 as shown in the red bars. In both cases (filter and projection), the rain-influenced period is considered to be from the beginning of the rainfall (July 22, 9 EST in Figure 4) to the end of the corresponding filter or projection period that “sees” the precipitation.

Recall that MESH is run in a continuous simulation mode for the period of June 2002 to November 2014, with a detailed analysis of the ensemble selection methodologies from June 1 to October 31, 2014. Within this time period, there are five significant precipitation events. The beginning and ending of the precipitation events are considered as follows:

- July 22, 14 UTC (9 EST) to 20 UTC (15 EST)
- August 11, 14 UTC (9 EST) to 20 UTC (15 EST)
- September 10, 08 UTC (3 EST) to September 11, 02 UTC (September 10, 21 EST)
- September 19, 20 UTC (15 EST) to September 20, 20 UTC (15 EST)
- October 3, 08 UTC (3 EST) to 20 UTC (15 EST)

For the rain-influenced and rain-free periods, the quality of the ensemble mean, distribution and skill are compared. The ~~skill is calculated with an unskilled reference forecast, which in this study is taken to be the measured streamflow at 00 UTC and 12 UTC each day as the forecast for the next 72 hours. This reference forecast is a persistence forecast, which assumes the streamflow is persistent for the forecast period. The~~ June 3 rain-influenced period is not assessed due to the fact that it was not a projection period in the preceding streamflow ~~and precipitation~~ filter.

2.8.3 Verification of the H-EPS

As with the earlier analysis when precipitation uncertainty is minimized, the mean error and CRPSS are calculated for streamflow for both rain-influenced and rain-free periods as determined by precipitation events in the basin. The overall mean error and CRPSS is also calculated.

The mean error, rank histograms and CRPS of the REPS precipitation ensemble mean are calculated using CaPA as the observation. The CRPSS is also calculated using the June to October CaPA “climatology” as the ~~unskilled~~ reference forecast. The mean error, CRPS and CRPSS are calculated above and below the ninetieth percentile of CaPA precipitation (0.42 mm/hr).

3 Results

The results are ordered according to ~~four filter configurations and then the H-EPS~~the ensemble selection methodologies. In all cases, the projection time period of interest is short-term, which is defined here as 3 days.

3.1 ~~Minimized Uncertainty Filter~~Optimal Hind-cast of 3-day Projections

In determining the effectiveness of the P-SEDA approach, it is necessary to see if the method has the possibility of succeeding with ~~accurate~~ prior knowledge of streamflow and precipitation data. ~~Figure ?? shows the match between the top 10 runs of~~

~~each 3-day period for the months of June 1 to October 31 for the years 2003 to 2008. The black lines represent the observed daily streamflow while the red lines represent an overlap of the top 10 matches to streamflow for every 3-day period beginning at 0 UTC and 12 UTC on each day.~~

~~These results indicate that the P-SEDA method using this particular model generally has the capability to simulate the observed streamflow for this basin if the optimal parameters are selected on each 3-day window. The notable exceptions are in early September, 2004 and much of June and July in 2007. A qualitative analysis of the actual precipitation was completed for early June, 2007 (not shown). Historical radar images show that it is possible that CaPA underestimates the precipitation in the basin for these time periods. The presence of streamflow that is not simulated in any of the 10,000 model runs indicates that it is quite likely that CaPA does not produce enough precipitation for this specific basin in the first half of June, 2007. We suspect that the reason for any obvious mis-match between the P-SEDA results and observations is due to limitations in CaPA, which are unavoidable due to the lack of precipitation observations in the basin.~~

~~To examine the P-SEDA approach in more detail at a higher temporal resolution, the method is also~~ For this purpose, the method is applied for June 1 to October 31, 2014 and compared with hourly ~~(rather than daily)~~ streamflow observations. Figure 7a shows precipitation from CaPA. Figure 7b shows the observed streamflow (black) and corresponding ~~top~~ optimal model runs (red). Figure 7c shows the corresponding basin-average water storage ~~values~~ state variables for each of the parameter-state pairs chosen by the ~~minimized uncertainty filter~~optimal hind-casts. These storage results will be discussed later.

For this study, the qualitative results in ~~Figures **??** and~~ Figure 7b) ~~illustrate~~ illustrates that CaPA precipitation cascades to reasonable streamflow values ~~most of the time. Since the objective of this study is to examine the effectiveness of the P-SEDA approach, we can simply ignore time periods where CaPA clearly fails to produce the appropriate precipitation for the basin. This allows us to focus our remaining analysis on the hourly results from the beginning of June to the end of October, 2014. Focusing on this time period also has the advantage of illustrating some results using the latest version of ECCC's meteorological REPS, which was implemented on December 4, 2013.~~

for the time period examined. A quantitative analysis compares these results to the other ~~filters~~ensemble selection methodologies, but a qualitative analysis is first performed for each of the ~~filters.~~

3.2 ~~Bulk Calibration Filter~~

~~This filter uses the more classical hydrological approach of bulk calibration (or parameter estimation), where optimal parameter sets are based on a long time-series of streamflow. In this case, the top parameter sets are filtered based on a calibration of the daily streamflow from June 1, 2003 to October 31, 2008, and then applied for the period of June 1 to October 31, 2014.~~

~~A rigorous calibration is not performed in favor of using the same LHS parameter sets to sample the parameter space. Other approachesto bulk calibration would certainly prove to be more fruitful if finding an ensemble of optimized runs from bulk calibration was the goal of this research, but using the same pool of parameter sets for all of the filters explored here allows for a fair comparison of different filter techniques for the P-SEDA method~~remaining approaches.

~~The top 10 RMSE simulations (from the 10,000 LHS runs) for the period of June 2, 2002 to June 1, 2009 are selected from the continuous simulations for analysis for the period of June 1 to October 31, 2014, the results of which are seen in Figure 7d.~~

~~The results are similar if any June to October period for 2002 to 2008 is used to determine the top 10 RMSE simulations (not shown). With the exception of early June, most calibrated runs overestimate the observed streamflow.~~

3.2 Preceding Streamflow Filter

~~Another~~ One manner in which to filter the parameter sets (and associated states) is to consider only the preceding streamflow. In this study, the best RMSE values from the preceding ~~3~~ days of streamflow are used to determine the parameter sets to use for the prediction of the subsequent 3 days of streamflow. This process is repeated twice daily at 0 UTC and 12 UTC (19 and 5 local time for the basin in question) for June 1 to October 31, 2014.

Figure 7~~e shows~~ d, e and f show the overall results. Qualitatively, the filter produces good results when there is negligible precipitation. However, the results degrade when it rains, particularly for the 3-day filter. To illustrate this aspect of the filter in more detail, Figures 8a and 8b show how the 3-day filter reacts for a single rain event on July 22, 2014. In Figure 8a, which is the equivalent of ABC6 in Figure 4, the filter period does not "see" the rain and the projected streamflows resulting from the filtered parameter sets overestimate the actual streamflow. In Figure 8b (equivalent to ABC7 in Figure 4), the rain event occurs during the filter period and the subsequent projected streamflows are much more closely aligned with the observations. This result is consistent with all significant precipitation events, with the filter choosing parameter sets that overestimate streamflow when the precipitation event is not "seen" by the filter. However, Figure 7e and f show that the impact of not seeing the precipitation event is reduced with a longer filter period.

3.3 Hindsight Parameter Constraint and Preceding 3-Day Streamflow Filter

The third filter explored here is one in which the top simulations are selected based on the preceding streamflow and parameter ranges that are proven to be important during the 2014 precipitation events. Figure 9 shows parameters that are particularly sensitive during six precipitation events. Each parameter is normalized between 0 and 1. Based on a subjective visual analysis of these box-plots, the 10,000 parameter sets are reduced to 91 parameter sets by confining the values of the normalized parameters as follows: KS1 < 0.1, WF_R2 > 0.6, CLAY11 > 0.5, CLAY12 > 0.5, SDEP1 > 0.2. Using the preceding streamflow filter with these 91 parameter sets to obtain the top 10 runs for each 3-day period yields Figure 10. With the exception of the June 3 precipitation event, these results are clearly much better than those found in ~~Figures 7cand 7e.~~ unconstrained 3-day filter shown in Figure 7c. Although this method clearly cannot be used in a forecasting context, the significance of these findings are examined in the discussion.

3.4 A Quantitative Comparison of Filters

Table 4 shows the mean error of the ensemble mean for the previously defined ~~rain-influenced and~~ rain-free and rain-influenced periods for the ~~3-day filter and the three projection methods, and the mean error for the~~ reference forecast, the optimal hind-cast, the ~~unskilled forecast. Most~~ preceding streamflow filter (with various lengths of time for the filter period), and the 3-day filter with constrained parameters. All of the methods, including the ~~unskilled method~~reference forecast, provide reasonable results for the rain-free periods. ~~The only exception is the bulk filter, which shows a slight positive bias in the predictions.~~ For the rain-influenced periods, which are the real periods of concern for this study, the ~~filter~~ optimal hind-cast is capable of finding parameter sets that have a low mean error~~at 24, 48 and 72 hours~~. The 3-day ~~projection period~~ filter performs the worst in terms of overpredicting streamflow in rain-influenced periods, ~~followed closely by the bulk calibration~~with results improving as the length of the filter period increases. The 3-day ~~projection~~ filter with constrained parameters performs close to the ~~3-day filter~~ optimal hind-cast with only a slight over-prediction of the observed flows.

~~Although not shown, the rank histograms illustrate that the 3-day filter is under-dispersive for the rain-free periods and over-predicts the rain-influenced periods. The bulk filter over-predicts both periods while the 3-day filter with constrained parameters over-predicts the rain-free periods and generally has the correct average spread for the rain-influenced periods.~~ The number of "top" runs selected (k_M) does not appear to have much influence over these mean error results. As a consequence the remainder of the analysis if performed with a value of $k_M = 10$.

Using the current streamflow as the reference forecast, Table 5 shows the skill of the ~~filter and projection methods using the current streamflow as the unskilled forecast. The filter exhibits a very~~ optimal hind-cast, 20-day filter and 3-day filter with constrained parameters. The optimal hind-cast exhibits a relatively high skill for rain-free ~~periods. Also for~~ and rain-influenced periods. For the rain-free periods, the ~~3-day projection~~ 20-day filter shows some skill for the 48 and 72 hour forecast, while the 3-day projection with constrained parameters ~~and the bulk projection show~~ shows no skill. For the rain-influenced periods, the only ~~projection~~ filter that shows any skill is the 3-day projection with constrained parameters. These results quantify the qualitative analysis shown in Figures 7 ~~a to 7c~~ and 10.

3.5 H-EPS

To address the question of how this data assimilation approach could be used in a forecasting context, a full H-EPS is used to force selected parameter-state ensemble members with ECCC's Meteorological Regional Ensemble Prediction System (REPS), as described in the methodology section. ~~Three~~ Two sets of parameter-state ensembles are selected to see how the REPS performs. The ensembles are based on 1) the ~~preceding streamflow filter,~~ optimal hind-cast of 3-day projections and 2) the ~~parameter and preceding streamflow filter, and 3) the minimized uncertainty filter~~ hindsight parameter constraint and preceding 3-day streamflow filter. These ensembles were selected because they were the only filters that showed any skill in the rain-influenced periods. Of course, neither of these ensembles can be used in operational forecasting, so they are used as a proxy for illustrative purposes assuming that the limitations of the preceding streamflow filter can be addressed as explored in the discussion.

Figure 11a) shows CaPA (reddish brown) and the 20 REPS precipitation members (blue). The resulting 200 streamflow ensembles ~~from each of the filters~~ (recall that $k_M = 10$) are shown in ~~Figures 11bto d~~Figure 11b, with the ~~grey lines coming from the preceding streamflowfilter~~black line representing the observed streamflow, the orange lines coming from the hindsight constrained parameter and preceding 3-day streamflow filter, and the green lines coming from the ~~minimized uncertainty filter. In all cases, the black line is the observed streamflow.~~ optimized hind-cast. Even for the ~~minimized uncertainty filter, that~~ optimal hind-cast, which shows near-perfect alignment with the observed streamflow when forced with GEM and CaPA, ~~these~~

~~few~~ the REPS members that overestimate the precipitation have an impact on the resulting ensemble of streamflows. ~~For the preceding streamflow filter, the parameter-state pairs are completely inappropriate for making streamflow projections when it rains (note the difference in scale for flow between Figure 11b to d.)~~

Table 6 shows the mean error for streamflow and Table 7 shows the CRPSS, for both rain-free and rain-influenced periods. The overall mean error and CRPSS are also calculated.

The mean error results show that the H-EPS ensemble mean overestimates streamflow in all cases. The CRPSS scores show that the H-EPS fails to show skill during key time periods for many of the ensembles when compared to using the current streamflow as the forecasted streamflow. This lack of skill will be ~~discussed~~ considered in the discussion. To examine these findings with respect to the precipitation; the mean error, rank histograms and CRPS of the REPS precipitation ensemble mean are calculated using CaPA as the observation. The CRPSS is also calculated using the June to October CaPA "climatology" as the ~~unskilled~~ reference forecast. The mean error, CRPS and CRPSS shown in Table 8 are calculated above and below the ninetieth percentile of CaPA precipitation (0.42 mm/hr). Below this threshold, the REPS mean precipitation over-estimates the CaPA precipitation. In the top 10 percent of CaPA precipitation values, however, the REPS mean under-estimates the CaPA precipitation. The rank histograms (not shown) indicate that the ensemble members tend to underestimate precipitation, although some REPS members do over-estimate the higher CaPA precipitation values. The CRPS shows the highest (worst) values, and the CRPSS shows the least skill, for the highest precipitation rates.

4 Discussion

The discussion is organized around three questions. The first question looks at whether-or-not the P-SEDA approach is capable of reproducing observed streamflow, which corresponds to the ~~minimized uncertainty filter~~optimized hind-cast. The second question considers the effectiveness of the remaining ~~three~~ filtering approaches. The third question revolves around the more realistic example of using the approach in a full H-EPS. Finally, advantages and limitations of the approach are discussed.

4.1 Given maximum data certainty, can the P-SEDA approach reproduce observed streamflow?

Although the P-SEDA filter could be applied to a hydrological model with few parameters, the Canadian MESH model is used with many parameters perturbed. This increases the dimensionality of the problem and is, to our knowledge, the first hydrological application of ABC in a short-term DA application with such a parameter-intensive model. Although much simpler models tend to dominate the operational hydrological modelling community, part of the motivation behind using a hydrologically-enhanced land-surface scheme in the case study is to begin laying some foundation for using such parameter-intensive models for operational ensemble hydrological forecasting.

One major limitation to the way in which MESH is applied in this study is the use of the relatively inefficient Latin Hypercube Sampling to determine the prior distribution of parameter sets to be used with the ABC approach. Despite this limitation, however, the results clearly show that the approach can, with confidence in the precipitation forcing and streamflow, find parameter-state sets that match the observed hydrograph over successive periods of a few days. ~~The exception is when the~~

~~streamflow clearly shows a signal that indicates that precipitation occurred in the basin, but the model is not forced with rain.~~ One possible way of dealing with the uncertainty in precipitation ~~for this basin~~ is to perturb the CaPA precipitation field as is examined by Carrera et al. (2015).

The widely varying nature of the simulated basin storage for the selected runs for each 3-day period also highlights a limitation with the study. This limitation is in only using streamflow as the state variable to determine the top parameters each time. Consider the following water balance equation for the basin: $P - E = R + \mathrm{d}S/\mathrm{d}t$, where P is precipitation, E is evapotranspiration, R is runoff and $\mathrm{d}S/\mathrm{d}t$ is the change in basin storage over time. Over the short time-periods of a few days in short-term hydrological prediction, E can generally be ignored, leaving only $P = R + \mathrm{d}S/\mathrm{d}t$. In the hindcasting exercise presented in this part of the study, P and R are considered to be known and the only remaining term is $\mathrm{d}S/\mathrm{d}t$. So why does the analysis show such a wide range of basin storage terms for the best matching assimilated streamflow? The answer lies in the fact that it is not the basin storage that balances the equation, but rather the change in storage over the time period of interest. The model is capable of releasing or storing the appropriate amount of water in both rain-influenced and rain-free scenarios, and the model determines $\mathrm{d}S/\mathrm{d}t$ based on the interaction of existing storage, model physics and parameters.

The issue of widely-varying simulated basin storage (Figure 7c) also highlights the issue of equifinality, which is defined here as the idea that many different model simulations can produce acceptable results (Beven, 1993). The model is able to find many parameter-state sets that fit the streamflow for short periods of time. If only streamflow observations are available, the selected simulations are equifinal. However, including the state of basin storage ~~in the assessment of equifinality~~ clearly shows that the parameter-state sets are not equal. If soil moisture observations are also available and used, then these simulations are not equifinal and the selected simulations can be further constrained.

One assumption in most environmental modelling exercises is that the parameters do not vary with time, or at least they vary slowly or if the system is disturbed in some way such as land-use change (Bard, 1974; Wagener et al., 2003; Liu and Gupta, 2007). Wagener et al. (2003) indicate that the inability of a single parameter set to simulate an entire streamflow record provides evidence of model structural error. It is incorrect to assume that MESH has a perfect model structure, so the results indicate that any model structural errors can be compensated for by the parameter sets. One can also presume that data errors can also be hidden by the selection of certain parameter sets. Clearly the model needs further constraints to give the results a more solid foundation. One of these constraints could be the assimilation of some aspects of storage in the model. One such possibility would be to examine the usefulness of the soil moisture and ocean salinity (SMOS) satellite ~~(Mecklenburg et al., 2012; Jackson et al., 2012)~~(Mecklenburg et al., 2012; Jackson et al., 2012; Ridler et al., 2014).

4.2 How well do different filters work?

The issues of parameter time-invariance and the most appropriate model structures are generally secondary considerations in DA. The focus in DA shifts from the exercise of improving the model and its parameterization to the exercise of making a more accurate prediction. The results presented from the various filters tested, however, indicate that some thought is required to determine the appropriate parameter sets at the appropriate times.

The only filter in this study that shows any ~~capability~~ skill in predicting streamflow when it rains is the ~~preceding streamflow filter with constrained parameters~~hindsight parameter constraint and preceding 3-day streamflow filter. The manner in which this filter is applied in this study reveals that 91 of the original 10,000 LHS parameter sets can be used effectively with the P-SEDA filter approach to perform short-term predictions in the basin for the months of July to October, 2014. Techniques other than LHS must be explored to obtain more appropriate parameter sets for the P-SEDA method to work in this type of situation.

The fact that constraining the parameter sets allows for the approach to produce reasonable results throughout the period provides some assurance that the method has the possibility of being able to predict streamflow ~~given a certain amount of precipitation~~with some skill. The key, at least in part, is expected to be in using a method other than LHS to determine the prior distribution of parameter sets. Alternative approaches could use algorithms such as Dynamically Dimensioned Search - Approximation of Uncertainty (DDS-AU) which have been shown to be more efficient than GLUE (Tolson and Shoemaker, 2008). The prior can also be obtained by looking for parameter sets that perform well for different hydrological signatures (e.g. Zhang et al., 2014; Shafii and Tolson, 2015) or different hydrological scenarios which might include streamflow responses to snow-melt, runoff over frozen ground, rain during wet conditions, rain during dry conditions, or whatever else can be considered a relevant hydrological event affecting streamflow.

As shown in this study, increasing the length of the filter period has a positive impact on the scores. The gains in mean error values do not improve after 20 day filter periods, indicating that there is a limit to the value of longer filter periods. In this study, the 20 day filter period allows the method to see the previous precipitation event in all cases examined. As such, the ability of the filter to capture important hydrological responses is critical to improving results. The downside to having a longer filter period, however, is that the forecaster must wait longer to apply the approach. This filter-period time limitation for the forecaster may not be true for basins where snow, ice and frozen ground are not dominant processes. We expect that different basins will have different optimal filter period lengths depending on the important hydrological processes in the basin.

If given more information about the state of the basin (other than streamflow), different hydrological scenarios could also be used in determining the appropriate parameter-state sets to filter. For example, if the SMOS satellite indicates that the basin is dry, the streamflow observation is relatively low and a certain amount of precipitation is expected in the near future, then past scenarios that fit this description could be used to filter the parameter sets. As a result, parameter sets that fit both the current state of the basin as well as the expected forcing could be filtered, if both the current basin state and expected precipitation has been previously experienced and observations are available.

Such an approach is very similar to the well-established k-nearest neighbor (k-~~nn~~-nn) bootstrap method as described by Lall and Sharma (1996). In its simplest form, the k-~~nn~~ -nn approach finds k similar patterns in the past data and uses this information to make a prediction about the next data point. The P-SEDA preceding streamflow filter essentially does the same thing, except that it looks for similar patterns in an ensemble of model runs rather than in a time series of data points. By including criteria beyond streamflow as suggested in the previous paragraph, one could (for example) look for past parameter sets that successfully simulated the streamflow when the basin exhibited a certain threshold of upper-layer soil moisture from SMOS, a given streamflow, and a specified amount of precipitation. This approach requires a relatively long time series of observational data with model simulations and could provide an interesting comparison between the model-centric P-SEDA filter and purely data-driven analogue methods.

4.3 How can this approach be used in a forecasting context (including precipitation uncertainty)?

The mean error results for the H-EPS ensemble mean streamflow, forced with the REPS (Table 6), are similar in nature to the mean ensemble streamflow forced with GEM and CaPA (Table 4). ~~The H-EPS with the REPS generally performs better in rain-free periods than rain-influenced periods, and the preceding streamflow projection during rain-influenced periods performs the worst.~~ However, an important finding is drawn from the CRPSS scores in Table 7. Overall, the ~~two projected parameter-state pair ensembles do~~ 3-day filter with constrained parameters does not show skill, ~~particularly for~~ with the exception being hour 72 for the rain-influenced ~~time periods. The H-EPS skill is lowest for the parameter and preceding streamflow filter, for both the overall and for the rain-free periods. But the lack of skill during the predominantly rain-free periods of this study is offset by the improved skill for the rain-influenced~~ periods. ~~In particular, the third day shows some skill for the ensembles resulting from the constrained parameters.~~

These findings illustrate that the H-EPS contains too much uncertainty to be used with any skill for this particular study. It is important to note that the same lack of skill may not be true for other time periods or different basins. For this particular study, it is not surprising that the REPS does not show any skill when compared to using the current streamflow as the forecast. For this basin and the time period considered, the streamflow is not very responsive to the precipitation input for much of the time. Situations when the river is not responsive to precipitation favor the approach of using the current streamflow as the forecast.

A resulting question is whether or not the lack of skill in the H-EPS is due to the uncertainty in the REPS precipitation, or the unresponsive behaviour of the streamflow to precipitation during this period. Looking more closely at the REPS precipitation mean error compared to CaPA (Table 8) indicates that the REPS tends to overestimate the bottom 90 percent, and underestimate the top 10 percent, of CaPA values, which are taken to be as close to observed as is possible in the basin. The only noticeable trend in time is that the underestimation in the top 10 percent of precipitation becomes more pronounced with time.

The unresponsive streamflow in this study is likely due to “fill-and-spill” dynamics (Spence, 2010). Being on the Precambrian Shield, and the starting point of many streams in the basin being small lakes, there are many parts of the basin that need to be filled-up before they contribute to streamflow. This physical process, especially with respect to the headwater lakes, is not represented in the version of MESH used in this study. Future work should focus on this aspect more closely.

Returning to the question of whether-or-not the lack of skill in the H-EPS is due to the uncertainty in the REPS precipitation, or the unresponsive behaviour of the streamflow to precipitation during this period, it seems that both factors contribute to the overall lack of skill. As Figure 11, shows, however, relatively small differences in precipitation result in large changes to streamflow, indicating that the land-surface physical processes (e.g. fill-and-spill) that determine the responsiveness of the streamflow to precipitation, are probably the more important of the two for this particular study.

~~The only other use of ECCC’s REPS as a part of an H-EPS is found in Abaza et al. (2013), in which the Canadian operational meteorological Global Ensemble Prediction System (GEPS) was compared with the REPS and the deterministic 15 km GEM NWP forcing of the province of Quebec’s operational streamflow forecasting system. The study found that both the GEPS and~~

~~REPS outperformed a deterministic run for eight watersheds ranging in size from 355 to 5820 km^2. The REPS was also found to be superior to the GEPS in terms of its ability to predict forecast uncertainty.~~

~~One issue highlighted in the conclusion of Abaza et al. (2013) is that the REPS was found to produce unusually high precipitation spikes. This issue of excessive precipitation was, in many cases, determined to be caused by the physics perturbation scheme that was used to generate the ensemble (Erfani et al., 2014) and was fixed in the version of the REPS that was officially released on December 4, 2013. The update to the REPS is one of the main reasons for focusing on 2014 as a period of interest.~~

4.4 Advantages and Challenges of the Approach

One key benefit of the P-SEDA filter is that it is conceptually straight-forward. In plain language, the idea is to setup a series of continuous simulations and draw the most appropriate runs from these simulations for making a projection or forecast. This concept is very easy to understand and implement. In an operational forecasting environment, in which DA approaches are fundamentally designed to support, this simplicity is desirable.

~~Another advantage is that the parameters and state variables are always consistent with one-another. This cannot be said for other approaches such as the dual Particle Filter or dual Ensemble Kalman Filter.~~

As the examples provided in this study have shown, the approach is also flexible. It can be used in the more traditional manner of hydrologic model calibration ~~,~~ by selecting multi-year filter periods, which hs not been shown here, or in other unique ways that have been examined and discussed throughout the paper. It can be seen as a more general approach to model calibration ~~, which is characterized in this study as the bulk calibration filter~~ and represents a hybrid approach to DA that blends the particle filter and variational DA methods.

Two challenges with the approach are 1) how to determine the prior parameter sets to run in continuous simulations, and 2) how to select the most appropriate runs for making a projection or forecast. This study uses a parameter-intensive H-LSS and deals with the first challenge by using ~~the relatively inefficient~~ LHS to determine the prior, and deals with the second challenge by comparing ~~three filters~~ different filter period lengths to select the appropriate runs. There are likely better ways of dealing with these challenges than have been explored here, and one proxy method (the hindsight parameter constrined and preceding 3-day streamflow filter) has been explored in lieu of these other potential methods.

Fortunately, there is an exhaustive body of research and a number of existing tools that can be used to overcome these challenges. Possible solutions to determine a better prior include: 1) selecting parameter sets based on more than just streamflow, 2) selecting parameter sets based on different hydrological signatures or aspects of the streamflow 3) using ~~k-nn~~ k-nn type approach of looking for parameter sets that worked in similar circumstances in the past, 4) using more efficient algorithms than LHS to determine the prior. Any or all of these methods can be used together to improve the determination of the prior. In terms of selecting the most effective particles once the prior has been established, one method that can be explored is to use more than streamflow to select the top particles with the ABC method. The length of the filter period is also a consideration that needs further exploration.

For both determining a better prior and selecting the most effective particles once the prior has been established, remote sensing offers such opportunities to gather information on the watershed state (e.g. soil moisture, snow) that can complement the limited information that streamflow provides. This approach would better constrain the model in the parameter and state estimation process~~, and thus help to reduce the equifinality issue~~. Using different hydrological signatures, or segmenting the hydrographs for different parameters (e.g. groundwater parameters during low flows), are also ideas worth exploring.

The effectiveness of these methods requires further study.

5 Conclusions

The main contribution of this work is the introduction of a new DA method (P-SEDA)~~that ensures the compatibility of parameters and states in the context of a H-EPS based on a parameter-intensive H-LSS~~. The DA method ~~simplifies~~ is a hybrid of the traditional particle filter ~~method by always returing~~ and variational DA. The method always returns to the initial particles and ~~removing~~ removs the need to resample the parameter space between each model run. The weighting of each particle from the original set of particles is then determined ~~using a rectangular filter,~~ by assigning each particle a value of zero or one. This ~~simplified PF~~ approach is the same as applying the ABC algorithm.

In this study, one filter is to use ABC ~~to select the top runs from a long time-period (bulk filter), a second filter is to apply ABC for only the preceding three~~ for the preceding days of streamflow every 12 hours (preceding streamflow filter)~~, and the third filter~~. It is shown that increasing the length of time for the filter period generally improves the results, up to a point (in this study example, 20 days). A second filter is the same as the ~~second~~ first filter with a parameter-constrained subset of the original 10,000 runs (preceding 3-day streamflow filter with parameter constraints). The parameter constraints are determined from an analysis of the filter-period results during rain-influenced periods.

The ~~preceding streamflow filter-period~~ optimal hind-cast results clearly show that the model and LHS method of sampling 10,000 prior parameter sets is capable of simulating the streamflow for any three-day period where the precipitation input is reasonable. The ~~three~~ methods tested to select the most appropriate runs, however, show that making a projection is more complicated. The only method that consistently shows reasonable projections in this work is the preceding streamflow filter with parameter constraints. The problem with this filter is that it is not immediately clear how such a filter can be used in a forecasting context. Something more is needed to provide better parameter estimates if the P-SEDA filter is to be useful in an operational forecasting setting. Fortunately, there are a number of approaches that can be explored to provide superior guidance on the parameters, either in pre-determining the prior or in selecting the most appropriate runs from the prior.

In addition to introducing P-SEDA, a fuller H-EPS is presented that includes forcing uncertainty from ECCC's REPS. For this particular basin and time-period, the resulting H-EPS is shown overall to be less skillful than using the current streamflow as the forecast for the future streamflow, likely due to model structural errors in MESH. This result is not generally applicable as one should expect the current streamflow to be a fairly good indicator of future streamflow when the stream is relatively unresponsive to precipitation inputs, as is the case in this study. It is expected that the REPS precipitation in an H-EPS would exhibit more skill in more responsive basins without the same fill-and-spill physical processes or for more responsive time periods in this basin.

Acknowledgements. We gratefully acknowledge Environment and Climate Change Canada, the Natural Sciences and Engineering Research Council (NSERC) and Hydro Quebec for funding this study. We also thank Ethan Johnson for his editorial support and the thesis committee for their comments on an earlier draft of this paper that is a part of the first author's PhD dissertation. Finally, we thank the reviewers and journal editor for their helpful comments.

References

Abaza, M., Anctil, F., Fortin, V., and Turcotte, R.: A comparison of the Canadian global and regional meteorological ensemble prediction systems for short-term hydrological forecasting, Monthly Weather Review, 141, 3462–3476, 2013.

Agriculture and Agri-Food Canada: National Ecological Framework, digital media. [Available online at http://sis.agr.gc.ca/cansis/nsdb/ecostrat/index.html.], 2015.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, Signal Processing, IEEE Transactions on, 50, 174–188, 2002.

Asch, M., Bocquet, M., and Nodet, M.: Data assimilation: methods, algorithms, and applications, vol. 11, SIAM, 2016.

Bard, Y.: Nonlinear parameter estimation, vol. 513, Academic Press New York, 1974.

Beck, M. B.: Water quality modeling: a review of the analysis of uncertainty, Water Resources Research, 23, 1393–1442, 1987.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in water resources, 16, 41–51, 1993.

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrological Processes, 6, 227–246, 1992.

Bi, H., Ma, J., Qin, S., and Zhang, H.: Simultaneous estimation of soil moisture and hydraulic parameters using residual resampling particle filter, Science China Earth Sciences, 57, 824–838, 2014.

Biau, G., Cérou, F., and Guyader, A.: New insights into Approximate Bayesian Computation, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 51, 376–403, 2015.

Brown, J. D., Demargne, J., Seo, D.-J., and Liu, Y.: The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations, Environmental Modelling & Software, 25, 854–872, 2010.

Carrera, M. L., Bélair, S., and Bilodeau, B.: The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study, Journal of Hydrometeorology, 16, 1293–1314, 2015.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC-MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation, Mon. Wea. Rev., 126, 1373–1395, 1998a.

Crins, W. J., Gray, P. A., Uhligh, P. W., and Wester, M. C.: The Ecosystems of Ontario, Part 1: Ecozones and Ecoregions, Tech. Rep. SIB TER IMA TR-01, Ontario Ministry of Natural Resources, Peterborough, Ontario, 2009.

Demargne, J., Brown, J., Liu, Y., Seo, D.-J., Wu, L., Toth, Z., and Zhu, Y.: Diagnostic verification of hydrometeorological and hydrologic ensembles, Atmospheric Science Letters, 11, 114–122, 2010.

Dingman, S. L.: Physical Hydrology, Prentice-Hall Inc., 2 edn., 2002.

Drécourt, J.-P., Madsen, H., and Rosbjerg, D.: Calibration framework for a Kalman filter applied to a groundwater model, Advances in Water Resources, 29, 719–734, 2006.

Erfani, A., Frenette, R., Gagnon, N., Charron, M., Beauregard, S., Giguère, A., and Parent, A.: The New Regional Ensemble prediction System (REPS) at 15 km horizontal grid spacing (from version 1.1.0 to 2.0.1), Tech. re., Meteorological Research Branch, National Predictions Development, and National Operations Divistions at the Canadian Meteorological Center, Environment Canada, 2014.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, 2008.

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., Seyfried, M., Moran, M. S., Goodrich, D. C., Kerr, Y. H., et al.: Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the US, Geoscience and Remote Sensing, IEEE Transactions on, 50, 1530–1543, 2012.

Kouwen, N., Mousavi, S., Singh, V., Frevert, D., et al.: WATFLOOD/SPL9 hydrological model & flood forecasting system., Mathematical models of large watershed hydrology, pp. 649–685, 2002.

Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, Journal of Hydrology, 249, 2–9, 2001.

Labarre, D., Grivel, E., Berthoumieu, Y., Todini, E., and Najim, M.: Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters, Signal Processing, 86, 2863–2876, 2006.

Lall, U. and Sharma, A.: A nearest neighbor bootstrap for resampling hydrologic time series, Water Resources Research, 32, 679–693, 1996.

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P., and Stadnyk, T.: Performance Evaluation of the Canadian Precipitation Analysis (CaPA), Journal of Hydrometeorology, 16, 2045–2064, 2015.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resources Research, 43, n/a–n/a, https://doi.org/10.1029/2006WR005756, 2007.

Mackey, B. G., McKenney, D. W., Yang, Y.-Q., McMahon, J. P., and Hutchinson, M. F.: Erratum: Site regions revisited: a climatic analysis of Hills' site regions for the province of Ontario using a parametric method, Canadian Journal of Forest Research, 26, 1112, 1996.

Mahfouf, J.-F., Brasnett, B., and Gagnon, S.: A Canadian precipitation analysis (CaPA) project: Description and preliminary results, Atmos.–Ocean, 45, 1–17, 2007.

Matott, L. S., Babendreier, J. E., and Purucker, S. T.: Evaluating uncertainty in integrated environmental models: A review of concepts and tools, Water Resources Research, 45, 2009.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979.

Mecklenburg, S., Drusch, M., Kerr, Y. H., Font, J., Martin-Neira, M., Delwart, S., Buenadicha, G., Reul, N., Daganzo-Eusebio, E., Oliva, R., et al.: ESA's soil moisture and ocean salinity mission: Mission performance and operations, Geoscience and Remote Sensing, IEEE Transactions on, 50, 1354–1366, 2012.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resources Research, 41, n/a–n/a, https://doi.org/10.1029/2004WR003604, 2005a.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state–parameter estimation of hydrological models using ensemble Kalman filter, Advances in Water Resources, 28, 135–147, 2005b.

Natural Resources Canada: Land Cover, circa 2000 - Vector, digital media. [Available online at http://wmsmir.cits.rncan.gc.ca/index.html/pub/geobase/official/lcc2000v_csc2000v/doc/Land_Cover.pdf.], 2015.

Nie, S., Zhu, J., and Luo, Y.: Simultaneous estimation of land surface scheme states and parameters using the ensemble Kalman filter: identical twin experiments, Hydrology and Earth System Sciences, 15, 2437–2457, 2011.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E., Caldwell, R., Evora, N., et al.: Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale, Hydrology and Earth System Sciences, 11, 1279–1294, 2007.

Qin, J., Liang, S., Yang, K., Kaihotsu, I., Liu, R., and Koike, T.: Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal, Journal of Geophysical Research: Atmospheres (1984–2012), 114, n/a–n/a, https://doi.org/10.1029/2008JD011358, 2009.

Rakovec, O., Weerts, A., Sumihar, J., and Uijlenhoet, R.: Operational aspects of asynchronous filtering for flood forecasting, Hydrology and Earth System Sciences, 19, 2911, 2015.

Ridler, M.-E., Madsen, H., Stisen, S., Bircher, S., and Fensholt, R.: Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark, Water Resources Research, 50, 8962–8981, 2014.

Shafii, M. and Tolson, B. A.: Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, Water Resources Research, pp. 3796–3814, https://doi.org/10.1002/2014WR016520, 2015.

Soulis, E., Snelgrove, K., Kouwen, N., Seglenieks, F., and Verseghy, D.: Towards closing the vertical water balance in Canadian atmospheric models: coupling of the land surface scheme CLASS with the distributed hydrological model WATFLOOD, Atmosphere-Ocean, 38, 251–269, 2000.

Soulis, E., Craig, J., Fortin, V., and Liu, G.: A simple expression for the bulk field capacity of a sloping soil horizon, Hydrological Processes, 25, 112–116, 2011.

Spence, C.: A paradigm shift in hydrology: Storage thresholds across scales influence catchment runoff generation, Geography Compass, 4, 819–833, 2010.

Sun, L., Seidou, O., Nistor, I., and Liu, K.: Review of the Kalman-type hydrological data assimilation, Hydrological Sciences Journal, 61, 2348–2366, 2016.

Sutcliffe, R.: Proterozoic geology of the Lake Superior area, in: Geology of Ontario, edited by Thurston, P., Williams, H., RH, S., and Stott, G., vol. 4 Part 1, pp. 627–658, Ontario Geological Survey, 1991.

Tolson, B. A. and Shoemaker, C. A.: Efficient prediction uncertainty approximation in the calibration of environmental simulation models, Water Resources Research, 44, n/a–n/a, https://doi.org/10.1029/2007WR005869, 2008.

Velázquez, J. A., Petit, T., Lavoie, A., Boucher, M.-A., Turcotte, R., Fortin, V., and Anctil, F.: An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting, Hydrology and Earth System Sciences, 13, 2221–2231, https://doi.org/10.5194/hess-13-2221-2009, http://www.hydrol-earth-syst-sci.net/13/2221/2009/, 2009.

Verseghy, D.: CLASS—The Canadian land surface scheme (Version 3.5). Technical Documentation (Version 1), 2011.

Verseghy, D., McFarlane, N., and Lazare, M.: CLASS—A Canadian land surface scheme for GCMs, II. Vegetation model and coupled runs, International Journal of Climatology, 13, 347–370, 1993.

Verseghy, D. L.: CLASS—A Canadian land surface scheme for GCMs. I. Soil model, International Journal of Climatology, 11, 111–133, 1991.

Vrugt, J. A. and Sadegh, M.: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, Water Resources Research, 49, 4335–4345, 2013.

Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resources Research, 41, n/a–n/a, https://doi.org/10.1029/2004WR003059, 2005.

Wagener, T., McIntyre, N., Lees, M., Wheater, H., and Gupta, H.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, Hydrological Processes, 17, 455–476, https://doi.org/10.1002/hyp.1135, 2003.

Xie, X. and Zhang, D.: A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter, Water Resources Research, 49, 7350–7365, 2013.

Zhang, Y., Vaze, J., Chiew, F. H., Teng, J., and Li, M.: Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall–runoff modelling, Journal of Hydrology, 517, 936 – 948, 2014.

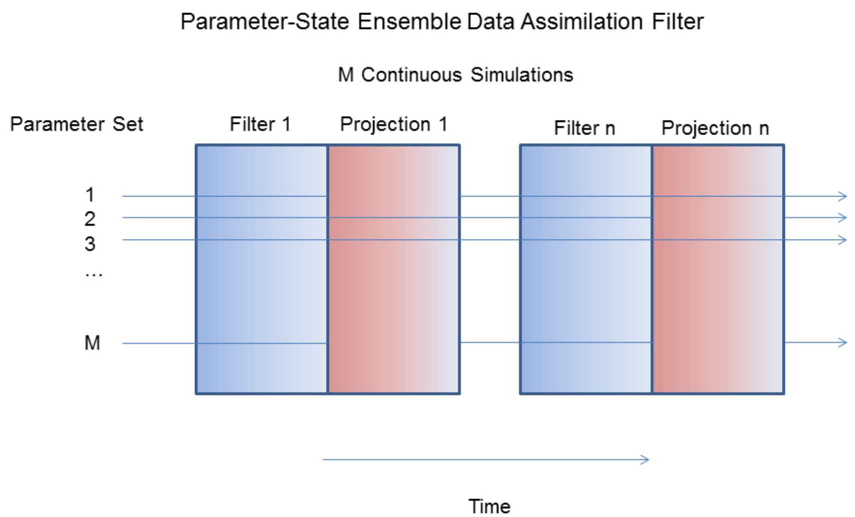

Figure 1. ~~Little Pic River basin near Coldwell, Ontario, Canada~~Schematic of the P-SEDA filter. M simulations are run continuously from a ~~) Location~~ model, of which the ~~basin and legend, b) basin outline with respect~~ filter chooses a number from which to ~~ecodistrict, c) river network and gauge location (02BA003), and d) landcover (MW~~ analyze a projection. The process is ~~Mixed Wood~~then repeated for subsequent filter periods, ~~CF is Coniferous Forest, BL is Broadleaf Forest, W is Water, and O is other)~~noting that the M simulations run continuously through the previous projection periods even though they are not all selected for the previous projection period analysis.

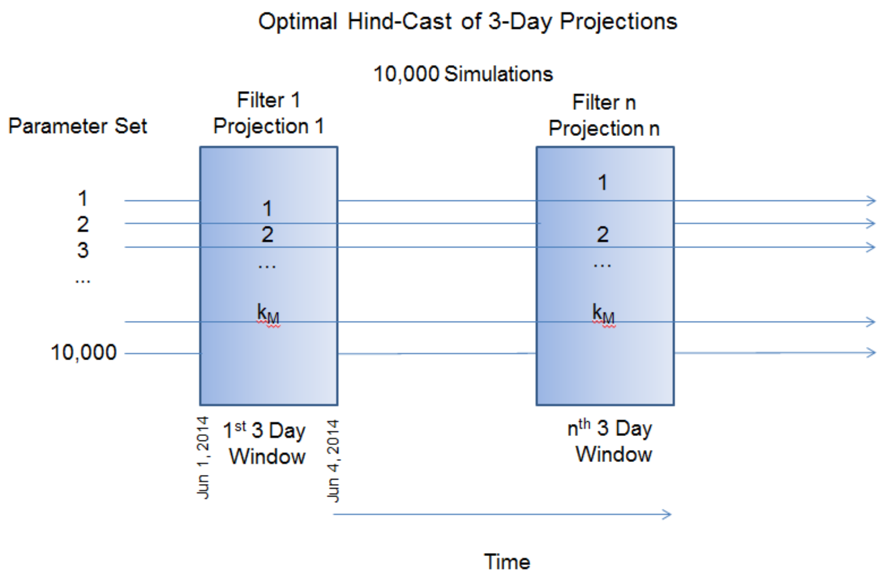

Figure 2. Schematic of the P-SEDA filter for the optimal hind-cast of 3-day projections used in this study. 10,000 simulations are run continuously through the MESH model, of which the filter chooses a number (k_M) for the hind-cast analysis. The process is then repeated for subsequent filter periods.

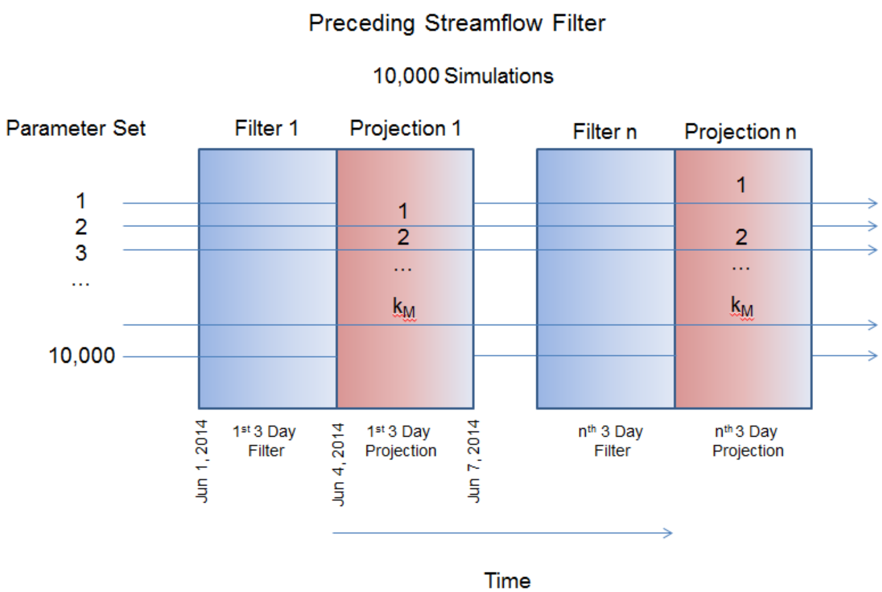

Figure 3. Schematic of the P-SEDA filter for the preceding streamflow filter. 10,000 simulations are run continuously through the MESH model, of which the filter chooses a number (k_M) from which to analyze a projection. The process is then repeated for subsequent filter periods, noting that the M simulations run continuously through the previous projection periods even though they are not all selected for the previous projection period analysis.

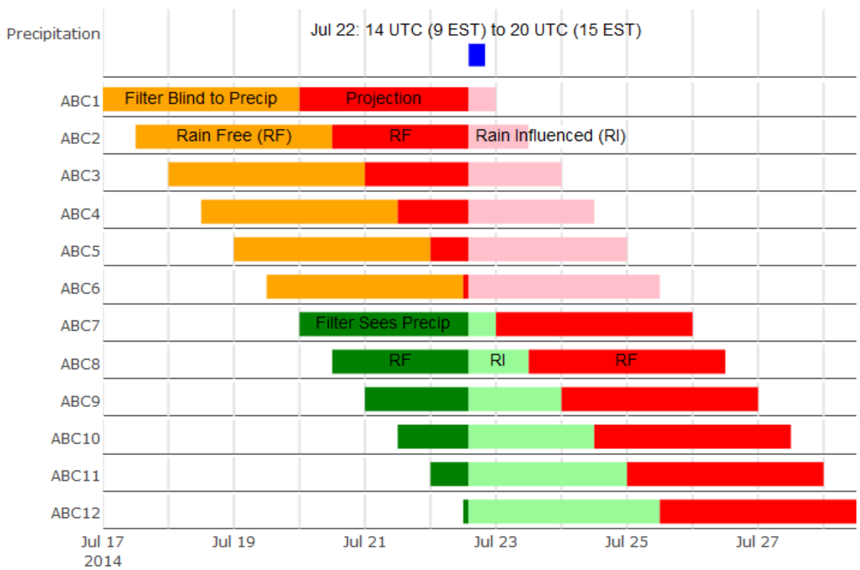

Figure 4. Preceding Streamflow Filter and Projection Periods for July 17 to July 28, 2014 using Hourly Streamflow. This Figure is fully explained in sections 2.7.2 and 2.8.2.

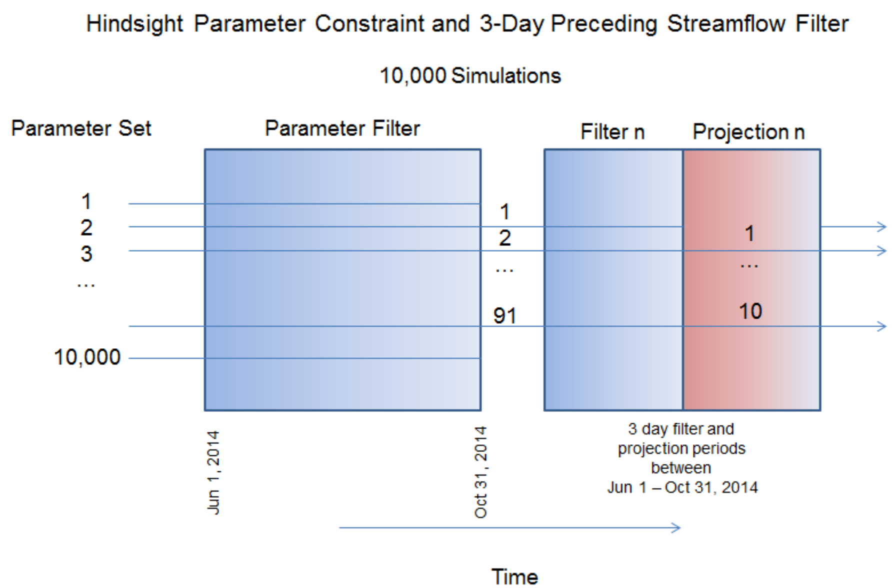

Figure 5. ~~Results~~ Schematic of the P-SEDA filter for ~~multiple~~ the hindsight parameter constraint and preceding 3-day streamflow filter ~~periods~~used in this study. ~~The black lines represent~~ 10,000 simulations are run continuously through the ~~observed daily streamflow while the red lines represent an overlap of the top~~ MESH model. The original 10~~matches~~,000 parameter sets are reduced to ~~streamflow for every 3-day period beginning at 0 UTC and 12 UTC~~ 91 parameter sets based on ~~each day. Where only~~ a ~~black line is seen~~hindsight analysis of parameters that are shown to be important during precipitation events between June 1 and October 31, ~~it simply covers~~ 2014. These remaining 91 parameter sets are then selected for analysis in the ~~red lines completely~~preceding streamflow filter as if $M = 91$ in Algorithm 2.

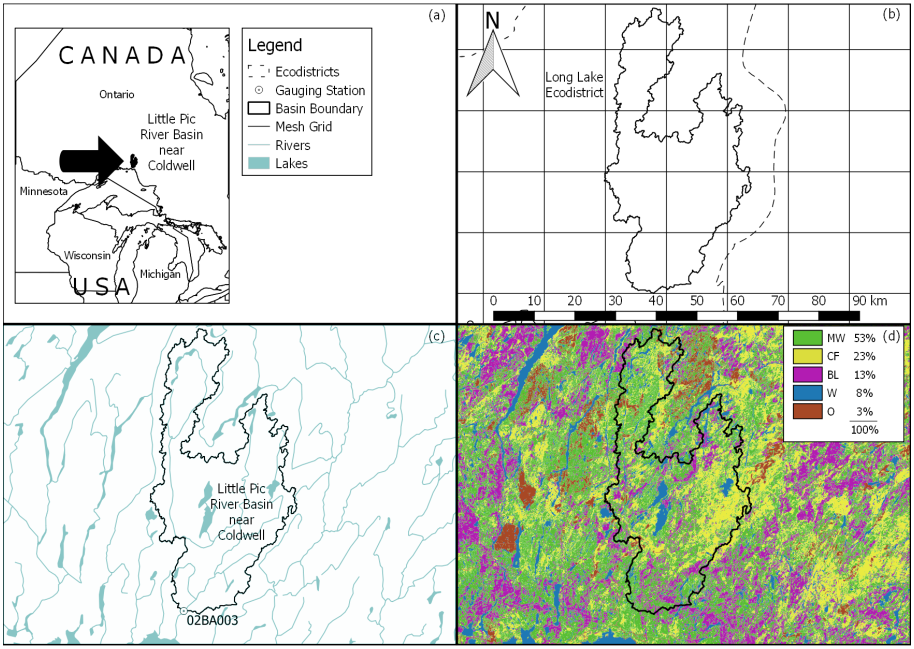

Figure 6. Little Pic River basin near Coldwell, Ontario, Canada. a) Location of the basin and legend, b) basin outline with respect to ecodistrict, c) river network and gauge location (02BA003), and d) landcover (MW is Mixed Wood, CF is Coniferous Forest, BL is Broadleaf Forest, W is Water, and O is other).

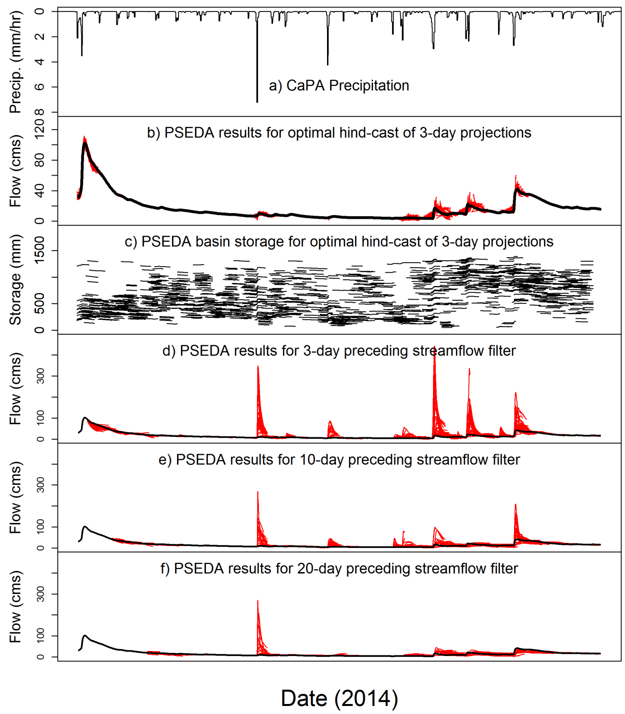

Figure 7. CaPA precipitation (a) for all simulations shown in this Figure. (b) Observed streamflow (black) and top 10 streamflow values (red) for the ~~minimized uncertainty filter~~optimal hind-cast of 3-day projections. (c) Corresponding basin-wide storage values for the ~~minimized uncertainty filter~~optimal hind-cast of 3-day projections. (d) ~~Results~~ Top 10 preceding streamflow filter projections for each of the ~~bulk calibration~~ 3-day filter periods. (e) Top 10 preceding streamflow filter projections ~~in~~ for each of the ~~3-day~~ 10-day filter periods. (f) Top 10 preceding streamflow filter projections for each of the 20-day filter periods. The black lines in (~~b~~d), (~~d~~e) and (~~e~~f) show observed streamflow with different y-axis scaling than in (b).

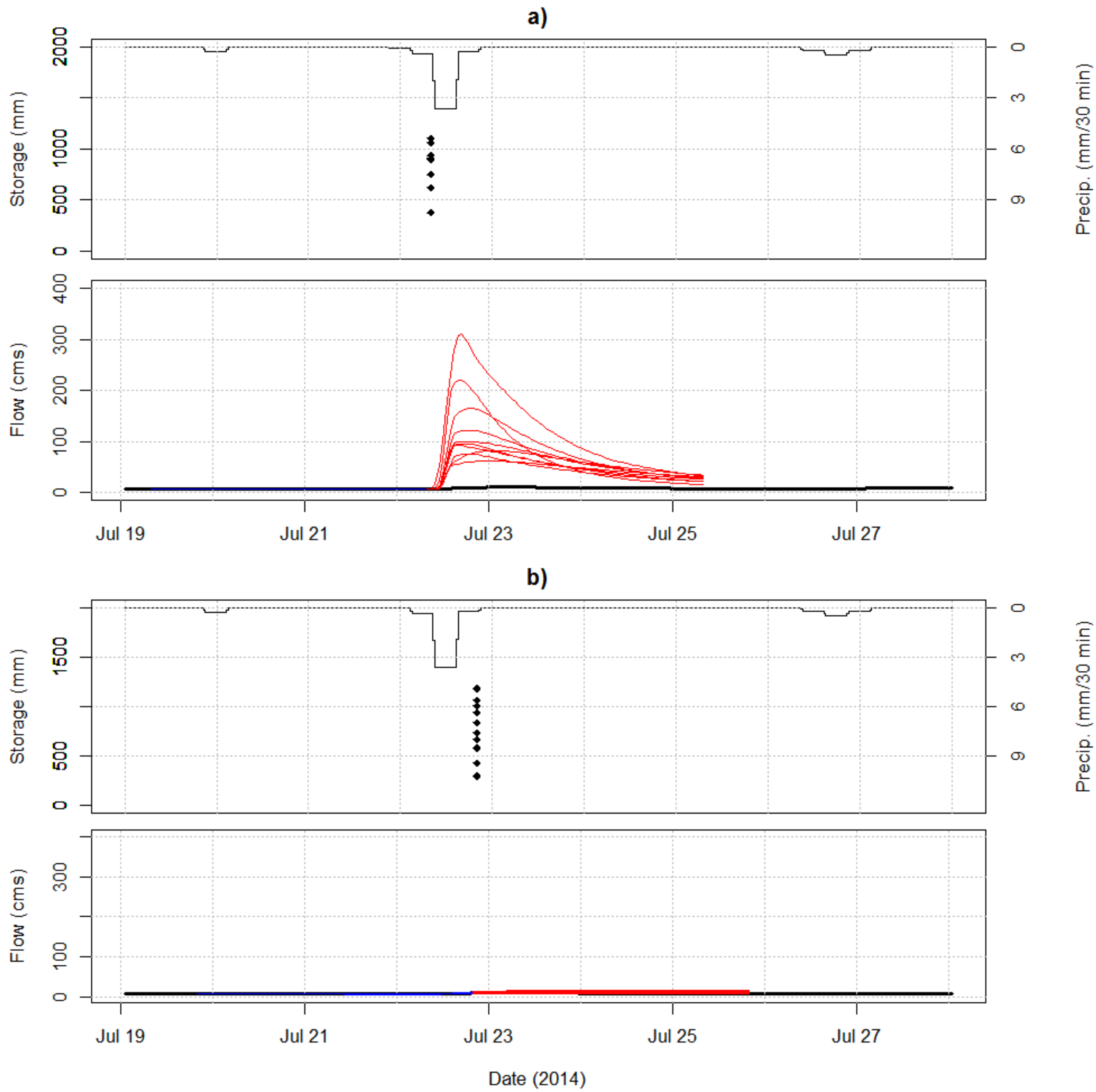

Figure 8. A single filter-projection period for two neighboring time periods. For the two sub-plots in a) the projection begins at 7:00 local time, July 22, 2014 (12 UTC). For the two sub-plots in b) the projection begins at 19:00 local time, July 22, 2014 (0 UTC, July 23). The upper plot of each sub-figure shows CaPA precipitation and instantaneous storage. The lower plot shows observed streamflow (black), the top 10 runs for the filter period (blue), and the corresponding streamflow projections (red).

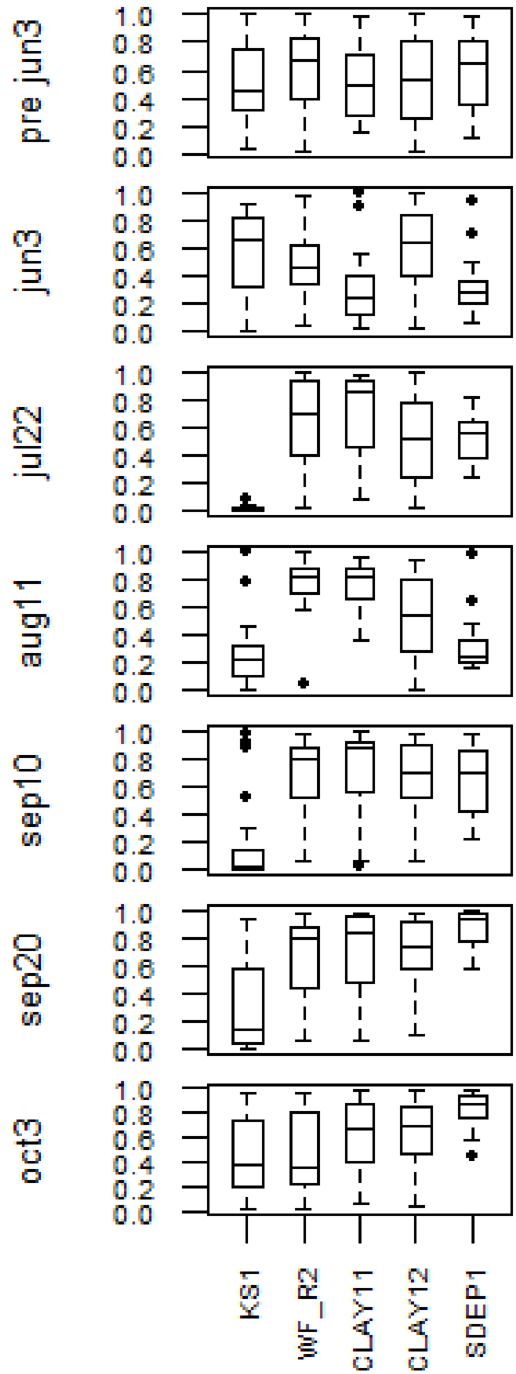

Figure 9. Importance of (normalized) parameters that see precipitation. The top set of box-plots shows that none of the top parameter sets have identifiable parameter values prior to the June 3, 2014 precipitation event. This result is similar to all parameter sets immediately prior to precipitation events that do not see the events. The remainder sets of box-plots show the parameter ranges for the top simulations during precipitation events.

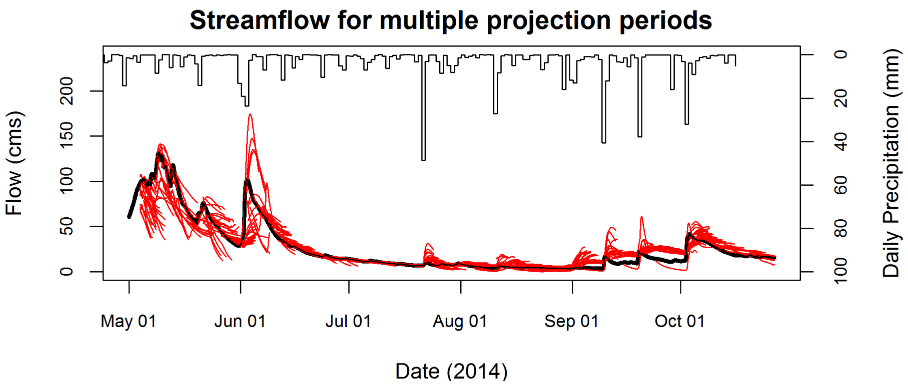

Figure 10. Projection period results after filtering based on parameter values and preceding streamflow.

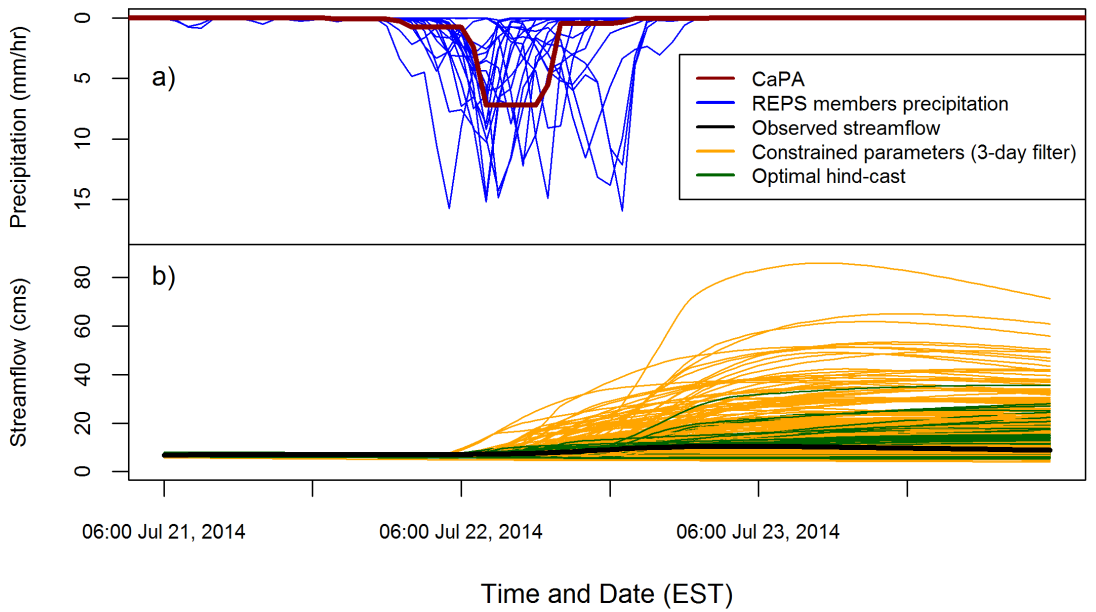

Figure 11. Results of the H-EPS for the 3-day period beginning at 6 Eastern Standard Time (EST) on July 21, 2014. The reddish brown line in sub-figure a) is the Canadian Precipitation Analysis (CaPA) while the blue lines represent the 20 Regional Ensemble Prediction System (REPS) precipitation traces. The single black line in ~~each of the~~ sub-figures b) ~~, c) and d) are~~ is the observed streamflow. The ~~grey,~~ orange and green lines show the 200 H-EPS streamflow traces for the projection periods of the ~~preceding streamflow filter,~~ constrained parameter ~~and~~ with a preceding 3-day streamflow filter, and the ~~minimized uncertainty filter, respectively~~optimized hind-cast.~~Note the difference in scale for streamflow between subfigures b, c and d.~~

Table 1. Landcover percentages based on LCC2000-V Landsat product.

Landcover	Percentage
Water	8
Coniferous Dense Forest	23
Broadleaf Dense Forest	13
Mixed Wood	53
Other	3

Table 2. Fixed landcover parameters.

Parameter Name	Description	Units	Value	Source
QA50 - NL	Reference value of incoming shortwave radiation used in stomatal resistance formula (Needleleaf)	[W m^{-2}]	30	Verseghy (2011)
QA50 - BL	Reference value of incoming shortwave radiation used in stomatal resistance formula (Broadleaf)	[W m^{-2}]	40	Verseghy (2011)
VPDA - NL	Vapour pressure deficit coefficient used in stomatal resistance formula (Needleleaf)	[ ]	0.65	Verseghy (2011)
VPDA - BL	Vapour pressure deficit coefficient used in stomatal resistance formula (Broadleaf)	[ ]	0.5	Verseghy (2011)
VPDB - NL	Vapour pressure deficit coefficient used in stomatal resistance formula (Needleleaf)	[ ]	1.05	Verseghy (2011)
VPDB - BL	Vapour pressure deficit coefficient used in stomatal resistance formula (Broadleaf)	[ ]	0.6	Verseghy (2011)
PSGA - NL	Soil moisture suction coefficient used in stomatal resistance formula (Needleleaf)	[ ]	100	Verseghy (2011)
PSGA - BL	Soil moisture suction coefficient used in stomatal resistance formula (Broadleaf)	[ ]	100	Verseghy (2011)
PSGB - NL	Soil moisture suction coefficient used in stomatal resistance formula (Needleleaf)	[ ]	5	Verseghy (2011)
PSGB - BL	Soil moisture suction coefficient used in stomatal resistance formula (Broadleaf)	[ ]	5	Verseghy (2011)
ROOT - NL	Root depth (Needleleaf)	[m]	0.05	User Selected
ROOT - BL	Root depth (Broadleaf)	[m]	0.05	User Selected
DDEN	Drainage density, equal to the length of the stream divided by area drained by the stream (Basin wide)	[km km^{-2}]	50	Dingman (2002)
XSLP	Average overland slope.	[rise/run]	grid-based	Calculated from Digital Elevation Model
GRKF	Ratio of saturated horizontal hydraulic conductivity at a depth of 1 metre to the saturated horizontal hydraulic conductivity at the surface (Basin wide)	[ ]	0.01	User defined

Table 3. Ranges for the perturbed parameters.

Parameter Name	Description	Units	Lower Limit	Upper Limit	Source
MANN	Manning's n for overland flow.	[m $s^{-1/3}$]	0.02	0.16	Dingman (2002)
KS	Saturated surface horizontal soil conductivity.	[m s^{-1}]	0.00001	0.1	User specified
ZSNL	Limiting snow depth below which coverage is less than one-hundred percent.	[m]	0.1	1	User specified
SDEP	Soil permeable depth, set to greater than model soil depth to simulate fully permeable soil.	[m]	0.1	4.2	User specified
WF-R2	River roughness factor that incorporates a channel shape and width to depth ratio as well as Manning's n.	[$m^{0.5}$ s^{-1}]	0.3	1	User specified
RSMN-NL	Minimum stomatal resistance (Needleleaf)	[s m^{-1}]	175	225	Verseghy (2011)
RSMN-BL	Minimum stomatal resistance (Broadleaf)	[s m^{-1}]	100	150	Verseghy (2011)
SAND-L1	Sand in soil layer 1.	[%]	35	58	Ecodistrict based
SAND-L2	Sand in soil layer 2.	[%]	35	58	Ecodistrict based
SAND-L3	Sand in soil layer 3.	[%]	35	58	Ecodistrict based
CLAY-L1	Clay in soil layer 1.	[%]	0	37	Ecodistrict based
CLAY-L2	Clay in soil layer 2.	[%]	0	37	Ecodistrict based
CLAY-L3	Clay in soil layer 3.	[%]	0	37	Ecodistrict based
LANZ0-NL	Natural log of roughness length (Needleleaf).	[ln(m)]	-0.7	1.1	Verseghy (2011)
LANZ0-BL	Natural log of roughness length (Broadleaf).	[ln(m)]	-0.7	1.1	Verseghy (2011)
ALVC-NL	Visible albedo (Needleleaf).	[ ]	0.02	0.09	Verseghy (2011)
ALVC-BL	Visible albedo (Broadleaf).	[ ]	0.02	0.09	Verseghy (2011)
ALIC-NL	Near infrared albedo (Needleleaf).	[ ]	0.1	0.5	Verseghy (2011)
ALIC-BL	Near infrared albedo (Broadleaf).	[ ]	0.1	0.5	Verseghy (2011)
LAMAX-NL	Maximum leaf area index (Needleleaf).	[ ]	1.8	2.2	Verseghy (2011)
LAMAX-BL	Maximum leaf area index (Broadleaf).	[ ]	4	10	Verseghy (2011)
LAMIN-NL	Minimum leaf area index (Needleleaf).	[ ]	1.4	1.8	Verseghy (2011)
LAMIN-BL	Minimum leaf area index (Broadleaf).	[ ]	0.2	4	Verseghy (2011)
MAXMASS-NL	Standing biomass density (Needleleaf).	[kg m^{-2}]	5	40	Verseghy (2011)
MAXMASS-BL	Standing biomass density (Broadleaf).	[kg m^{-2}]	5	40	Verseghy (2011)
ZPLS	Maximum water ponding depth for snow-covered areas.	[m]	0.1	0.5	User specified
ZPLG	Maximum water ponding depth for snow-free areas.	[m]	0.1	0.5	User specified
DRN	Drainage index, set to 1.0 to allow the soil physics to model drainage or to a value between 0.0 and 1.0 to impede drainage.	[m]	0	1	User specified

Table 4. Mean error ($m^3\ s^{-1}$) as an ~~assessement~~ assessment of the ensemble mean streamflow ~~from~~ for the reference forecast, the optimal hind-cast, and various configurations of the P-SEDA filter~~(10 members) for rain-influenced~~. The value of k_M from Algorithms 1 and 2 varies from 5 to 50 as shown. Rain-influenced and rain-free periods from June to October, ~~2014.~~ 2014 as described in the text.

	Rain Free						Rain Influenced					
	k_M						k_M					
	~~24~~ 5	~~48~~ 10	~~72~~ 20	~~24~~ 30	~~48~~ 40	~~72~~ 50	5	10	20	30	40	50
~~unskilled~~ Reference Forecast	~~0~~ 2	~~1~~ 2	2	~~-5~~ 2	~~-7~~ 2	~~-7~~ 2	-8	-8	-8	-8	-8	-8
~~3-day filter~~ Optimal hind-cast	0	0	0	0	0	0	3	3	3	4	4	4
3-day ~~projection~~ filter	1	2	2	2	2	2	41	42	45	46	46	46
10-day filter	1	1	~~61~~ 2	~~53~~ 2	~~42~~ 2	2	11	10	12	12	12	13
~~bulk projection~~ 20-day filter	1	1	1	2	2	2	4	5	7	8	9	9
30-day filter	2	1	2	2	3	3	4	6	8	9	9	9
40-day filter	2	2	3	3	3	4	6	~~5~~ 7	~~25~~ 8	~~30~~ 9	~~31~~ 9	9
3-day ~~projection~~ filter with constrained parameters	~~2~~ 3	~~2~~ 4	4	4	5	5	2	~~1~~ 4	~~3~~ 4	~~3~~ 5	5	6

Table 5. Mean continuous ranked probability skill score (CRPSS) as an ~~assessement~~ assessment of the ensemble skill from the P-SEDA filter for rain-influenced and rain-free periods from June to October, 2014. The configurations considered here are the optimal hind-cast, the 3-day projection for the 20-day filter, and the 3-day projection for the 3-day filter with constrained parameters. The reference ~~low-skill~~ forecast is the measured streamflow at 00 UTC and 12 UTC each day as the forecast for the next 72 hours. The value of $k_m = 10$ from Algorithms 1 and 2 in all cases.

	Rain Free		
	Forecast Hour		
	24	48	72
~~3-day filter~~ Optimal hind-cast	0.85	~~0.91~~ 0.92	~~0.89~~ 0.90
~~3-day projection~~ 20-day filter	~~0.07~~ -0.09	~~0.09~~ 0.27	~~0.16~~ 0.42
~~bulk projection -3.97 -1.95 -1.06 -1.90 -1.71 -1.89~~ 3-day ~~projection~~ filter with constrained parameters	~~-1.06~~ -0.74	~~-0.60~~ -0.22	~~-0.39~~ -0.02

Table 6. Mean error (mean H-EPS streamflow - observed) ($m^3\ s^{-1}$) as an ~~assessement~~ assessment of the ensemble mean streamflow from the H-EPS (200 members) for rain-influenced, rain-free and overall periods from June to October, 2014.

	Rain Free			Rain Influenced			Overall		
	Forecast Hour			Forecast Hour			Forecast Hour		
	24	48	72	24	48	72	24	48	72
~~3-day filter~~ optimized hind-cast	1	2	3	1	2	7	1	2	3
3-day ~~projection 1 2 3 34 36 34 3 5 7 3-day projection~~ filter with constrained parameters	2	3	3	1	2	4	2	3	3

Table 7. Mean continuous ranked probability skill score (CRPSS) as an ~~assessement~~ assessment of the ensemble skill from the H-EPS for rain-influenced, rain-free and overall periods from June to October, 2014. The reference low-skill forecast is the measured streamflow at 00 UTC and 12 UTC each day as the forecast for the next 72 hours.

	Rain Free			R
	Forecast Hour			F
	24	48	72	24
~~3-day filter~~ optimized hind-cast	0.77	0.76	0.67	-0.3
3-day ~~projection 0.21 0.29 0.31 -8.33 -4.15 -2.44 -0.23 -0.09 -0.02 3-day projection~~ filter with constrained parameters	-0.77	-0.31	-0.14	-1.2

Table 8. Mean error (mean REPS precipitation - CaPA), CRPS and CRPSS (with JJASO, 2014 "climatology" as reference forecast) for June 1 to October 31, 2014.

	Mean Error			CRPS			CRPSS		
Threshold	Forecast Hour			Forecast Hour			Forecast Hour		
	24	48	72	24	48	72	24	48	72
$Pr \leq 0.9(0.42 \text{ mm h}^{-1})$	0.05	0.06	0.09	0.03	0.03	0.04	0.30	0.19	0.08
$Pr > 0.9(0.42 \text{ mm h}^{-1})$	-0.14	-0.21	-0.19	0.45	0.49	0.56	0.38	0.32	0.22
all	0.03	0.03	0.06	0.08	0.08	0.10	0.36	0.28	0.18