Responses to Referee #1

We thank referee #1 for showing interest in our work and the constructive comments. The issues raised are interesting and we (the authors) will address them in the revised version of this manuscript. Here we provide our response to the comments (in italics). Please note that all citations herein refer to the papers in the discussion paper.

Major comments:

1. *The authors claim that the described complementary conceptual and data-driven (error) models is a new approach. However, as stated in Lines 4-5, Page 12067, "Several example applications can be found in the scientific literature on using conceptual and data driven models complementarily", similar works have been found in the previous studies. Furthermore, the HBV model for conceptual model and autoregressive (AR) model for error model are both very mature models in hydrology. Therefore it is hard to find the new contribution or improvement in this paper.*

   We agree with referee #1 that complementary use of conceptual and data driven models is not new to the hydrology community. As referee #1 correctly pointed out, we have cited and reviewed the pioneer works since the 1990s that applied a similar principle. We appreciate this comment as it persuades us of the need to highlight the application aspect and methodological contribution of the present work. In this regard, the paper presents application of the principles of complementary modelling for forecasting inflows into hydropower reservoirs over extended lead times. The new approach we present in the paper deals with the structure the conceptual model fails to capture. It should be noted that earlier works that apply complementary modelling often deal with the bias and persistence structure in the residual series. This paper, however, recognizes that heteroscedasticity seen in the residuals from the conceptual model reflects the failure in the perceptual model, and is important in defining the manner the residual series is dealt with. Accordingly, as outlined in section 2.1.2 of the discussion paper (P12069-P12072), the present study examines the bias, persistence and heteroscedasticity in the residuals and employs an iterative algorithm for estimating parameters of the AR model as well as the transformation parameters.

   The contribution of this work and what distinguishes it from pioneer works will be clearly stated in the revised paper.

2. *Actually, there are many error models at present, e.g. autoregressive model, autoregressive threshold model, fuzzy autoregressive threshold model, ARIMA based error models and artificial neural network models, and so on. This paper selected the autoregressive model to describe the error processes. The reason or additional statement should be given to be clear to the readers. More error models should be used and compared to obtain more reasonable and high accuracy results.*

   The suggestion by referee #1 to conduct comparative assessment of different error models would be an interesting work. Conclusions from previous research works (reviewed in the discussion paper: P12067 L10-L19) that investigated performance of 4 to 8 error-forecast models influenced the selection of the error model. Xiong and O'Connor (2002), in particular, affirm that AR model's longstanding popularity is deservedly right and further emphasize effectiveness of a very parsimonious model such as AR model for error forecasting.

We will clearly state this in the revised manuscript.

3. *This paper attempted to produce probabilistic inflow forecasts through a complementary modelling framework. However, it is known to all that the Bayesian forecasting system (BFS) and generalized likelihood uncertainty estimation (GLUE) may be the two most popular and widely used frameworks to produce probabilistic inflow forecasts. Comparisons of the results of the proposed method and the two methods mentioned above are necessary to verify whether the proposed method are more effective and reliable or not?*

   It is true that the Bayesian forecasting system (BFS), the generalized likelihood uncertainty estimation (GLUE) and the Bayesian recursive estimation (BaRE) are popular methods for producing probabilistic forecasts. In this study, the probabilistic inflow forecasts were produced based on deterministic forecasts, in which we attempted to mimic the operational forecasting method employed in the Norwegian hydropower industry. As demonstrated by Smith et al. (2012), performance of the probabilistic forecasts was assessed based on the fraction of observations contained in a given confidence interval (see Table A1 and Fig. A1) and comparison with a deterministic metric (see Table A2). We believe that this assessment is adequate to evaluate the probabilistic forecasts for the present purpose but agree with referee #1 that intercomparison of the probabilistic forecasts (using the present and the above mentioned techniques) would lead to identifying the more effective and reliable method, and would be an interesting topic for further analysis.

Minor comments:

1. *As shown in Fig. 8, the unit of inflow should be transformed to international unit "m3/s".*

   This can be converted in the revised manuscript if necessary.

2. *Some indexes in the following references can help identify and evaluate the quality of prediction interval, such as the percentage of coverage (POC), the average relative width (ARW) etc.*

   We thank reviewer #1 for the suggested references. We will use the terms coined in those literatures to present our results in the revised manuscript.

Table A1. Fraction of observations bracketed in the 95 % prediction interval for selected forecast lead-times.

Autumn							Winter					
2	6	9	12	18	24		2	6	9	12	18	24
99.9	97.8	97.8	94.5	90.1	89	06/07	99.9	96.7	95.6	95.6	92.2	90
99.9	97.8	97.8	94.5	87.9	83.5	07/08	99.9	97.8	96.7	93.4	95.6	94.5
99.9	98.9	95.6	95.6	93.4	90.1	08/09	99.9	98.9	97.8	97.8	94.4	95.6
99.9	96.7	94.5	91.2	93.4	90.1	09/10	99.9	99.9	99.9	98.9	98.9	97.8
99.9	97.8	97.8	95.6	94.5	91.2	10/11	99.9	96.7	96.7	96.7	95.6	94.4
Spring							Summer					
2	6	9	12	18	24		2	6	9	12	18	24
99.9	95.7	89.1	89.1	88	83.7	06/07	99.9	99.9	99.9	99.9	98.9	97.8
99.9	99.9	98.9	98.9	94.6	94.6	07/08	99.9	98.9	98.9	98.9	98.9	98.9
99.9	98.9	97.8	96.7	95.7	92.4	08/09	99.9	99.9	98.9	98.9	98.9	98.9
99.9	97.8	94.6	93.5	91.3	90.2	09/10	99.9	99.9	99.9	99.9	99.9	98.9
98.9	95.7	92.4	90.2	85.9	82.6	10/11	99.9	98.9	98.9	96.7	96.7	95.7

Table 3. Fraction of observations at Welsh bridge bracketed by the estimated 95 % prediction intervals during calibration. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time periods, high levels (> 2m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.98 (*0.95*,**0.98**)	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.88*,**0.96**)	0.93 (*0.86*,**0.94**)	0.92 (*0.84*,**0.92**)	0.91 (*0.81*,**0.92**)	0.89 (*0.79*,**0.90**)	0.87 (*0.78*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.88*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.89*,**0.95**)	0.94 (*0.89*,**0.95**)	0.94 (*0.90*,**0.95**)
AR	All	0.98 (*0.95*,**0.98**)	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.88*,**0.96**)	0.93 (*0.87*,**0.94**)	0.92 (*0.86*,**0.92**)	0.91 (*0.84*,**0.92**)	0.89 (*0.82*,**0.90**)	0.87 (*0.80*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.88*,**0.96**)	0.95 (*0.89*,**0.95**)	0.95 (*0.90*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.90*,**0.95**)
LLT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.96**)
	High	0.94 (*0.88*,**0.95**)	0.90 (*0.84*,**0.92**)	0.89 (*0.82*,**0.90**)	0.88 (*0.81*,**0.89**)	0.87 (*0.79*,**0.88**)	0.86 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.94 (*0.88*,**0.95**)	0.94 (*0.89*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
DLLT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)
	High	0.94 (*0.89*,**0.95**)	0.92 (*0.83*,**0.92**)	0.90 (*0.81*,**0.90**)	0.89 (*0.80*,**0.89**)	0.88 (*0.80*,**0.88**)	0.87 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.95 (*0.88*,**0.95**)	0.94 (*0.90*,**0.94**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
RWD	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)
	High	0.93 (*0.87*,**0.95**)	0.90 (*0.84*,**0.92**)	0.89 (*0.82*,**0.90**)	0.88 (*0.81*,**0.89**)	0.87 (*0.79*,**0.88**)	0.86 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.94 (*0.88*,**0.95**)	0.94 (*0.89*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
IRW	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)
	High	0.94 (*0.89*,**0.95**)	0.92 (*0.83*,**0.92**)	0.90 (*0.81*,**0.90**)	0.89 (*0.80*,**0.89**)	0.88 (*0.80*,**0.88**)	0.87 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.95 (*0.88*,**0.95**)	0.94 (*0.90*,**0.94**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
SRW	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.87*,**0.96**)	0.92 (*0.86*,**0.94**)	0.91 (*0.85*,**0.92**)	0.90 (*0.83*,**0.91**)	0.89 (*0.81*,**0.90**)	0.88 (*0.79*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.87*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.95**)
DT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.87*,**0.96**)	0.92 (*0.86*,**0.94**)	0.91 (*0.85*,**0.92**)	0.90 (*0.83*,**0.91**)	0.89 (*0.81*,**0.90**)	0.88 (*0.79*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.87*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.95**)
SLLT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.88*,**0.96**)	0.92 (*0.87*,**0.94**)	0.91 (*0.86*,**0.92**)	0.90 (*0.83*,**0.91**)	0.90 (*0.82*,**0.90**)	0.88 (*0.81*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.88*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.90*,**0.95**)	0.95 (*0.91*,**0.95**)	0.95 (*0.91*,**0.95**)

Figure A1. Screenshot of Table 3 from Smith et al. 2012.

Table A2. Relative RMSE reductions (%) in the inflows forecast for selected forecast lead-time (designates relative RMSE reduction of <0).*

Autumn							Winter					
2	6	9	12	18	24		2	6	9	12	18	24
79.3	52.3	41.7	31.6	16.6	10	06/07	87.9	49.3	31.3	17.5	1.3	*
84.4	62.2	50.7	41.4	19.5	*	07/08	81.9	54.1	40	30.8	17.5	18.4
87.9	65.9	54.1	39.4	70	11.5	08/09	83.9	68.4	49.5	36	12.8	5.8
83.2	59.1	47.2	38.1	23.3	18.4	09/10	91.4	78.8	65.9	46.9	32.2	22.2
84.9	57	47.5	40.3	26.2	10.9	10/11	88.7	64.9	52.3	36.9	90	11.1
Spring							Summer					
2	6	9	12	18	24		2	6	9	12	18	24
88.2	66.3	52.3	43.1	30	23.2	06/07	90	76.3	67.4	63.7	49.6	34.4
93.3	79	70.4	64.9	54	46	07/08	81.4	55.6	45.4	39.4	29.3	23.2
90.4	73.7	65.7	58.7	44.7	31.6	08/09	94.4	78.2	52.9	36.7	18.6	11.9
87.7	64.9	50.6	42.5	32.8	28.3	09/10	84.8	71.5	61	51.9	39.3	30
88.6	63.4	48.7	41.7	29.8	23.7	10/11	88.7	64.4	49.8	39.8	28.6	25