

1 **Reviews of paper “Simple Box-Cox probabilistic models for hourly streamflow predictions”**

2 We thank Reviewer 1 for dedicating his/her time and expertise to review our manuscript, and his/her
3 constructive comments.

4 The remainder of this document provides detailed responses to each of R1’s comment and proposes actions to
5 address his/her comments. The reviewer comments are in black and our responses are in *red italics*.

6

7 **Reviewer 1**

8 The paper “Simple Box-Cox probabilistic models for hourly streamflow Predictions” by Prieto et al. aims at
9 improving hourly streamflow prediction by identifying practical methods for quantifying predictive
10 uncertainty, with a focus on how to represent heteroscedasticity and temporal persistence in streamflow errors.

11 I find the paper interesting and relevant for the readers of the journal, and I recommend its publication after
12 minor revisions.

13 *Thank you for your positive feedback and constructive comments.*

14

15 R1.1. A general comment is that the authors appear to assume familiarity with several concepts and
16 methodological choices that would benefit from being briefly clarified within the manuscript. For example, I
17 suggest specifying whether the classification of the catchments (e.g., arid, semi-arid, etc.) is based on indicators
18 such as the P/PET and Q/R ratios, in order to improve clarity for the reader.

19 *Catchment’s classification is based on the aridity index (PET/P) from Arora et al. 2002. We agree this
20 information will be helpful to the readers and will add it to the relevant section of the manuscript.*

21

22 R1.2. Additionally, I recommend removing references to equation numbering from other papers, such as in the
23 sentence: “The volumetric bias metric is also taken from Equation 34 of McInerney et al. (2017).” Instead, the
24 equation should be explicitly reported in the manuscript, and the reference to the original numbering should
25 be removed. This will make the paper more self-contained and easier to follow.

26 *This comment refers to the equations defining the performance metrics. The current manuscript does not state
27 these equations directly and instead refers to the equation number in the original publication, namely Equation
28 23a, 23b from Renard et al. 2010 for reliability; equation 33 from McInerney et al. for sharpness; and equation
29 34 from McInerney et al. 2017 for volumetric Bias. This choice was made to avoid interrupting the flow of the
30 manuscript with detailed equations already published in earlier work. As noted by Reviewer 1, this choice
31 does come at the cost of making the manuscript less self-contained.*

32 *As part of the revisions, we will compare a few options for stating the equations without distracting the readers
33 from newer aspects in this work. For example these equations might be added in an Appendix – this will be
34 evaluated during the revisions*

35

36 R1.3. I kindly suggest reorganizing the manuscript so that all methodological aspects are presented within a
37 single section, rather than being distributed across different parts of the paper. In particular, the content related
38 to methodological procedures starting from Section 3.2 should be moved from the Case Study section to the
39 Methodology section to improve clarity and readability.

40 *The logic behind the current structure is as follows.*

41 *The Theory section presents the central theoretical concepts of the manuscript, i.e., the general form of the
42 probabilistic model in Section 2.1 and the general residual error models in Section 2.2, including the Box-Cox
43 models in Section 2.2.1 and general Autoregressive Moving Average (ARMA) models in Section 2.2.2. These
44 sections provide the theoretical basis of the modelling methods investigated in our study.*

45 *In contrast, the Case Study section lists the more specific choices made in our case study. Notably, the choice
46 of GR4H as the deterministic model, the specific choice of Box-Cox power (λ) values, the order of the
47 ARMA models, as well as the choice of performance metrics, are in effect specific to our Case Study
48 implementation, and in our opinion, it would be confusing to present them in the more general Theory section.*

49 *As such, a reader may wish to apply the probabilistic approaches we propose here with a different choice of*
50 *deterministic model, in other catchments, and evaluate their performance using other metrics. Indeed they may*
51 *also apply different values of Box-Cox lambdas and/or different ARMA models. We will amend the manuscript*
52 *to better explain to the organization of the paper. As part of revisions, we will also double-check to ensure the*
53 *above principle is implemented consistently.*

54

55 R1.4. Additionally, I recommend improving the visual quality and formatting of figures and tables. Figures
56 should use color-blind-friendly palettes to ensure accessibility and clarity for all readers.

57 *Thank you for raising this issue. We will modify the figures to make them color-blind-friendly so they are*
58 *accessible to all readers.*

59

60 R1.5. Line 278: the reference is missing.

61 *Thank you, we will add reference to Section 5.1*

62

63 R1.6. line 347: "Table 1 lists the best metric residual error models and the best-median residual" I think you
64 meant "Table 2"

65 *Thank you, we will correct to Table 2*

66

67 R1.7. lines 382-383: " The hydrographs in Fig. 2 show that the SLS transformation has tighter predictive limits
68 for high flows (i.e. lower uncertainty) and wider predictive limits for low flows (i.e. higher uncertainty)." this
69 is not clear from figure 2 and the low flows appear to be better predicted than high flow. Can you please
70 elaborate on that?

71 *This sentence was indeed confusing, as it was intended in relation to the other transformations.*

72 *We will reword to "The hydrographs in Fig. 2 show that the SLS transformation (i.e. $\lambda=1$) has narrower*
73 *predictive limits high flows (i.e. lower uncertainty) and wider predictive limits for low flows (i.e. higher*
74 *uncertainty) compared with Log, BC02 and BC05".*

	<i>SLS: streamflow Lasarte 01/02/2018-01/03/2018</i>		
<i>thresholds</i>	<i><0.2 mm/h</i>	<i>>1 mm/h</i>	<i>>1.5 mm/h</i>
<i>average width</i>	<i>0.22</i>	<i>0.23</i>	<i>0.24</i>

75 e.g. for Log

	<i>Log: streamflow Lasarte 01/02/2018-01/03/2018</i>		
<i>thresholds</i>	<i><0.2 mm/h</i>	<i>>1 mm/h</i>	<i>>1.5 mm/h</i>
<i>average width</i>	<i>0.18</i>	<i>0.96</i>	<i>0.44</i>

76

77 R1.8. Fig. 4 and following: please note that dots may overlap, so please kindly use different markers, so that
78 the reader can see both or explain where the superposition is.

79 *We will modify figure to improve readability/clarity.*

80

81 R1.9. Line 441: "if it is also is reliable" please remove the second "is"

82 *Thank you, we will correct*

83 R1.10. Line 446 and following: "For volumetric bias, our Figure 1 is slightly different from Figure 2 in
84 McInerney et al. [2017]. McInerney et al. [2017] show that BC0 to BC05 are better, whereas in our study none
85 of the transformations are statistically significantly worse than the others. SLS has the best median
86 performance and in McInerney et al. [2017], SLS has the worst median performance between Log, BC02,
87 BC05 and SLS". Can you please elaborate on the reason behind it?

88 *Thank you, we agree the paragraph is confusing and indeed, for Volumetric Bias, the best-median performance*
89 *is achieved by the BC05 scheme rather than by the SLS scheme.*

90 *We will correct this paragraph as follows:*

91 *For volumetric bias, the findings in Figure 1c are generally similar to the findings reported by McInerney et*
92 *al. [2017]. Notably, the BC05 scheme achieves the best (least) volumetric bias. However note a subtle*
93 *difference: in our hourly study the other schemes do not perform statistically significantly worse, whereas in*
94 *the daily study of McInerney et al [2017] the BC05 scheme was the single best-metric scheme.*

95

96

97 R1.11. Table 3, the caption reports that the values are estimated across catchments, can you please explain
98 how catchments are represented in the tables? I wonder if it would be possible to organize the tables in a better
99 way as they are multiple, however just one caption is provided and it is not fully explanatory

100 *The table represents mean (average) values across catchments. We will add equations to explain this*

101

102 *We will also add individual captions for each table:*

103

104 *"Table 3. Mean reliability and precision across catchments for hourly, daily and monthly scales, for the Log*
105 *scheme"*

106

107 *"Table 4. Mean reliability and precision across catchments for hourly, daily and monthly scales, for BC05*
108 *scheme"*

109

110 *"Table 5 Mean reliability and precision across catchments for hourly, daily and monthly scales, for SLS*
111 *scheme"*

112

113 *In addition, we will also modify the color usage as per comment R1.4.*

114