

1 **Streamflow prediction improvement using deterministic and stochastic wavelet-**
2 **based frameworks with Extreme Learning Machine**

3

7 Abstract

8 Accurate and reliable streamflow forecasting is important in hydrology and water
9 resources planning and management. In the present work, wavelet-based direct (DF)
10 and multi-component (MF) forecast methods performed by the à trous algorithm (AT)
11 are proposed for both deterministic and stochastic monthly streamflow prediction
12 improvement. They are developed in the case of the one-month lead streamflow
13 prediction of the East River basin in China, and then compared with the benchmarks
14 that are implemented without wavelet transform so as to evaluate the effectiveness for
15 forecasting accuracy improvement. An existing blueprint that is flexible and practical
16 to incorporate various sources of forecast uncertainty is extended to generate the
17 stochastic probability prediction of streamflow. Partial mutual information is adopted
18 for predictors selection, and six kinds of Extreme learning machine (i.e. one linear
19 ELM and five common nonlinear kinds) are separately used as the learning algorithms
20 coupled with the wavelet-based forecast methods to conduct a comprehensive
21 performance evaluation. The comparison results indicate that both DF and MF can
22 effectively increase the point prediction accuracy of monthly streamflow under
23 deterministic and stochastic forecasting conditions, while MF performs better than DF.
24 For stochastic prediction, it is much more reasonable to consider both parameter and
25 model error uncertainties than just to consider only parameter uncertainty, and with
26 the reasonable setting MF method can significantly improve the probabilistic interval
27 prediction by greatly improving the forecast sharpness. It can be concluded that the
28 approach using AT wavelet-based DF or MF could provide a feasible way for
29 streamflow prediction improvement.

30 **Keywords:** Streamflow prediction improvement; Wavelet-based forecast method;

31 Probabilistic interval prediction; Extreme learning machine

32 1 Introduction

33 Owing to the complex physical and natural characteristics (e.g. chaotic and
34 stochastic disturbances, non-stationary and non-linear behaviours) associated with
35 river flow systems, the forecasting of the characteristic pattern of a river system and
36 its streamflow is a challenging task (Adnan et al. 2019; Yaseen et al. 2019). While
37 streamflow prediction has been conducted using a variety of hydrologic models
38 (which are knowledge-driven with physical or semi-physical representation of the
39 runoff yield and concentration processes), taking the data-driven Artificial
40 Intelligence (AI) technique as a complement or replacement scheme has received
41 considerable attention in the last twenty-odd years (Yaseen et al. 2015; Yaseen et al.
42 2019; Zhang et al. 2018). The primary motivation for using data-driven models for
43 streamflow prediction is that such models are able to capture nonlinear relationships
44 between predictors and predictand, mainly relying on historical observations, rather
45 than the many requirements for the development of knowledge-driven models such as
46 intensive computation, identification of initial and boundary conditions and large
47 amount of measurement data.

48 Data-driven AI techniques, such as artificial neural network, support vector
49 machine and random forest, can well fit the existing nonlinear relationships in
50 hydrological process modelling with both flexible and practical characteristics;
51 however, they are limited in dealing with the nonstationary behaviour of streamflow
52 signals. In this regard, hybrid forecast framework coupling data pre-/post-processing
53 methods (e.g. singular spectrum analysis, empirical mode decomposition and
54 principal component analysis) with data-driven models are usually used to overcome
55 the shortcomings of a single model (Hu et al. 2019; Meng et al. 2019; Tayfur et al.
56 2013; Unnikrishnan and Jothiprakash, 2018). Recently, wavelet-based data-driven

57 models have received much attention owing to their advantage of being able to extract
58 dynamic and multi-scale features from nonstationary hydrological time series. Many
59 studies have found that the wavelet decomposes the original signal into various
60 resolution levels to capture useful information such as periodic, volatile and trend
61 behaviours, thereby improving the model prediction performance (Nourani et al. 2014;
62 Yaseen et al. 2015).

63 In spite of the growing popularity of wavelet-based hydrological forecast methods,
64 how to select the correct wavelet and develop the appropriate forecast framework has
65 not always been scrutinised. It should be noted that the discrete wavelet
66 transformation (DWT) performed with the non-redundant Mallat's algorithm, which
67 is popular in hydrological and water resources forecasts, is not suitable for real-world
68 applications (Du et al. 2017; Zhang et al. 2015). This is because the non-redundant
69 algorithm will lead to two undesirable characteristics: (1) time-shift sensitivity (i.e.
70 decomposing the new added input data changes the already decomposed inputs) and
71 (2) the boundary effect (i.e. future target/input data to be forecast are involved during
72 current computing). In order to avoid these drawbacks, two redundant algorithms, i.e.
73 maximal overlap discrete wavelet transform (MODWT) and the à trous algorithm
74 (AT), can be adopted to perform the time-shift invariant DWT (Hang and Nabney,
75 2010; Mouatadid et al. 2019; Quilty and Adamowski, 2018; Zhou et al. 2020).
76 Whereas MODWT can only be used in the direct forecast method (referred to as *DF*,
77 which only decomposes predictors and then forecasts the predictand directly), AT can
78 be used in both *DF* and the multi-component forecast method (*MF*), which
79 decomposes both predictors and predictand and acquires the forecast predictand via
80 additive reconstruction. To date, the AT-based forecast framework has been
81 successfully applied in hydrological and water resources systems such as the

82 prediction of drought, urban water demand, water quality and runoff, even though
83 there are limited research cases available for reference (Belayneh et al. 2013;
84 Maheswaran and Khosa, 2012; Quilty and Adamowski, 2018; Zhou et al. 2020). As
85 such, it is worth exploring the potential of AT-based methods. This study
86 simultaneously evaluates the effectiveness of *DF* and *MF* in runoff prediction (which,
87 to our knowledge, has not been previously conducted) since these two forecast
88 methods may yield different performances for a given data set; therefore, it is
89 necessary to take both into consideration to identify the most appropriate (Hang and
90 Nabney, 2010).

91 In general, streamflow prediction based on either a knowledge-driven or data-
92 driven model always involves uncertainties arising from various sources, including
93 input data, model parameters, model structure and initial conditions, which need to be
94 estimated and presented by means of a probabilistic prediction interval in order to (1)
95 make forecasts reliable and (2) carry out a comprehensive performance evaluation of
96 the forecast model (Samadi et al. 2017; Zhang et al. 2011). In addition, the prediction
97 interval can be easily converted to a risk assessment, which is significant in the
98 operation and decision-making process of water resources management. To serve
99 these goals, much effort has been spent on developing and innovating various
100 probabilistic prediction methods. Recently, Montanari and Koutsoyiannis (2012)
101 proposed a blueprint to convert a deterministic hydrological model into a stochastic
102 model that is able to explicitly or implicitly incorporate various sources of uncertainty
103 and does not require model likelihood computation or impose any restrictions to
104 model complexity. The authors also pointed out that the proposed conversion
105 framework is applicable to any deterministic scheme, and the assumptions in the
106 framework can be flexibly adjusted according to practical conditions. Subsequently,

107 Quilty et al. (2019) and Quilty and Adamowski (2020) successfully extended the
108 original blueprint to focus on deterministic data-driven models, developing a *DF*-
109 based stochastic data-driven forecast framework for urban water demand. In view of
110 its flexibility, extensibility and practicality, the blueprint is worthy of further
111 development for application to additional scenarios and in the present study it was
112 utilized to formulate the stochastic wavelet-based forecast methods for streamflow.

113 The main objective of this study was to develop and evaluate the AT wavelet-based
114 models for deterministic and stochastic streamflow prediction improvement, while
115 both *DF* and *MF* methods were explored and the blueprint of Montanari and
116 Koutsoyiannis (2012) is extended to both *DF* and *MF* to build corresponding
117 stochastic forecast frameworks of *SDF* and *SMF*. It should be noted that the blueprint
118 requires the data-driven models to make repeated and extremely extensive
119 calculations since the bootstrapping technique is used to generate abundant
120 probabilistic prediction ensemble members. To counter this potential drawback, the
121 extreme learning machine (ELM), which possessed an extremely rapid learning speed
122 and acceptable generalisation skill, was adopted as the data-driven learning algorithm.
123 The monthly streamflow prediction in the East River basin of China was analysed as a
124 case study.

125 2 Study area and data

126 The East River basin is situated in southern China and has a drainage area of
127 35,340 km². The basin is also an important sub-basin of the Pearl River basin; the
128 Pearl River has the second largest annual runoff amount in China. Under the influence
129 of the monsoon-dominated climate, its annual mean temperature is approximately
130 20.8°C–22.8°C, while the annual mean precipitation varies between 1500 and 2400
131 mm. With the rapid increase in population and booming economic development in the

132 urban agglomerations of the Pearl River Delta, the East River has become an
133 important source of water resources for agriculture, commerce, living and industry
134 (Zhang et al. 2013). The Boluo hydrological station was selected as a case, which is
135 the downstream control station of the main stream of the East River, controlling
136 25,325 km² (71.7%) of basin area.

137 Three kinds of potential input variables were considered for monthly streamflow
138 prediction: antecedent streamflow (since a hydrological time series usually has self-
139 memory), antecedent point precipitation (which is the driving factor directly affecting
140 runoff) and climate phenomenon indices (which may indirectly cause runoff variation
141 by affecting meteorological factors such as precipitation, temperature and
142 evaporation). These data were obtained from Boluo hydrological station, Huiyang
143 precipitation station (Fig. 1) and the National Oceanic and Atmospheric
144 Administration Earth System Research Laboratory (NOAA-ESRL)
145 (<https://psl.noaa.gov/data/climateindices/list/>). The length of the target monthly
146 streamflow time series to be calibrated and tested was over 50 years, i.e. from January
147 1958 to December 2008. As for the potential explanatory climate variables, 19 climate
148 phenomenon indices were selected with the criterion based on temporal coverage, in
149 which each selected index was required to cover the period from the 1950s to
150 December 2008 and those with limited temporal coverage or missing values were not
151 considered. Table 1 gives detailed information about the selected potential climate
152 phenomenon indices.

153 **【Insert Figure 1】**

154 **【Insert Table 1】**

155 Several statistical correlation analysis methods were used to determine the lead
156 time windows for the potential input candidates. Through autocorrelation analysis, we

157 found that the current monthly streamflow series had the highest correlation with the
158 1-month lead streamflow series (correlation coefficient = 0.6). Subsequently, the
159 correlation coefficient quickly decreased to less than 0.1 as the lead time was
160 extended to 3 months. With Pearson and mutual information correlation analyses
161 (Cover and Allen, 1991), the antecedent point precipitation series with lead times of
162 1–3 months were confirmed to be positively correlated with the current monthly
163 streamflow series, and the correlation greatly weakened as the lead time increased.
164 Additionally, each of the antecedent climate phenomenon index series had the largest
165 correlation coefficient with the current streamflow series within the lead time range of
166 1–6 months. Accordingly, the lead time windows of antecedent streamflow,
167 precipitation and the climate phenomenon index were adopted as 2, 3 and 6 months,
168 respectively, resulting in a total of 119 potential input variables. To avoid variables
169 with a large range from being dominant in the data pattern, all data inputs to the ELM
170 models were scaled in the range [−1, 1].

171 For the model development, the complete dataset was divided into three parts, i.e.
172 model training, validation and testing (in general, the training and validation phases
173 are collectively termed the calibration phase), which contain the time series 1958–
174 1988, 1989–2000 and 2001–2008, respectively. Table 2 displays the statistical
175 characteristics of streamflow and precipitation including the complete, training,
176 validation and testing data span, from which it is evident that the streamflow training
177 set fully covers the validation or testing set.

178 **【Insert Table 2】**

179 **3 Methodologies**

180 Fig. 2 presents the overall research roadmap for the study objectives. Partial mutual

181 information (PMI) was used to select inputs to feed an ELM. Both linear and
182 nonlinear ELMs were adopted as the learning algorithms for coupling with the nine
183 forecast methods, which can be categorised as deterministic and stochastic forecast
184 methods. Each kind of deterministic method (i.e. *OF* (ordinary forecast method), *DF*
185 and *MF*) was transformed into two corresponding stochastic methods based on the
186 blueprint of Montanari and Koutsoyiannis (2012) and considering different sources of
187 uncertainty. Specifically, stochastic methods involving parameter uncertainty
188 consisted of *SOF_P*, *SDF_P* and *SMF_P*, which were further converted into *SOF_PM*,
189 *SDF_PM* and *SMF_PM*, respectively, when involving both parameter and model error
190 uncertainties. Details of the setting of the methodologies involved and how they were
191 determined are given below. For more detailed description of the algorithms, the
192 reader is referred to Appendix A.

193 **【Insert Figure 2】**

194 3.1 Input variable selection and partial mutual information
195 Input variable selection (IVS) is a fundamental consideration in developing a
196 reasonable data-driven model for prediction. Its purpose is to identify the relevant and
197 irredundant variables to interpret the behaviour of the model output (May et al. 2011).
198 For data-driven modelling, automatic algorithms for IVS can be categorised broadly
199 as (Galelli et al. 2014; Guyon and Elisseeff, 2003): (1) filter methods based on
200 statistical analysis and (2) wrapper or embedded methods based on a learning
201 algorithm. Filter methods measure the dependence between variables via the intrinsic
202 properties of data and tend to be computationally simple, although the selected inputs
203 may not optimise the model performance. In this study, the PMI-based selection
204 technique, which is categorised as filter method and has been widely used in water

205 resources and environmental modelling, was adopted.

206 The most common statistical analysis methods used to select input variables are
207 linear correlation-based filter methods such as Pearson correlation and partial
208 correlation, which are efficient when the underlying assumption of linearity is valid
209 (Galelli et al. 2014). However, as the product of complex land–atmosphere
210 interactions, streamflow is usually characterised by high complexity, non-stationarity
211 and non-linearity (Dietrich, 1987; Farmer, 2016). In order to effectively measure the
212 nonlinear relationships among hydro-climatic variables, the PMI-based selection
213 algorithm was proposed by Sharma (2000) and further developed by Bowden et al.
214 (2005b) and May et al. (2008).

215 An effective termination criterion is a key consideration of the IVS algorithm
216 because it affects the quantity of the final input variables and has a marked impact on
217 model performance. In general, a termination criterion is chosen based on either the
218 search procedure or the evaluation function (Dash and Liu, 1997); the latter strategy is
219 typically adopted, which includes (1) whether addition or deletion of any variable
220 leads to a better subset (e.g. model validation error or model selection criteria such as
221 Akaike's information criterion) or (2) whether one potential input is selected
222 according to certain evaluation criteria (e.g. using statistical tests and resampling such
223 as the F-test and bootstrapping) (Dash and Liu, 1997; Galelli et al. 2014; He et al.
224 2011; May et al. 2008). In the present study, the termination criterion for the PMI-
225 based selection iteration is defined as that at which the *NSE* indicator of the model
226 validation set reaches the maximum. This was chosen since it determines appropriate
227 inputs based on the model performance and enables the performance of the testing set
228 to be optimal.

229 3.2 Extreme Learning Machine

230 An ELM is a learning algorithm for the generalised single-hidden layer feed
231 forward neural network (SLFN) (Huang et al. 2012), which consists of three neuron
232 layers, i.e. the input layer, the hidden node layer and the output layer. The input layer
233 receives inputs without making any transformation, and its neuron nodes are equal in
234 number to the inputs. In this study, the output layer is a linear transfer function
235 without bias; this is a common setting in hydrological forecasting (Deo and SAhin,
236 2015; Yaseen et al. 2016). The two layers are linked through the hidden node layer,
237 which performs computations with activation functions.

238 The choice of activation function usually affects the ability of the model to
239 transform data information. There are a variety of activation functions available, of
240 which the six types, ‘*lin*’, ‘*sigm*’, ‘*tanh*’, ‘*rbf_l1*’, ‘*rbf_l2*’ and ‘*rbf_linf*’, collected
241 from <https://hpeelm.readthedocs.io/en/latest/> were used for the ELM model. The
242 sigmoid and hyperbolic tangents are popular functions used for neural networks,
243 although the latter is a variant of the former. Another common type is the Gaussian
244 radial basis function (RBF), which uses distances to centroids as inputs to the hidden
245 layer rather than a linear transformation (i.e. $w_i \cdot x_t + b_i$ in Eq. (A.9)) (Chen et al.
246 1991). Any function of distances between samples and centroids is theoretically
247 available, such as L1, L2 or L^∞ norms. All of the above-described types are nonlinear
248 functions, which can greatly benefit the learning ability of an ELM when the output is
249 nonlinearly dependent on the inputs. In addition to a nonlinear function, linear hidden
250 neurons, which have no transformation function and learn the linear dependency
251 between input data and the target directly (Akusok et al. 2015), were also used to
252 build ELM models as a reference in this study.

253 Random initialisation and non-iteration of weights and biases enables ELM to learn

254 quickly. However, its parameters easily become trapped in a local optimal solution,
255 leading to differences in two model outputs when using the same input variables. As a
256 matter of fact, even with the automatic optimisation technique, this situation might
257 still occur. To deal with this problem, ensemble strategy can be used to acquire the
258 final output. The strategy is performed by repeatedly running a sufficient number of
259 models with the same inputs and then taking their mean value. Considering both the
260 computation time and the stability, the operation is repeated 100 times since we found
261 that with this setting the mean value is close to the expectation and becomes
262 sufficiently stable.

263 3.3 Wavelet transformation

264 MODWT and AT are redundant wavelets that can be directly used (without
265 boundary processing required) for actual prediction applications. However, it should
266 be noted that the coefficients obtained by MODWT cannot be directly added to obtain
267 the original series (as in Eq. (A.18)) because MODWT uses multiresolution analysis
268 to first reconstruct the detail and approximation coefficients, which leads to the
269 boundary effect of using future data. This difference makes MODWT unsuitable for
270 use in the multi-component wavelet forecast framework (which follows the building
271 process of ‘decomposition–prediction–reconstruction’ and was studied in the present
272 work). Therefore, we selected AT to develop wavelet-based forecast models.

273 The choice of wavelet filters can affect the performance of a wavelet-based
274 framework. The Haar filter has the shortest support length among all wavelet filters,
275 which gives it sufficient ability to describe the local details and makes it suitable for
276 use in time series with complex volatility. Hence, the filter was applied for AT
277 transformation in the present investigation. The decomposition level is another key
278 element of wavelet transformation that should be optimised to ensure model

279 performance. Many previous studies have used empirical equations to determine the
280 decomposition level. The following empirical equation suggested by Nourani et al.
281 (2009) was adopted to determine the level (L) in the present paper:

282
$$L = \text{int}[\log(N)], \quad (1)$$

283 where N is the size of the time series to be decomposed. Hence, three levels for
284 decomposition were determined herein since the total number of series samples was
285 up to 612. The four coefficients, $D_{1,t}^a, D_{2,t}^a, D_{3,t}^a$ and $A_{3,t}^a$ (referred to as D1, D2, D3 and
286 A3, respectively), obtained from the three-level decomposition were used as potential
287 input candidates in the wavelet-based forecast framework.

288 3.4 Deterministic wavelet-based forecast method

289 In DF , only the input variables are decomposed and selected; then, those selected
290 are fed directly into a single forecast algorithm, following the building process of
291 ‘decomposition–prediction’. Different from DF , both the input and target variables
292 should be decomposed in MF . Each of the decomposed components from the target
293 variable is then forecast, followed by reconstruction with the already forecast
294 components. Since the original inputs might also provide available information, both
295 the original and wavelet-decomposed inputs are adopted as potential inputs to be
296 filtered via the PMI based selection technique, which correspond to the ‘*Single-hybrid*’
297 and ‘*Across-hybrid*’ methods of WDDFF in Quilty and Adamowski (2018). The
298 detailed building processes for these two frameworks can be seen in Fig. 3 (a) and (b).

299 **【Insert Figure 3】**

300 3.5 Stochastic wavelet-based forecast method

301 Montanari and Koutsoyiannis (2012) assumed that the forecast uncertainty comes
302 from the input data, the model parameter and the model error uncertainties, whereas

303 the uncertainty of the input data related to the monitoring approaches is hard to access.
304 In this study, each developed stochastic forecast model produced outputs under two
305 conditions: (1) considering only parameter uncertainty and (2) considering both
306 parameter and model error uncertainties, for the purpose of exploring the differences
307 when considering different sources of uncertainties and identifying the most
308 appropriate condition. For condition (1) (i.e. SOF_P , SDF_P and SMF_P), a
309 bootstrapping size of $n = 100$ is adopted to obtain stable results. For condition (2) (i.e.
310 SOF_PM , SDF_PM and SMF_PM), the forecast results will be affected by the
311 interactive component of the bootstrapping size (which is denoted as b_size and
312 equals the resampling parameter of m/n) and the quantity of nearest neighbours
313 (which should be selected for KNN fitting and is denoted as $n_neighbors$). The
314 response surfaces of the forecast performance metrics are discussed in Section B.1 for
315 the selection of appropriate b_size and $n_neighbors$. To facilitate this analysis, the
316 parameters m and n are set to the same value.

317 3.5.1 Determination of parameters

318 Section B.1 showed that the point prediction is less affected by $n_neighbors$ or
319 b_size , whereas the interval prediction is significantly affected by $n_neighbors$. When
320 $n_neighbors = 1$, the corresponding $PICP$, $CRPS$ and coverage probability plot all
321 indicate the most reliable interval prediction. In this study, SOF_PM , SDF_PM and
322 SMF_PM showed identical characteristic, as in the case shown in Fig. B.1. Quilty et
323 al. (2019) found that the reliability of estimating the target pdf of $f_Q(Q)$ was sufficient
324 for a bootstrapping size of 50 or greater. This study uses 100 resamples to estimate the
325 pdf and obtain a stable point and interval prediction for SOF_PM , SDF_PM and
326 SMF_PM . While reliable uncertainty estimation does not always indicate useful
327 forecasts since it can lack sharpness, poor reliability will always indicate poor results

328 for the probabilistic prediction (Humphrey et al. 2016). Therefore, $n_neighbors = 1$
329 was adopted for the KNN bootstrapping step to obtain acceptable forecast reliability.

330 3.6 Model evaluation metrics

331 The deterministic performance measures adopted for evaluating model point
332 prediction include the Nash–Sutcliffe efficiency (*NSE*), the mean absolute error
333 (*MAE*) and the root mean square error (*RMSE*). These are all commonly used metrics
334 in hydrology and water resources (Legates and McCabe, 1999; Moriasi, 2007);
335 therefore, their calculations are not described in detail here. *NSE* can be used to
336 measure the degree of proximity between observed and predicted time series plots,
337 ranging from $-\infty$ to 1.0 and with $NSE = 1$ corresponding to the perfect match.
338 *RMSE* shows the discrepancy between the observed and predicted values, which
339 would increase from zero for perfect estimates to infinite positive values as the
340 discrepancy grows. The function of *MAE* is similar to that of *RMSE* but it is less
341 sensitive to large predicted errors.

342 The probabilistic evaluation metrics used for model interval prediction include the
343 continuous ranked probability score (*CRPS*), the prediction interval coverage
344 probability (*PICP*) and the average prediction interval width (*AW*) (Quilty et al.
345 2019). *PICP* measures forecast reliability and reflects the proportion of the actual
346 observation values falling within the confidence interval of prediction. *AW* is a
347 measure of forecast sharpness, representing the average width of the prediction
348 interval under a certain confidence level and reflecting the degree of forecast
349 uncertainty. *CRPS* simultaneously measures forecast sharpness and reliability,
350 although it is not straightforward to quantify forecast reliability or sharpness
351 (Gneiting et al. 2007). In order to calculate *PICP* and *AW*, the confidence level needs

352 to be specified. In this study, we selected a 95% confidence level, corresponding to
353 optimal *PICP* and *AW* scores of 95%.

354 To provide a more informative picture of forecast reliability, the coverage
355 probability plot is also considered (Laio and Tamea, 2007), which is able to depict the
356 degree of over-/under-estimation of forecast uncertainty, i.e. the degree of forecast
357 bias.

358 4 Results and discussion

359 4.1 Deterministic prediction improvement with *DF* and *MF*

360 Table 3 presents the testing-set performance measures of ELM models with various
361 activation functions using deterministic forecast methods (i.e. *OF*, *DF* and *MF*). The
362 improvements in the performance metrics for *DF* and *MF* are as follows:

- 363 ● *DF*: *NSE* (+1.4% to +11.3%), *MAE* (-2.0% to -4.8%), *RMSE* (-0.8% to -6.6%);
364 ● *MF*: *NSE* (+4.5% to +16.1%), *MAE* (-5.1% to -10.1%), *RMSE* (-2.4% to
365 -9.5%).

366 The results intuitively indicate the effectiveness of the wavelet-based framework for
367 forecast accuracy improvement. While ELMs are found to be improved by *DF* to
368 varying degrees, they can be further improved by *MF*. For example, the *NSE* of
369 *ELM_rbf_11* with *DF* increased from 0.530 to 0.590, whereas the *MAE* and *RMSE*
370 decreased from 6.609 to 6.446 and 11.019 to 10.293, respectively. It can be seen from
371 Table 3 that the *NSE*, *MAE* and *RMSE* of *ELM_rbf_11* are further improved to 0.615,
372 5.939 and 9.972, respectively. Although *ELM_rbf_12* with *MF* performs slightly
373 worse than that with *DF* in terms of *NSE* and *RMSE* (i.e. *NSE* values are, respectively,
374 0.591 and 0.587 while *RMSE* values are 10.279 and 10.321), the ELMs using other
375 activation functions all perform better with *MF*. On the whole, *MF* performs better

376 than DF , while DF is better than OF . In addition, Table 3 also shows that the linear
377 ELM performs worse than any nonlinear ELM in terms of any of the three
378 deterministic methods. This result demonstrates the inability of the linear ELM to fit
379 the nonlinear relationship between streamflow and its driving factors.

380 **【Insert Table 3】**

381 4.2 Stochastic prediction improvement with DF and MF

382 Table 4 presents the testing-dataset performances of ELMs with stochastic methods
383 considering parameter uncertainty. In terms of NSE , MAE , $RMSE$ and $CRPS$
384 indicators, each ELM with SMF_P is observed to outperform those with SDF_P ,
385 while the two methods both perform better than SOF_P . The metric improvements
386 with SDF_P and SMF_P are as follows:

- 387 • SDF_P : NSE (+1.7% to +12.3%), MAE (-1.1% to -6.2%), $RMSE$ (-1.0% to
388 -6.7%), $CRPS$ (-1.2% to -5.6%);
389 • SMF_P : NSE (+3.5% to +21.2%), MAE (-4.7% to -11.7%), $RMSE$ (-2.3% to
390 -11.9%), $CRPS$ (-5.0% to -13.8%).

391 Overall, SMF_P improves the above indicators more than does SDF_P . However, the
392 models in Table 4 show no regular changes after using SDF_P or SMF_P in terms of
393 $PICP$ and AW . For example, the $PICP$ of ELM_sigm increases from 23.958% to
394 33.333% and 27.083%, while AW increases from 4.634 to 5.322 and 4.876. However,
395 for ELM_lin , the $PICP$ decreases from 57.292% to 45.833% and 52.083%, while AW
396 varies from 9.724 to 10.333 and 9.279. Therefore, there is no evidence that the
397 wavelet-based method can effectively improve the forecast reliability and sharpness
398 of probabilistic prediction taking only parameter uncertainty into account.
399 Nevertheless, the improvements in NSE , MAE , $RMSE$ and $CRPS$, which measure

400 point prediction performance and interval prediction error, can still demonstrate the
401 advantage of both *DF* and *MF*. Additionally, from the *AW* values in Table 4, it can be
402 seen that the prediction interval of the linear ELM is much wider than that of the
403 nonlinear ELM. Correspondingly, the coverage of the linear ELM indicated by *PICP*
404 is also larger than that of the nonlinear ELM. A similar result is also shown in Table 5.
405 These findings indicate that the linear ELM has greater uncertainty than the non-linear
406 ELM, which is related to the fact that the outputs of the former are more sensitive to
407 disturbance.

408 **【Insert Table 4】**

409 The improvements in *SDF_PM* and *SMF_PM* for *SOF_PM* are as follows:

- 410 ● *SDF_PM*: *NSE* (+3.0% to +12.5%), *MAE* (-0.5% to -10.3%), *RMSE* (-2.0% to
411 -6.8%), *CRPS* (-2.1% to -7.7%);
412 ● *SMF_PM*: *NSE* (+3.7% to +19.7%), *MAE* (-3.7% to -12.1%), *RMSE* (-2.5% to
413 -11.0%), *CRPS* (-2.9% to -12.0%).

414 From the evaluation metrics in Table 5, it can be concluded that in terms of *NSE*,
415 *MAE*, *RMSE* and *CRPS*, *SMF_PM* performs better than *SDF_PM*, which is consistent
416 with the conclusion drawn from Table 4. However, the difference is that *SMF_PM*
417 can greatly improve forecast sharpness at the expense of only a slight reduction in
418 forecast reliability. For example, in terms of *PICP*, *SDF_PM* has no obvious change
419 compared with *SOF_PM*, with the variation ranging between -2.2% and +2.4%. At
420 the same time, *AW* variations in ELM with different activation functions are as
421 follows: -12.5%, +4.5%, +2.3%, +14.4%, +3.9% and -1.5%. These findings indicate
422 that on the whole, *SDF_PM* cannot effectively improve forecast reliability or
423 sharpness. For *SMF_PM*, the ranges in *PICP* variation are -8.2% to +1.2% and -7.1%
424 to -1.1% compared with *SOF_PM* and *SDF_PM*, respectively, showing non-serious

425 degradation in forecast reliability. However, it can be found from Table 4 that the *AW*
426 values of *SMF_PM* are all significantly smaller than those of *SOF_PM* and *SDF_PM*,
427 with the percentage decreases reaching -37.6% to -20.2% and -37.0% to -25.2% ,
428 respectively. Meanwhile, the prediction interval of *SMF_PM* in Fig. 4 covers most of
429 the streamflow observations, although its prediction interval is apparently narrower
430 than the other two stochastic methods. The result confirms that *SMF_PM* can
431 effectively reduce forecast uncertainty while maintaining a relatively reliable
432 probability coverage interval.

433 **【Insert Table 5】**

434 **【Insert Figure 4】**

435 4.3 Prediction with different sources of uncertainty

436 The point prediction of a stochastic forecast method is obtained by averaging the
437 probabilistic ensemble members. In terms of the point prediction performance
438 indicated by *NSE*, *MAE* and *RMSE* in Table 4, no prediction method is found to be
439 absolutely dominant in the comparison between the deterministic and stochastic
440 methods or between the two types of stochastic method.

441 From both the quantitative indicators and performance graphs, the stochastic
442 models considering only parameter uncertainty showed obvious defects in forecast
443 uncertainty estimation. It is apparent that the *PICP* values of the models in Table 4 are
444 all low (most are not more than 50%), which is far from the expected 95%. In the
445 prediction process diagram (Fig. 4), the probabilistic prediction interval that reflects
446 only parameter uncertainty is rather narrow, leading to many observations not being
447 covered. According to the coverage probability plot in Fig. 5 (a), the quantile–quantile
448 curves of the models all present an obvious ‘S’ shape and their intersections with the

449 bisector lie in the upper middle, above the average, indicating that the model results
450 greatly underestimate forecast uncertainty and that the overall forecast values are
451 somewhat high. When both parameter and model error uncertainties are involved, the
452 interval prediction is significantly improved. For example, compared with *SOF_P*, the
453 *CRPS*, *PICP* and *AW* indicators of *ELM_tanh* with *SOF_PM* change from 6.057,
454 30.208% and 5.208, respectively, to 5.267, 85.417% and 31.202, respectively, with
455 the *PICP* of each ELM model exceeding 80%. Moreover, each prediction interval in
456 Fig. 4 covers most of the observations, except that several local maximum values fall
457 outside the interval. The quantile–quantile curve of each stochastic model in Fig. 5 (b)
458 approaches the 1:1 bisector, indicating that there is no underestimation or
459 overestimation of forecast uncertainty. Although the forecast uncertainty indicated by
460 *AW* increases greatly, the significant improvement in forecast reliability avoids an
461 invalid interval prediction.

462 **【Insert Figure 5】**

463 4.4 Discussion

464 4.4.1 Rationality of selected *n_neighbors* parameter

465 It is shown that as *n_neighbors* increases, the probabilistic prediction interval
466 becomes narrower, with forecast reliability deteriorating and forecast error increasing.
467 This is mainly because *n_neighbors* can affect the homogeneity of the KNN simulated
468 error. If the parameter is too small, KNN will be susceptible to the noise in the
469 searched neighbours and consequently the variance of the simulated error will be
470 considerable; however, too large a parameter may greatly reduce the variance
471 (Sikorska-Senoner et al. 2014). From the simulated error distribution diagrams of Fig.
472 6 (a)–(c), it is evident that the distribution of each forecast method appears narrower

473 and taller with larger $n_neighbors$ values, indicating that the simulated error values
474 are concentrated within a smaller range. Since the KNN bootstrapping step is operated
475 via resampling from the simulated error distribution, the numeric range of the
476 bootstrapping model error will also be more concentrated (Fig. 6 (d)–(f)). Finally,
477 each bootstrapping model error series with small variance is individually added into
478 the corresponding model output to generate the probabilistic ensemble prediction,
479 which is supposed to have a narrower prediction interval than those series with larger
480 variance. This explains the phenomenon that the probabilistic prediction interval
481 decreases as $n_neighbors$ increases.

482 **【Insert Figure 6】**

483 In this research, for the sake of achieving acceptable forecast reliability,
484 $n_neighbors$ was adjusted to its minimum. We can further infer that setting the
485 parameter to a value that could obtain reasonable forecast reliability depends on the
486 accuracy that the point prediction can achieve. As shown in Fig. 5 (a)–(c), many
487 observations have considerable forecast errors in the point prediction results for
488 *SOF_PM*, *SDF_PM* and *SMF_PM*. This leads to the need for a wider interval to
489 obtain effective coverage; as such, the parameter should be adjusted to a small value.
490 However, for a streamflow forecast scenario with high point prediction accuracy
491 (contrary to the case considered in this study), too small an $n_neighbors$ value can
492 decrease forecast sharpness markedly. It also lacks the potential to improve forecast
493 reliability as the prediction interval could cover most of the observations; therefore, it
494 does not require a wider interval.

495 4.4.2 Rationality of selected predictors

496 For data-driven forecast modelling in environmental and water resources research,
497 the selected predictors generally require certain causal relationships with the
498 predictand (Bowden et al. 2005a; Liu et al. 2015; Lu et al. 2010; Maier and Dandy,
499 1999; Weerts et al. 2016). In this study, the top 20 variables, which are ranked by PMI
500 and are from the potential predictor set including no decomposed variables, consist of
501 antecedent runoff, precipitation and the climate indices of Nino 1+2, Nino 3, AO, NP,
502 MEI, ONI and SOI (as shown in Fig. 7 (a)). Nino 1+2, Nino 3, MEI, ONI and SOI are
503 usually used to describe ENSO activities (Glantz and Ramirez, 2020; Lu et al. 2010;
504 Wolter and Timlin, 2011).

505 **【Insert Figure 7】**

506 Some researchers have implied that when an ENSO event occurs, the sea surface
507 temperature in the central and eastern equatorial Pacific heats up abnormally, leading
508 to intensified atmospheric convection activity in the tropical central Pacific; due to the
509 teleconnection, the circulation in the northern hemisphere at mid-high latitudes is
510 abnormal, which affects regional climate (Tian et al. 2016; Zhang et al. 2017). In
511 some existing studies, several climate indices and the ENSO signal have been found
512 to be closely related to precipitation variation in the East River basin or South China.
513 For example, Zhang et al. (2013) found that sea surface temperature anomalies (SSTA)
514 of Nino 1+2, Nino 3, Nino 4, Nino 3.4 and SOI can serve as good indicators of the
515 areal precipitation in the East River basin from January to March. In a study
516 conducted to examine inter-decadal variations in the early summer monsoon rainfall
517 over South China, Chan and Zhou (2005) concluded that the variations are related to
518 both ENSO and PDO. Based on the data analysis techniques of wavelet analysis,
519 principal component analysis and rank correlation method, Niu (2013) identified
520 ‘...many in-phase relations between the precipitation of the Pearl River basin and

521 teleconnection patterns of ENSO and IOD'. Additionally, both NP and AO are
522 demonstrated to have significant influences on regional climates in China (Gong et al.
523 2001; He et al. 2016; Li et al. 2004; Li and Li, 2000). Gu et al. (2014) conducted a
524 frequency analysis for the annual maximum discharge series (AMS) in the East River
525 Basin from 1954 to 2009, showing that NP had linear impacts on AMS at all research
526 sites. In research aiming to reveal the relationship between AO and climate over
527 South China, Yang (2011) found a significant influence of the December AO on
528 January precipitation and temperature. These confirm that the PMI-based method is
529 able to identify predictors with physical meaning. Moreover, in the modelling process,
530 each ELM model coupled with *OF* involves at least four types of predictors (Fig. 7),
531 namely antecedent runoff, Nino 1+2, NP and AO, indicating that each developed
532 model could represent the statistical relationship between predictors and predictand
533 based on causal links.

534 4.4.3 Performance of deterministic and stochastic wavelet-based methods

535 In Section 4.1, *DF* is found to outperform *OF* under the deterministic condition. In
536 fact, the superiority of the direct wavelet-based forecast method has also been
537 identified in recent studies (Ghaemi et al. 2019; Mouatadid et al. 2019; Quilty and
538 Adamowski, 2018; Rahman et al. 2020; Zhou et al. 2020). From the perspective of
539 model development, the difference between *DF* and *OF* lies only in the phase of
540 feeding input variables, wherein *DF* involves the additional step of data pre-
541 processing with wavelet decomposition. However, this additional step provides
542 predictors that are more conducive to prediction for the ELM model. It can be seen
543 from Fig. 7 (c) that the cumulative PMI values of *DF* are not less than those of *OF*
544 with equal input variables quantities, and that their cumulative values gradually
545 increase the gap when the number of inputs is greater than eight. This can be

546 explained by the fact that the predictor set containing decomposed variables carries
547 more available information for prediction, which is provided by the decomposed
548 components of the climate indices. Specifically, the difference in the top 20 ranked
549 variables between *DF* and *OF* lies in the climate indices (Fig. 7 (a)–(b)), where it can
550 be found that in contrast with *OF*, *DF* excludes the ONI and SOI variables but
551 contains many decomposed climate indices. For instance, in terms of a certain type of
552 index, *OF* involves Nino 1+2_t-3 and Nino 1+2_t-4, whereas *DF* involves Nino
553 1+2_D2_t-4, Nino 1+2_D1_t-4 and Nino 1+2_D2_t-3. Likewise, in a study of
554 estuarine salinity prediction, Zhou et al. (2020) found that the direct wavelet
555 framework improves model performance, which may be related to the fact that further
556 information is exposed via wavelet decomposition, since the correlation between
557 predictors and predictand increased. On the other hand, there are more (or at least an
558 equivalent amount of) selected predictors for each ELM model with *DF*, which has
559 also been found in research on short-term electricity demand and gas price forecasts
560 (Hang and Nabney, 2010). In the present study, when increasing the quantity of
561 predictors for *OF* to make it equal to that for *DF*, the model performances were not
562 found to be effectively improved. On the contrary, some worsened (e.g. *NSE* values
563 were calculated as 0.500, 0.541, 0.528, 0.533, 0.551 and 0.545, corresponding in
564 order to the ELM models with different activation functions in Table 3). This is due to
565 the fact that using irrelevant variables as predictors reduces the prediction accuracy of
566 the models. The above discussion implies that the wavelet pre-processing of *DF* can
567 provide appropriate input variables with more information to benefit streamflow
568 forecast.

569 When constructing *MF*, the original time series is first decomposed into several
570 high-frequency and low-frequency components before they are predicted. As shown

571 in Table 6, the forecast accuracy of each model improves as the component frequency
572 decreases from D1 to D3/A3. Ultimately, after the component integration is
573 completed, this ‘divide-and-conquer’ method has the opportunity to yield an
574 improved performance. However, it should be noted that the performance of *MF* is
575 influenced by several factors such as object characteristics, model complexity and
576 specific wavelet decomposition operation (e.g. the decomposition level and wavelet
577 basis function). Thus, it has a certain degree of randomness and should not always
578 outperform *DF*. This is supported by several existing studies, although there are few
579 such studies on a comparison between the two methods. For example, Hang and
580 Nabney (2010) and Zhou et al. (2020) both reported the superiority of the multi-
581 component method, whereas Quilty and Adamowski (2018) found that the direct
582 method of MODWT-single-hybrid_strategy performs best in urban water demand
583 forecasting across the different daily lead times (1, 3, 5, 7 and 14 d). In the present
584 study, the accuracy of *MF* is found to generally be greater than that of *DF* and *OF*.
585 Therefore, the potential of *MF* in streamflow prediction is demonstrated, although it
586 has no absolute advantage over *DF* in principle.

587 **【Insert Table 6】**

588 Regarding the probabilistic prediction, one noteworthy aspect in this research is
589 that *SMF_PM* can effectively improve forecast sharpness. This is related to the
590 smaller variances of the KNN simulated and bootstrapping errors. As presented in the
591 boxplots of Fig. 8 (a) and (b), the variances of the KNN simulated error series are
592 overall significantly smaller than those of *SOF_PM* or *SDF_PM* (Fig. 8 (a)). The
593 variances of the bootstrapping error series (which are resampled from the simulated
594 error series) are correspondingly smaller (Fig. 8 (b)). Here, we can recall that the
595 smaller variance of the KNN bootstrapping error specifically causes a narrower

596 prediction interval when a larger $n_neighbors$ value is adopted. The reason why
597 SMF_PM has a smaller error variance is due to the overall smaller simulated error, i.e.
598 we can observe from Fig. 8 (c)–(e) that the minimum, median and maximum of the
599 simulated error are all close to the zero point, of which the maximum is markedly less
600 than that of SOF_PM or SDF_PM .

601 **【Insert Figure 8】**

602 In terms of point prediction metrics (i.e. NSE , MAE and $RMSE$), MF performs
603 better than DF , while DF outperforms OF under the stochastic condition, which is in
604 consistent with the results obtained under the deterministic condition. This is easily
605 explained by the fact that the point prediction of a stochastic forecast model
606 represents the average performance of all the deterministic point prediction ensemble
607 members. Another point that needs to be mentioned is that the point prediction of a
608 stochastic model is the result of which bias has been eliminated after considering the
609 uncertainty (Quilty and Adamowski, 2020). Therefore, taking different uncertainties
610 into account will result in various point prediction performances, which explains the
611 existing differences in evaluation metrics between the deterministic and stochastic
612 methods and between the two types of stochastic method (see Section 4.3).

613 **5 Conclusions**

614 In this study, the potential of an AT wavelet coupled with an ELM under
615 deterministic and stochastic conditions for streamflow prediction is explored. Nine
616 forecast methods are built and compared for predicting monthly streamflow of the
617 East River basin in China, using more than 50 years of historical runoff records, point
618 precipitation, climate phenomenon indices and the lead-time information of predictors.
619 Based on the experiments conducted, the following conclusions can be drawn:

- 620 1. Under both deterministic and stochastic conditions, use of the wavelet method can
621 effectively improve point prediction accuracy, wherein *MF* performs better than
622 *DF*. In detail, using the *OF*, *SOF_P* and *SOF_PM* forecast methods without
623 wavelet transformation as benchmarks, the maximum increases in the *NSE* value
624 after using *DF* are 11.3%, 12.3% and 12.5%, respectively, while they are 16.1%,
625 21.2% and 19.7% after using *MF*. Appropriate input variables with the
626 information, which benefit the streamflow forecast, are provided via the wavelet
627 pre-processing of *DF*.
- 628 2. Considering both parameter and model error uncertainties can provide much more
629 reasonable interval predictions than considering only parameter uncertainty,
630 wherein most *PICP* values of the latter are less than 50%, i.e. markedly different
631 to the expected value of 95%. The *SMF_PM* proposed is able to greatly improve
632 the forecast sharpness of *SOF_PM*. In detail, *AW* values indicating forecast
633 sharpness decreased by 20.2–37.6% with *SMF_PM*.
- 634 3. The interval prediction performances of stochastic models that consider two
635 sources of uncertainties are mainly and significantly affected by *n_neighbors*,
636 wherein the KNN simulated and bootstrapping error values are concentrated
637 within a smaller range with a larger *n_neighbors* value, resulting in a narrower
638 probabilistic prediction interval.
- 639 4. Linear ELM performs worse than any nonlinear ELM for all three deterministic
640 methods (i.e. *OF*, *DF* and *MF*) owing to the inability of the linear ELM to fit the
641 nonlinear relationship between streamflow and its driving factors.
- 642

643 Appendix A

644 A.1. PMI-based selection algorithm

645 PMI is based on the mutual information (MI) method of information theory (Cover
646 and Allen, 1991). MI is a type of correlation analysis method that makes no
647 assumptions regarding the relationships among variables; therefore, it can be used to
648 detect both linear and nonlinear dependencies in a multivariate system. Moreover, MI
649 has shown insensitivity to noise and data transformations, indicating that it is robust
650 for practical application. It can be defined in the following mathematical form:

$$651 \quad I(X; Y) = \iint p(x, y) \ln \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (\text{A.1})$$

652 where X is the input variable and Y the output variable; $I(X; Y)$ represents the mutual
653 information between X and Y ; $p(x)$ and $p(y)$ are, respectively, the marginal
654 probability density functions (pdf) of X and Y and $p(x, y)$ is the joint pdf. It can be
655 further shown in discrete form for practical use:

$$656 \quad I(X; Y) \approx \frac{1}{n} \sum_{i=1}^n \ln \left[\frac{f(x_i, y_i)}{f(x_i)f(y_i)} \right], \quad (\text{A.2})$$

657 where f denotes the estimated probability density based on the n observed sample
658 pairs of x_i and y_i . Kernel density estimation is typically used for estimating f :

$$659 \quad f = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i), \quad (\text{A.3})$$

660 where K_h represents the kernel function and h denotes the kernel bandwidth. A
661 common choice for K_h is the Gaussian kernel:

$$662 \quad K_h = \frac{1}{(\sqrt{2\pi h})^d \sqrt{|\Sigma|}} \exp \left(\frac{-\|x - x_i\|}{2h^2} \right), \quad (\text{A.4})$$

663 where d is the number of dimensions of variable X , Σ is the sample covariance matrix
664 and $\|x - x_i\|$ is the Mahalanobis distance metric. The choice of bandwidth h has an
665 influence on the calculated MI value. Too small a h value may result in an
666 overestimated MI due to sensitivity to noise in the data sample, whereas too large a h

667 value that over-smooths the pdf may underestimate MI. Given its simplicity and
668 computational efficiency (Scott, 1992; Silverman, 1986), the Gaussian reference
669 bandwidth that was empirically recommended by Sharma (2000) was adopted in the
670 present study.

671 While MI is a useful measure of dependence between X and Y , it cannot account
672 for redundancy in the input variable pool. Analogous to the partial correlation
673 coefficient, PMI can handle the redundancy since it measures the partial dependence
674 between two variables, which is conditional on any input variables that have already
675 been selected. The PMI-based input selection algorithm generally operates as follows:
676 (i) define potential input candidates based on statistical analysis, prior knowledge,
677 available data etc.; (ii) initialise the candidate set C and the selected input variable set
678 S ; (iii) calculate the residual output u and residual v of each input existing in
679 candidate C by estimating the relationship with currently selected inputs:

$$680 \quad u = Y - E[y|Z], \quad (\text{A.5})$$

$$681 \quad v = X - E[x|Z], \quad (\text{A.6})$$

682 where u and v , respectively, represent the residual information of Y and X , as the
683 already selected variable Z has been considered. The conditional expectation E of y
684 on Z with observations z_i can be written as

$$685 \quad E[y|Z = z_i] = \frac{1}{n} \frac{\sum_{i=1}^n y_i K_h(z - z_i)}{\sum_{i=1}^n K_h(z - z_i)}, \quad (\text{A.7})$$

686 then, $E[x|Z]$ can be similarly estimated. In step (iv) the input variable that maximises
687 $I(Z; Y|X)$ is moved from C to S :

$$688 \quad I(Z; Y|X) = I(v; u). \quad (\text{A.8})$$

689 Steps (iii)–(iv) iterate until the termination criterion is satisfied, after which the final
690 input variable set can be determined.

691 A.2. ELM algorithm

692 A general mathematical description of an ELM can be given as follows: assuming a
 693 set of N training samples $(x_t, y_t, t = 1, 2, \dots, N)$ with $x_t, y_t \in R^m, R^n$, an SLFN model
 694 with Λ number of hidden neurons can be expressed as

$$695 \quad \sum_{i=1}^{\Lambda} \beta_i f(w_i \cdot x_t + b_i) = y_t, \quad (\text{A.9})$$

696 where $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ are the input weights, b_i is the bias and $\beta_i =$
 697 $[\beta_{i1}, \beta_{i2}, \dots, \beta_{in}]^T$ are the output weights. The input layer weights w_i and bias b_i are
 698 both initialised randomly and then fixed. The output layer weights β_i are thus
 699 independent and can be solved directly without an iteration process. This explains
 700 why ELM generally performs well with an extremely fast learning speed compared
 701 with traditional learning algorithms (e.g. the back-propagation technique). Here $w_i \cdot$
 702 $x_t + b_i$ could also be expressed as WX , where X represents the inputs. The activation
 703 function f in a hidden neuron transforms WX into a new matrix. After the
 704 transformation in the hidden layer, the activated matrix is used to solve the output
 705 weights β_i .

706 In practice, an ELM can be written compactly in matrix form by gathering outputs
 707 of all the hidden neurons into a matrix H . One may estimate the matrix β directly with
 708 the input and output dataset based on a system of linear equations (Huang et al. 2006):

$$709 \quad Y = H\beta, \quad (\text{A.10})$$

710 where

$$711 \quad H = f(WX) = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{bmatrix} = \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \cdots & f(w_\Lambda \cdot x_1 + b_\Lambda) \\ \vdots & \ddots & \vdots \\ f(w_1 \cdot x_N + b_1) & \cdots & f(w_\Lambda \cdot x_N + b_\Lambda) \end{bmatrix}_{N \times \Lambda} \quad (\text{A.11})$$

712 and

$$713 \quad \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_\Lambda^T \end{bmatrix}_{\Lambda \times n} \quad \text{and} \quad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times n}. \quad (\text{A.12})$$

714 From Eqs. (A.11) and (A.12), an ELM can be viewed as two projections with a
715 transformation between them, and the number of hidden neurons determines the size
716 of matrices H , β and W . For determined and under-determined instances ($N \leq \Lambda$), an
717 ELM should use regularisation to prevent poor generalisation performance (Huang et
718 al. 2012). In most situations, the number of training samples is greater than that of
719 hidden neurons, which is an overdetermined problem. A unique solution for this
720 problem is to deduce the output weights by inverting the hidden layer matrix using the
721 Moore–Penrose generalised inverse function (H^+) (Huang et al. 2006):

722
$$\hat{\beta} = H^+Y, \quad (\text{A.13})$$

723 where $\hat{\beta}$ represents the deduced output weights.

724 For the learning stage, appropriate optimisation for the number of neurons in the
725 hidden layer is significant because it prevents the ELM from learning noise data,
726 which may result in overfitting and a reduction in model performance. To determine
727 the optimised structure, many workers have gradually increased the number of
728 neurons and observed changes in model performance (Deo and SAhin, 2015;
729 Mouatadid and Adamowski, 2016; Yaseen et al. 2016). However, this approach may
730 not work efficiently; accordingly, automatic optimisation methods have been
731 developed. Miche et al. (2010) presented the optimally pruned ELM for modelling
732 both classification and regression problems, which can add and remove hidden
733 neurons or automatically rank the parameters by relevance to the problem. The
734 optimally pruned algorithm for ELM models is executed in three steps: 1) ELM
735 model building, 2) ranking of the hidden neurons and 3) decision-making on how
736 many neurons are pruned. Detailed information about the principles and application of
737 the optimally pruned algorithm has been presented previously (Huang et al. 2006;
738 Miche et al. 2010). Since this automatic optimisation technique is considered to train

739 ELM efficiently while maintaining its accuracy and robustness (Miche et al. 2010), it
 740 was adopted in the present study for ELM model structure optimisation.

741 A.3. Wavelet transformation algorithm

742 The wavelet ($W_{j,t}$) and scale ($V_{j,t}$) coefficients can be decomposed by MODWT:

743
$$W_{j,t} = \sum_{l=0}^{L-1} h_l V_{j-1,t-2^{j-1}l \bmod N} \quad (\text{A.14})$$

744 and

745
$$V_{j,t} = \sum_{l=0}^{L-1} g_l V_{j-1,t-2^{j-1}l \bmod N}, \quad (\text{A.15})$$

746 where N is the sample size and \bmod refers to the modulo operator; $W_{j,t}$ and $V_{j,t}$

747 represent the j th level wavelet and scaling coefficients, respectively, at time t ; $V_{0,t}$

748 represents the original time series; $g_l \equiv \tilde{g}_l/\sqrt{2}$, $h_l \equiv \tilde{h}_l/\sqrt{2}$ and \tilde{g}_l (\tilde{h}_l) are scaling

749 (wavelet) filters and L is the length of the filter. Unlike Mallat's algorithm, when

750 MODWT convolves the data at a certain time it does not need to perform non-

751 redundant resampling; rather, it simply uses the current and previous values.

752 Accordingly, the sequences generated by MODWT have time-shift invariance.

753 Besides, the right side of a subsequence will not be distorted, which can be

754 conveniently applied in an actual prediction. As for AT, it uses the scale coefficient

755 generated by MODWT as an approximation coefficient ($A_{j,t}^a$) and then directly

756 subtracts the original series from the obtained approximation coefficient to obtain the

757 detail coefficient ($D_{j,t}^a$):

758
$$A_{j,t}^a = V_{j,t} \quad (\text{A.16})$$

759 and

760
$$D_{j,t}^a = A_{j-1,t}^a - A_{j,t}^a. \quad (\text{A.17})$$

761 Here, the original time series decomposed by AT can be regained by adding detail and

762 approximation coefficients:

763 $A_{0,t}^a = \sum_{j=1}^J D_{j,t}^a + A_{J,t}^a.$ (A.18)

764 A.4. Stochastic forecast algorithm

765 The stochastic conversion blueprint in Montanari and Koutsoyiannis (2012) begins
 766 with the simple deterministic hydrologic model formulation:

767
$$Q = S(\boldsymbol{\Theta}, \mathbf{X}) + e,$$
 (A.19)

768 where Q is the true series of the target to be forecast, S represents the deterministic
 769 model structure that transforms inputs into outputs, \mathbf{X} represents the input variables, $\boldsymbol{\Theta}$
 770 signifies the model parameters and e reflects a deviation from the deterministic
 771 transform. Through several formula derivation steps (shown as Eqs. (3)–(6) in
 772 Montanari and Koutsoyiannis (2012)), the deterministic model in Eq. (A.19) is
 773 converted to the corresponding stochastic form:

774
$$f_Q(Q) = \int_{\boldsymbol{\Theta}} \int_{\mathbf{X}} f_e(Q - S(\boldsymbol{\Theta}, \mathbf{X}) | \boldsymbol{\Theta}, \mathbf{X}) f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}) f_{\mathbf{X}}(\mathbf{X}) d\boldsymbol{\Theta} d\mathbf{X},$$
 (A.20)

775 where $Q - S(\boldsymbol{\Theta}, \mathbf{X})$ is the model error e incorporating all uncertainties not explicitly
 776 accounted for in Eq. (A.20), such as the uncertainty due to initial conditions or model
 777 structure; $f_{\boldsymbol{\Theta}}(\boldsymbol{\Theta})$ and $f_{\mathbf{X}}(\mathbf{X})$ are the pdfs explicitly representing uncertainties of model
 778 parameter $\boldsymbol{\Theta}$ and input data \mathbf{X} , respectively, which are assumed to be independent of
 779 each other; $f_e(Q - S(\boldsymbol{\Theta}, \mathbf{X}) | \boldsymbol{\Theta}, \mathbf{X})$ is the conditional pdf of the model error e ,
 780 conditioned on $\boldsymbol{\Theta}$ and \mathbf{X} and $f_Q(Q)$ is the pdf of the true target variable to be forecast,
 781 which quantifies uncertainty in the true target time series. In Eq. (A.20), several
 782 assumptions and definitions are also made (Montanari and Koutsoyiannis, 2012), i.e.
 783 (1) the forecast uncertainty is the uncertainty in the forecast time series and is formed
 784 by the input data, the model parameter and the model error uncertainties; (2) the input
 785 data uncertainty is mostly related to observation methods and networks, while the
 786 model parameter contains uncertainties due to the model structure, calibration

787 schemes and dataset consistency; (3) for the forecast process, forecast uncertainty is
788 defined using the prediction interval of the forecast results, which determines a range
789 containing the true value of the target variable with probability equal to the nominal
790 confidence level; (4) as for the pdf estimation, the resampling method of
791 bootstrapping can be used to avoid the definition of the likelihood function, which is a
792 difficult task in hydrology and may cause a lack of fit while the inappropriate
793 uncertainty estimation propagates.

794 A.4.1. Stochastic OF and DF

795 Quilty et al. (2019) and Quilty and Adamowski (2020) extended the original
796 blueprint to focus on deterministic data-driven models with slight modification,
797 developing both ordinary stochastic and direct wavelet-based stochastic data-driven
798 forecast frameworks. In fact, they took into account the input selection uncertainty
799 instead of the input data uncertainty because the latter is difficult to measure
800 practically. According to Quilty and Adamowski (2020), the pdf taking both the
801 parameters and model error uncertainties into consideration can be denoted as

802
$$f_Q(Q) = \int_{\Theta} f_e(Q - S(\Theta, X) | \Theta) f_{\Theta}(\Theta) d\Theta. \quad (\text{A.21})$$

803 The steps for using the bootstrapping method to estimate the forecasted pdf $f_Q(Q)$ of
804 *SOF_PM/SDF_PM* are as follows:

- 805 ● Step 1: A set of model parameters is resampled from the pdf of $f_{\Theta}(\Theta)$;
806 ● Step 2: Using the sampled parameters, a point series is forecast via $S(\Theta, X)$;
807 ● Step 3: The k nearest neighbour method (KNN) introduced in Sikorska-Senoner
808 et al. (2014) is used to simulate the model error $Q - S(\Theta, X)$ for the testing set,
809 while the model error of the validation set is taken as the searched neighbour;

- 810 ● Step 4: A certain number (m) of error series is resampled and picked up from the
 811 simulated error series in Step 3 and then separately added to $S(\boldsymbol{\theta}, \mathbf{X})$ generated in
 812 Step 2 to form an m point forecast series;
 813 ● Step 5: Steps 1–4 are repeated n times, resulting in $m \times n$ different point forecast
 814 series (which are also called probabilistic ensemble members), from which the
 815 forecast pdf $f_Q(Q)$ is realised.

816 Hereinafter, Steps 3 and 4 are collectively termed the KNN bootstrapping step
 817 (Sikorska-Senoner et al. 2014). By skipping this step, the stochastic framework will
 818 convert to *SOF_P/SDF_P*, considering only the parameter uncertainty.

819 A.4.2. Stochastic MF

820 Taking the building process of *SOF_PM/SDF_PM* as a reference, this study
 821 proposes the following method to build *SMF_PM*, which involves both parameters
 822 and model error uncertainties:

$$823 \quad f_Q(Q) = \sum_{\omega \in \mathcal{C}} \int_{\boldsymbol{\theta}^\omega} f_e(Q^\omega - S(\boldsymbol{\theta}^\omega, \mathbf{X}^\omega) | \boldsymbol{\theta}^\omega) f_{\boldsymbol{\theta}^\omega}(\boldsymbol{\theta}^\omega) d\boldsymbol{\theta}^\omega, \quad (\text{A.22})$$

824 where Q^ω represents the ω th component in the decomposed component set \mathcal{C} of the
 825 target variable, while $\{D_{1,t}^a, D_{2,t}^a, \dots, D_{J,t}^a, A_{J,t}^a\} \in \mathcal{C}$ and $\sum_{j=1}^J D_{j,t}^a + A_{J,t}^a = A_{0,t}^a$ (see
 826 Section A.3); \mathbf{X}^ω and $\boldsymbol{\theta}^\omega$ are, respectively, the input variable and model parameter set
 827 for forecasting the corresponding Q^ω . The steps for estimating $f_Q(Q)$ of Eq. (A.22)
 828 are as follows:

- 829 ● Step 1: For each Q^ω , a set of random model parameters is resampled from the
 830 corresponding pdf of $f_{\boldsymbol{\theta}^\omega}(\boldsymbol{\theta}^\omega)$;
 831 ● Step 2: With $S(\boldsymbol{\theta}^\omega, \mathbf{X}^\omega)$, $J + 1$ of the forecast series of the target decomposed
 832 components is obtained;

833 ● Step 3: $J + 1$ of model error series $Q^\omega - S(\boldsymbol{\theta}^\omega, \mathbf{X}^\omega)$ is simulated via KNN and
834 then $J + 1$ of the sampled error series is randomly chosen from the corresponding
835 simulated model error series;
836 ● Step 4: The sampled series generated in Step 3 are individually added to the
837 corresponding point series generated by $S(\boldsymbol{\theta}^\omega, \mathbf{X}^\omega)$ in Step 2 to obtain $J + 1$ of
838 the new forecast decomposed components, which are added together to acquire
839 one target point forecast series;
840 ● Step 5: Steps 3 and 4 are repeated m times to generate an m point forecast series;
841 ● Step 6: Steps 1–5 are repeated n times, resulting in $m \times n$ different forecast series
842 of Q to realise $f_Q(Q)$.
843 To build SMF_P , Steps 3–5 are replaced with the step of adding together the $J + 1$
844 values of the forecast decomposed components resulting from Step 2 to output the
845 target forecast series.

846 Appendix B

847 B.1. Influence of parameters b_size and $n_neighbors$ on probabilistic prediction
848 **【Insert Figure B.1】**
849 **【Insert Figure B.2】**
850 Figure B.1 shows a representative analysis case, presenting the response surfaces of
851 point and interval prediction performance metrics influenced by b_size and
852 $n_neighbors$ for the testing set of ELM_sigm with SOF_PM . The b_size in each
853 response space ranges from 20 to 100, whereas $n_neighbors$ varies from 1 to 20. From
854 Fig. B.1 (a)–(c), it can be seen that the response of each deterministic indicator to
855 b_size or $n_neighbors$ is slight but irregular throughout the entire response space. On

856 the contrary, the response surface of each interval prediction indicator in Fig. B.1 (d)–
857 (f) shows large variation with a regular changing trend. The largest variations in
858 *CRPS*, *PICP* and *AW* are 12.1%, 110% and 421.5%, respectively, in the response
859 spaces, while they are merely 2.8%, 4.0% and 1.8% for *NSE*, *MAE* and *RMSE*,
860 respectively.

861 Although the overall variations in *CRPS*, *PICP* and *AW* are marked, their response
862 surfaces in Fig. B.1 (d)–(f) are observed to vary with *n_neighbors*. This implies that
863 the interval prediction performances are mainly affected by *n_neighbors*. The same
864 conclusion can be drawn from the coverage probability plot in Fig. B.2. For instance,
865 when *b_size* is fixed to 100, the curve of forecast quantiles versus theoretical quantiles
866 in Fig. B.2 (a) changes significantly with the increase of *n_neighbors*. However, for a
867 different *b_size* but the same *n_neighbors* (i.e. with *b_size* varying from 20 to 100
868 when *n_neighbors* is fixed to 1, 8 and 20 in Fig. B.2 (b)–(d), respectively) each curve
869 shows only slight changes. This result shows that the influence of *n_neighbors* on the
870 interval prediction is much greater than that of *b_size*.

871 The influence of *n_neighbors* on interval prediction was also found to be regular. In
872 Fig. B.2 (a), it can be seen that the curve is rather close to the 1:1 bisector while
873 *n_neighbors* is 1. As it increases from 1 to 20, the curve tends to become more curved,
874 indicating a growing underestimation of forecast uncertainty. The quantitative
875 indicator *PICP* in Fig. B.1 (e) will be closer to the expected level of 95% as
876 *n_neighbors* decreases, indicating an improvement in forecast reliability. In the
877 meantime, *AW* increases significantly, implying a deteriorating forecast sharpness.
878 This is a normal result since the prediction interval indicated by *PICP* gets wider.
879 Additionally, *CRPS*, which measures the forecast error between interval prediction
880 and observation in Fig. B.1 (d), decreases continually as *n_neighbors* decreases.

881

882 Acknowledgments

883 This study was supported by the Innovation Fund of Guangzhou City water science
884 and technology (GZSWKJ-2020-2), the National Natural Science Foundation of
885 China (Grant Nos. 51879289), and Guangdong Basic and Applied Basic Research
886 Foundation (2019B1515120052).

887

888 Compliance with Ethical Standards

889 The authors declare that they have no conflict of interest. This article does not
890 contain any studies with human participants or animals performed by any of the
891 authors, and ethics approval was not required for this research. Informed consent was
892 obtained from all individual participants included in the study.

893

894 References

- 895 Adnan, R.M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Kisi, O., 2019. Daily
896 streamflow prediction using optimally pruned extreme learning machine. Journal of
897 Hydrology, 577: 123981.
898 Akusok, A., Bjork, K.M., Miche, Y., Lendasse, A., 2015. High-Performance Extreme
899 Learning Machines: A Complete Toolbox for Big Data Applications. Access IEEE, 3:
900 1011-1025.
901 Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2013. Long-term SPI drought
902 forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression
903 models. Journal of Hydrology, 508.
904 Bowden, G., Dandy, G., Maier, H., 2005a. Input determination for neural network models in
905 water resources applications. Part 1 - Background and methodology. Journal of Hydrology,
906 301: 75-92.
907 Bowden, G.J., Maier, H.R., Dandy, G.C., 2005b. Input determination for neural network
908 models in water resources applications. Part 2. Case study: forecasting salinity in a river.
909 Journal of Hydrology, 301(1): 93-107.
910 Chan., J.C.L., Zhou., W., 2005. PDO, ENSO and the early summer monsoon rainfall over
911 South China. Geophysical Research Letters, 32.
912 CHEN, S., CHNG, E.S., ALKADHIMI, K., 1991. Orthogonal least squares learning
913 algorithm for radial basis function networks. IEEE Trans Neural Netw, 64(5): 829-837.

- 914 Cover, T.M., Allen, T., 1991. Elements of Information Theory, Wiley Series in
 915 Telecommunications. Tsinghua University Press.
- 916 Dash, M., Liu, H., 1997. Feature selection for classification. *Intell. Data Anal.*, 1: 131-156.
- 917 Deo, R.C., SAhin, M., 2015. Application of the extreme learning machine algorithm for the
 918 prediction of monthly Effective Drought Index in eastern Australia. *Atmospheric*
 919 *Research*, 153: 512-525.
- 920 Dietrich, W.E., 1987. Mechanics of flow and sediment transport in River bends. *River*
 921 *channels*: 179-227.
- 922 Du, K., Zhao, Y., Lei, J., 2017. The wrong usage of singular spectral analysis and discrete
 923 wavelet transform in hybrid models to predict hydrological time series. *Journal of*
 924 *Hydrology*, 552.
- 925 Farmer, W., 2016. Ordinary kriging as a tool to estimate historical daily streamflow records.
 926 *Hydrology and Earth System Sciences*, 20: 2721-2735.
- 927 Galelli, S. et al., 2014. An evaluation framework for input variable selection algorithms for
 928 environmental data-driven models. *Environmental Modelling & Software*, 62: 33–51.
- 929 Ghaemi, A., Rezaie-Balf, M., Adamowski, J., Kisi, O., Quilty, J., 2019. On the applicability
 930 of maximum overlap discrete wavelet transform integrated with MARS and M5 model
 931 tree for monthly pan evaporation prediction. *Agricultural and Forest Meteorology*, 278.
- 932 Glantz, M., Ramirez, I., 2020. Reviewing the Oceanic Niño Index (ONI) to Enhance Societal
 933 Readiness for El Niño's Impacts. *Int J Disaster Risk Sci*, 11: 394–403
- 934 Gneiting, T., Balabdaoui, F., Raftery, A., 2007. Probabilistic forecasts, calibration and
 935 sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69:
 936 243-268.
- 937 Gong, D.-Y., Wang, S.-W., Zhu, J.-H., 2001. East Asian Winter Monsoon and Arctic
 938 Oscillation. *Geophysical Research Letters*, 28: 2073-2076.
- 939 GU, X., ZHANG, Q., CHEN, X., JIANG, T., 2014. Nonstationary Flood Frequency Analysis
 940 Considering the Combined Effects of Climate Change and Human Activities in the East
 941 River Basin (in Chinese with English abstract). *Tropical Geography*, 34 (6): 746-757.
- 942 Guyon, I., Elisseeff, A., 2003. An Introduction of Variable and Feature Selection. *J. Machine*
 943 *Learning Research Special Issue on Variable and Feature Selection*, 3: 1157-1182.
- 944 Hang, T.N., Nabney, I.T., 2010. Short-term electricity demand and gas price forecasts using
 945 wavelet transforms and adaptive models. *Energy*, 35(9): 3674-3685.
- 946 He, J., Valeo, C., Chu, A., Neumann, N.F., 2011. Prediction of event-based stormwater runoff
 947 quantity and quality by ANNs developed using PMI-based input selection. *Journal of*
 948 *Hydrology*, 400(1-2): 10-23.
- 949 He, S., Gao, Y., li, F., Wang, H., He, Y., 2016. Impact of Arctic Oscillation on the East Asian
 950 climate: A review. *Earth-Science Reviews*, 164.
- 951 Hu, J., Liu, B., Peng, S., 2019. Forecasting salinity time series using RF and ELM approaches
 952 coupled with decomposition techniques. *Stochastic Environmental Research & Risk*
 953 *Assessment*, 33(4-6): 1117-1135.
- 954 Huang, G.-B., Zhu, Q.-Y., Siew, C., 2006. Extreme Learning Machine: Theory and
 955 Applications. *Neurocomputing*, 70.
- 956 Huang, G.B., Zhou, H., Ding, X., Rui, Z., 2012. Extreme Learning Machine for Regression
 957 and Multiclass Classification. *IEEE Transactions on Systems Man & Cybernetics Part B*,
 958 42(2): 513-529.
- 959 Humphrey, G., Gibbs, M., Dandy, G., Maier, H., 2016. A hybrid approach to monthly
 960 streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial
 961 neural network. *Journal of Hydrology*, 540.
- 962 Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous
 963 hydrological variables. *Water Resources Research*, 11: 1267-1277.
- 964 Legates, D., McCabe, G., 1999. Evaluating the Use Of “Goodness-of-Fit” Measures in
 965 Hydrologic and Hydroclimatic Model Validation. *Water Resources Research*, 35: 233-241.
- 966 Li, C., Jinhai, H., Jinhong, Z., 2004. A review of decadal/interdecadal climate variation
 967 studies in China. *Advances in Atmospheric Sciences*, 21: 425-436.

- 968 Li, C., Li, G., 2000. The NPO/ NAO and interdecadal climate variation in China. Advances in
969 Atmospheric Sciences, 17: 555-561.
- 970 Liu, Z., Zhou, P., Zhang, Y., 2015. A Probabilistic Wavelet-Support Vector Regression
971 Model for Streamflow Forecasting with Rainfall and Climate Information Input. Journal
972 of Hydrometeorology, 16: 2209-2229.
- 973 Lu, A. et al., 2010. El Nio-Southern Oscillation and water resources in Headwaters Region of
974 the Yellow River: links and potential for forecasting. Hydrology & Earth System Ences
975 Discussions, 7(5): 1273-1281.
- 976 Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic
977 forecasting. Computers Geoences, 46(none): 284-295.
- 978 Maier, H., Dandy, G., 1999. Empirical comparison of various methods for training feed-
979 Forward neural networks for salinity forecasting. Water Resources Research, 35.
- 980 May, R., Dandy, G., Maier, H., 2011. Review of Input Variable Selection Methods for
981 Artificial Neural Networks. InTech.
- 982 May, R.J., Maier, H.R., Dandy, G.C., Fernando, T.M.K.G., 2008. Non-linear variable
983 selection for artificial neural networks using partial mutual information. Environmental
984 Modelling & Software, 23(10-11): 1312-1326.
- 985 Meng, E. et al., 2019. A robust method for non-stationary streamflow prediction based on
986 improved EMD-SVM model. Journal of Hydrology, 568: 462-478.
- 987 Miche, Y. et al., 2010. OP-ELM: Optimally-Pruned Extreme Learning Machine. IEEE
988 Transactions on Neural Networks, 21(1): 158-162.
- 989 Montanari, A., Koutsoyiannis, D., 2012. A blueprint for process-based modeling of uncertain
990 hydrological systems. Water Resources Research, 48(9): 9555.
- 991 Moriasi, D., 2007. Model evaluation guidelines for systematic quantification of accuracy in
992 watershed simulations. Transactions of the Asabe, 50.
- 993 Mouatadid, S., Adamowski, J., 2016. Using extreme learning machines for short-term urban
994 water demand forecasting. Urban Water Journal, 14: 1-9.
- 995 Mouatadid, S., Adamowski, J., C, M.K.T., B, J.M.Q., 2019. Coupling the maximum overlap
996 discrete wavelet transform and long short-term memory networks for irrigation flow
997 forecasting. Agricultural Water Management, 219: 72-85.
- 998 Niu, J., 2013. Precipitation in the Pearl River basin, South China: Scaling, regional patterns,
999 and influence of large-scale climate anomalies. Stochastic Environmental Research and
1000 Risk Assessment, 27: 1253-1268.
- 1001 Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid
1002 wavelet-Artificial Intelligence models in hydrology: A review. Journal of Hydrology, 514:
1003 358-377.
- 1004 Nourani, V., Komasi, M., Mano, A., 2009. A Multivariate ANN-Wavelet Approach for
1005 Rainfall–Runoff Modeling. Water Resources Management, 23(14): 2877.
- 1006 Quilty, J., Adamowski, J., 2018. Addressing the incorrect usage of wavelet-based
1007 hydrological and water resources forecasting models for real-world applications with best
1008 practices and a new forecasting framework. Journal of Hydrology: 336-353.
- 1009 Quilty, J., Adamowski, J., 2020. A stochastic wavelet-based data-driven framework for
1010 forecasting uncertain multiscale hydrological and water resources processes.
1011 Environmental Modelling & Software, 130.
- 1012 Quilty, J., Adamowski, J., Boucher, M.A., 2019. A Stochastic Data-Driven Ensemble
1013 Forecasting Framework for Water Resources: A Case Study Using Ensemble Members
1014 Derived From a Database of Deterministic Wavelet-Based Models. Water Resources
1015 Research, 55(1): 175-202.
- 1016 Rahman, A.T.M.S., Hosono, T., Quilty, J.M., Das, J., Basak, A., 2020. Multiscale
1017 Groundwater Level Forecasting: Coupling New Machine Learning Approaches with
1018 Wavelet Transforms. Advances in Water Resources, 141: 103595.
- 1019 Samadi, V., Tufford, D., Carbone, G., 2017. Estimating Hydrologic Model Uncertainty in the
1020 Presence of Complex Residual Error Structures. Stochastic Environmental Research and
1021 Risk Assessment, 32.

- 1022 Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. In:
 1023 Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., New York.
- 1024 Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water
 1025 supply management: Part 1 — A strategy for system predictor identification. *Journal of*
 1026 *Hydrology*, 239: 232-239.
- 1027 Sikorska-Senoner, A.E., Montanari, A., Koutsoyiannis, D., 2014. Estimating the Uncertainty
 1028 of Hydrological Predictions through Data-Driven Resampling Techniques. *Journal of*
 1029 *Hydrologic Engineering*, 20.
- 1030 Silverman, B., 1986. Density estimation for statistic and data analysis. In: *Mongraphs on*
 1031 *Statistics and Applied Probability*. Chapman and Hall, London.
- 1032 Tayfur, G., Karimi, Y., Singh, V., 2013. Principle Component Analysis in Conjunction with
 1033 Data Driven Methods for Sediment Load Prediction. *Water Resources Management*, 27,
- 1034 Tian, Q., Prange, M., Merkel, U., 2016. Precipitation and temperature changes in the major
 1035 Chinese river basins during 1957-2013 and links to sea surface temperature. *Journal of*
 1036 *Hydrology*, 536.
- 1037 Unnikrishnan, P., Jothiprakash, V., 2018. Daily rainfall forecasting for one year in a single
 1038 run using Singular Spectrum Analysis. *Journal of Hydrology*: S0022169418302828.
- 1039 Weerts et al., 2016. ENSO- conditioned weather resampling method for seasonal ensemble
 1040 streamflow prediction. *Hydrology & Earth System Sciences*, 20: 3277-3287.
- 1041 Wolter, K., Timlin, M., 1998. Measuring the strength of ENSO events: How does 1997/98
 1042 rank? *Weather*, 53: 315-324.
- 1043 Wolter, K., Timlin, M., 2011. El Nino-Southern Oscillation behaviour since 1871 as
 1044 diagnosed in an extended multivariate ENSO index (MEI.ext). *International Journal of*
 1045 *Climatology*, 31: 1074-1087.
- 1046 Yang, H., 2011. The Significant Relationship between the Arctic Oscillation (AO) in
 1047 December and the January Climate over South China. *Advances in Atmospheric Sciences*
 1048 - ADV ATMOS SCI, 28: 398-407.
- 1049 Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial Intelligence
 1050 based models for stream-flow forecasting: 2000-2015. *Journal of Hydrology*, 530: 829-
 1051 844.
- 1052 Yaseen, Z.M. et al., 2016. Stream-flow forecasting using extreme learning machines: a case
 1053 study in a semi-arid region in Iraq. *Journal of Hydrology*, 542: 603-614.
- 1054 Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.W., 2019. An enhanced extreme learning
 1055 machine model for river flow forecasting: State-of-the-art, practical applications in water
 1056 resource engineering area and future research direction. *Journal of Hydrology*, 569: 387-
 1057 408.
- 1058 Zhang, D. et al., 2018. Modeling and simulating of reservoir operation using the artificial
 1059 neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*,
 1060 565.
- 1061 Zhang, Q., Li, J., Singh, V., Xu, C.-Y., Deng, J., 2013. Influence of ENSO on precipitation in
 1062 the East River basin, South China. *Journal of Geophysical Research (Atmospheres)*, 118:
 1063 2207-2219.
- 1064 Zhang, R., Min, Q., Su, J., 2017. Impact of El Niño on atmospheric circulations over East
 1065 Asia and rainfall in China: Role of the anomalous western North Pacific anticyclone.
 1066 *Science China Earth Sciences*, 60.
- 1067 Zhang, X., Liang, F., Yu, B., Zong, Z., 2011. Explicitly Integrating Parameter, Input, and
 1068 Structure Uncertainties into Bayesian Neural Networks for Probabilistic Hydrologic
 1069 Forecasting. *Journal of Hydrology*, 409: 696-709.
- 1070 Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data
 1071 preprocessing techniques suitable for monthly streamflow forecasting? Some experiment
 1072 evidences. *Journal of Hydrology*, 530.
- 1073 Zhou, F., Liu, B., Duan, K., 2020. Coupling wavelet transform and artificial neural network
 1074 for forecasting estuarine salinity. *Journal of Hydrology*, 588: 125127.
- 1075
 1076

Figure Captions

Fig. 1 Boluo hydrological station and Huiyang precipitation station in the East River basin.

Fig. 2 The overall research roadmap of this study.

Fig. 3 Building process of (a) direct and (b) multi-component wavelet-based forecast methods.

Fig. 4 Prediction process diagrams of testing set for different stochastic wavelet-based forecast methods.

Fig. 5 Coverage probability plots of testing-set interval prediction for different stochastic wavelet-based forecast methods.

Fig. 6 Probability density distribution of KNN simulated and bootstrapping error with the $n_neighbors$ of 1, 10 and 20.

Fig. 7 The heat maps of PMI values for top 20 ranked variables of OF and DF , and the curves of cumulative PMI values of top n ranked variables.

Fig. 8 Statistical characteristic boxplots of KNN simulated and bootstrapping error.

Fig. B.1 Response surfaces of testing-set point and interval prediction performances influenced by b_size and $n_neighbors$.

Fig. B.2 Coverage probability plots of testing-set interval prediction for ELM_sigm with SOF_PM .

Table Captions

Table 1 19 selected potential climate phenomenon indices for monthly streamflow forecast.

Table 2 Statistical characteristics of monthly streamflow and precipitation series.

Table 3 Testing-set performances of different deterministic forecast methods.

Table 4 Testing-set performances of different stochastic forecast methods considering only parameter uncertainty.

Table 5 Testing-set performance of different stochastic forecast methods considering both parameter and model error uncertainties.

Table 6 NSE values of testing set for each forecast component of MF .

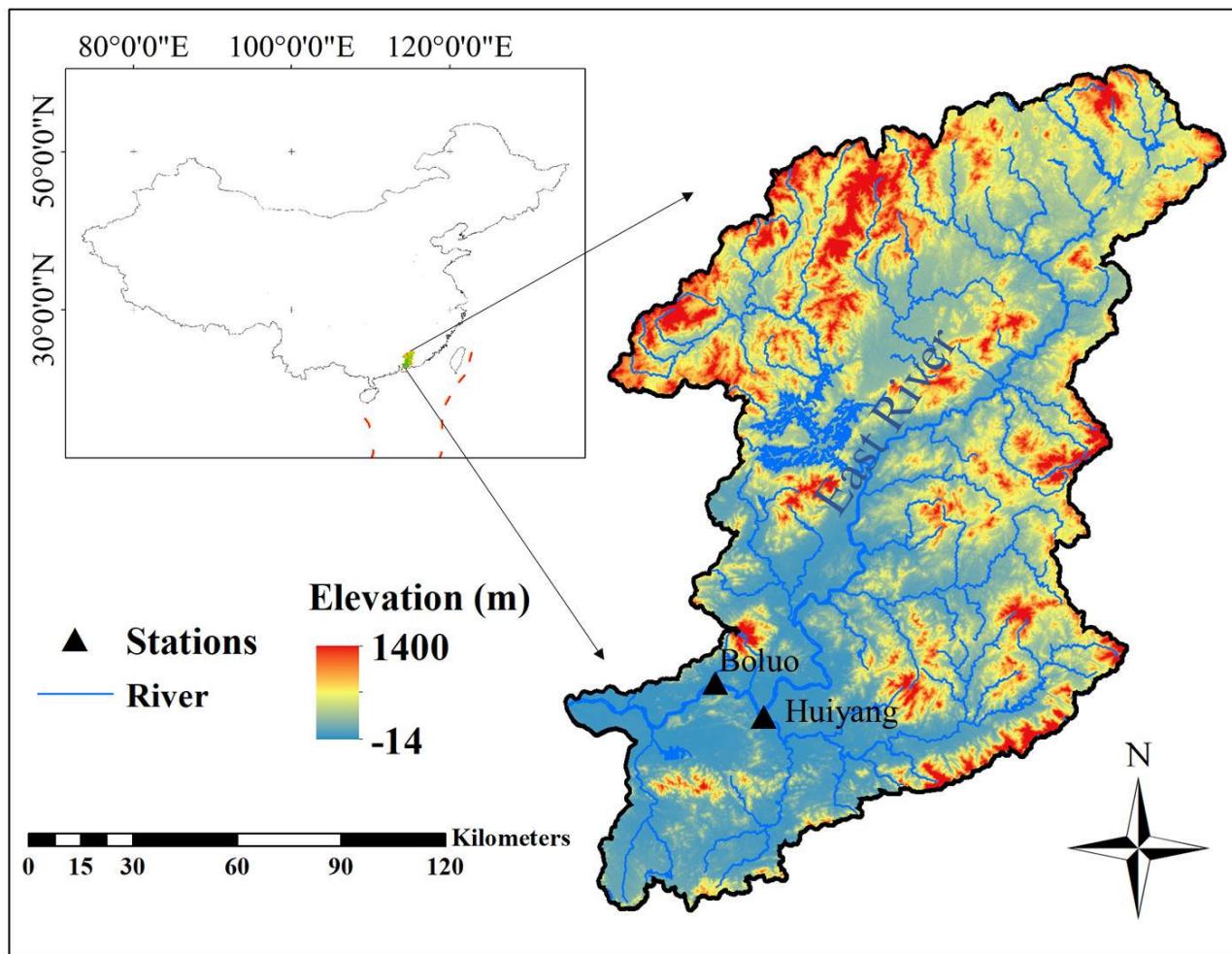


Fig. 1 Boluo hydrological station and Huiyang precipitation station in the East River basin.

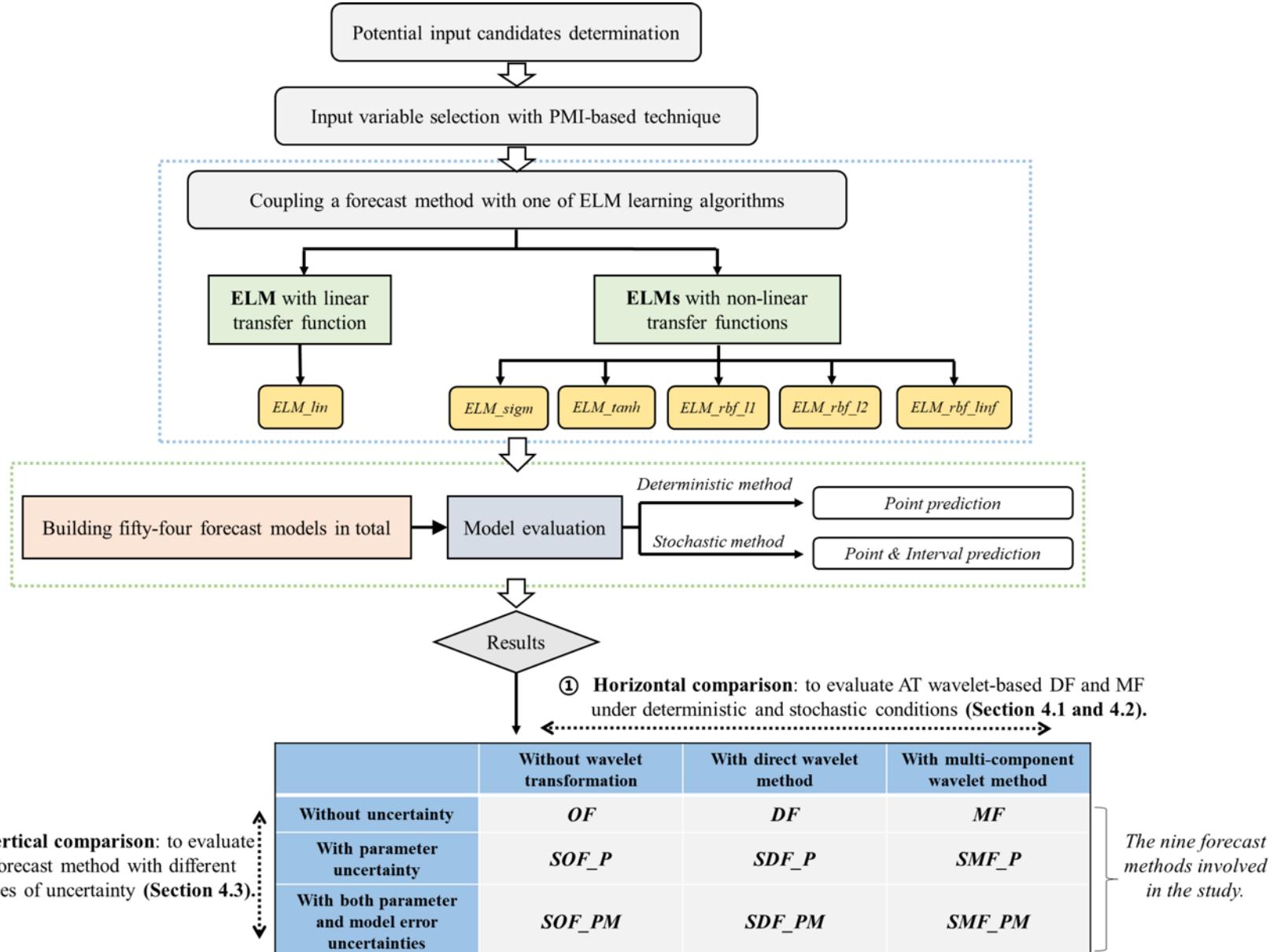


Fig. 2 The overall research roadmap of this study.

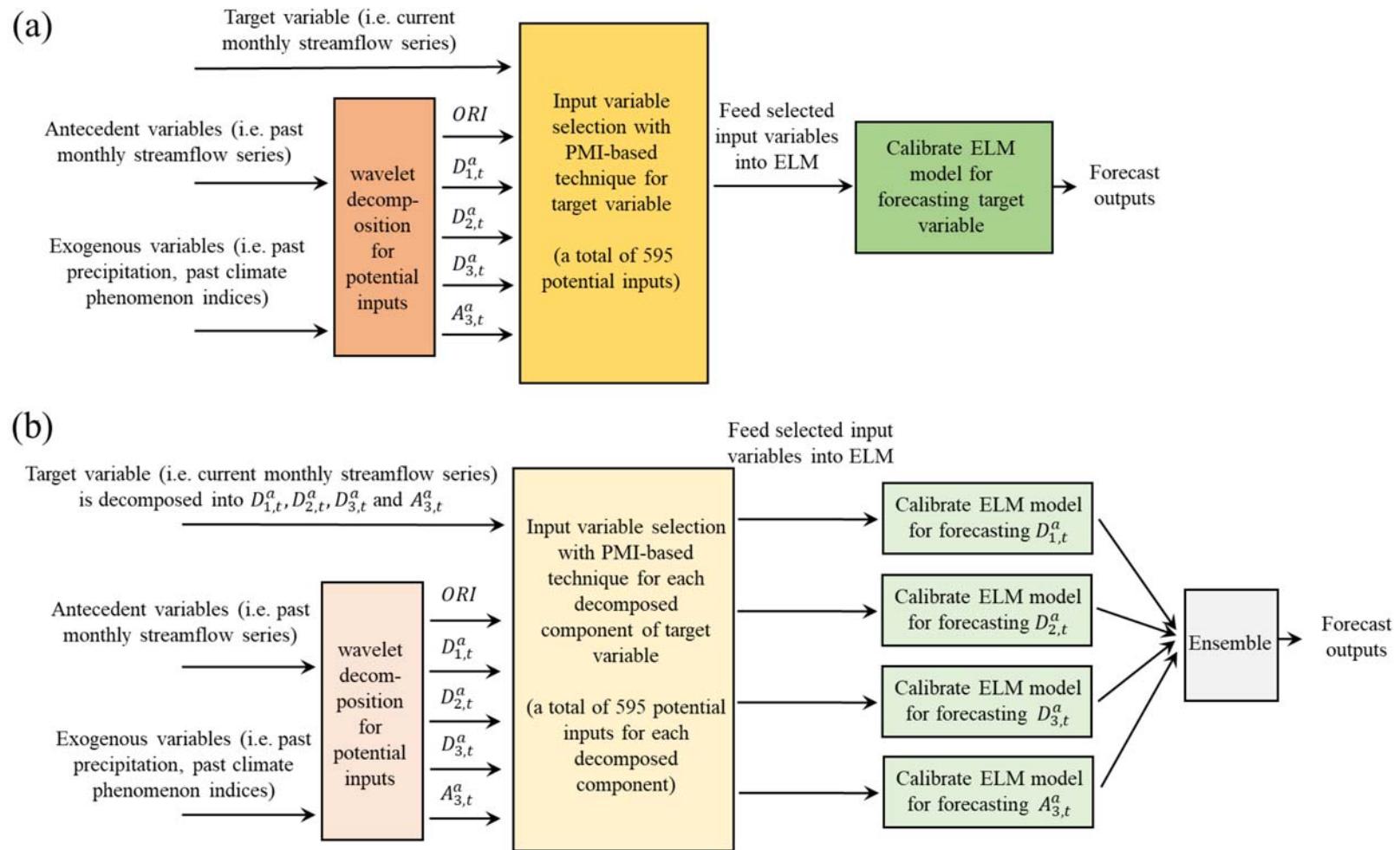


Fig. 3 Building process of (a) direct and (b) multi-component wavelet-based forecast methods.

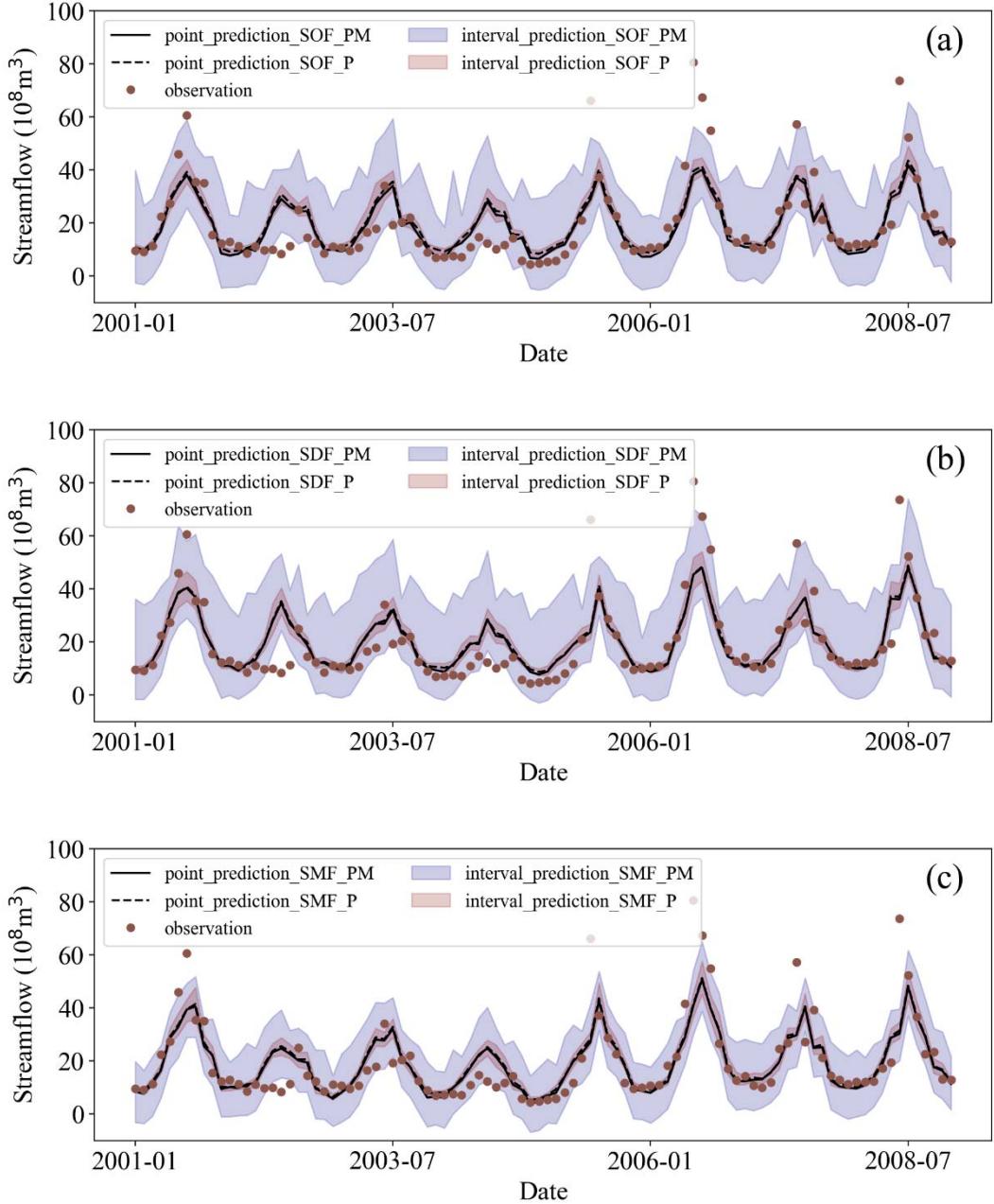


Fig. 4 Prediction process diagrams of testing set for different stochastic wavelet-based forecast methods. The point prediction is obtained by averaging the probabilistic ensemble members. Note that *ELM_sigm* is taken as the case here.

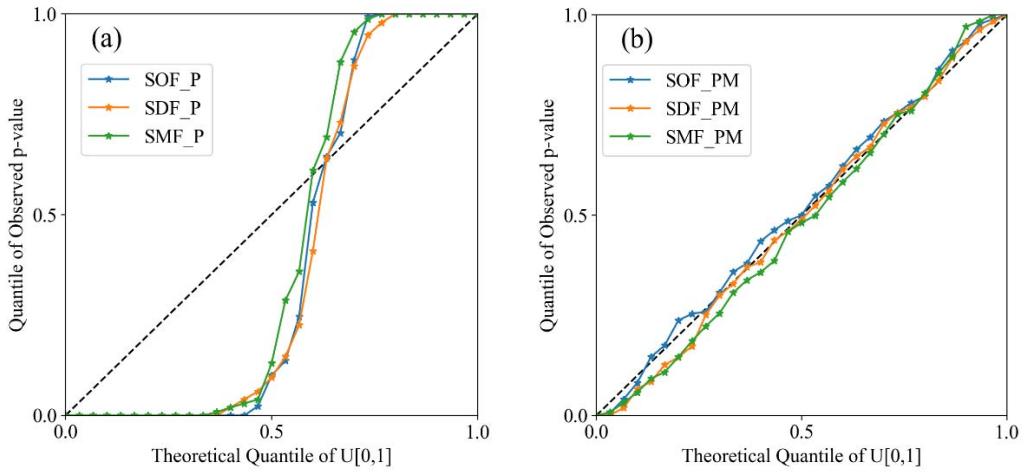


Fig. 5 Coverage probability plots of testing-set interval prediction for different stochastic wavelet-based forecast methods. Note that *ELM_sigm* is taken as the case here.

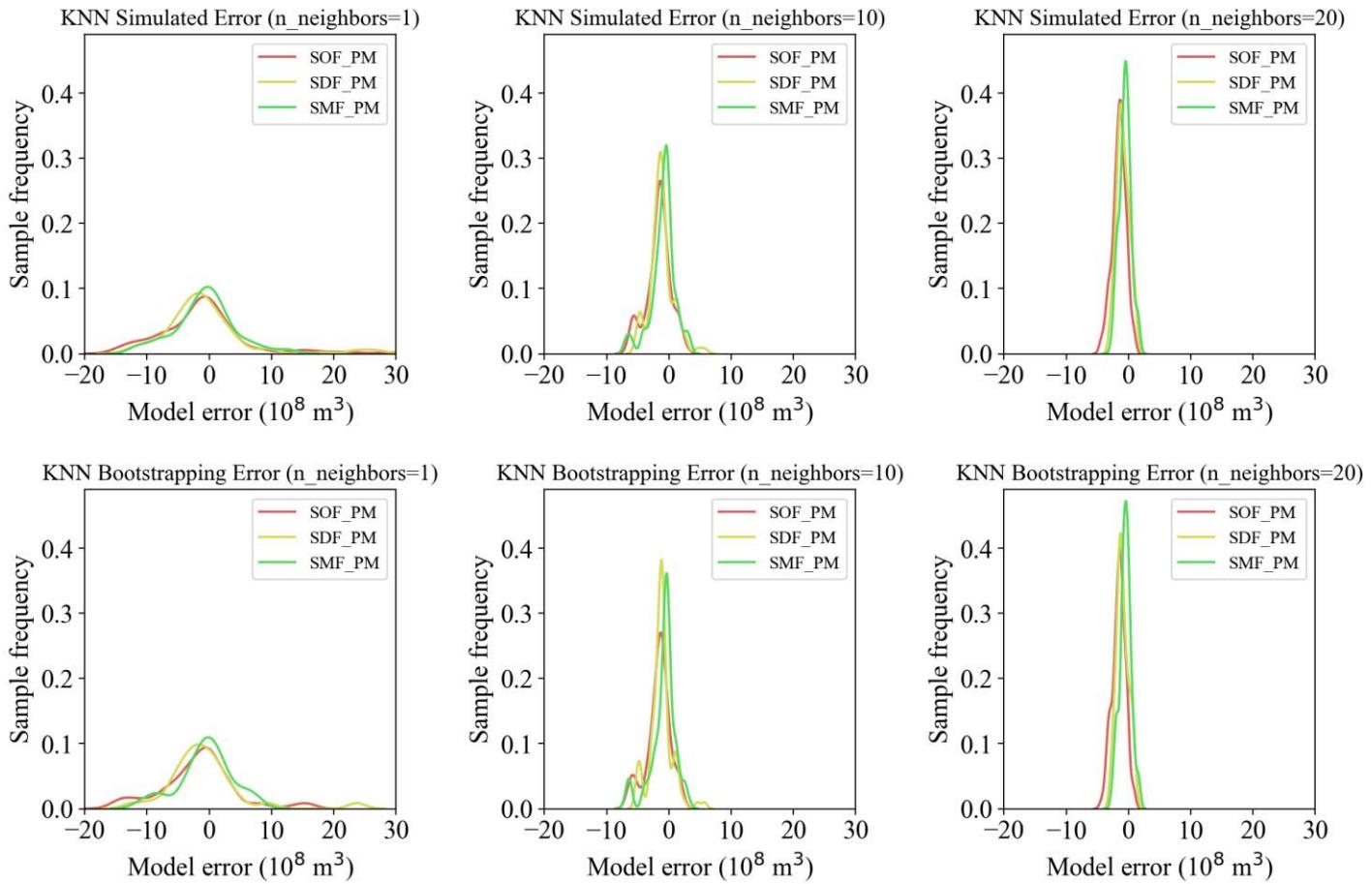


Fig. 6 Probability density distribution of KNN simulated and bootstrapping error with the $n_{neighbors}$ of 1, 10 and 20. Note that each distribution presents one representative error series, and *ELM_sigm* is adopted as the analysis case here.

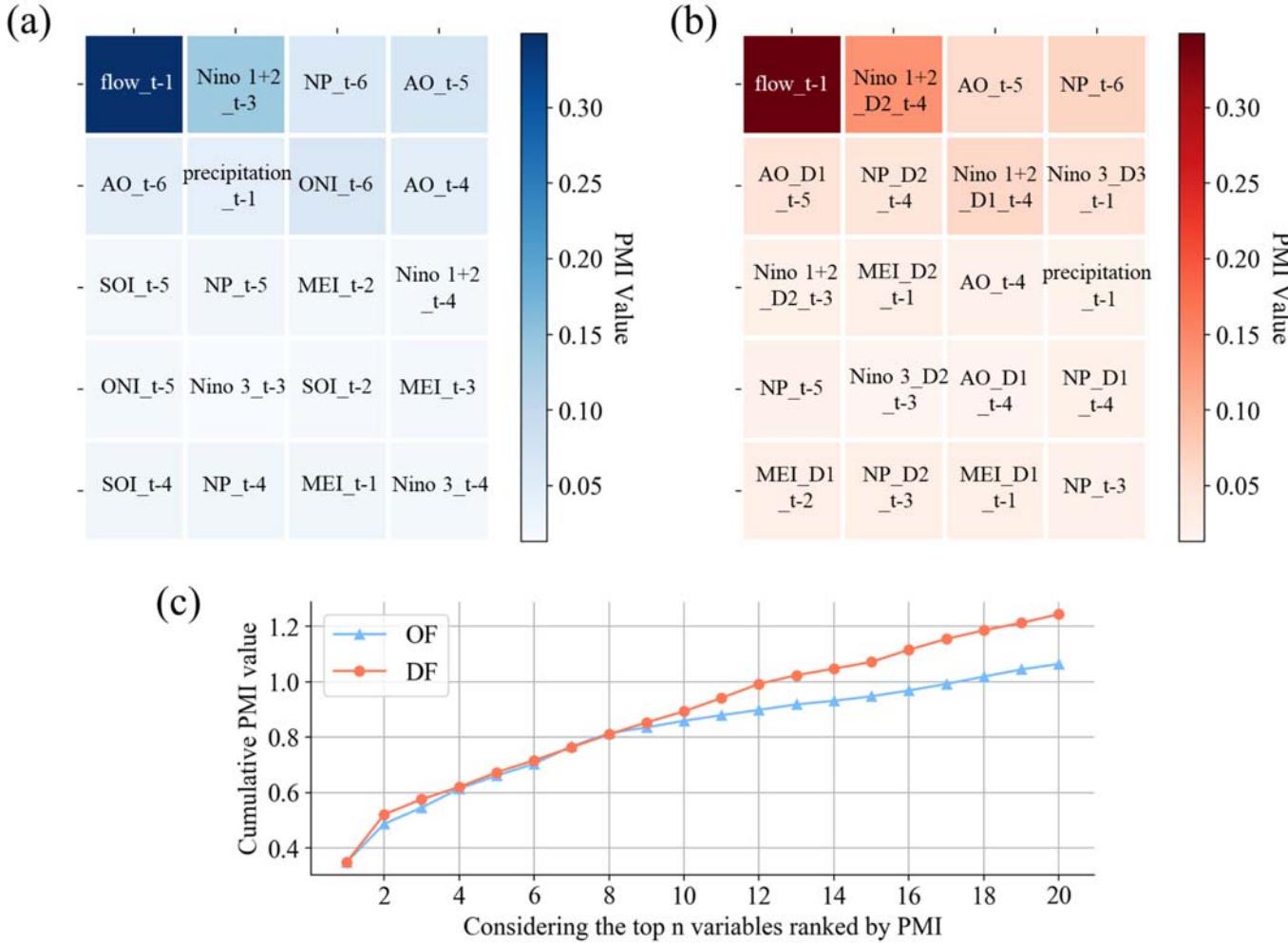


Fig. 7 Subfigure (a) and (b) represent the heat maps of PMI values for top 20 ranked variables of *OF* and *DF*, respectively. In the two subfigures, the ranking of variables decreases from left to right and from top to bottom. Annotation “*t-x*” in each variable represents the lead times of *x* months. ELMs with *lin*, *sigm*, *tanh*, *rbf_l1*, *rbf_l2* and *rbf_linf* select top 15, 7, 7, 4, 7 and 7 variables for *OF*, respectively, while top 17, 9, 10, 5, 7 and 10 variables are correspondingly fed for *DF*.

Subfigure (c) reflects the curves of cumulative PMI values of top *n* ranked variables shown in (a) and (b).

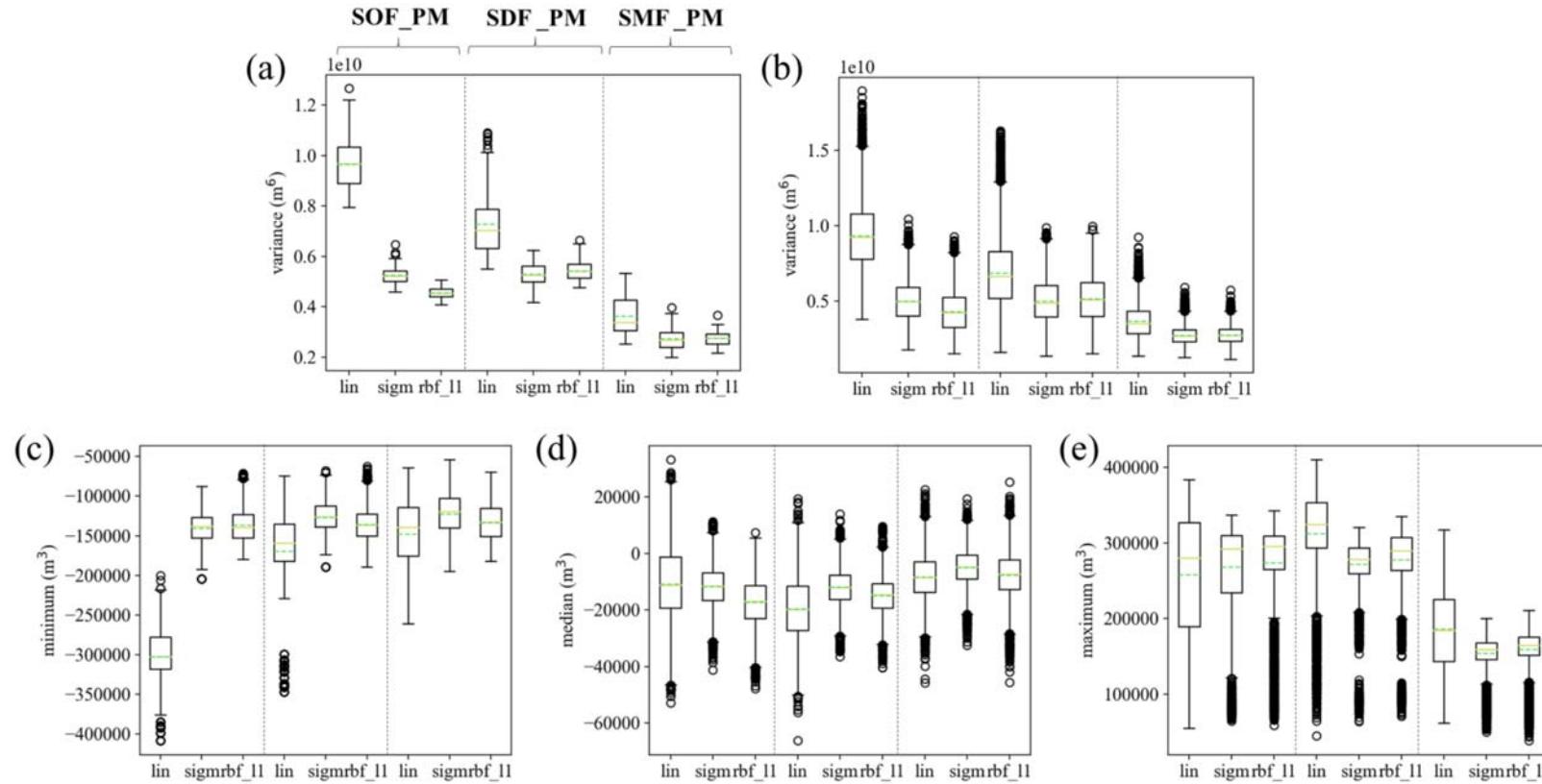


Fig. 8 Statistical characteristic boxplots of KNN simulated and bootstrapping error. Subfigure (a) and (b) respectively show the variance boxplots of KNN simulated and bootstrapping error. Subfigure (c)-(e) respectively show the minimum, median and maximum boxplots of KNN bootstrapping error. Hereinto, the yellow solid and green dotted line in each boxplot are respectively median and mean. In each subfigure of (a)-(e), the left, middle and right panels respectively represent *SOF_PM*, *SDF_PM* and *SMF_PM*, and ELMs with *lin*, *sigm* and *rbf_11* are taken as cases.

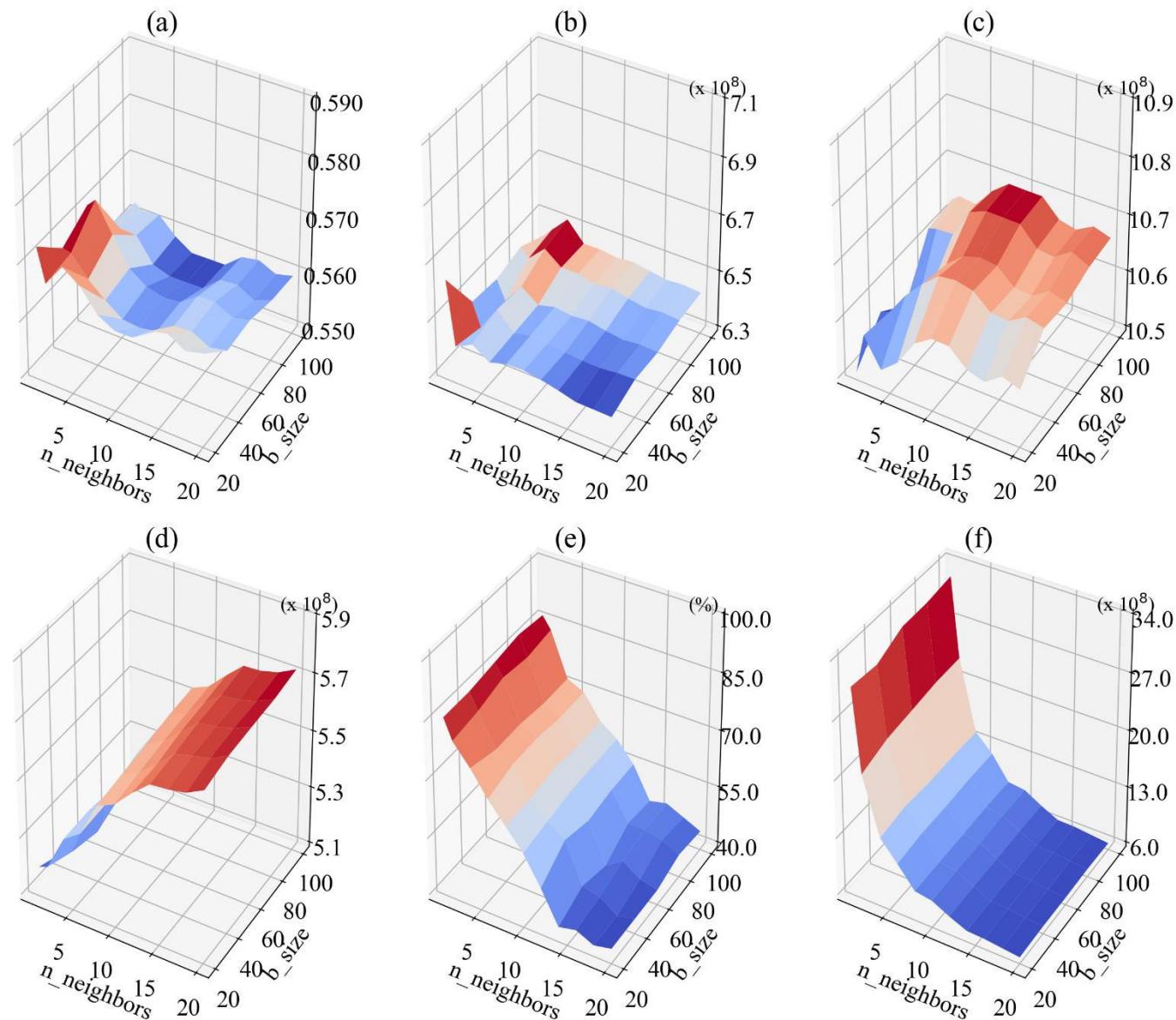


Fig. B.1 Response surfaces of testing-set point and interval prediction performances influenced by *b_size* and *n_neighbors*. *ELM_sigm* with *SOF_PM* is taken as the analysis case. Subfigure (a)-(f) respectively reflect the metrics of *NSE*, *MAE*, *RMSE*, *CRPS*, *PICP* and *AW*.

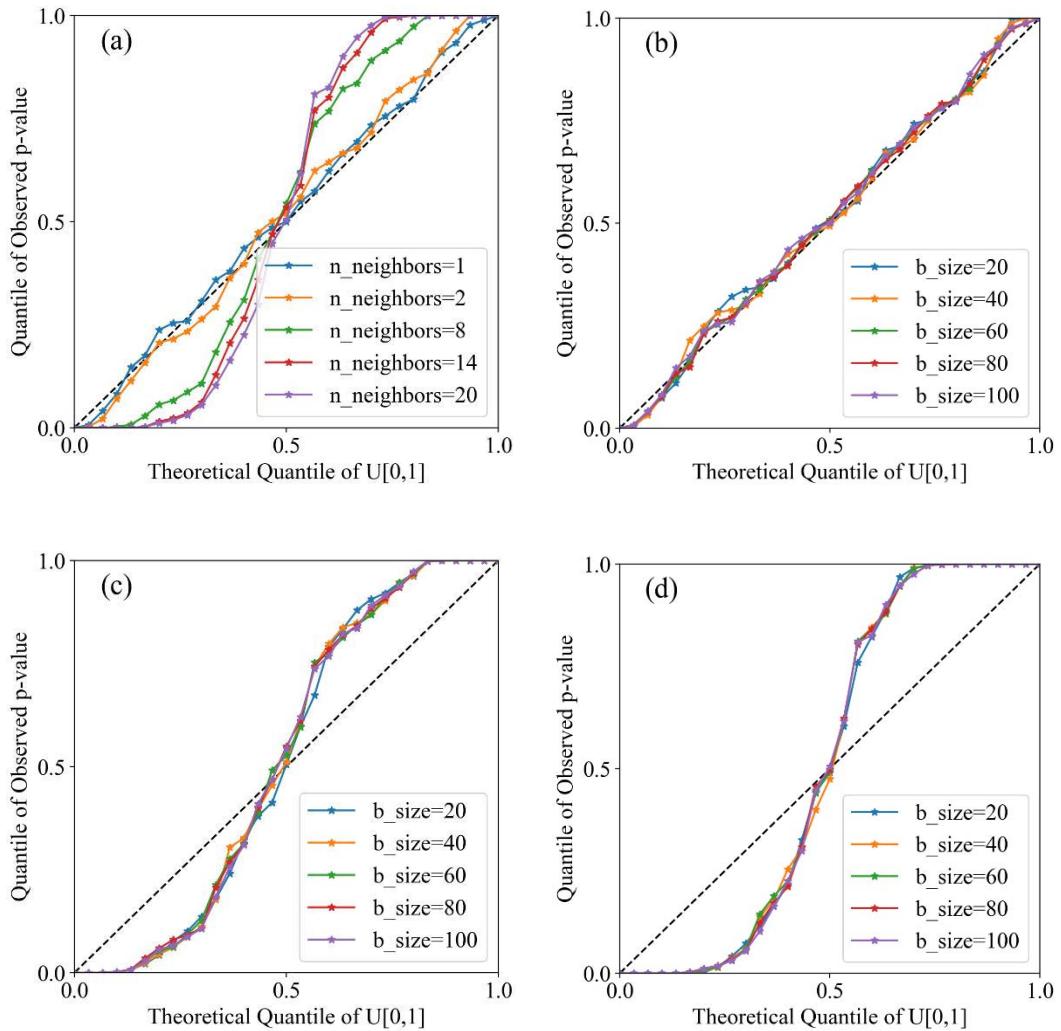


Fig. B.2 Coverage probability plots of testing-set interval prediction for *ELM_sigm* with *SOF_PM*.

The *b_size* of Subfigure (a) is 100, while the *n_neighbors* of Subfigure (b)-(d) are 1, 8 and 20,

respectively

Table 1 19 selected potential climate phenomenon indices for monthly streamflow forecast.

Index	Name	Source
AO	Antarctic Oscillation.	NOAA Climate Prediction Center (CPC)
EAWR	Eastern Asia/Western Russia	CPC
MEI	Multivariate ENSO Index	https://psl.noaa.gov/enso/mei.old/table.html (Wolter and Timlin, 1998)
NAO	North Atlantic Oscillation	CPC
Nino 1+2	Extreme Eastern Tropical Pacific SST *(0-10S, 90W-80W)	CPC
Nino 3	Eastern Tropical Pacific SST (5N-5S,150W-90W)	CPC
Nino 4	Central Tropical Pacific SST *(5N-5S) (160E-150W)	CPC
Nino 34	East Central Tropical Pacific SST* (5N-5S)(170-120W)	CPC
NP	North Pacific pattern	CPC
ONI	Oceanic Niño Index	CPC
PDO	Pacific Decadal Oscillation	CPC
PNA	Pacific North American Index	CPC
QBO	Quasi-Biennial Oscillation	CPC
SOI	Southern Oscillation Index	CPC
TNA	Tropical Northern Atlantic Index	CPC
TNI	Indices of El Niño evolution	CPC
TSA	Tropical Southern Atlantic Index	CPC
WHWP	Western Hemisphere warm pool	CPC
WP	Western Pacific Index	CPC

Table 2 Statistical characteristics of monthly streamflow and precipitation series.

Partition	Station	Sample size	Statistical characteristics				
			max	min	mean	median	st.dev
Monthly streamflow (10^8 m^3)	Boluo	612	127.3	1.9	19.8	15.3	14.4
		372	127.3	1.9	19.9	15.6	14.8
		144	70.2	5.5	19.5	15.4	11.7
		96	80.6	4.4	20.1	12.8	16.1
Monthly precipitation (mm)	Huiyang	612	782.6	0.0	145.7	98.8	144.5
		372	738.2	0.0	145.1	102.8	140.5
		144	559.9	0.0	143.1	102.4	132.9
		96	782.6	0.0	151.6	78.1	173.7

Table 3 Testing-set performances of different deterministic forecast methods.

Learning algorithm	NSE			MAE (10^8 m^3)			RMSE (10^8 m^3)		
	<i>OF</i>	<i>DF</i>	<i>MF</i>	<i>OF</i>	<i>DF</i>	<i>MF</i>	<i>OF</i>	<i>DF</i>	<i>MF</i>
<i>ELM_lin</i>	0.514	0.532	0.537	7.521	7.294	7.031	11.206	10.990	10.936
<i>ELM_sigm</i>	0.539	0.573	0.591	6.773	6.524	6.372	10.912	10.494	10.272
<i>ELM_tanh</i>	0.535	0.542	0.595	6.751	6.615	6.270	10.963	10.876	10.231
<i>ELM_rbf_l1</i>	0.530	0.590	0.615	6.609	6.446	5.939	11.019	10.293	9.972
<i>ELM_rbf_l2</i>	0.551	0.591	0.587	6.748	6.422	6.317	10.770	10.279	10.321
<i>ELM_rbf_linf</i>	0.563	0.595	0.606	6.559	6.338	6.227	10.624	10.230	10.081

Table 4 Testing-set performances of different stochastic forecast methods considering only parameter uncertainty.

Method	Learning algorithm	NSE	MAE (10^8 m^3)	RMSE (10^8 m^3)	CRPS (10^8 m^3)	PICP (%)	AW (10^8 m^3)
<i>SOF_P</i>	<i>ELM_lin</i>	0.514	7.462	11.203	6.293	57.292	9.724
	<i>ELM_sigm</i>	0.569	6.584	10.548	5.972	23.958	4.634
	<i>ELM_tanh</i>	0.545	6.714	10.841	6.057	30.208	5.208
	<i>ELM_rbf_ll</i>	0.514	6.834	11.204	6.274	32.292	4.532
	<i>ELM_rbf_l2</i>	0.575	6.593	10.476	5.949	26.042	4.912
	<i>ELM_rbf_linf</i>	0.559	6.609	10.674	6.019	26.042	4.542
<i>SDF_P</i>	<i>ELM_lin</i>	0.555	7.002	10.714	5.939	45.833	10.333
	<i>ELM_sigm</i>	0.583	6.493	10.379	5.787	33.333	5.322
	<i>ELM_tanh</i>	0.554	6.637	10.735	5.982	32.292	5.149
	<i>ELM_rbf_ll</i>	0.577	6.643	10.456	5.978	27.083	4.988
	<i>ELM_rbf_l2</i>	0.585	6.496	10.353	5.781	34.375	5.460
	<i>ELM_rbf_linf</i>	0.590	6.404	10.293	5.703	35.417	5.366
<i>SMF_P</i>	<i>ELM_lin</i>	0.597	6.660	10.203	5.508	52.083	9.279
	<i>ELM_sigm</i>	0.610	6.227	10.033	5.571	27.083	4.876
	<i>ELM_tanh</i>	0.620	6.067	9.909	5.423	31.250	4.848
	<i>ELM_rbf_ll</i>	0.623	6.037	9.870	5.408	32.292	5.091
	<i>ELM_rbf_l2</i>	0.595	6.283	10.230	5.649	33.333	5.137
	<i>ELM_rbf_linf</i>	0.621	6.078	9.893	5.464	28.125	4.636

Table 5 Testing-set performance of different stochastic forecast methods considering both parameter and model error uncertainties.

Method	Learning algorithm	NSE	MAE (10^8 m^3)	RMSE (10^8 m^3)	CRPS (10^8 m^3)	PICP (%)	AW (10^8 m^3)
<i>SOF_PM</i>	<i>ELM_lin</i>	0.521	7.352	11.124	5.559	93.750	44.629
	<i>ELM_sigm</i>	0.562	6.426	10.630	5.152	86.458	32.421
	<i>ELM_tanh</i>	0.534	6.582	10.965	5.267	85.417	31.202
	<i>ELM_rbf_ll</i>	0.513	6.593	11.217	5.388	85.417	29.630
	<i>ELM_rbf_l2</i>	0.569	6.462	10.550	5.146	88.542	32.732
	<i>ELM_rbf_linf</i>	0.555	6.525	10.724	5.215	87.500	32.700
<i>SDF_PM</i>	<i>ELM_lin</i>	0.561	6.598	10.653	5.129	91.667	39.039
	<i>ELM_sigm</i>	0.585	6.343	10.346	5.001	87.500	33.892
	<i>ELM_tanh</i>	0.556	6.546	10.702	5.153	85.417	31.922
	<i>ELM_rbf_ll</i>	0.577	6.514	10.449	5.146	87.500	33.887
	<i>ELM_rbf_l2</i>	0.586	6.374	10.334	5.038	87.500	34.021
	<i>ELM_rbf_linf</i>	0.591	6.390	10.279	4.967	87.500	32.197
<i>SMF_PM</i>	<i>ELM_lin</i>	0.597	6.465	10.207	4.938	90.625	27.842
	<i>ELM_sigm</i>	0.605	6.174	10.101	4.876	84.375	23.394
	<i>ELM_tanh</i>	0.612	6.049	10.010	4.787	84.375	23.864
	<i>ELM_rbf_ll</i>	0.614	5.942	9.980	4.743	86.458	23.635
	<i>ELM_rbf_l2</i>	0.590	6.222	10.284	4.995	81.250	21.425
	<i>ELM_rbf_linf</i>	0.617	5.969	9.951	4.764	83.333	22.165

Table 6 NSE values of testing set for each forecast component of *MF*.

Learning algorithm	D1	D2	D3	A3
<i>ELM_lin</i>	0.176	0.878	0.974	0.965
<i>ELM_sigm</i>	0.337	0.885	0.973	0.971
<i>ELM_tanh</i>	0.362	0.883	0.970	0.970
<i>ELM_rbf_l1</i>	0.368	0.886	0.972	0.962
<i>ELM_rbf_l2</i>	0.320	0.882	0.968	0.970
<i>ELM_rbf_linf</i>	0.345	0.888	0.974	0.972