



*Supplement of*

**Refining marine net primary production estimates: advanced uncertainty quantification through probability prediction models**

**Jie Niu et al.**

*Correspondence to:* Yanqun Lu (yanqunlv@163.com) and Pan Wu (pwu@gzu.edu.cn)

The copyright of individual parts of the supplement might differ from the article licence.

## Text S1

The CDF curve highlights the cumulative difference across the distribution, while the scatterplot reflects the bias at the points. The differences between the model output and the original values are compared for the three cases where NPP is used as an output. Among them, when CAFE and CbPM are used as outputs, the CDF curves of the original values and predicted means (Fig. 8, Fig. S5) are more similar to each other relative to VGPM (Fig. S2); however, observing Fig. S4, the model-predicted mean fluctuates more flatly when CbPM is used as an output, which is not able to capture its fluctuation well, and the model-predicted mean is able to show almost the same fluctuation trend as that of the CAFE when CAFE is used as an output (Fig. 7), which is much closer to the original values. The fluctuation trend is almost the same as the original value, and the model predicted mean is closer to the original value.

To better understand the model's ability to reproduce extreme values, this article removed the seasonal signals from the original CAFE values and the predicted means of the two probabilistic prediction models and plotted the abnormal time series graphs (Figs. S7 and S8). From Fig. S7, it can be seen that the NN predicted mean values overlap more with the original values, better reflecting the fluctuation size of the original CAFE values, and is superior to Bayes in reproducing extreme values. Fig. S8 compares the prediction means of NN and Bayes when removing seasonal signals. As can be seen from the figure, when the models are applied to the NPP forecast from 2007 to March 2018, the average predictions of the two models are mostly close, but the NN output results fluctuate more significantly, better reflecting the complexity of the actual data.

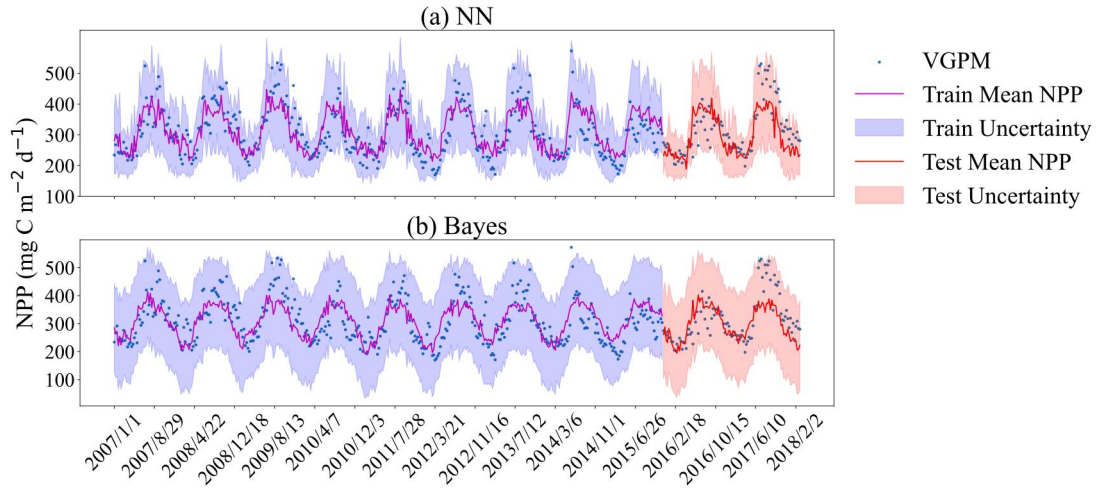


Fig. S1. Comparison of VGPM and predicted mean values at an 8-day temporal resolution within a 95% confidence interval. (a) Probabilistic prediction results are based on neural networks; (b) Bayesian probabilistic prediction results are based on empirical distributions. The dashed lines represent the mean values of the probabilistic predictions. The purple and red shaded areas illustrate the uncertainty ranges for the training and the test sets, respectively. Blue dots signify observed data points. All predictions and observations are presented in chronological sequence.

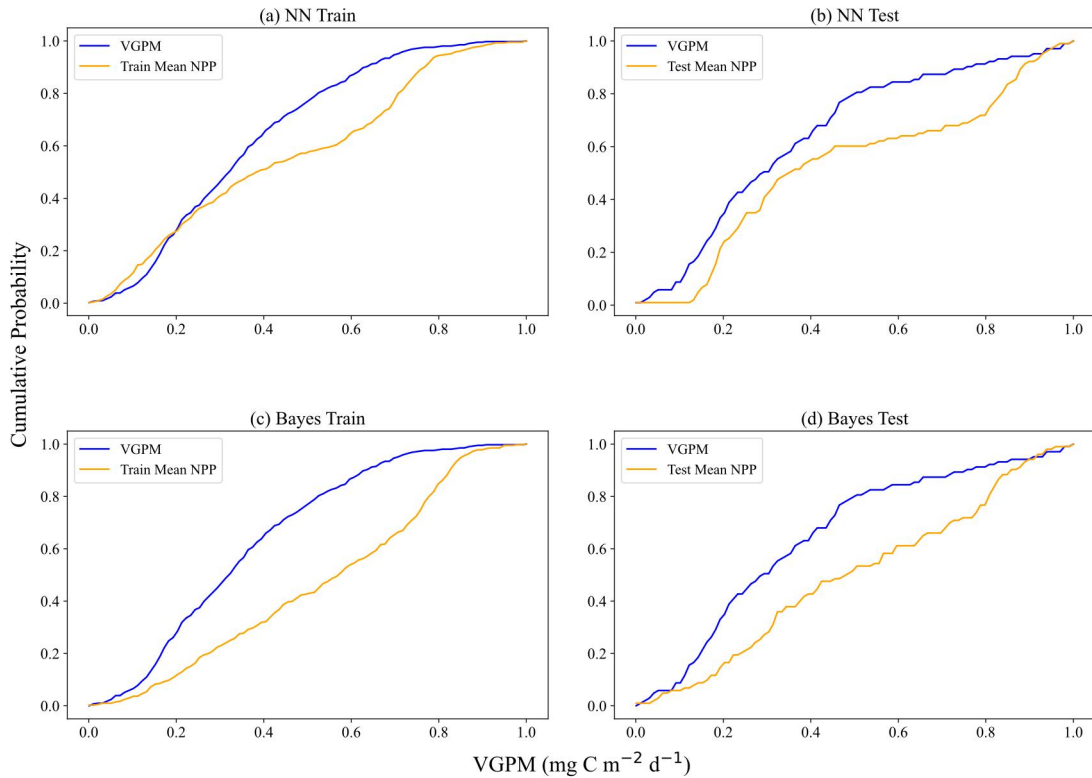


Fig. S2. Comparison of VGPM and predicted mean CDF. Panels (a) and (b) display the performance of the training and test sets, respectively, in the neural network-based probabilistic prediction model. Panels (c) and (d) illustrate the performance of the training and test sets, respectively, in the empirical distribution-based Bayesian probabilistic prediction model. The data has been normalized to a scale of 0–1 to ensure consistency across metrics and facilitate direct comparison between the two models. In each panel, the blue curves represent the CDFs of the

CAFE values, while the yellow curves depict the CDFs of the model's predicted mean values.

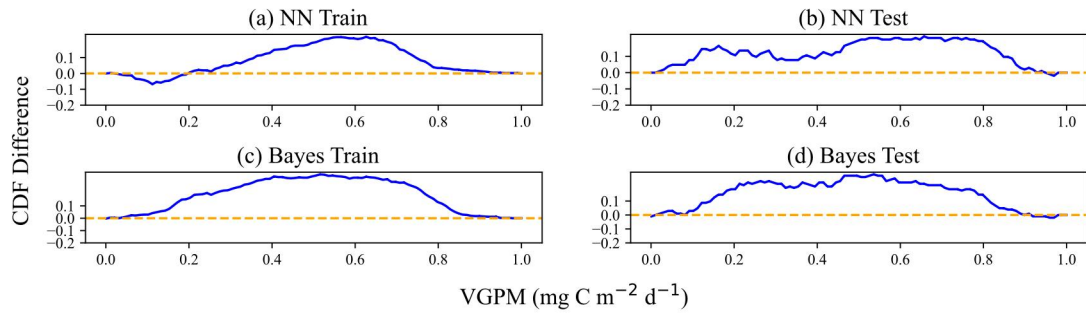


Fig. S3. Difference between the VGPM CDF and predicted mean CDF of model predictions. Panels (a) and (b) represent the performance of the training set and test sets, respectively, in the neural network-based probabilistic prediction model. Panels (c) and (d) showcase the performances of the training set and test sets, respectively, in the empirical distribution-based Bayesian probabilistic prediction model. The residuals are expressed in normalized units (0–1), enabling consistent assessment of model performance across different NPP ranges. The blue curves in each panel indicate the differential magnitude of the CDFs. Instances, where the blue curves align with the yellow lines, denote zero discrepancy between the input data CDF and the model's predicted mean CDF.

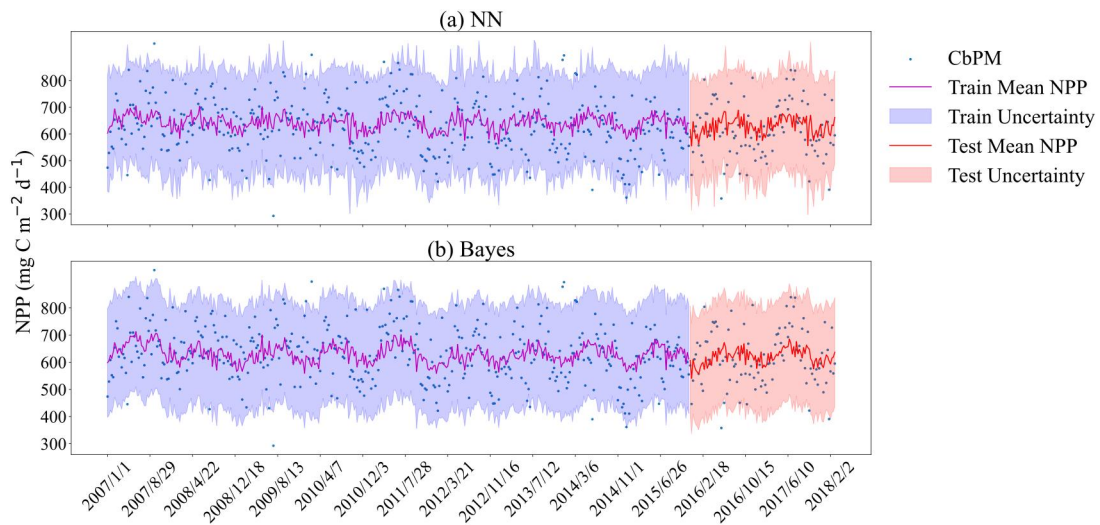


Fig. S4. Comparison of CbPM and predicted mean values shown at an 8-day temporal resolution within a 95% confidence interval. (a) Probabilistic prediction results are based on neural networks; (b) Bayesian probabilistic prediction results are based on empirical distributions. The dashed lines represent the mean values of the probabilistic predictions. The purple and red shaded areas illustrate the uncertainty ranges for the training and the test sets, respectively. Blue dots signify observed data points. All predictions and observations are presented in chronological sequence.

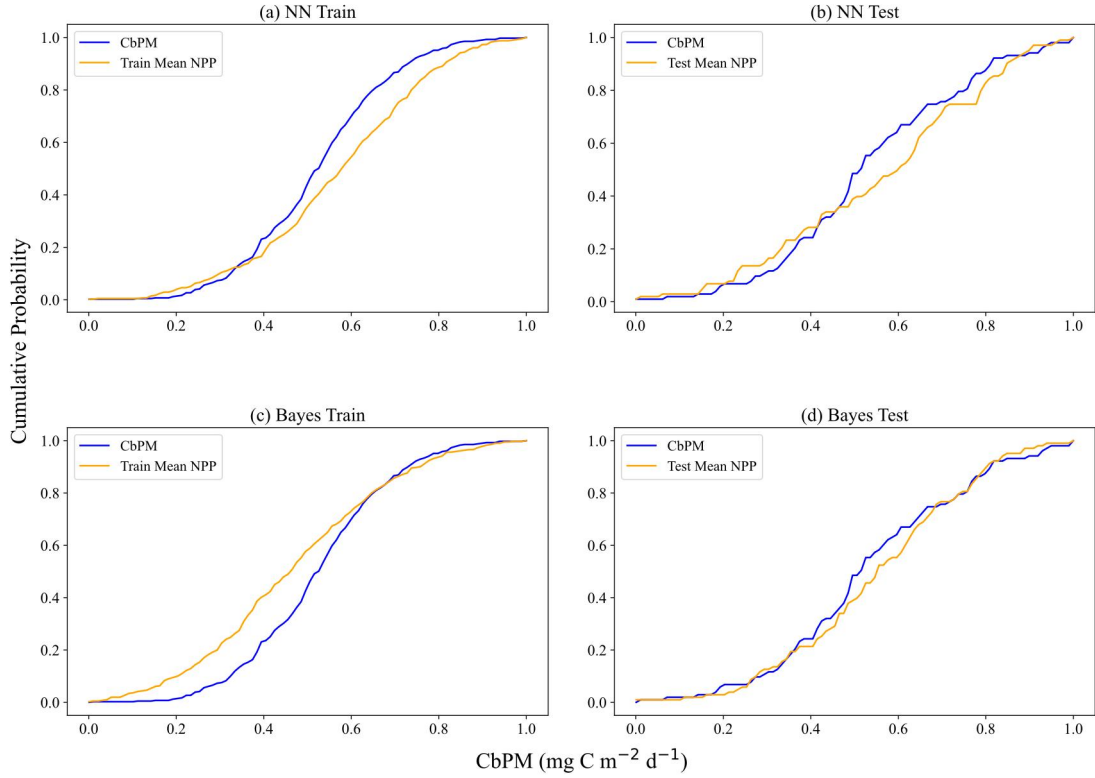


Fig. S5. Comparison of CbPM and predicted mean CDF. Panels (a) and (b) display the performance of the training and test sets, respectively, in the neural network-based probabilistic prediction model. Panels (c) and (d) illustrate the performance of the training and test sets, respectively, in the empirical distribution-based Bayesian probabilistic prediction model. The data has been normalized to a scale of 0–1 to ensure consistency across metrics and facilitate direct comparison between the two models. In each panel, the blue curves represent the CDFs of the CAFE values, while the yellow curves depict the CDFs of the model's predicted mean values.

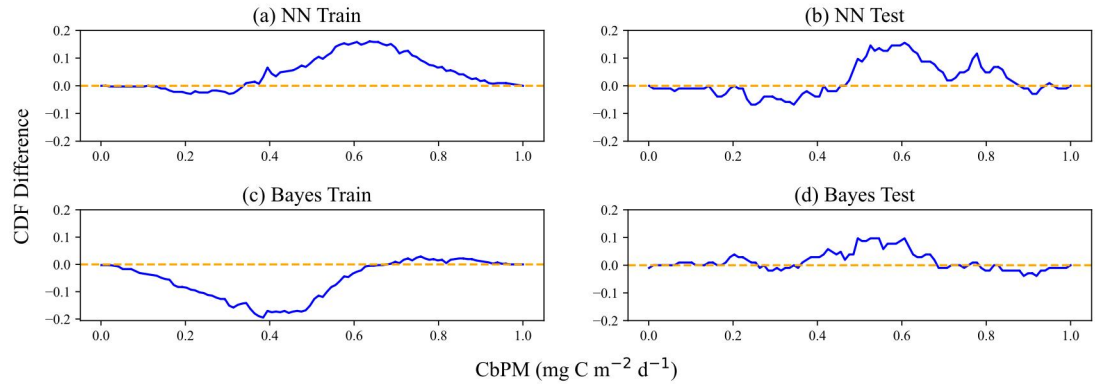


Fig. S6. Difference between the CbPM CDF and predicted mean CDF of model predictions. Panels (a) and (b) represent the performance of the training set and test sets, respectively, in the neural network-based probabilistic prediction model. Panels (c) and (d) showcase the performances of the training set and test sets, respectively, in the empirical distribution-based Bayesian probabilistic prediction model. The residuals are expressed in normalized units (0–1), enabling consistent assessment of model performance across different NPP ranges. The blue curves in each panel indicate the differential magnitude of the CDFs. Instances, where the blue curves align with the yellow lines, denote zero discrepancy between the input data CDF and the model's predicted mean CDF.

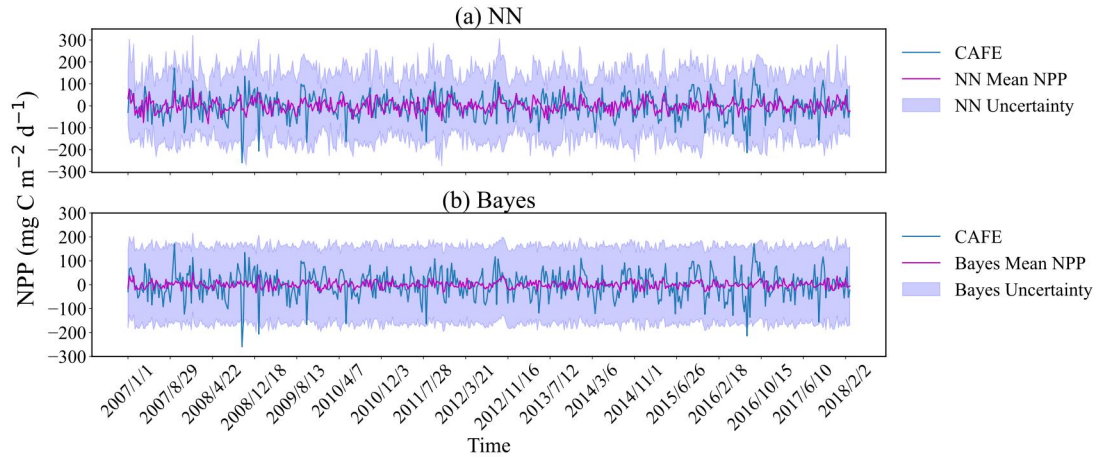


Fig. S7. Comparison of CAFE and predicted mean values shown at an 8-day temporal resolution within a 95% confidence interval. In this case, the seasonal signals have been removed from the original data and the predicted mean values to form an anomalous time series.

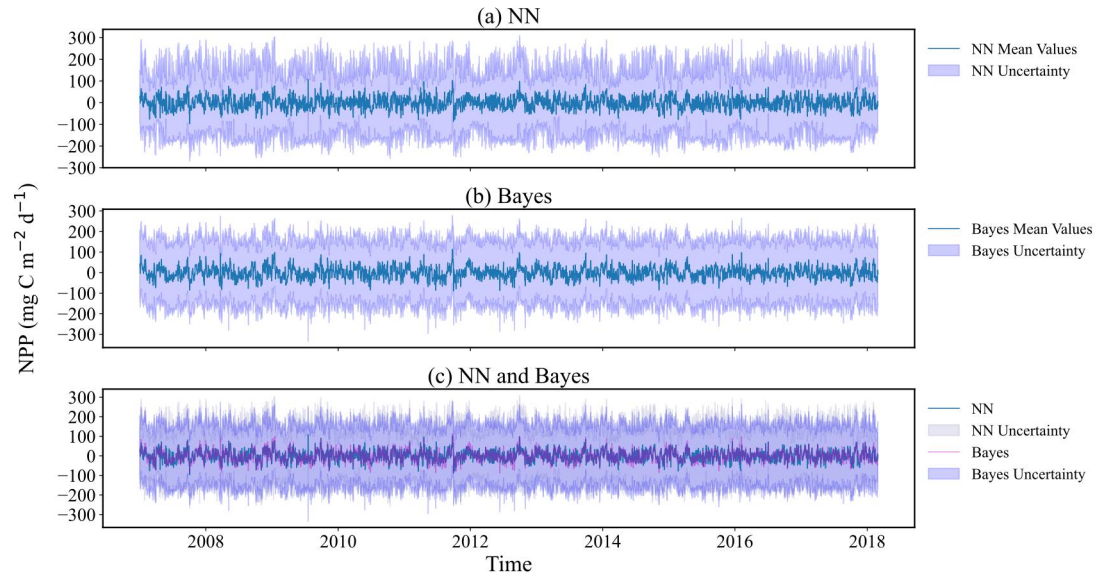


Fig. S8. Time series plots of daily probabilistic NPP predictions in Weizhou Island (2007 – March 2018). In this case, the seasonal signals have been removed from the predicted mean values to form an anomalous time series.