Assessment of the NOAA Operational Short-Range Streamflow forecast: A Case Study for February 2019 events in San Francisco Bay Area

Jungho Kim[1,2], Lynn Johnson[1,2], Rob Cifelli[2]

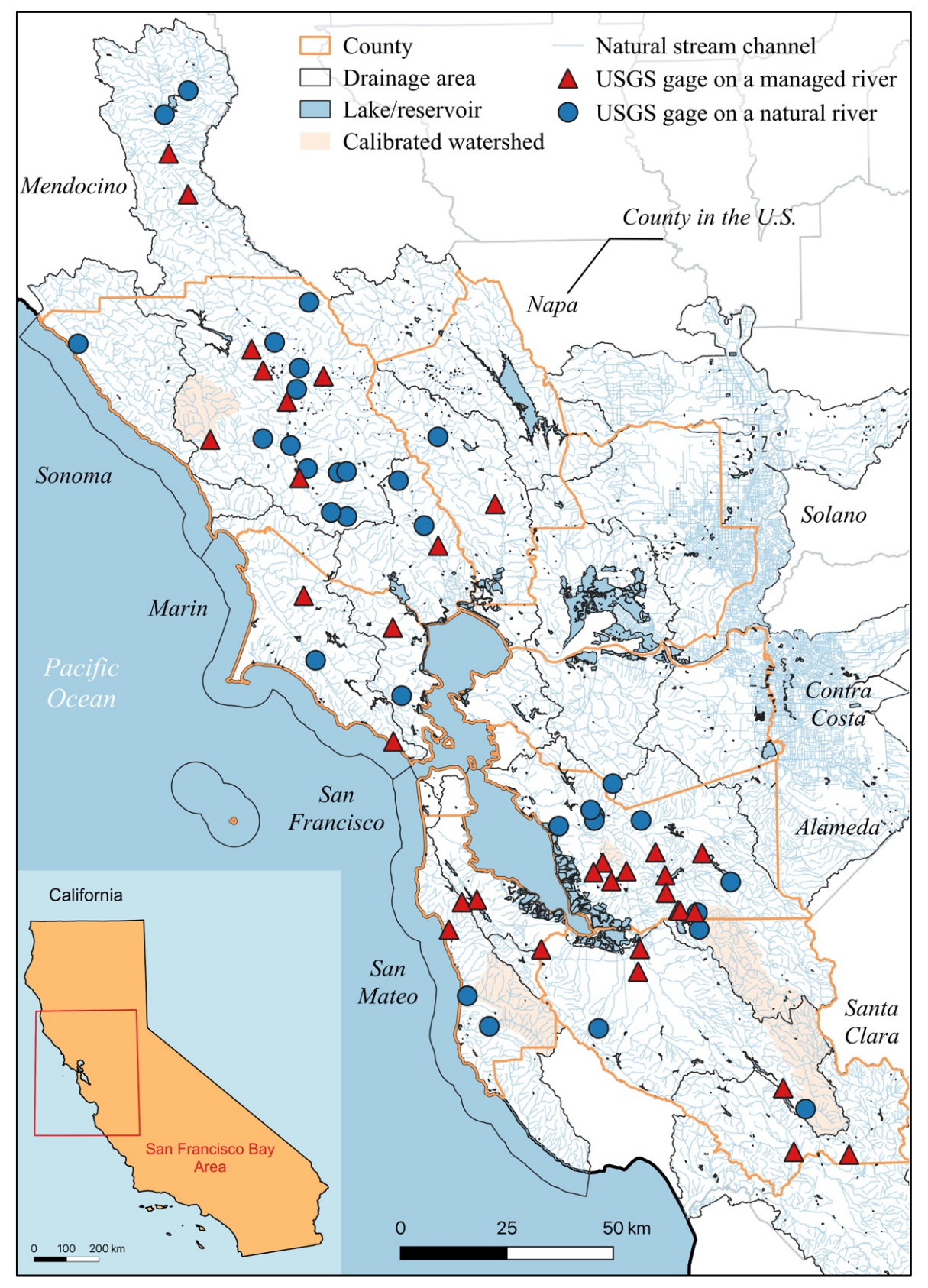
[1] Cooperative Institute for Research in the Atmosphere, Colorado State University, Ft. Collins, CO / [2] Physical Science Division, NOAA Earth System Research Laboratory, Boulder, CO

1 Introduction

- Verifying forecast skill of an operational-hydrologic prediction system is extremely beneficial to stakeholders for decision-making in flood-related issues.
- This study aims to address a preliminary assessment of the operational streamflow forecasts of the NOAA National Water Model (NWM) and answers the following overarching questions: 1) What is the forecast reliability dependent on lead time? and 2) What are the strengths and weaknesses of the forecasts compared to persistence forecasts?
- This study focuses on the short-range (out to 18 hours) streamflow forecasts for flood forecasting purposes.

2 Materials and Methods

- The San Francisco (SF) Bay area was selected as the application domain.
- The wet season from October to March in 2018-2019 was selected as the application period.
- The application period includes a series of storms that occurred in January and February 2019 and produced over $150 million in flood related damages, resulting in a presidential disaster declaration.



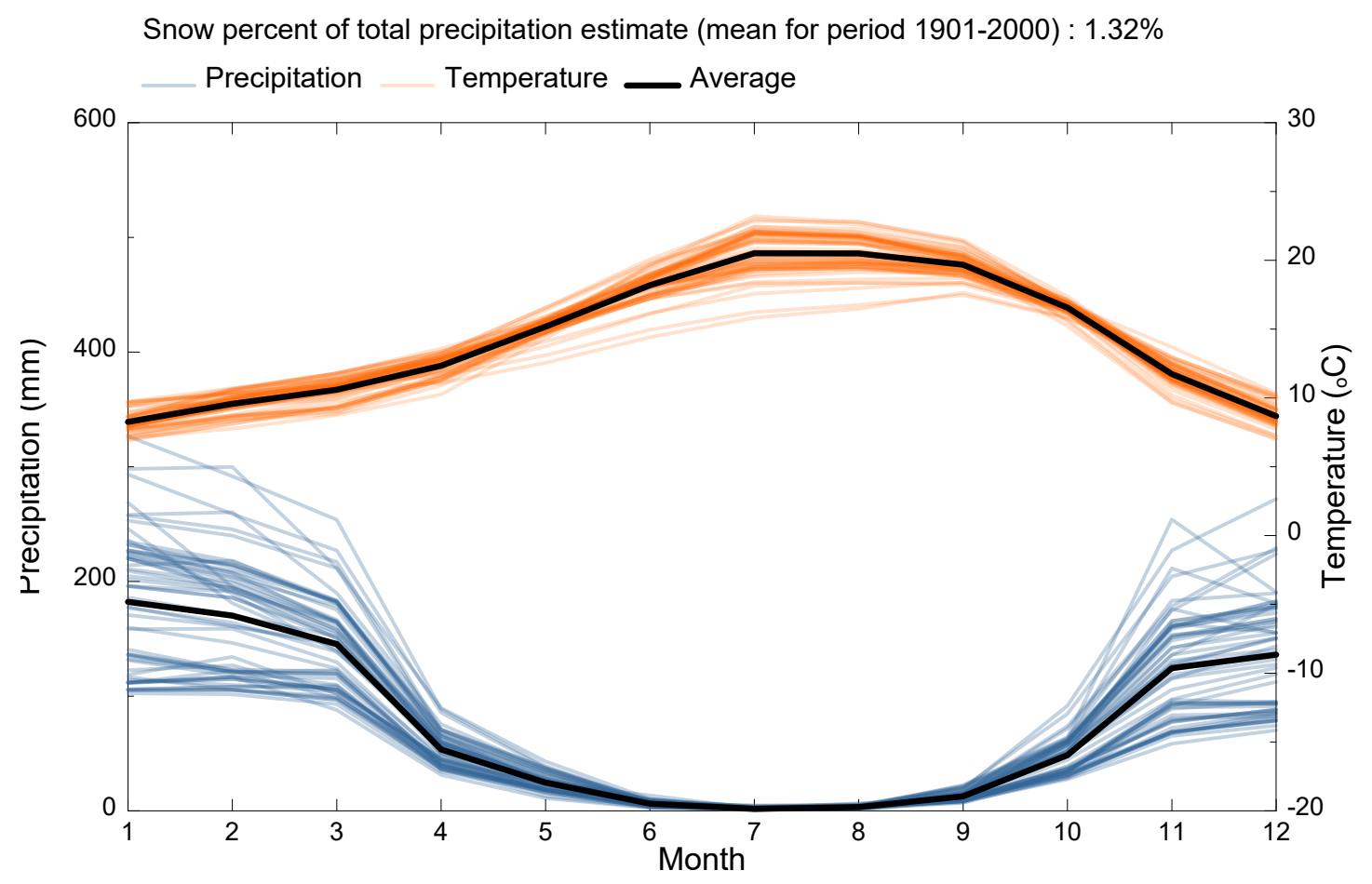


Figure 1: Climatology of precipitation and temperature in the SF Bay area.

Figure 2: The SF Bay area with the USGS streamflow monitoring gages.

- The wet season is from October to March in the SF Bay area (Fig. 1). Monthly mean temperature is higher than 5 °C and most precipitation falls as rain.
- 33 (triangle in fig. 2) USGS gages are located in managed river systems that are affected by anthropogenic behaviors for a flood control and water resources management. 32 (circle in fig. 2) USGS gages are located in natural river systems.

Error Indices	Acronym	Equation
Correlation Coefficient	CC	$\frac{\sum(Q_{sim}-\overline{Q_{sim}})(Q_{obs}-\overline{Q_{obs}})}{\sqrt{\sum(Q_{sim}-\overline{Q_{sim}})^2}\sqrt{\sum(Q_{obs}-\overline{Q_{obs}})^2}}$
Relative Bias	RB	$\sum Q_{sim} / \sum Q_{obs}$
Nash-Sutcliffe efficiency	NSE	$1-\frac{\sum(Q_{sim}-Q_{obs})^2}{\sum(Q_{obs}-\overline{Q_{obs}})^2}$
Root Mean Square Error	RMSE	$\sqrt{\frac{\sum(Q_{sim}-\overline{Q_{obs}})^2}{N}}$
Peak time error	-	T_{sim}- T_{obs}
Peak flow error	-	$(Peak\ Q_{sim} - Peak\ Q_{obs})\ /\ Peak\ Q_{obs} \times 100(\%)$
RMSE-Skill Score	RMSE-SS	$1-\frac{RMSE\ (forecast)}{RMSE\ (reference)}$

Table 1: Forecast verification metrics used in this study

- RMSE-SS>0: The NWM forecast skill is better than the reference forecast, persistence as a benchmark forecast.
- RMSE-SS<0: The opposite meaning of the RMSE-SS>0.
- RMSE-SS=0: The NWM forecast performs as like as the reference forecast.

3 Forecast Skill for a Wet Season

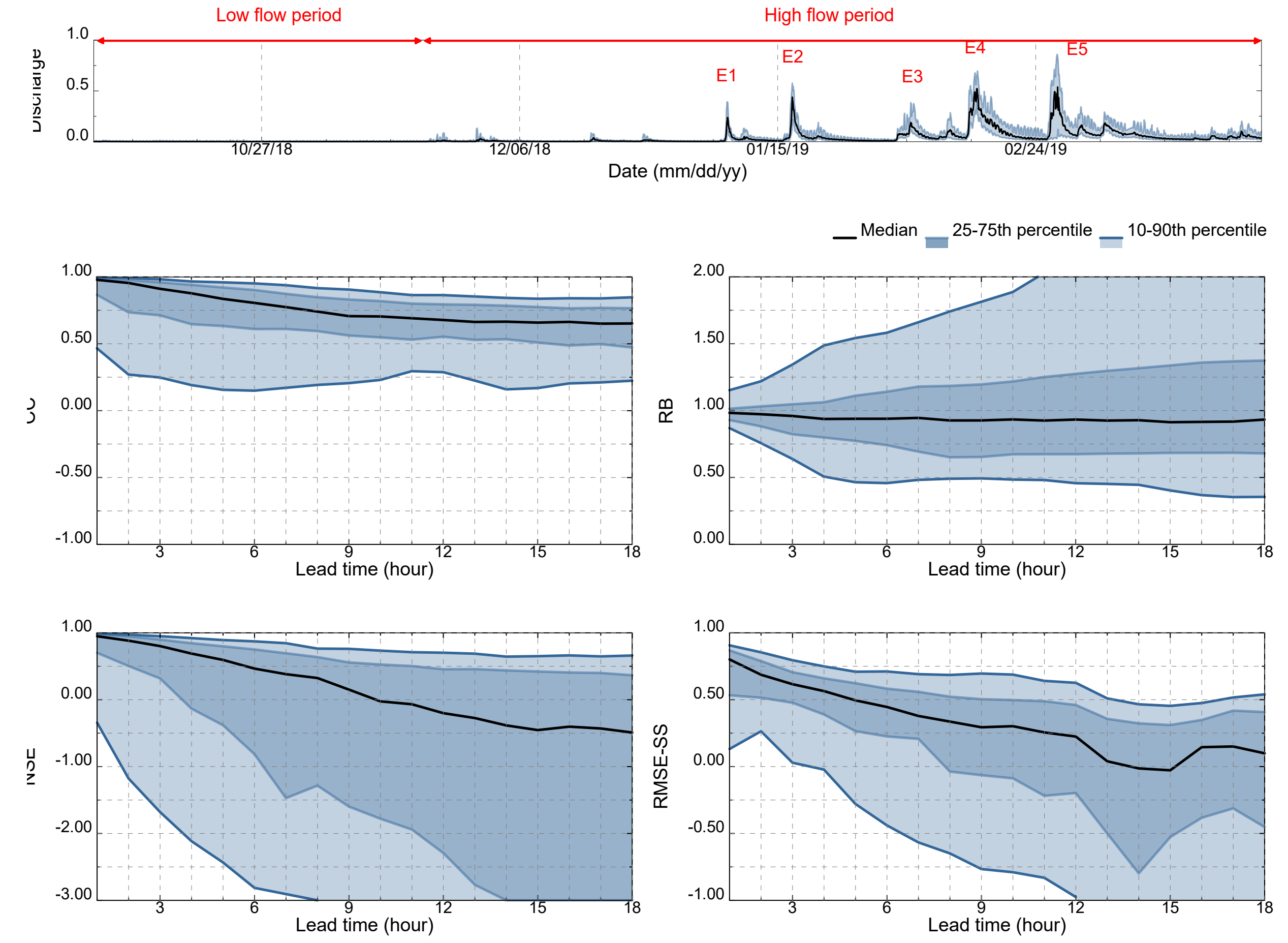


Figure 3: Time series of the observed discharge and forecast of the NWM by lead time for the entire period from Oct. 2018 to Mar. 2019.

- Correlation between the NWM forecast and the observed gradually decreased as lead time increased. 75% of watersheds was higher than 0.5 (CC) and 50% was higher than 0.7 in all lead times.
- The overall range of RB broadened as lead time increased, the median RB remained close to 1.0 (no bias) throughout the forecast.
- NSE exhibited a broad range that is similar with other metrics, decreased over time. However, the median remained above zero for 10 hours.
- The useful lead time (ULT, when RMSE-SS>0) of the NWM forecast was 13 hours based on the median performance.

4 Forecast Skill in Different Flow Regimes

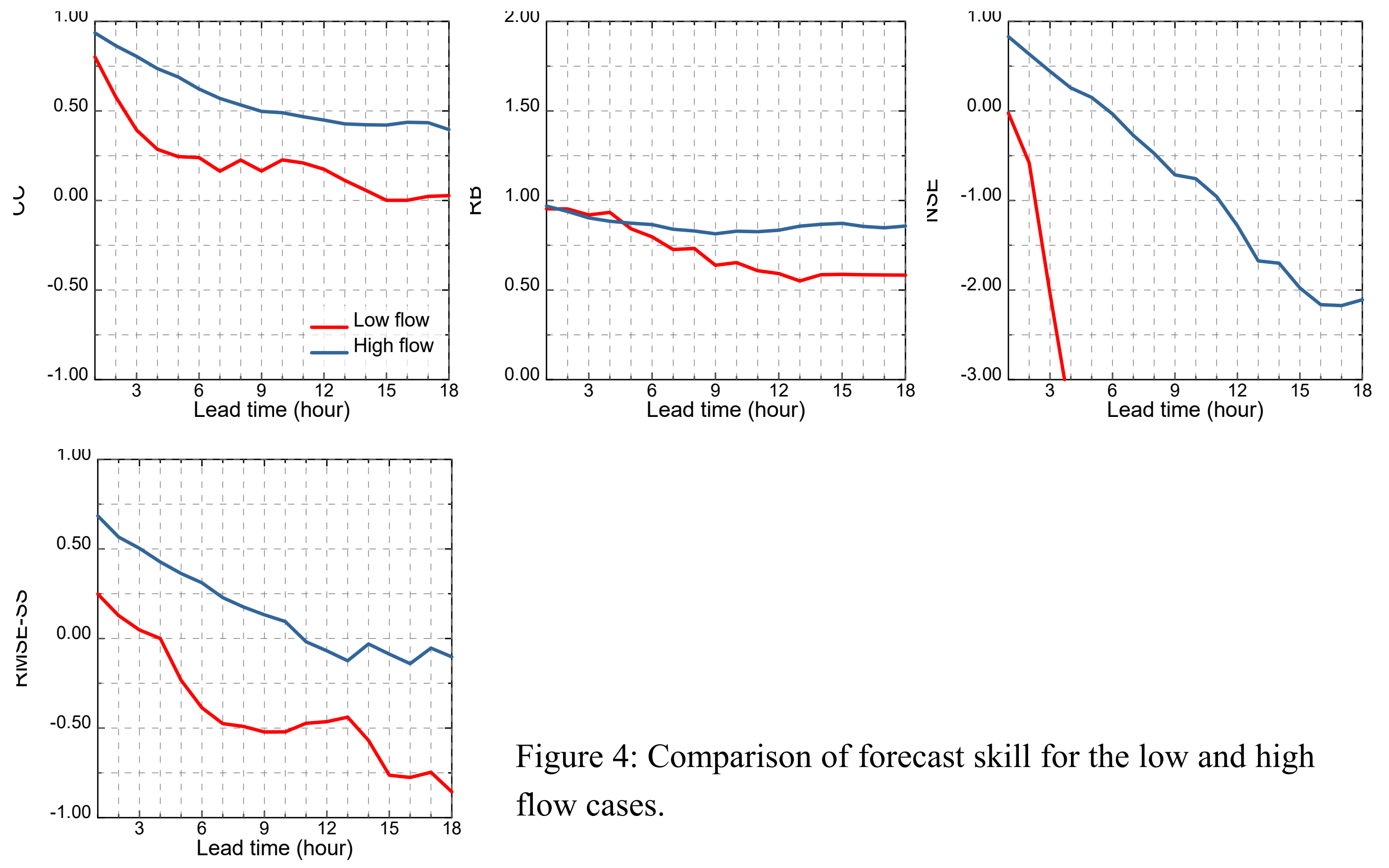


Figure 4: Comparison of forecast skill for the low and high flow cases.

- Forecast skill at lead time=1hr (initial forecast skill) was better in the high flow case than in the low flow case.
- A decrease in forecast skill by lead time was at least 3.0 times steeper in the low flow case than in the high flow case.
- In the low flow case, the RMSE-SS kept staying in above 0.0 by 4 hours of lead time and dramatically decreased.
- In the high flow case, the RMSE-SS smoothly decreased and kept staying in a positive value by 10 hours of lead time.
- The high flow and low flow cases had 10 and 4 hours of ULT respectively.

5 Forecast Skill for Managed and Unmanaged Watersheds

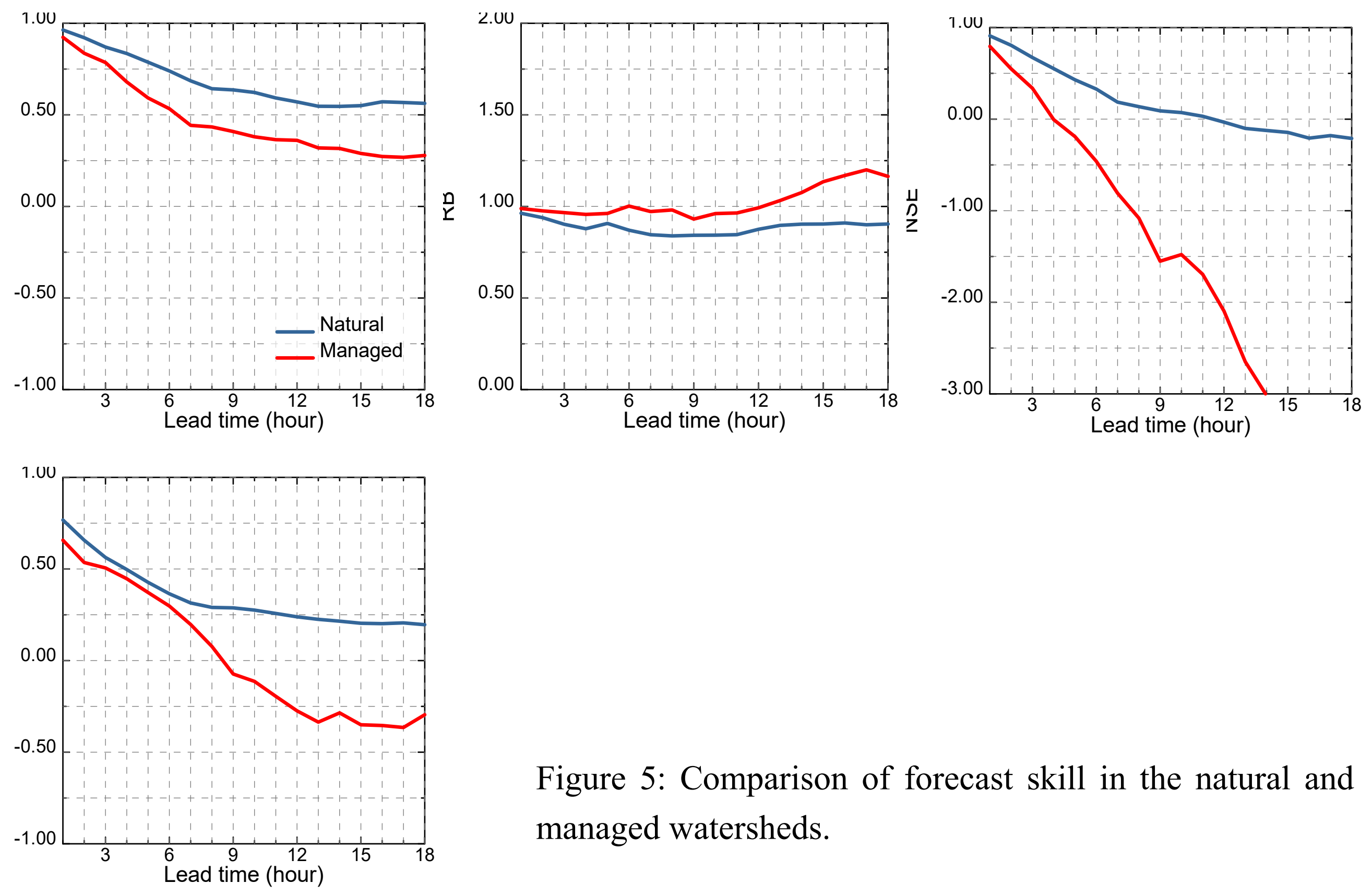


Figure 5: Comparison of forecast skill in the natural and managed watersheds.

- Forecast skill by lead time was different in the natural and managed watersheds, in terms of an initial forecast skill, decreasing trend, and ULT.
- The NWM tended to underpredict streamflows overall.
- The natural watersheds had 18 hours of ULT while the managed watersheds had ULT at 8 hours.
- This is likely due to the relatively simplistic representation of reservoirs in the NWM.

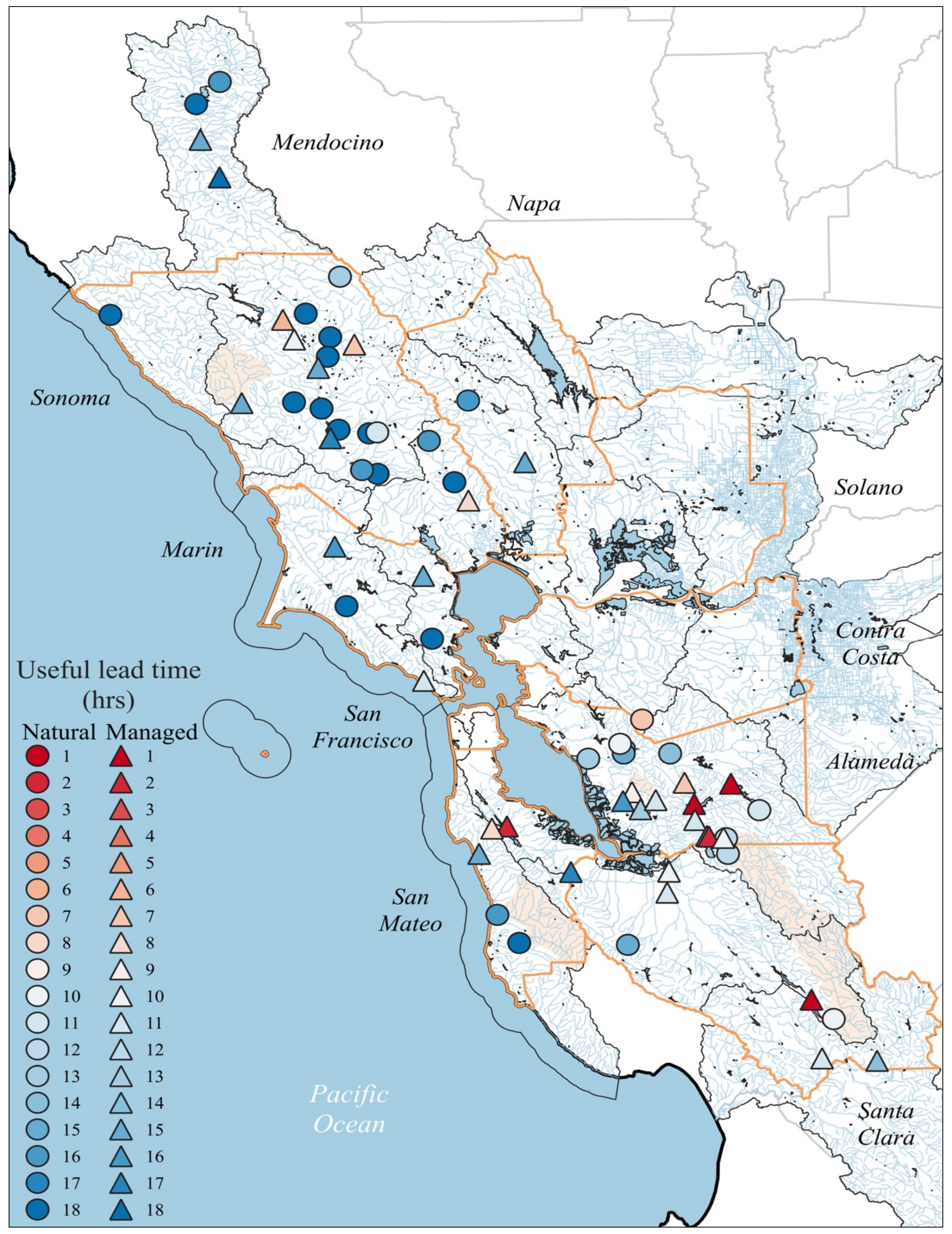


Figure 6: A map of useful lead time (ULT) in the SF bay area. ULT is defined as a lead time that is more useful than a reference forecast, persistence.

- The North (Marin, Sonoma, and Napa) and South (San Mateo, Santa Clara, and Alameda) SF Bay areas had 15 and 10 hours of ULT respectively.

6 Forecast Skill for Peak Values

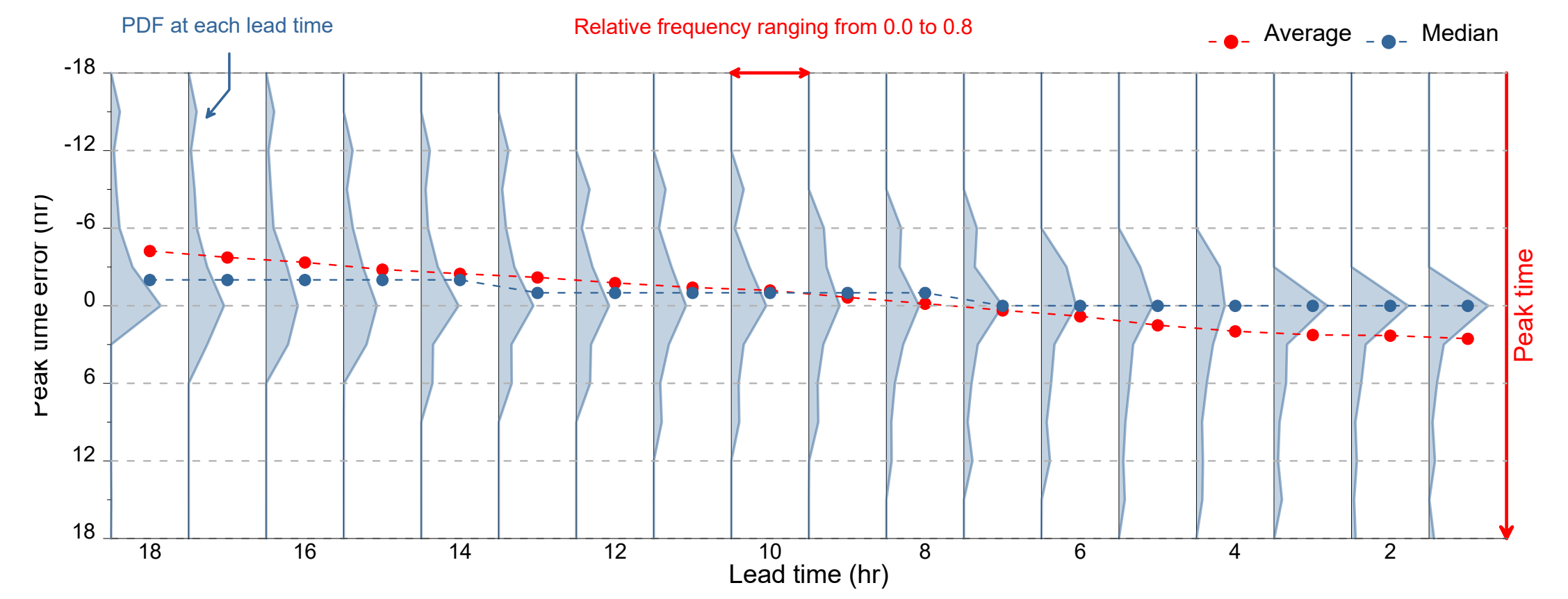


Figure 7: Probability density functions (PDF) of peak time error by lead time. Relative frequency constantly ranges from 0.0 to 0.8. Negative peak time error means 'Predicted peak time is earlier than observed' and a positive case indicates the opposite.

- The PDFs were formed in a standard distribution having mean near 0.0 in all lead times.
- A fraction of +/- 3 hour peak time error was more than 60% in all lead times. The median and average values of peak time error were within a range of +/- 3 hour peak time error since 12 hours lead time.
- As expected, the NWM forecasted peak flows in managed watersheds had relatively higher error than the natural watersheds.

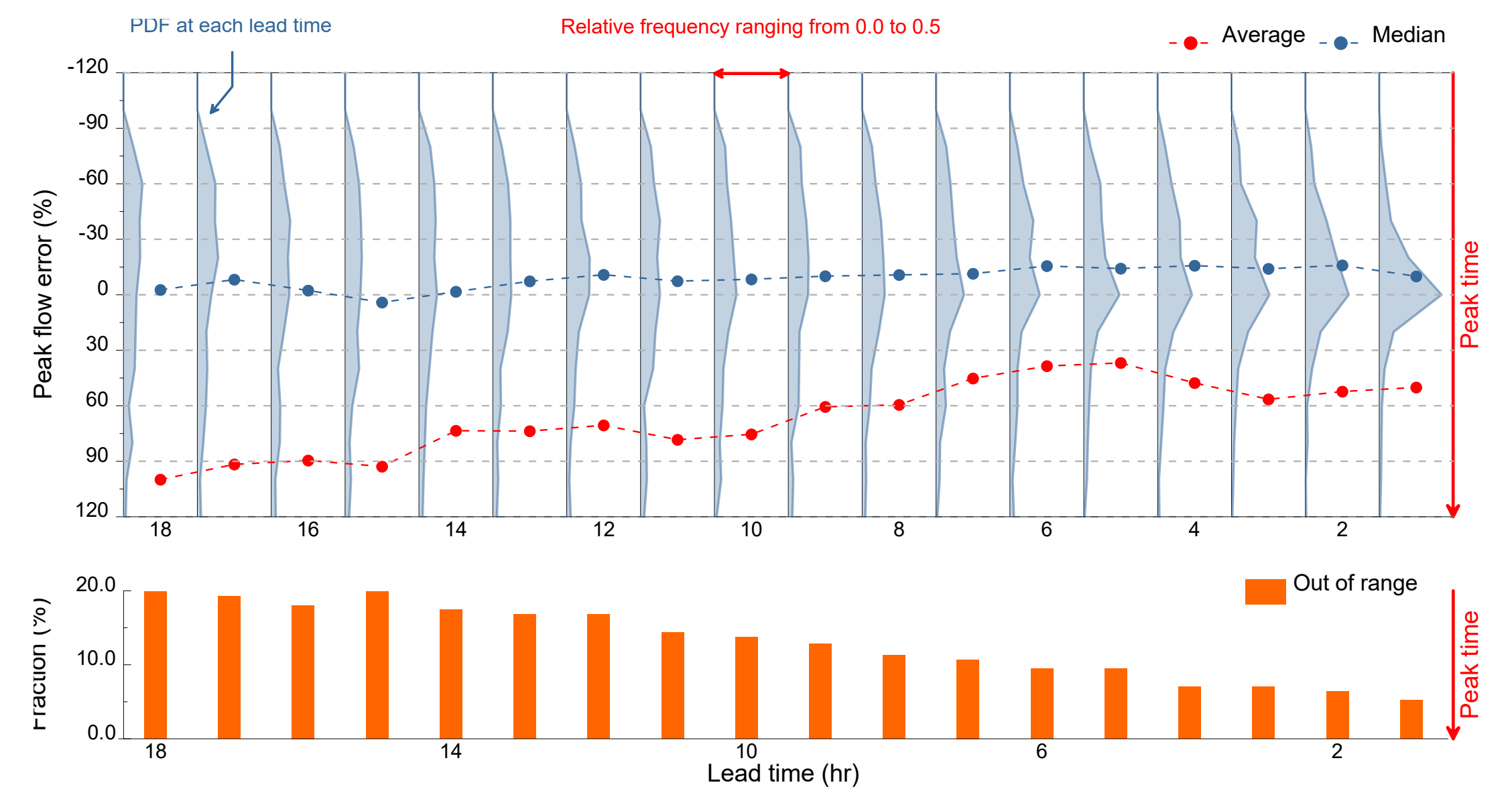


Figure 8: Results of peak flow error by lead times: this figure includes PDFs for each lead time with average and median values, relative frequency constantly ranging from 0.0 to 0.6 and a fraction (%) of out of range that is over +/- 120% of peak flow error. Negative peak flow error means 'Peak flow is under-predicted against observed' and a positive case is the opposite meaning of that.

- The median error was around -15% in all lead times.
- Highly over-predicted peak flows that are greater than +120% of peak flow error were identified. From the bottom panel, a fraction of the over-predicted peak flow errors decreased as lead time decreased.

7 Conclusions

- In the SF Bay area where a fraction of calibrated areas is only 11.2 %, the NWM performed well in predicting overall streamflow.
- Forecast skill of the NWM was significantly better in the high flow regime compared to streamflow in the low flow regime.
- The NWM forecasts were considered useful in the mostly natural watersheds and good at predicting time to peak at all lead times.
- Considering that many of the streams in the SF Bay area currently do not have forecast information and are prone to flash flooding, the results presented here are encouraging and suggest that the NWM may have value for water management decision making.

Acknowledgment

This work was supported by the California Department of Water Resources.