Supplement of Ocean Sci., 21, 1761–1785, 2025
https://doi.org/10.5194/os-21-1761-2025-supplement

Ocean Science
Open Access

Supplement of

Assessing subseasonal forecast skill for use in predicting US coastal inundation risk

John R. Albers et al.

Correspondence to: John R. Albers (john.albers@noaa.gov)

S1 Introduction

The following provides: (S1) details regarding how the IFS and CNRM climatologies were created; (S2) figures comparing the anomaly correlation skill of IFS, CNRM, and damped persistence hindcasts (raw and linearly detrended) at CONUS coastal gauge stations for forecast leads from Weeks 1-6; and (S3) difference in anomaly correlation skill between IFS hindcasts with and without the VLM included for forecast leads from Week 1-6 for Alaska, the West Coast, the Gulf Coast, and the East Coast.

S2 Supplement S2

Climatology creation

For the IFS, removing the previous 20 years from each forecast is trivial because the hindcast initializations are twice weekly for the same set of calendar days for each of the 20 years in the hindcast, thus the climatology is calculated by averaging each reforecast day over the 20 years of the reforecast period. However, for the CNRM, the hindcasts are initialized once a week and do not occur on the same calendar day over the course of the 25 year hindcast period, thus, for example, if a forecast is initialized on January 1, there are not 25 January 1 hindcast days to build a climatology from; this means that an additional few steps are required to compute a climatology. First, for each grid point and lead time, all days between each hindcast initialization are calculated via linear interpolation (in time) to create a daily time series 1993-2017. Next, a 365-day (plus one day for leap years) climatology is calculated by averaging over all years, and extracting the first four harmonics (plus the mean) via Fourier transform, yielding the final climatology, which is then removed from the hindcasts. To test how well building a climatology in this way works, the above method was applied to GLORYS SSH data using all daily data and by leaving out the identical days that are not included in the CNRM hindcast data. When the GLORYS climatology was computed in both ways and then compared, the global mean linear correlation between the two methods was 0.96, which we deem sufficient for our purposes here.

For the IFS, a separate climatology is created for each hindcast year using a leave-one-out method, which means that each climatology includes 19 years of data so that the mean bias correction is out-of-sample. For the CNRM, only one climatology was created, so that the mean bias correction is in-sample.

S3 Supplement S3

Anomaly correlation skill of IFS, CNRM, and damped persistence (Weeks 1-6):

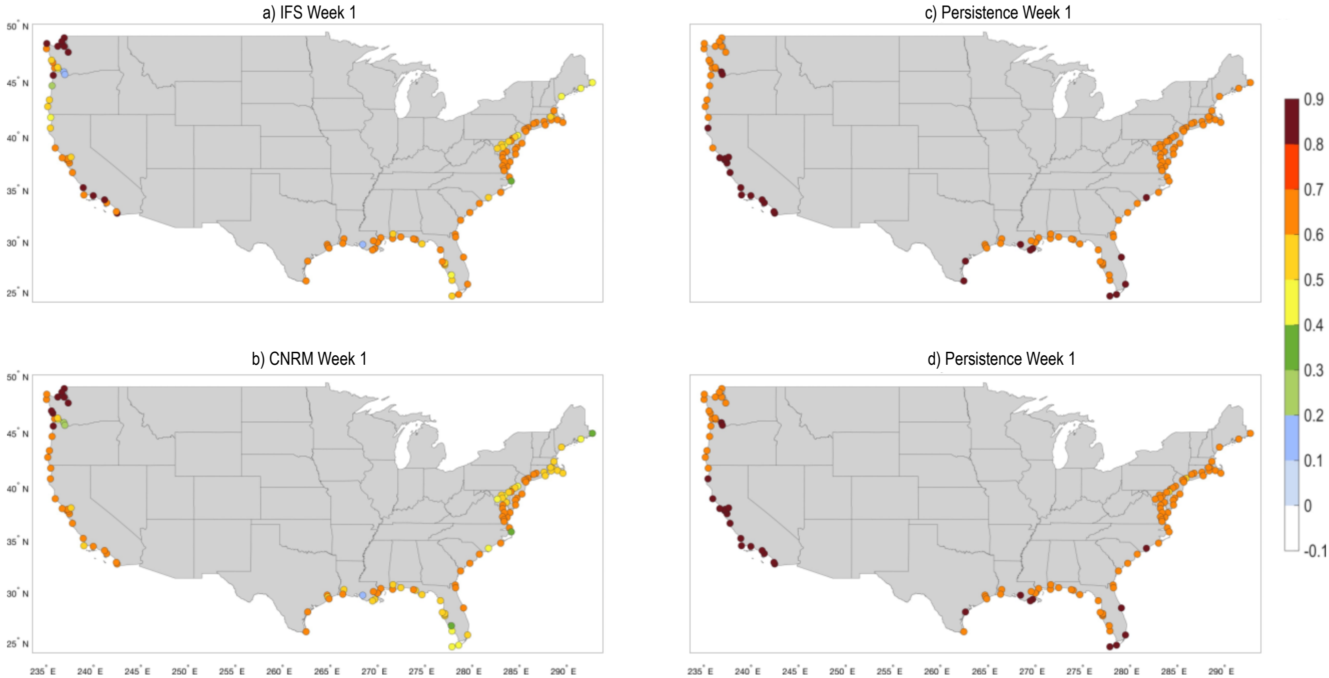

Figure S1. Year-round (2000-2017) week 1 anomaly correlation skill between (**a**) IFS and (b) CNRM hindcast SSH anomalies and NOAA gauge station water level anomalies, and persistence skill for IFS and CNRM hindcast days, (**c**) and (**d**), respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels (**c**) and (**d**) suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

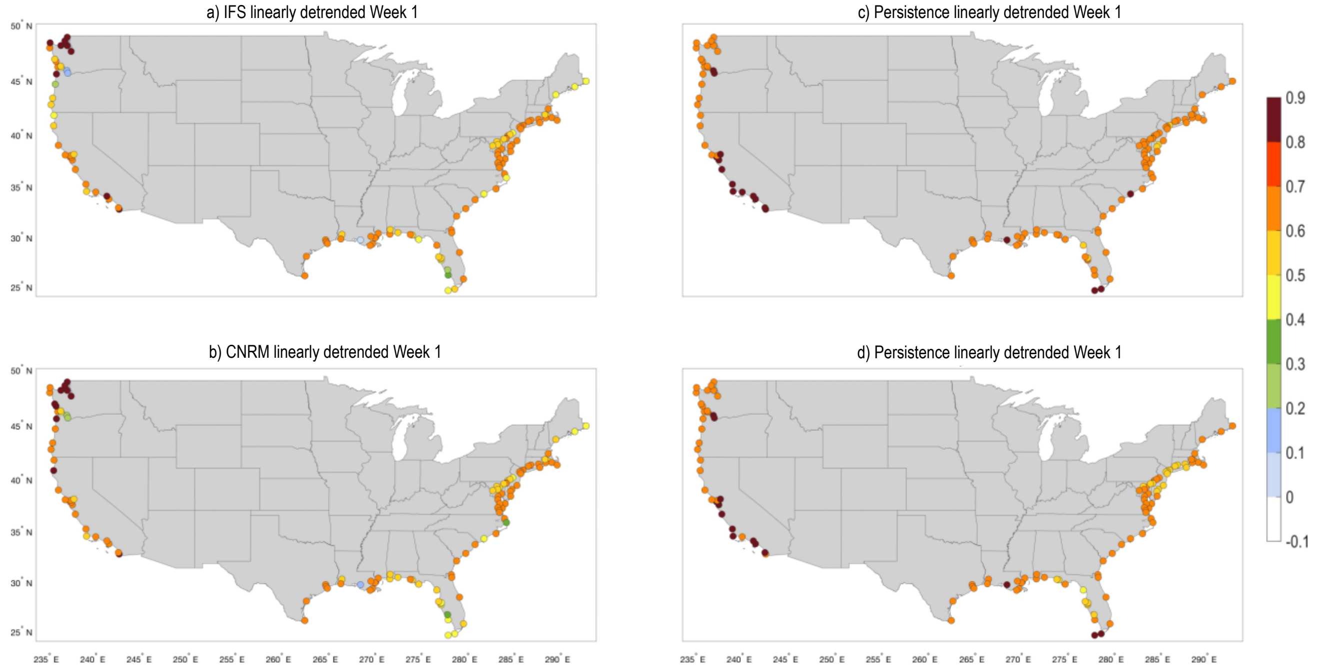

Figure S2. Year-round (2000-2017) week 1 anomaly correlation skill between linearly detrended **(a)** IFS and **(b)** CNRM hindcast SSH anomalies and linearly detrended NOAA gauge station water level anomalies, and linearly detrended persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

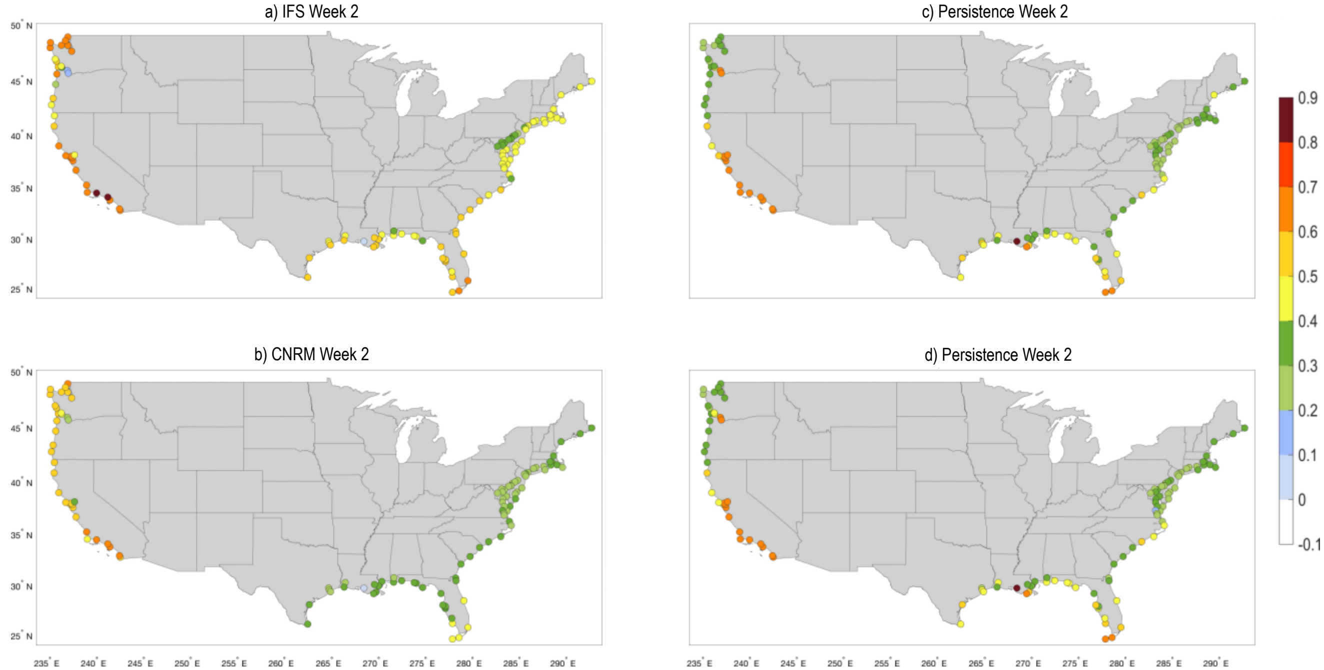

Figure S3. Year-round (2000-2017) week 2 anomaly correlation skill between **(**a) IFS and **(**b**)** CNRM hindcast SSH anomalies and NOAA gauge station water level anomalies, and persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

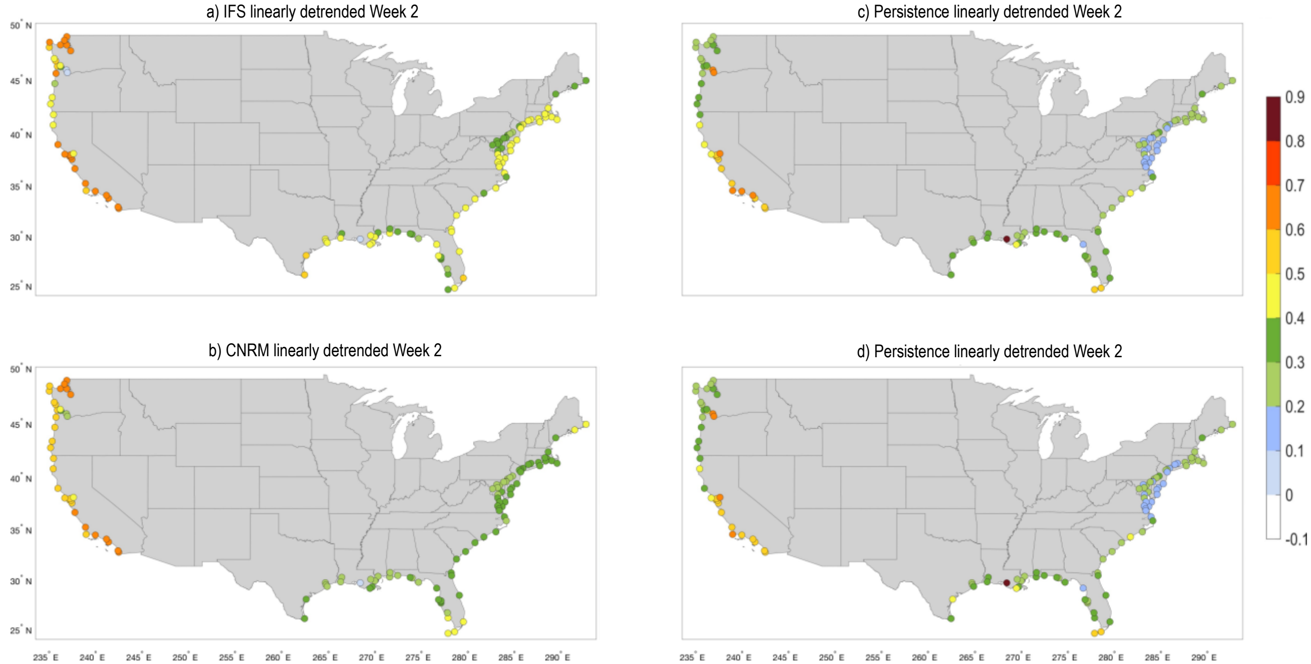

Figure S4. Year-round (2000-2017) week 2 anomaly correlation skill between linearly detrended **(a)** IFS and **(b)** CNRM hindcast SSH anomalies and linearly detrended NOAA gauge station water level anomalies, and linearly detrended persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

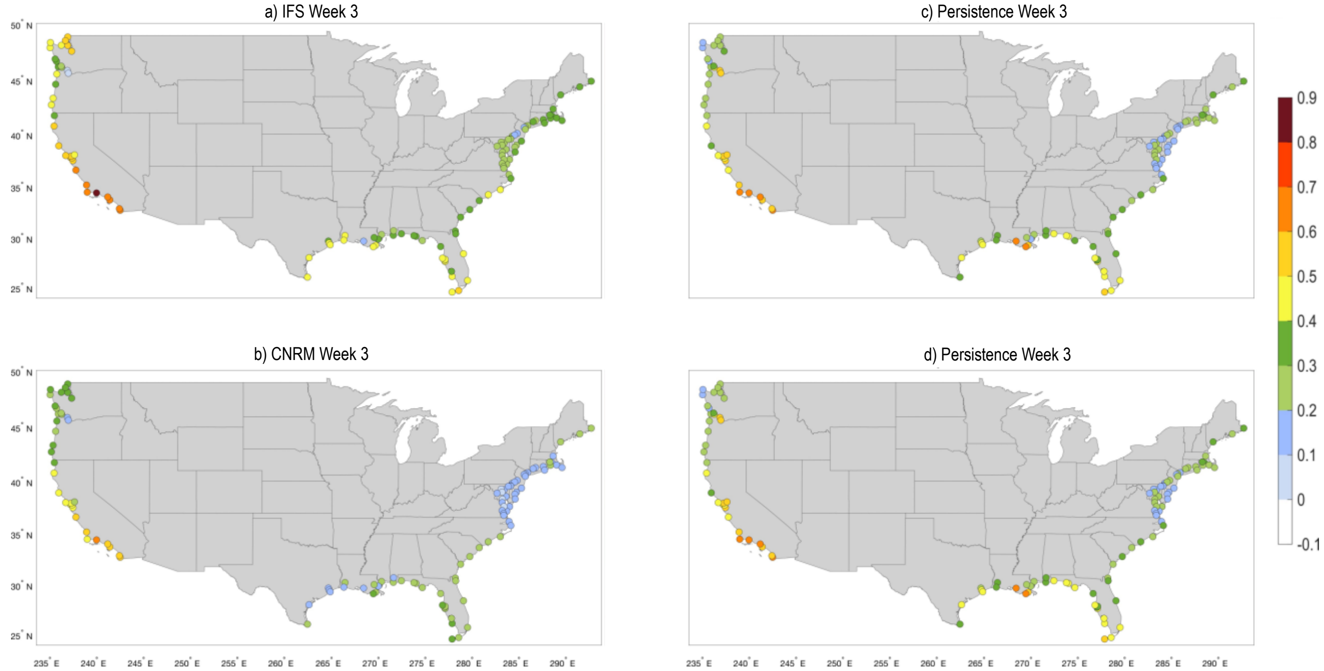

Figure S5. Year-round (2000-2017) week 3 anomaly correlation skill between **(a**) IFS and **(b)** CNRM hindcast SSH anomalies and NOAA gauge station water level anomalies, and persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

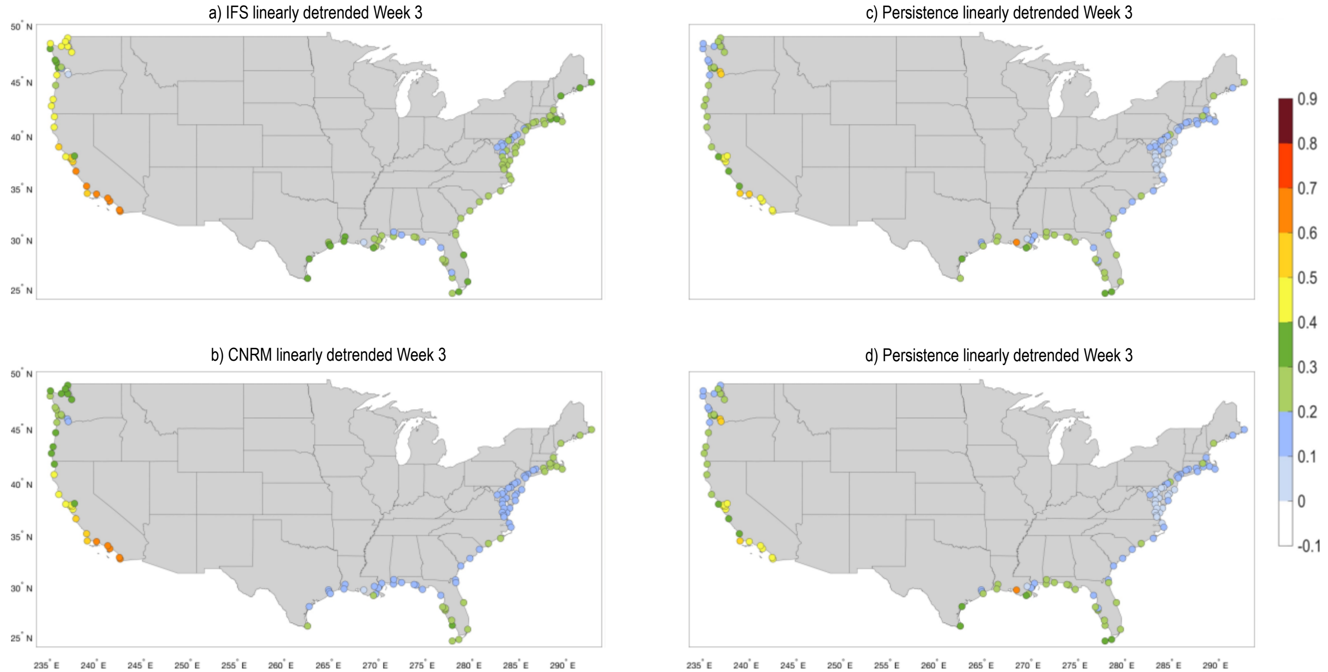

Figure S6. Year-round (2000-2017) week 3 anomaly correlation skill between linearly detrended **(a)** IFS and **(b)** CNRM hindcast SSH anomalies and linearly detrended NOAA gauge station water level anomalies, and linearly detrended persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

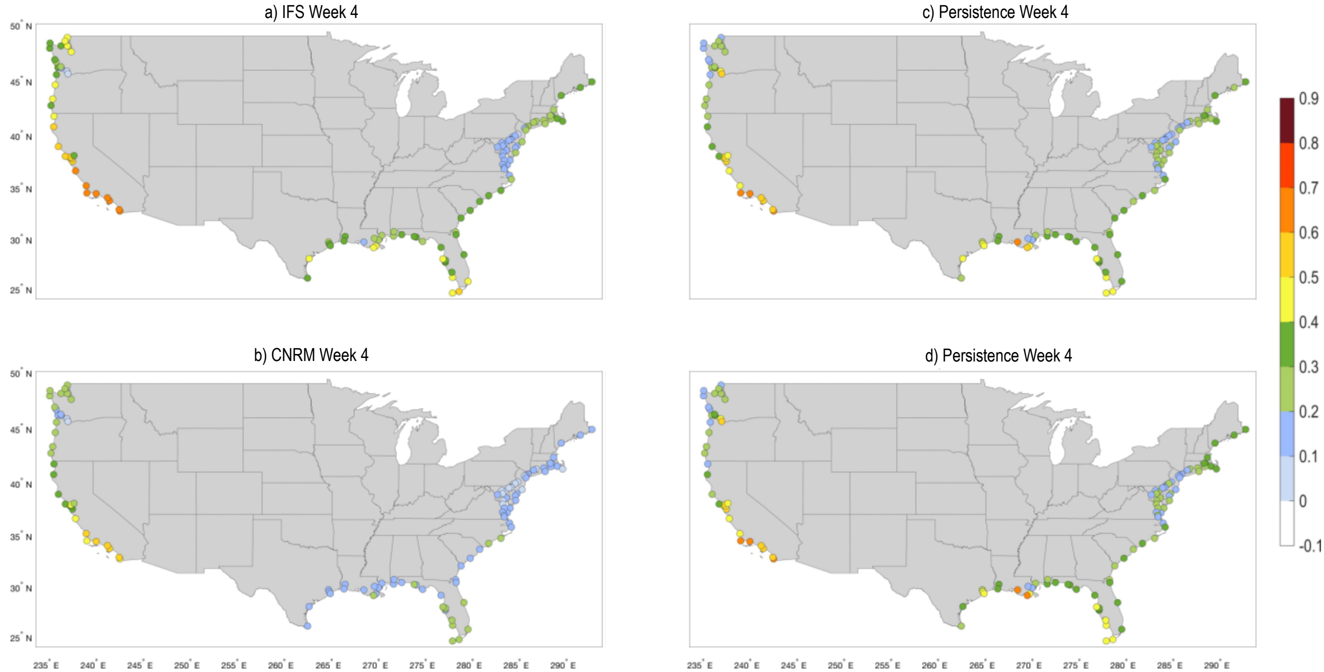

Figure S7. Year-round (2000-2017) week 4 anomaly correlation skill between **(a**) IFS and **(b)** CNRM hindcast SSH anomalies and NOAA gauge station water level anomalies, and persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

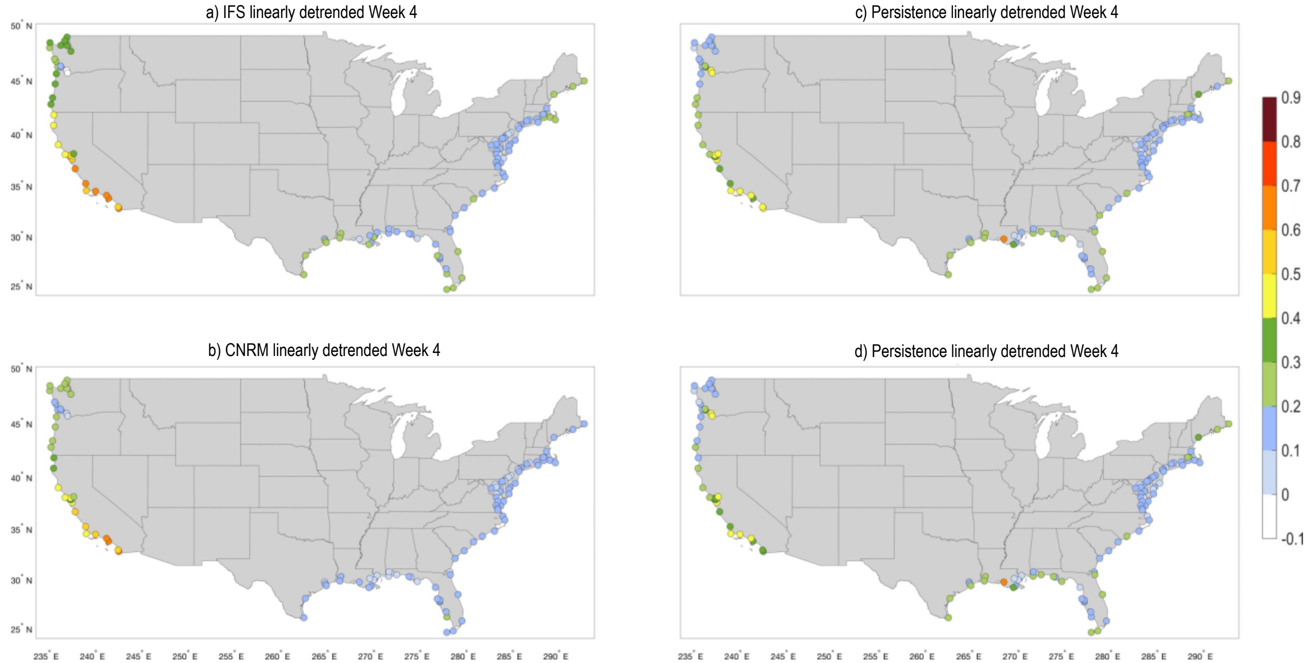

Figure S8. Year-round (2000-2017) week 4 anomaly correlation skill between linearly detrended **(a)** IFS and **(b)** CNRM hindcast SSH anomalies and linearly detrended NOAA gauge station water level anomalies, and linearly detrended persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

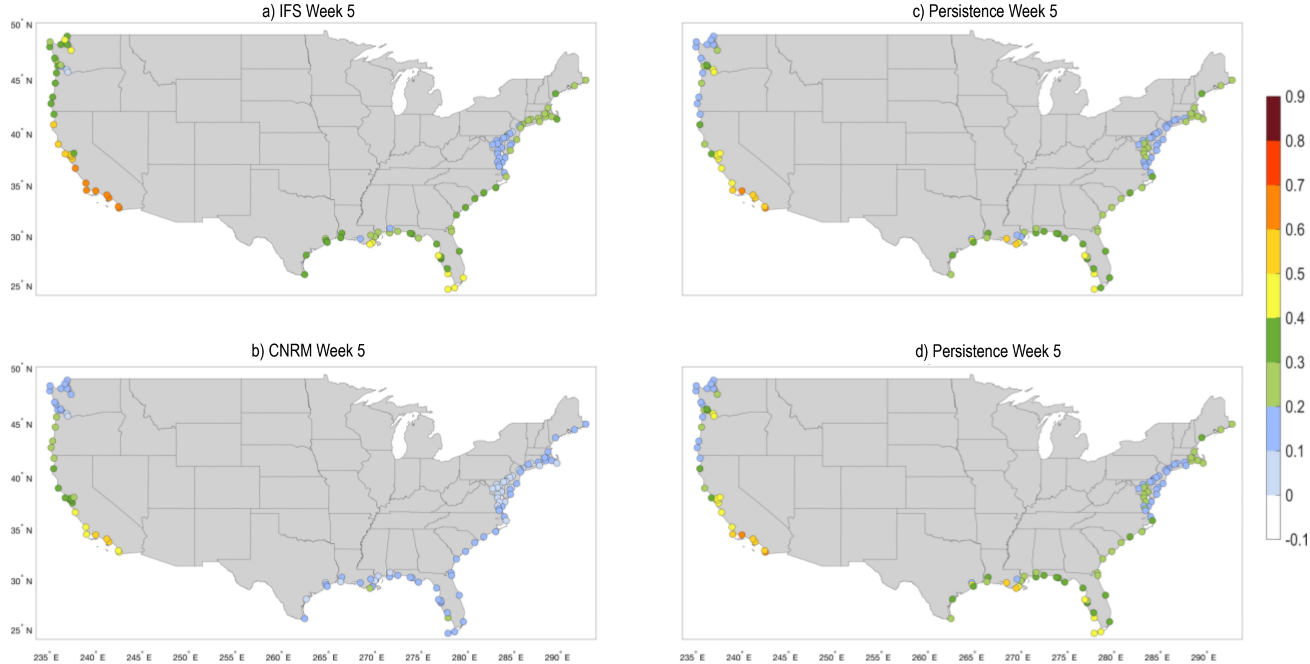

Figure S9. Year-round (2000-2017) week 5 anomaly correlation skill between **(a**) IFS and **(b)** CNRM hindcast SSH anomalies and NOAA gauge station water level anomalies, and persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

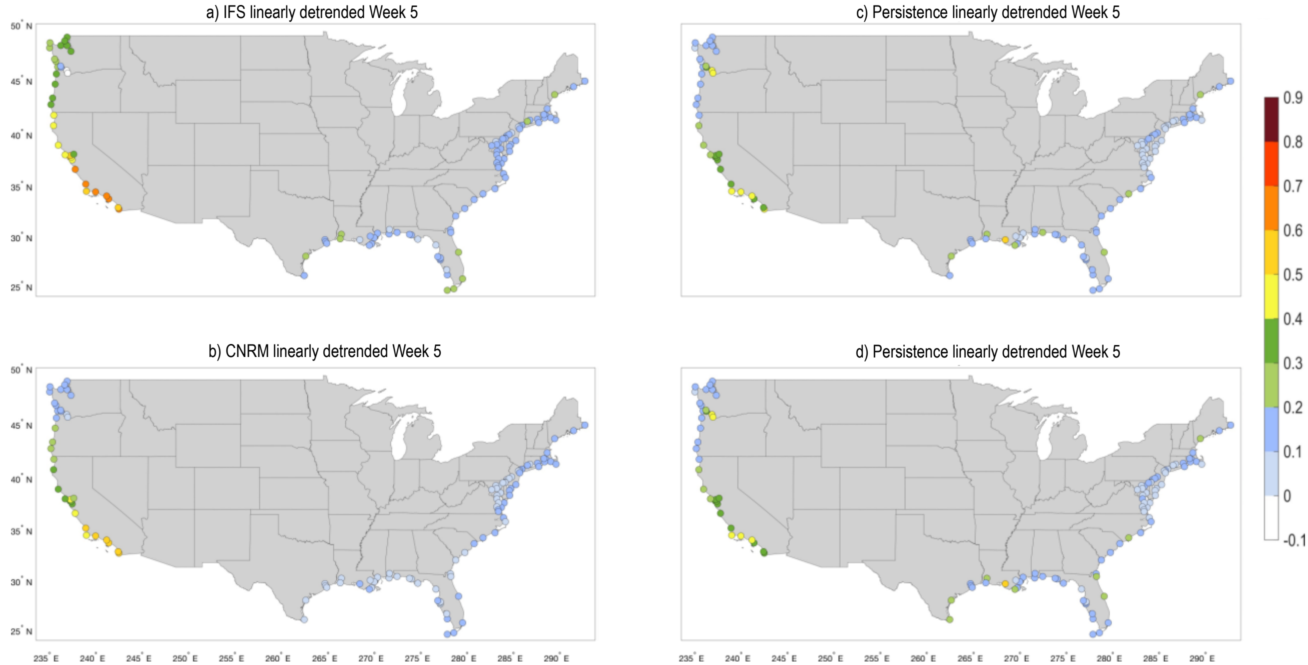

Figure S10. Year-round (2000-2017) week 5 anomaly correlation skill between linearly detrended **(a)** IFS and **(b)** CNRM hindcast SSH anomalies and linearly detrended NOAA gauge station water level anomalies, and linearly detrended persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

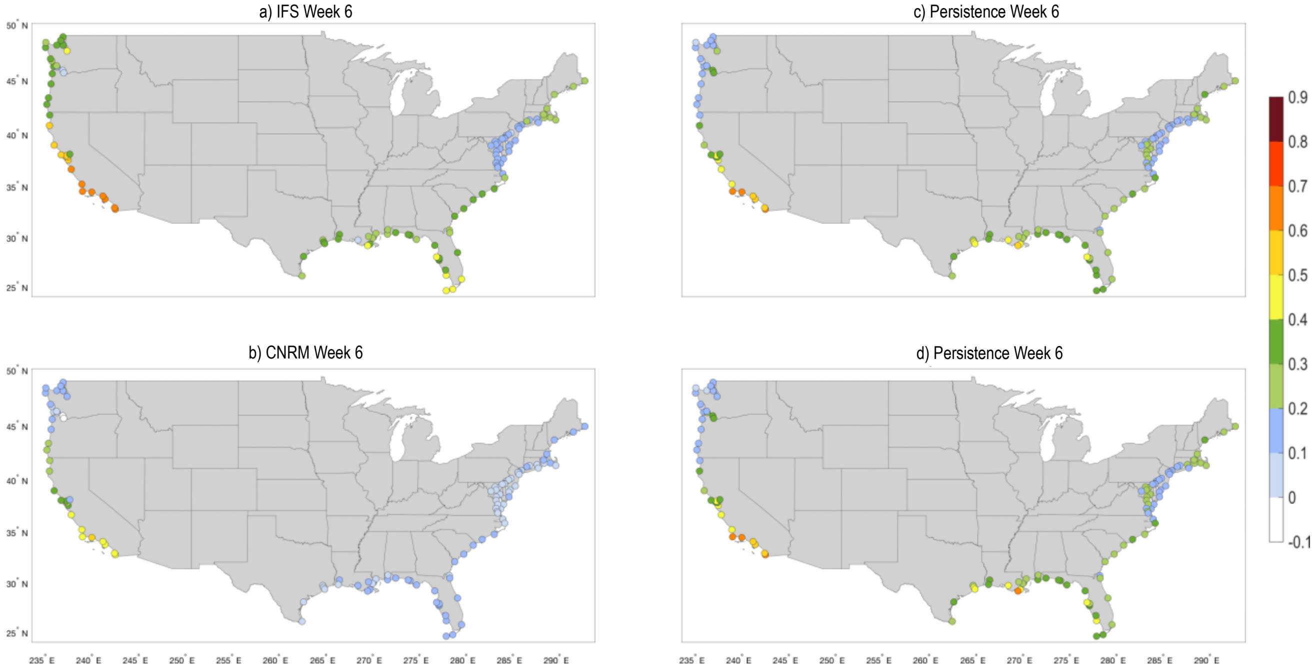

Figure S11. Year-round (2000-2017) week 6 anomaly correlation skill between **(a**) IFS and **(b)** CNRM hindcast SSH anomalies and NOAA gauge station water level anomalies, and persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

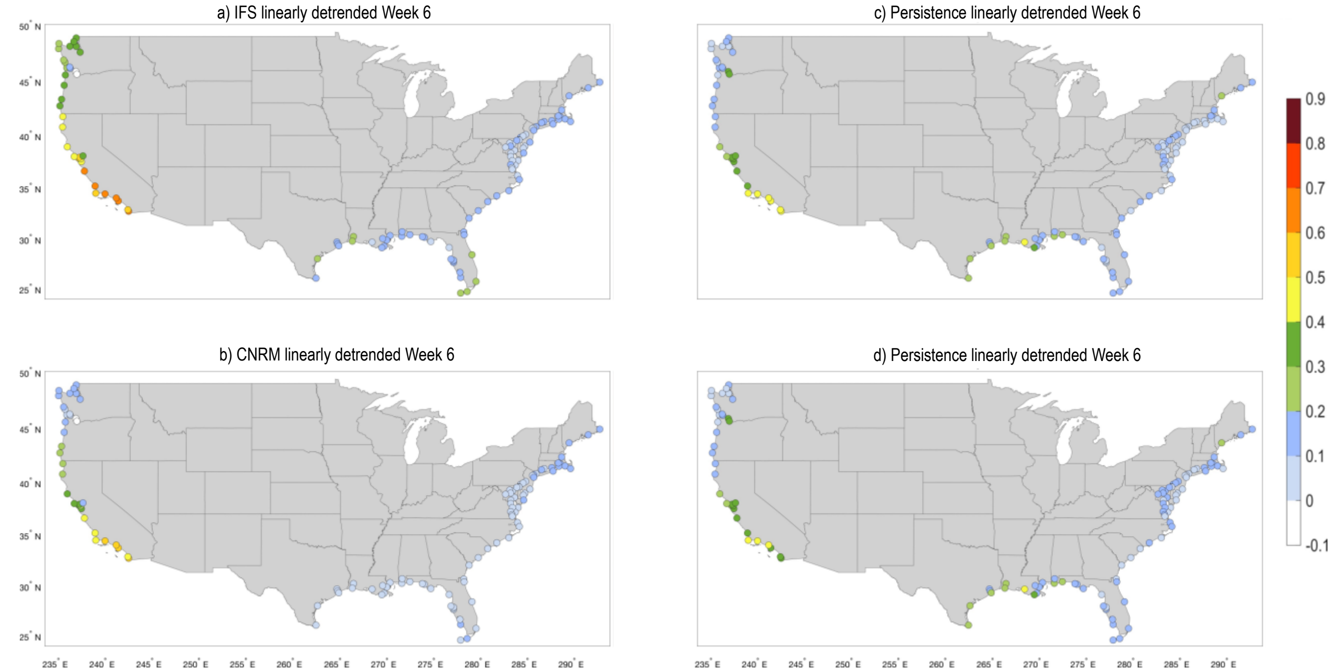

Figure S12. Year-round (2000-2017) week 6 anomaly correlation skill between linearly detrended **(a)** IFS and **(b)** CNRM hindcast SSH anomalies and linearly detrended NOAA gauge station water level anomalies, and linearly detrended persistence skill for IFS and CNRM hindcast days, **(c)** and **(d)**, respectively. While the sample years used in the anomaly correlation calculation (2000-2017) are the same, the dates are slightly different because of different initialization dates. All dates are used in both datasets, regardless of whether they overlap with the other dataset, because otherwise there are too few samples. The high similarity between the damped persistence forecasts in panels **(c)** and **(d)** suggests that the slightly different sampling days does not strongly affect the assessment of skill for the 2000-2017 period, and thus the IFS and CNRM hindcasts for the 2000-2017 shown here can be fairly compared.

S4 Supplement S4

<u>Difference between anomaly correlation skill of IFS (SSH + IBE + VLM) and IFS (SSH + IBE only (Weeks 1-6):</u>

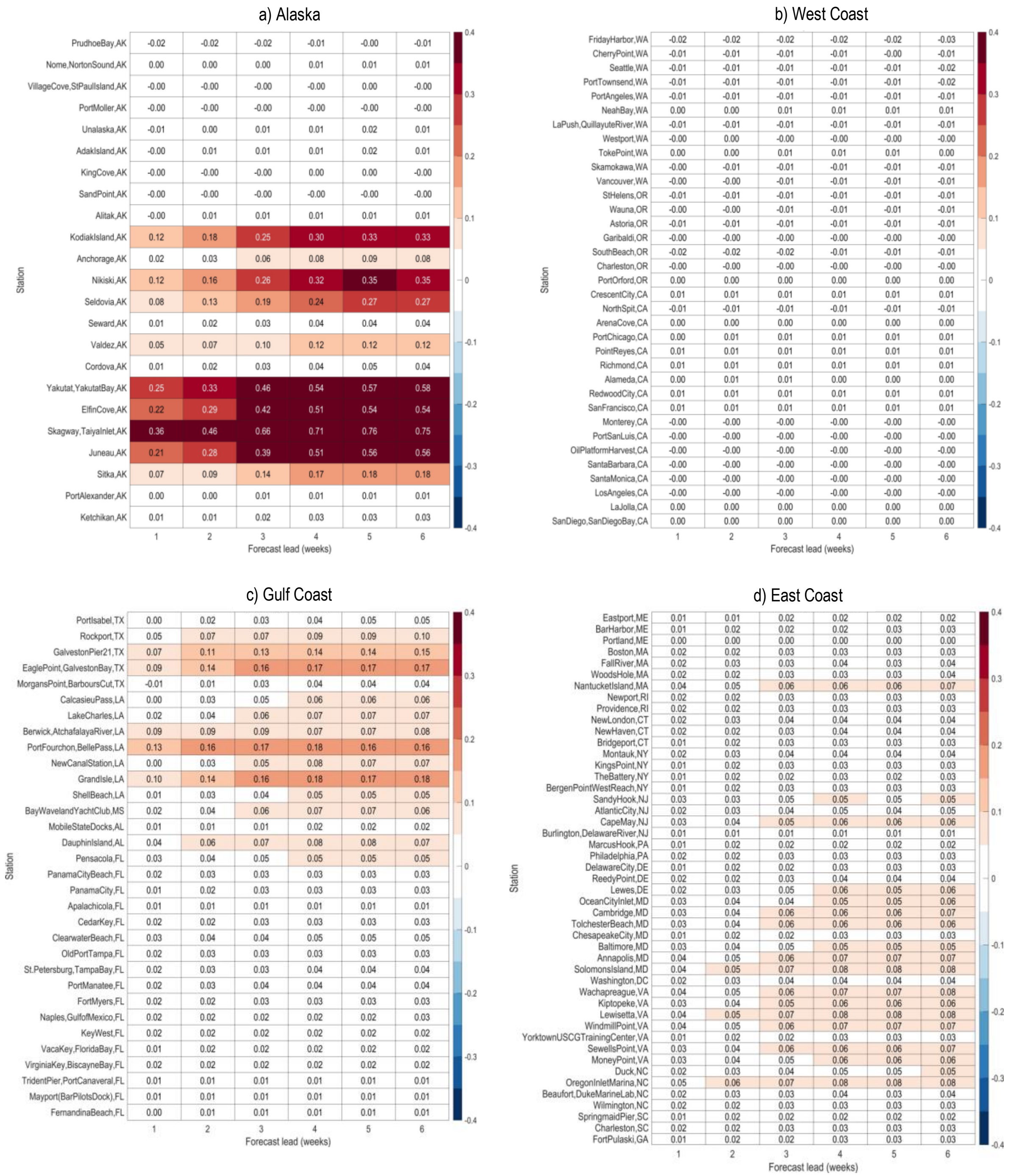

Figure S13. Difference between anomaly correlation skill of $\widehat{\boldsymbol{NTR}}$ (SSH + IBE + VLM) and IFS-only (SSH + IBE) for **(a)** Alaska, **(b)** the West Coast, **(c)** the Gulf Coast, and **(d)** the East Coast.