Subseasonal Climate Prediction in the Western US using Bayesian Spatial Models (Supplementary Material)

Vishwak Srinivasan[1] **Justin Khim**[3] **Arindam Banerjee**[2] **Pradeep Ravikumar**[1]

[1]Machine Learning Department, Carnegie Mellon University,
[2]Department of Computer Science, University of Illinois at Urbana-Champaign,
[3]Amazon

A FURTHER RELATED WORK

In this section, we discuss more related work. There are two primary areas of related work: weather and climate forecasting problems and spatial statistics.

We start with climate and weather forecasting. The subseasonal problem itself is considered by Cohen et al. [2019], DelSole and Banerjee [2017], Hwang et al. [2019], Raff et al. [2017], Totz et al. [2017], White et al. [2017]. The methods given by Hwang et al. [2019] include a joint-stepwise procedure based on cosine similarity for variable selection at different spatial locations, a nearest-neighbor predictor, and an ensemble of the two. The prediction is over states in the western US as defined by the Rodeo, and we use the same states but not the same grid of locations for which to predict temperature. Second, Totz et al. [2017] considers a hierarchical clustering algorithm for predicting monthly precipitation in the Mediterranean. Specifically, spatial patterns of precipitation over a given time are clustered, and the cluster identities can be thought of as estimated latent variables. The latent variables are also assumed to affect the covariates linearly, and so given a new prediction problem, predictions are simply a weighted average of the clusters with weights determined by the covariates. Thus, the hierarchical clustering considered by Totz et al. [2017] is quite dissimilar from our use. Finally, another related problem is considered by Gonçalves et al. [2016], who attempt to aggregate the output from multiple Earth System Models to predict average monthly temperature in North and South America.

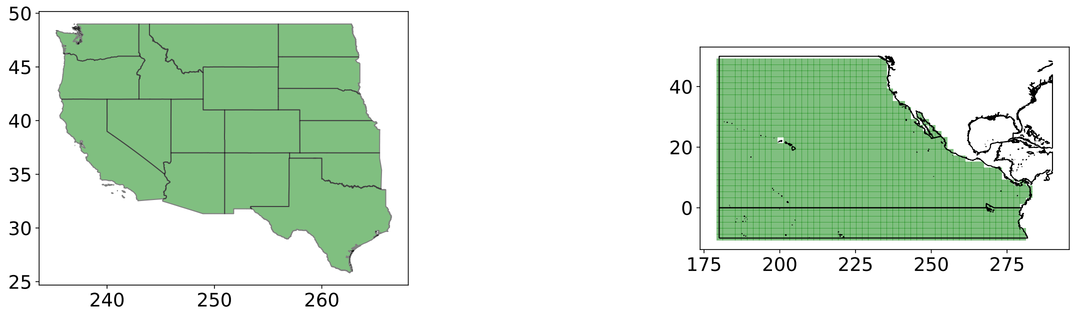
Now, we discuss the more relevant aspects of spatial statistics; for more comprehensive references, see Banerjee et al. [2014], Cressie and Wikle [2011], Gelfand et al. [2010] and the references therein. Specifically, we consider spatial statistics related to climate and then spatial statistics more broadly. In general, spatial statistics is concerned with building models to predict quantities at different locations and, in the case of spatio-temporal statistics, times. This may be cumbersome because the location and time need not lie within the observed data locations and times but merely near them, which necessitates good spatio-temporal covariance functions. Alternatively, data is often missing for reasons as simple as cloud cover that obscures satellite observations, and so models that permit interpolation are helpful. On the other hand, our problem is much simpler since we do not need to interpolate.

First, there has been a variety of statistical work on climate problems [Cox and Isham, 1988, Brillinger, 1997, Gneiting and Guttorp, 2010], and many of these models try to incorporate simplified versions of climate dynamics. Even within statistical approaches, the use of Bayesian methods is far from recent [Epstein, 1985]. Berliner et al. [2000] and Wikle et al. [2001] use Bayesian hierarchical models to predict Pacific sst months in advance and to provide high-resolution estimates for Pacific wind speeds. Additionally, the models allow for time-varying regression coefficients to capture the non-linearity of the underlying processes. Wikle and Anderson [2003] consider spatial Bayesian point process models for tornado counts in the US. Milliff et al. [2011] model surface vector winds over the Mediterranean Sea up to 10 days into the future with a Bayesian hierarchical model to quantify the uncertainty in forecasts.

Of course, spatial models have found widespread use outside of climate prediction; two examples are epidemiology and crimonology. In epidemiology, one is interested in better understanding the spatial structure of disease [Anderson et al., 2014, Feng et al., 2016]. Generally, one seeks to identify regions where individuals are at a greater risk of contracting disease. In crimonology, one tries to predict trends in crime over time and where crimes are more likely to occur by neighborhood

Climate Variable	Location	Spatial Resolution	Temporal Range
`tmp2m`	Land	$2^\circ \times 2^\circ$	1981–2018
`hgt500` `rhum.sig995`	Land, Sea	$0.5^\circ \times 0.5^\circ$ (Land) $2^\circ \times 2^\circ$ (Sea)	1981–2018
`sst` `slp`	Sea	$2^\circ \times 2^\circ$	1981–2018

Table 1: Climate variable data details. We use lower-resolution data for variables defined on the seas since the area is larger.



Western US

Northern Pacific

Figure 1: Land and sea area from which data was used in our analysis (shaded in green).

[Balocchi and Jensen, 2019, Balocchi et al., 2019, Flaxman et al., 2019], and these are helpful for quantifying how cities are changing and for predictive policing. Additionally, there are many more applications under the umbrella of econometrics [Pace and LeSage, 2010].

B DATA AND PREPROCESSING

Table 1 contains the different climate variables, the types of locations over which they are measured, the latitude-longitude resolution of the measurements, and the time periods over which they are defined.

B.1 STANDARDIZATION

Here, we describe out standardization procedure. Before we discuss the details, we make the note of the notations used. s denotes a location and t denotes a date - represented as an integer. For a climate variable m, denote $c^{(m)}(s,t)$ denote the measurement at location s and date t. Let $\text{days}(t)$ denote the set of set of dates for the same day of the year. For example, $\text{days}(\text{December 31, 1999})$ is the set of all December 31$^{\text{st}}$s from 1981 to 2010.

The first step is to obtain the rolling average every two weeks, since the targets are averaged over two weeks. Define the temporal rolling average:

$$\overline{c^{(m)}}(s,t) = \frac{1}{14} \sum_{t'=t-13}^{t} c^{(m)}(s,t).$$

With the definition of the rolling averages, we proceed to standardize these. Define the average for the date t or the *climatology* for date t by

$$\overline{c^{(m)}}_{\text{clim}}(s,t) = \frac{1}{\text{days}(t)} \sum_{t' \in \text{days}(t)} \overline{c^{(m)}}(s,t').$$

Analogously, define the standard deviation for the date t by

$$\overline{c^{(m)}}_{\text{dev}}(s,t) = \sqrt{\frac{1}{|\text{days}(t)|-1} \sum_{t' \in \text{days}(t)} \left(\overline{c^{(m)}}(s,t') - \overline{c^{(m)}}_{\text{clim}}(s,t')\right)^2}.$$

The centered two-week anomaly is thus given by

$$a^{(m)}(s,t) = \overline{c^{(m)}}(s,t) - \overline{c^{(m)}}_{\text{clim}}(s,t).$$

The centered two-week standardized anomaly or two-week z-score is given by

$$z^{(m)}(s,t) = \frac{\overline{c^{(m)}}(s,t) - \overline{c^{(m)}}_{\text{clim}}(s,t)}{\overline{c^{(m)}}_{\text{dev}}(s,t)}.$$

The covariates and targets used in the clustering algorithms are matrices of these standardized anomalies. In the subseasonal forecasting regime, the targets are defined to be

$$y^{(m)'}(s,t) = z^{(m)}(s,t+28).$$

However, for prediction, we are interested in the centered anomaly rather than the standardized anomaly. For this, we re-scale our predictions using $\overline{c^{(m)}}_{\text{dev}}$. To be precise, given a prediction $\widehat{y^{(m)}}'(s,t) = \hat{z}^{(m)}(s,t+28)$, we obtain the prediction for the centered anomaly as

$$\widehat{y^{(m)}}(s,t) = \widehat{a^{(m)}}(s,t+28) = \widehat{y^{(m)}}'(s,t) \cdot \overline{c^{(m)}}_{\text{dev}}(s,t).$$

B.2 SUBSAMPLING AND MISSING DATA

In the course of our analysis, we find it necessary to make decisions regarding subsampling and missing data.

First, we discuss subsampling. Here, we specifically refer to subsampling with respect to dates. We choose to have 14 days between predictions, e.g., we would predict on the 1$^{\text{st}}$ and then on the 16$^{\text{th}}$ of a month, to limit dependence between samples.

Next, for missing data, we note that there are missing values for both the covariates and targets. We navigate this issue by filling these missing instances by forward-filling. Specifically, if data corresponding to a location s and timestamp t for climate variable m is missing, then we fill this missing entry with the latest value available before timestamp t.

$$x^{(m)}(s,t) := x^{(m)}\left(s, \operatorname*{argmax}_{t' \le t} \mathsf{AVAIL}(s,t',m)\right)$$

where $\mathsf{AVAIL}(s,t',m) \in \{0,1\}$ checks if data corresponding to a location s and timestamp t' for a climate variable m is available and returns 1 or 0 otherwise.

C HYPERPARAMETER TUNING FOR THE SPATIAL MODEL

In our regression analysis, we have a number of hyperparameters to tune. The first govern the scale of the variance of the observed data, given by τ_ϵ, and are denoted by a_ϵ and b_ϵ. This is one variable for which we used empirical Bayes tuning, which was performed as follows. We compute the empirical variance of the anomalies on the 132 `tmp2m` locations from the training set, which gives 0.906. Since tau_ϵ^2 governs the variance in the noise, we set $b_\epsilon = 0.906(a_\epsilon - 1)$. This ensures the prior mean is 0.906, and now we only need to specify a_ϵ. We simply set $a_\epsilon = 32$ so the prior variance would be reasonably small.

The remaining hyperparameters are a_j and b_j for $j = 1, \dots, d$ corresponding to the Inverse Gamma parameters for each τ_j^2. For this too, we use an empirical Bayes approach. Specifically, we obtain a vanilla multiple regression problem and check the variance of each of j rows of parameters. We find this to be particularly small and almost invariant across covariates i.e., j, so we set $a_j = 100$ and $b_j = 0.01$. This ensures that the prior mean of τ_j^2 is roughly 10^{-4}, and the high value of a_j ensures that the variance of the prior is small too.

C.1 IMPLEMENTATION SPECIFIC DETAILS

We use 1000 samples for the warmup phase and 4000 samples after that, which are our prediction samples. For all of the unobserved variables, we see that the Gelman-Rubin statistic is 1, which indicates that the method has converged. Furthermore, we verify that the implementation is stable by running it several times.

D FURTHER NUMERICAL RESULTS

Year	Spatial-regression	XGBoost	Multi-task Lasso	Neural Network
2011	$3.53 \cdot 10^{-1}$	$2.44 \cdot 10^{-2}$	$\mathbf{3.69 \cdot 10^{-1}}$	$3.62 \cdot 10^{-1}$
2012	$3.78 \cdot 10^{-1}$	$\mathbf{4.79 \cdot 10^{-1}}$	$3.58 \cdot 10^{-1}$	$3.66 \cdot 10^{-1}$
2013	$-4.93 \cdot 10^{-2}$	$-8.78 \cdot 10^{-2}$	$\mathbf{-3.92 \cdot 10^{-2}}$	$-3.95 \cdot 10^{-2}$
2014	$\mathbf{2.00 \cdot 10^{-1}}$	$7.19 \cdot 10^{-2}$	$1.93 \cdot 10^{-1}$	$1.93 \cdot 10^{-1}$
2015	$5.12 \cdot 10^{-2}$	$\mathbf{3.88 \cdot 10^{-1}}$	$2.48 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$
2016	$\mathbf{4.09 \cdot 10^{-1}}$	$4.08 \cdot 10^{-1}$	$3.99 \cdot 10^{-1}$	$4.03 \cdot 10^{-1}$
2017	$2.38 \cdot 10^{-1}$	$\mathbf{2.75 \cdot 10^{-1}}$	$2.26 \cdot 10^{-1}$	$2.23 \cdot 10^{-1}$
2018	$\mathbf{2.43 \cdot 10^{-1}}$	$7.02 \cdot 10^{-2}$	$2.42 \cdot 10^{-1}$	$2.32 \cdot 10^{-1}$

Table 2: Variation of cosine similarity annually on the test set. Note the considerable heterogeneity across years.

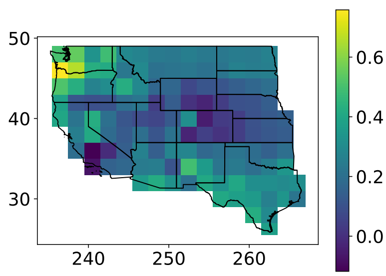
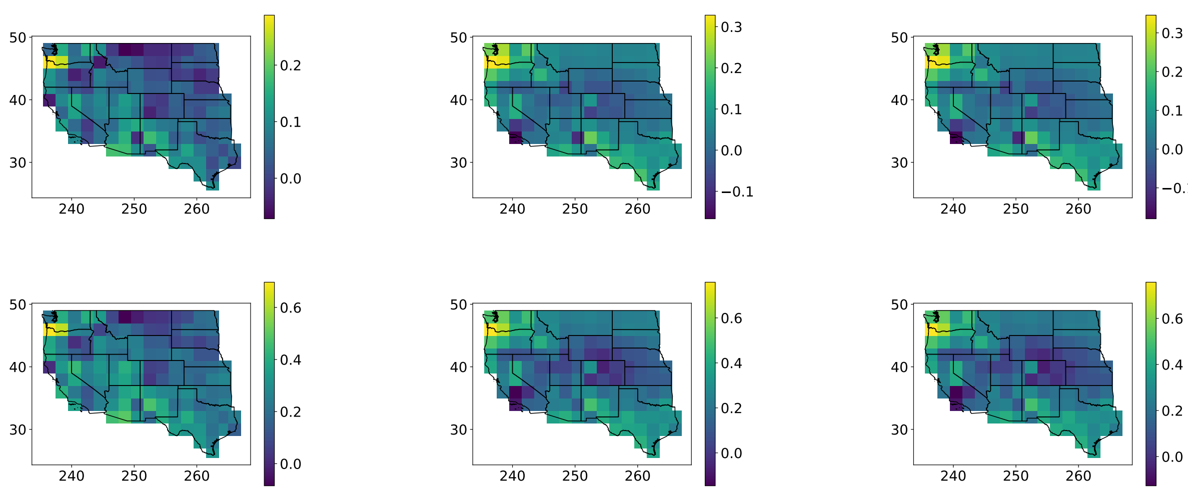


Figure 2: Variation of cosine similarity by location. Note the variance across locations, especially those closer to the coast.

In this section, we provide additional numerical results that we had alluded to earlier in Section 5.

First, we present the variation in cosine similarity annually. This is available in Table 2. Note that the trends are similar to that observed in Table 2: we see poor performance in a majority of the models in 2013 and 2015, which we hypothesize was due to the extreme winter and the El Ñino event that occurred in the respective years.

Next, we present the variation of cosine-similarity spatially for the predictions from our Bayesian spatial model. Again, we notice that the variation is very similar to skill as shown in Figure 1.



XGBoost Multi-Task Lasso Neural Network

Figure 3: Top row: Variation of skill by location. Bottom row: Variation of cosine-similarity by location.

In Figure 3, we depict the variation of the skill and cosine similarity of the baseline models spatially. Note that XGBoost has average performance throughout, and in contrast to the other models, performs relatively badly in the northern states.

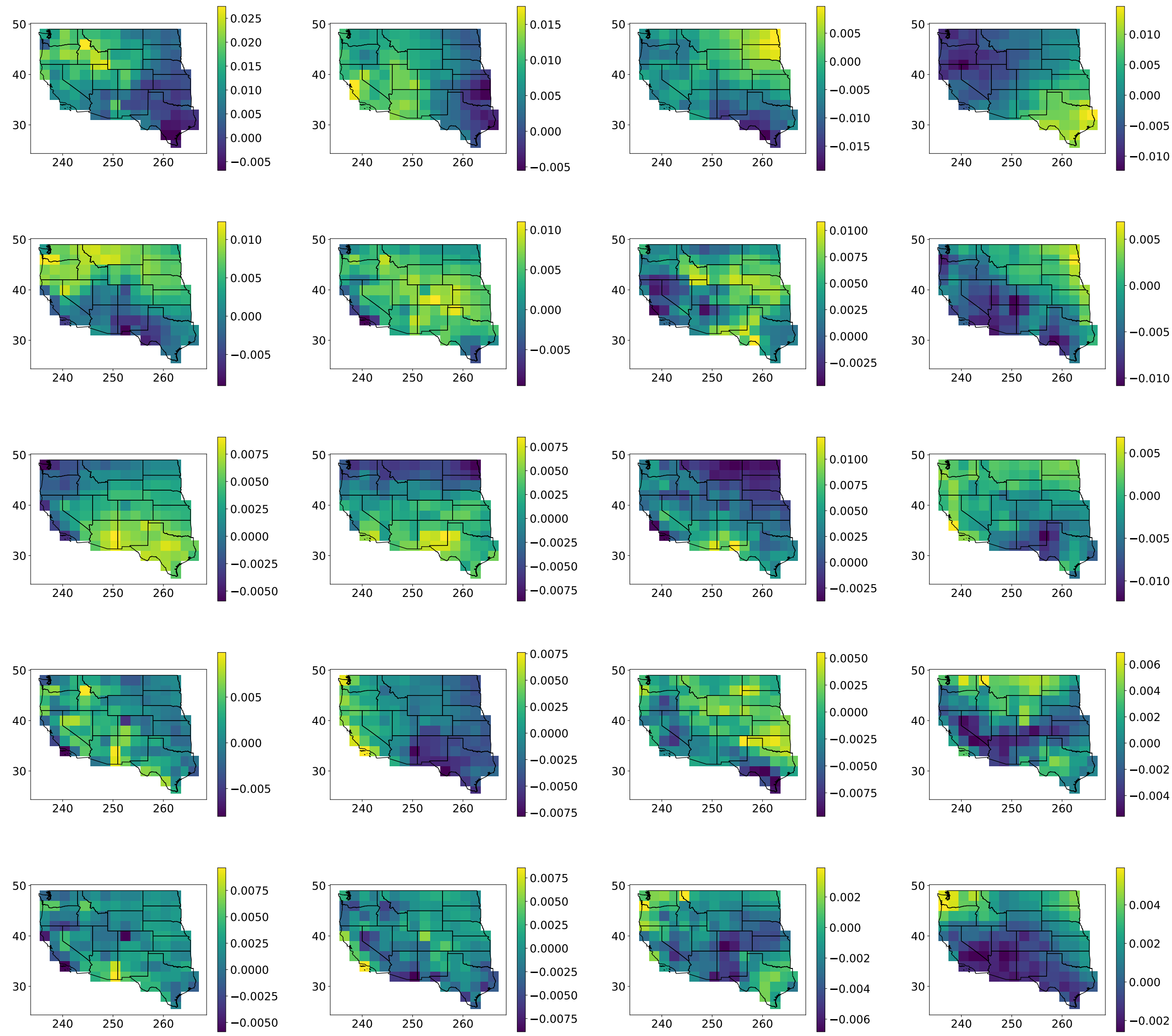


Figure 4: The posterior means of regression coefficients by spatial location for all covariates.

We next show spatial variation of the posterior means of the regression coefficients corresponding to the remaining covariates. For brevity we include the 4 shown in Figure 2. All of the regression coefficients are shown in Figure 4.

D.1 HYPERPARAMETER CHOICES

Finally, we specify the manner in which we performed model selection for XGBoost, Multi-Task Lasso and the nonlinear neural network. We split the train set into two parts: we use data from years 1981 – 1995 to train the model, and 1996 – 2010 to test the model. We then pick the hyperparameter configuration that yields the lowest error. For each of the models, we highlight the hyperparameters tuned and the choices used in Table 3. The chosen hyperparameters are in bold.

Model	Hyperparameter	Choices
XGBoost	learning rate	$\{0.001, \mathbf{0.01}, 0.1\}$
	maximum depth	$\{\mathbf{4}, 8\}$
	number of estimators	$\{50, \mathbf{100}, 200\}$
Multi-task Lasso	regularization	$\{0.001, 0.002, 0.005, 0.01, 0.02, 0.02, 0.05, 0.1, 0.2, \mathbf{0.5}\}$
Neural Network	architecture	`20 - 76 - 132`, `20 - 51 - 102 - 132`, **`20 - 38 - 76 - 114 - 132`**
	activation	`ReLU`, `TanH`, **none**

Table 3: Hyperparameters

References

Craig Anderson, Duncan Lee, and Nema Dean. Identifying clusters in Bayesian disease mapping. *Biostatistics*, 15(3): 457–469, 2014.

Cecilia Balocchi and Shane T Jensen. Spatial modeling of trends in crime over time in Philadelphia. *Annals of Applied Statistics*, 13(4):2235–2259, 2019.

Cecilia Balocchi, Sameer K Deshpande, Edward I George, and Shane T Jensen. Crime in Philadelphia: Bayesian clustering with particle optimization. *arXiv preprint arXiv:1912.00111*, 2019.

Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC press, 2014.

L Mark Berliner, Christopher K Wikle, and Noel Cressie. Long-lead prediction of pacific ssts via bayesian dynamic modeling. *Journal of Climate*, 13(22):3953–3968, 2000.

David R Brillinger. An application of statistics to meteorology: estimation of motion. In *Festschrift for Lucien Le Cam*, pages 93–105. Springer, 1997.

Judah Cohen, Dim Coumou, Jessica Hwang, Lester Mackey, Paulo Orenstein, Sonja Totz, and Eli Tziperman. S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdisciplinary Reviews: Climate Change*, 10(2), 2019.

David Roxbee Cox and Valerie Isham. A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 415(1849):317–328, 1988.

Noel Cressie and Christopher K Wikle. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 2011.

Timothy DelSole and Arindam Banerjee. Statistical seasonal prediction based on regularized regression. *Journal of Climate*, 30(4):1345–1361, 2017.

Edward S Epstein. *Statistical Inference and Prediction in Climatology: a Bayesian Approach*. American Meteorological Society, 1985.

Wenning Feng, Chae Young Lim, Tapabrata Maiti, and Zhen Zhang. Spatial regression and estimation of disease risks: A clustering-based approach. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(6):417–434, 2016.

Seth Flaxman, Michael Chirico, Pau Pereira, Charles Loeffler, et al. Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the nij "real-time crime forecasting challenge". *The Annals of Applied Statistics*, 13(4):2564–2585, 2019.

Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of Spatial Statistics*. CRC press, 2010.

Tilmann Gneiting and Peter Guttorp. Continuous parameter spatio-temporal processes. In Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes, editors, *Handbook of Spatial Statistics*, chapter 23. CRC press, 2010.

André R Gonçalves, Fernando J Von Zuben, and Arindam Banerjee. Multi-task sparse structure learning with gaussian copula models. *The Journal of Machine Learning Research*, 17(1):1205–1234, 2016.

Jessica Hwang, Paulo Orenstein, Judah Cohen, Karl Pfeiffer, and Lester Mackey. Improving subseasonal forecasting in the western us with machine learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2325–2335, 2019.

Ralph F Milliff, Alessandro Bonazzi, Christopher K Wikle, Nadia Pinardi, and L Mark Berliner. Ocean ensemble forecasting. part i: Ensemble mediterranean winds from a bayesian hierarchical model. *Quarterly Journal of the Royal Meteorological Society*, 137(657):858–878, 2011.

R. Kelley Pace and James LeSage. Spatial econometrics. In Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes, editors, *Handbook of Spatial Statistics*, chapter 15. CRC press, 2010.

David Raff, Kenneth Nowak, Robert Cifelli, Levi D Brekke, and Robert Stabler Webb. Sub-seasonal climate forecast rodeo. In *AGU Fall Meeting*, 2017.

Sonja Totz, Eli Tziperman, Dim Coumou, Karl Pfeiffer, and Judah Cohen. Winter precipitation forecast in the european and mediterranean regions using cluster analysis. *Geophysical Research Letters*, 44(24):12–418, 2017.

Christopher J White, Henrik Carlsen, Andrew W Robertson, Richard JT Klein, Jeffrey K Lazo, Arun Kumar, Frederic Vitart, Erin Coughlan de Perez, Andrea J Ray, Virginia Murray, et al. Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteorological applications*, 24(3):315–325, 2017.

Christopher K Wikle and Christopher J Anderson. Climatological analysis of tornado report counts using a hierarchical bayesian spatiotemporal model. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.

Christopher K Wikle, Ralph F Milliff, Doug Nychka, and L Mark Berliner. Spatiotemporal hierarchical bayesian modeling tropical ocean surface winds. *Journal of the American Statistical Association*, 96(454):382–397, 2001.