Title
Graph-Guided Regularized Regression of Pacific Ocean Climate Variables to Increase Predictive Skill of Southwestern U.S. Winter Precipitation.

Permalink

Journal

ISSN

Authors
Stevens, Abby
Willett, Rebecca
Mamalakis, Antonios
et al.

Publication Date

DOI

Peer reviewed

Graph-Guided Regularized Regression of Pacific Ocean Climate Variables to Increase Predictive Skill of Southwestern U.S. Winter Precipitation

Abby Stevens[a], Rebecca Willett[a,b], Antonios Mamalakis[c], Efi Foufoula-Georgiou[c,d], Alejandro Tejedor[e], James T. Randerson[d], Padhraic Smyth[f,g], Stephen Wright[h]

[a]Department of Statistics, University of Chicago, Chicago, Illinois

[b]Department of Computer Science, University of Chicago, Chicago, Illinois

[c]Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, California

[d]Department of Earth System Science, University of California, Irvine, Irvine, California

[e]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

[f]Department of Computer Science, University of California, Irvine, Irvine, California

[g]Department of Statistics, University of California, Irvine, Irvine, California

[h]Computer Sciences Department, University of Wisconsin–Madison, Madison, Wisconsin

Abstract

Understanding the physical drivers of seasonal hydroclimatic variability and improving predictive skill remains a challenge with important socioeconomic and environmental implications for many regions around the world. Physics-based deterministic models show limited ability to predict precipitation as the lead time increases, due to imperfect representation of physical processes and incomplete knowledge of initial conditions. Similarly, statistical methods drawing upon established climate teleconnections have low prediction skill due to the complex nature of the climate system. Recently, promising data-driven approaches have been proposed, but they often suffer from overparameterization and overfitting due to the short observational record, and they often do not account for spatiotemporal dependencies among covariates (i.e., predictors such as sea surface temperatures). This study addresses these challenges via a predictive model based on a graph-guided regularizer that simultaneously promotes similarity of predictive weights for highly correlated covariates and enforces sparsity in the covariate domain. This approach both decreases

the effective dimensionality of the problem and identifies the most predictive features without specifying them a priori. We use large ensemble simulations from a climate model to construct this regularizer, reducing the structural uncertainty in the estimation. We apply the learned model to predict winter precipitation in the southwestern United States using sea surface temperatures over the entire Pacific basin, and demonstrate its superiority compared to other regularization approaches and statistical models informed by known teleconnections. Our results highlight the potential to combine optimally the space–time structure of predictor variables learned from climate models with new graph-based regularizers to improve seasonal prediction.

Keywords

Precipitation; Seasonal forecasting; Climate models; Dimensionality reduction; Machine learning; Regression

1. Introduction

Seasonal prediction of regional hydroclimate is typically based on deterministic physical models or statistical techniques, yet both approaches exhibit limited predictive ability (Wang et al. 2009; National Academies of Sciences, Engineering, and Medicine 2016). Precipitation predictions based on deterministic physical models (regional climate models) exhibit high uncertainty due to imperfect physical conceptualizations, sensitivity to initial and boundary conditions, and variations in model physics and grid resolutions (e.g., Chang et al. 2000). On the other hand, predictive statistical approaches (e.g., Wu et al. 2009; Schepen et al. 2012; Peng et al. 2014; Tao et al. 2017), which exploit historically and physically established climate teleconnections between regional hydroclimate and large-scale modes of climate variability [e.g., El Niño–Southern Oscillation (ENSO); see, e.g., Ropelewski and Halpert 1986; Bradley et al. 1987; Redmond and Koch 1991; McCabe and Dettinger 1999; Dai 2013], also exhibit limited predictive skill. The main reason is that the complex and nonstationary interactions between large-scale dynamics and regional hydroclimate cannot be captured sufficiently well with a limited number of prespecified climate indices (regions used for computing sea surface temperature anomalies) as predictors, even when sophisticated statistical schemes are used (nonlinear statistical schemes, Bayesian techniques, etc.).

Recognizing the limitations above, the community has been increasingly embracing the application of methods that aim to learn from both climate models and statistical schemes in order to improve seasonal predictive skill. These methods range from weighted multimodel averaging techniques (Raftery et al. 2005; Luo et al. 2007; Schepen and Wang 2013; Cheng and AghaKouchak 2015) or methods that directly combine predictions from climate models and statistical schemes (Coelho et al. 2004; Schepen et al. 2014; Madadgar et al. 2016) to data-driven approaches based on machine learning, in a setting where predictor variables are not prespecified but rather are guided by the data or climate model outputs (Quan et al. 2006; DelSole and Banerjee 2017; Hewitt et al. 2018; Ham et al. 2019; Willard et al. 2020; He et al. 2020). In the former category of methods, the prediction skill depends strongly on the skill of each of the models considered, thus making such techniques prone to all the

aforementioned limitations. In contrast, the use of machine learning has potential since climate information from the entire globe can in principle be used to inform the prediction. However, these techniques also face important practical limitations. First, because of the short record of observations and the large number of predictors, the number of degrees of freedom of the problem is vast, significantly increasing the risk of overfitting (Ham et al. 2019). Second, strong spatiotemporal dependences among the predictor variables, which are certainly present in climate applications, need to be taken into consideration for imposing structure in the predictor space to reduce the dimensionality of the problem and improve physical interpretability.

In this study, we aim to address the challenges discussed above by introducing a regularized regression scheme that accounts explicitly for spatiotemporally correlated predictors. Regularization is an established technique in statistics, machine learning, and signal processing that can mitigate the challenges of many degrees of freedom relative to the amount of data. The key idea is that rather than simply finding the model that best fits the data according to some loss function, we instead minimize the sum of the loss and a regularization function, where the latter reflects some prior belief about which models are better than others. Sparsity regularization [e.g., the least absolute shrinkage and selection operator (LASSO) regularization; Tibshirani (1996)] has already been explored in the context of precipitation downscaling and data assimilation (Ebtehaj et al. 2012; Ebtehaj and Foufoula-Georgiou 2013) and climate forecasting [see the recent studies of DelSole and Banerjee (2017) and He et al. (2020)], but suffers from ignoring the spatiotemporal dependencies among predictors. To respect the embedded space–time structure of the climate system and enforce sparsity, we use a "graph total variation" (GTV) regularizer (i.e., constraint) that promotes similarity of weights (i.e., regression coefficients) for highly correlated predictors. The GTV is a graph-based regularizer, based on the graph formed by the covariance matrix of the predictors, that was recently introduced by Li et al. (2018). To address the issue of the short observational record, and to estimate robustly the covariance matrix of the predictor variables, we make use of a large ensemble of climate model outputs. Using climate model outputs in the training of machine learning (ML) models is a subcase of the general category of techniques that aim to integrate physical knowledge and machine learning (Willard et al. 2020), and it has recently been shown to be highly efficient in increasing predictive skill on seasonal to interannual time scales (DelSole and Banerjee 2017; Ham et al. 2019). Although our study differs from these studies in that it uses the climate model outputs not to train the ML model per se, but only to compute the covariance matrix of the predictors used as a GTV regularizer, it adds to this important new line of research in synergistically leveraging both climate models and observations with the goal of improving prediction skill.

We explore the prediction skill of our methodology for the case study of predicting precipitation over the southwestern United States (SWUS), focusing on the winter season (specifically, November–March), when the majority of precipitation occurs. Despite the increasing attention that it has received over the years (Schonher and Nicholson 1989; McCabe and Dettinger 1999; Gershunov and Cayan 2003; Schubert et al. 2016; Madadgar et al. 2016; Liu et al. 2018; Hao et al. 2018; Zhang et al. 2018; Mamalakis et al. 2018; Pan et al. 2019), early and accurate prediction of winter precipitation in SWUS remains a

challenge, with significant implications for the region's population and economy (Howitt et al. 2014, 2015; AghaKouchak et al. 2015; Medellín-Azuara et al. 2016). Traditional climatic drivers of SWUS precipitation (e.g., ENSO) explain just a small fraction of the interannual variability of precipitation totals, which in some cases are determined by a small number of winter storms (Dettinger et al. 2011; Dettinger and Cayan 2014). Moreover, it is known that the ENSO relationship with SWUS climate undergoes multidecadal fluctuations (McCabe and Dettinger 1999; Yu et al. 2012), with many recent studies pointing out that it has been losing strength in the recent decades, while the western Pacific climatic state is gaining in importance (Wang et al. 2014; Baxter and Nigam 2015; Teng and Branstator 2017; Seager et al. 2017; Swain et al. 2017; Myoung et al. 2018; Mamalakis et al. 2018; Lee et al. 2018). The special difficulty of this problem also arises because the SWUS lies within a transition zone between the subtropics and the midlatitudes (i.e., 30°–40°N). In fact, the latter is among the reasons that the effect of climate change on future precipitation trends over the SWUS is highly uncertain, with midlatitude regions expected to become wetter and subtropical regions drier (Allen and Luptowitz 2017). Because of its intrinsic complexity, this region offers an excellent case study for exploring and bench-marking data-driven predictive methods.

As predictor variables, we use late summer and early fall (July, August, September, and October) sea surface temperature (SST) over the entire Pacific basin. Note that although there are studies indicating the importance of Atlantic Ocean temperatures as drivers of SWUS precipitation as well (Enfield et al. 2001; McCabe et al. 2004), our focus here is only on the Pacific Ocean, as a first step. We cast the prediction problem as an estimation problem in which predictors are not specified in advance, but rather emerge from the data by minimizing an appropriate loss function. We first demonstrate the increased predictive skill of the proposed GTV model when the covariance matrix that defines the GTV regularizer is computed from a large ensemble of a climate model, rather than the single realization of observations. Second, we benchmark the GTV model against two different classes of predictive models: 1) other regularized regression methods (LASSO and fused LASSO; Tibshirani et al. 2005) and 2) simple ordinary least squares using known teleconnection indices as predictors. Our analysis shows that constraining the predictive model by the spatiotemporal covariance of the predictors via a GTV regularization outperforms all the other considered models and substantially increases the seasonal precipitation predictive skill. Last, we show that both the GTV performance and the emerged predictors of precipitation are quite robust to perturbations in the covariance matrix that is used to define the GTV regularization term.

The structure of the paper is as follows. In section 2, we describe the prediction problem and the data used. We introduce the proposed methodology in section 3 and discuss the advantages of using a graph-based regularizer in which the graph is based on covariance information from a climate model, instead from the limited observations. In section 4, we present results on the performance of our proposed model and compare its skill to other methods. Moreover, we study the emergent predictors, aiming to gain physical insight about the drivers of SWUS precipitation. We also perform a sensitivity analysis of our results to gain more confidence in the predictive performance and the emergent predictors. Conclusions and directions for future research are discussed in section 5.

2. Prediction problem and data/models used

The SWUS (California, Nevada, Utah, and Arizona) is composed of 25 climate divisions, for each of which precipitation series are available at https://www.ncdc.noaa.gov/cag/time-series/us (Vose et al. 2014). The season we aim to predict precipitation is November–March, when the majority of the annual precipitation occurs; note also that winter precipitation, especially that which is stored as snowpack, is necessary for sustaining the water supply through relatively dry summers (Mote et al. 2005; Shukla et al. 2015; Liu et al. 2018). As shown in Fig. 1a, the northwestern part of the region receives much higher precipitation than the rest of the SWUS, which is generally considered a fairly dry area. However, the interannual variability in the central and southern part is high, generally higher than 40%–50% of the mean precipitation (coefficient of variation of the order of 0.4–0.5 or higher), compared to the northern part of the SWUS, where the coefficient of variation is 0.3 or lower (Fig. 1b). To distinguish between the two different hydroclimatologies of the northern part and the central/southern part of the SWUS, previous studies have used different approaches, such as focusing on the area below a certain latitude (Liu et al. 2018) or considering only the climate divisions for which a specific predictor (e.g., the Niño-3.4 index, an ENSO index) exhibits a significant relation with precipitation (Mamalakis et al. 2018). Here, we distinguish between the two precipitation regimes using an area-weighted principal component (PC) analysis. As can be seen from Fig. 1c, PC1 (which explain about 64% of the total precipitation variability in the SWUS) is more strongly associated with the central and southern part of the SWUS than with the northern part. Based on PC1 and PC2 (which explain about 22% of the total precipitation variability in the SWUS; not shown), we have selected the region of focus for our analysis, composed of 18 climate divisions, which is shown in Fig. 1d, together with the series of the area-weighted average precipitation.

As predictor variables, we use late summer and early fall (July–October) SSTs over the Pacific basin, which is defined as the area in 60°S–60°N, 80°–280°E. Historical time series of SST (monthly series on a $1° \times 1°$ grid; see Hirahara et al. 2014) are obtained from https://www.esrl.noaa.gov/psd/data/gridded/data.cobe2.html. At that resolution, the number of predictor variables is very large (roughly $120 \times 200 \times 4 = 96\,000$) making the problem highly ill posed. Thus, we upscale the original SST field by simple areal averaging into grid boxes of size $10° \times 10°$ over the Pacific basin, to reduce the dimensionality of the problem. After removing the boxes over land, we end up with roughly 900 predictor variables in total (i.e., 4 months over 12 boxes in latitude and 20 boxes in longitude).

The analysis is performed for the years 1940/41–2018/19, since SST records are not trustworthy before the 1940s, due to the limited availability of observations over the Pacific basin and globally (Deser et al. 2010). In particular, we use years 1940/41–1989/90 as a training period, and years 1990/91–2018/19 as a test period. All individual SST series are linearly detrended and standardized (zero mean and unit variance) before they are used in the analysis.

To reduce the uncertainty in the estimation of the spatiotemporal dependency of the predictors (covariance matrix of SST predictors, used to define the GTV regularizer), we also examined simulations from the Community Earth System Model-Large Ensemble

project (CESM-LENS; Kay et al. 2015), which can be found at http://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html. Specifically, we use monthly series (on a $1.25° \times 0.9°$ grid) of surface temperatures over the Pacific basin, which we also upscale to $10° \times 10°$ grids to match the grids used for the observed SSTs. We note that the CESM-LENS project consists of 40 ensemble members, each one corresponding to the same model physics but different initial conditions in the atmosphere. CESM-LENS relies on historical boundary conditions for the period 1920–2005, and the representative concentration pathway 8.5 (RCP8.5) is applied as forcing for years 2006–2100. Here we used only archives of simulation output from 1940 to 2005 to build our model covariance matrices, as the focus of our analysis is on improving prediction for the contemporary period. Note that the 40 ensemble members constitute independent but equally probable trajectories of the Earth system with historical forcing [see Kay et al. (2015) for more information].

3. Methodology

Let $y_r^{(i)}$ denote the winter precipitation in year i and climate division r. We hypothesize that $y_r^{(i)}$ can be predicted from climate variables at different locations over the Pacific Ocean and different lag times (e.g., months ahead of the winter period), with a model of the form

$$y_r^{(i)} = \sum_{j=1}^{p} x_j^{(i)} \beta_{j,r} + \varepsilon_r^{(i)}, \tag{1}$$

where $\varepsilon_r^{(i)}$ is a Gaussian noise $N(0, \sigma^2)$ term. Writing (1) in a matrix form and dropping the index r for convenience results in

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{y} = \left(y^{(1)}, y^{(2)}, ..., y^{(n)}\right)^{\mathrm{T}} \in \mathbb{R}^n$ is the vector of winter precipitation over n years, $\mathbf{X} = \left[\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)}\right]^{\mathrm{T}} \in \mathbb{R}^{n \times p}$ is the matrix of climate variables (i.e., SSTs over the Pacific Ocean and in the four different months preceding the winter season, namely July, August, September, and October), $\boldsymbol{\beta} = \left(\beta_1, \beta_2, ..., \beta_p\right)^{\mathrm{T}} \in \mathbb{R}^p$ is the vector of weights corresponding to p predictors, and $\boldsymbol{\varepsilon} = \left(\varepsilon^{(1)}, \varepsilon^{(2)}, ..., \varepsilon^{(n)}\right)^{\mathrm{T}} \in \mathbb{R}^n$ is a Gaussian noise vector. We clarify that $\mathbf{x}^{(i)}$ is a p-dimensional vector in year i of SSTs arranged by moving along all longitudes and latitudes of the Pacific Ocean and for the four months of July, August, September, and October. Thus, $p = 900$ in our case, while the number of available years is $n = 79$ (i.e., we use 50 years for training and 29 years for testing; from 1940/41 to 1989/90 and from 1990/91 to 2018/19, respectively). Obviously, this problem is highly underdetermined, since $n \ll p$. To solve for $\boldsymbol{\beta}$, we reduce the effective dimension of the problem by adding regularization terms, leading to the formulation

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|_2^2 + \lambda R(\boldsymbol{\beta}) \right\}, \tag{3}$$

where $R(\boldsymbol{\beta})$ is a regularization term, chosen to impose structure and sparsity on $\boldsymbol{\beta}$, and $\lambda > 0$ is the regularization parameter. A popular choice for $R(\boldsymbol{\beta})$ is the LASSO regularizer (Tibshirani 1996), that is, $R(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1 = \sum_{i=1}^{p} |\beta_i|$, which yields minimizers of (3) for which $\boldsymbol{\beta}$ is sparse; that is, there are only a few spatiotemporal variables that are truly predictive while the rest are conditionally independent of the response $\mathbf{y}$. The value of λ, typically estimated using cross-validation, reflects the weight given to the sparsity constraint. However, the LASSO regularization does not take into account that predictors might have a significant spatiotemporal dependence structure that, if included, might further constrain and improve the prediction.

a. Graph total variation

To overcome this problem, we propose a regularizer that accounts explicitly for the spatiotemporal covariance of the predictors. The central idea of this regularizer is that if covariates $\mathbf{x}_j$ and $\mathbf{x}_k$ are highly correlated with one another, then they should receive similar weights $\hat{\beta}_j$ and $\hat{\beta}_k$. This approach helps us select highly correlated *collections* of covariates that serve as strong predictors of precipitation. In contrast, the LASSO estimator would generally select either $\hat{\beta}_j$ or $\hat{\beta}_k$, but not both, and the selected covariate would be very sensitive to any noise in the data. We form a graph to represent the correlations between pairs of covariates, and select a set of weights $\boldsymbol{\beta}$ that is "aligned" with the graph. This regularization scheme, known as graph total variation (GTV), was introduced in Li et al. (2018). Although graph-based regularizers have been explored before (e.g., fused LASSO, edge LASSO, graphtrend filtering), Li et al. (2018) developed theoretical guarantees for the GTV regularizer for highly correlated covariates, and showed how imposing additional structure on $\boldsymbol{\beta}$ to encourage "alignment" with the covariance graph can lead to optimal solutions. This property is important in our problem since $\mathbf{X}$ contains highly correlated columns, resulting for example from dependence between SST anomalies at nearby locations for small time lags or at distant locations but lagged in time.

Let $\widehat{\boldsymbol{\Sigma}}$ be an estimate of the covariance matrix of $\mathbf{X}$ and let $\hat{s}_{j,k} = \mathrm{sign}\left(\widehat{\Sigma}_{j,k}\right)$. The GTV estimator is given by

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\mathrm{argmin}} \left\{ \frac{1}{n} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_{\mathrm{TV}} \sum_{j,k} |\widehat{\Sigma}_{j,k}|^{1/2} |\beta_j - \hat{s}_{j,k}\beta_k| + \lambda_1 \|\boldsymbol{\beta}\|_1 \right\}, \qquad (4)$$

where λ_1 and λ_{TV} are regularization parameters chosen through cross-validation. Here we use a standard fivefold cross-validation approach applied to the training data (i.e., a total of 50 years) to estimate the optimal $\left(\lambda_1^*, \lambda_{\mathrm{TV}}^*\right)$ combination. Specifically, we split the training dataset into five nonoverlapping, random 10-yr sets; for each of the five sets, we train our model (i.e., estimate $\widehat{\boldsymbol{\beta}}$) using the other four sets and compute the prediction error on the held out fifth set. This is repeated for each candidate tuning parameter pair $(\lambda_1, \lambda_{\mathrm{TV}})$. The optimal $\left(\lambda_1^*, \lambda_{\mathrm{TV}}^*\right)$ combination is the one that, on average, minimizes the prediction error across the five different holdout sets. Note that choosing the value of λ_{TV} that determines the importance of the GTV term via cross-validation can mitigate the effect of any

systematic biases reflected in the estimate $\widehat{\mathbf{\Sigma}}$, since if $\widehat{\mathbf{\Sigma}}$ were not informative at all, the optimal λ_{TV} would be close to zero.

We can interpret the estimator in Eq. (4) from a graph perspective by defining a covariance graph based on $\widehat{\mathbf{\Sigma}}$. Let $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$ be an undirected weighted graph with vertices $\mathbf{V} = \{1, 2, \ldots, p\}$, edges $\mathbf{E} := \left\{(j, k) : \left|\widehat{\Sigma}_{j,k}\right| > \theta, \quad j \neq k\right\}$, and weight matrix $\mathbf{W}$ with $w_{j,k} = \left|\widehat{\Sigma}_{j,k}\right|^{1/2}$. That is, each predictor variable (e.g., SST at a particular place and time) is associated with one of the nodes of the graph, and edges reflect the pairs of predictors that are correlated. A threshold parameter θ can be applied to the covariance matrix (Bickel and Levina 2008) for assessing which edges (i.e., links between covariates) will be used in the GTV term (see further discussion about parameter θ in section 3c).

The expression in Eq. (4) may be rewritten using new notation that highlights connections with previous methods and known software for solving the optimization problem. Specifically, let $\mathbf{\Gamma} \in \mathbb{R}^{|\mathbf{E}| \times p}$ be the *weighted edge incidence matrix* of $\mathbf{G}$, where each row l represents a pair of connected vertices (j_l, k_l):

$$\begin{aligned} \Gamma_{l,j_l} &= w_{j_l, k_l}, \\ \Gamma_{l,k_l} &= -\widehat{s}_{j_l,k_l} w_{j_l,k_l}. \end{aligned} \tag{5}$$

Then, we can write (4) as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\mathrm{argmin}}\left\{\frac{1}{n}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_{TV}\|\mathbf{\Gamma}\boldsymbol{\beta}\|_1 + \lambda_1\|\boldsymbol{\beta}\|_1\right\}. \tag{6}$$

As mentioned above, GTV promotes estimates of $\boldsymbol{\beta}$ that contain sparse clusters of coefficients, each cluster corresponding to a highly correlated set of variables. That is, the stronger the correlation between $\mathbf{x}_j$ and $\mathbf{x}_k$, the more similar $\widehat{\beta}_j$ and $\widehat{\beta}_k$. The $\|\boldsymbol{\beta}\|_1$ term promotes overall sparsity. We note that (6) can be viewed as a generalized LASSO estimator (Tibshirani and Taylor 2011), for which a number of efficient implementations exist.

b. Other regularization methods

This work sits alongside a growing body of literature of structured estimation in high dimensions with a specific application to climate data (e.g., Goncalves et al. 2017). A variety of regularization schemes have recently shown promise in improving predictive skill by imposing structure and sparsity on the predictors. Chatterjee et al. (2012) proposed using the sparse group lasso (SGL), in which the regularizer is given by

$$R(\boldsymbol{\beta}) = f\|\boldsymbol{\beta}\|_1 + (1 - f)\|\boldsymbol{\beta}\|_{1,G},$$

where $G = \{G_1, G_2, \ldots, G_M\}$ are M groups of variables across multiple locations and times. This scheme yields solutions in which variables at certain locations and times are simultaneously selected or else zeroed out. Using this scheme to predict monthly temperature and precipitation showed significant improvement over LASSO. He et al. (2019) proposed a weighted LASSO given by

$$R(\beta) = \sum_{i=1}^{p} w_i |\beta_i|,$$

where the weights are chosen to be proportional to the distance between the location of the feature and the location of its response. This penalizes predictors that are far away from the region of interest and it is not appropriate for our problem, in which long-distance climate teleconnections play an important role.

Finally, we note that GTV is a special case of the fused LASSO estimator (Tibshirani et al. 2005). While these estimators have significant theoretical support, the theory relies on the assumption that $\mathbf{X}$ is full rank and does not consider the role of correlations among columns of $\mathbf{X}$. Furthermore, the edges included in the fusion penalty are assumed to be highly structured (i.e., only direct spatial or temporal neighbors), and the theory does not generalize to the types of unstructured covariance graphs that arise in many applications. In climate and other domains, there are known long-range correlation patterns that would not be captured by a direct neighbor penalty. We will, however, benchmark GTV against the fused LASSO, which has the regularization term

$$R(\beta) = \sum_{j,k \in \mathcal{N}} |\beta_j - \beta_k| + \lambda_1 \|\beta\|_1,$$

$$\mathcal{N} := \left\{ (j,k) \mid \left(\mathbf{x}_j, \mathbf{x}_k \right) \text{ are spatially adjacent} \right\} \cup \left\{ (j,k) \mid \left(\mathbf{x}_j, \mathbf{x}_k \right) \text{ are temporally adjacent} \right\}.$$

c. Using climate model outputs to compute the covariance of the SST predictors

The theoretical guarantees of GTV depend on a sufficiently accurate estimate of the covariance matrix of the Pacific SSTs $\mathbf{\Sigma} := \mathbb{E}\left(\mathbf{X}^T \mathbf{X}\right)$. However, for our problem where $n \ll p$, the sample covariance $(1/n)\mathbf{X}^T \mathbf{X}$ is a highly uncertain estimate of $\mathbf{\Sigma}$ (Bickel and Levina 2008; Cai et al. 2016). Thus, we propose to explore the use of ensemble simulations from climate models, the size of which is several times larger than that of observations, in order to reduce the uncertainty of the covariance matrix estimate and improve the performance of the GTV regularized regression. While we acknowledge that climate models might not accurately capture all multiscale space–time variability of SSTs in the Pacific (e.g., de Szoeke and Xie 2008; Kim et al. 2014; Bellenger et al. 2014; Li and Xie 2014; Wang and Miao 2018), we assert that leveraging their information content to improve high-dimensional data-driven predictive methods offers great potential and deserves careful examination. In a recent study by Ham et al. (2019), climate model outputs were used in the context of "physics-guided initialization" [the term is adopted from Willard et al. (2020)]. Particularly, the adopted ML model was first trained using climate model outputs, so some initial estimates of the weights were obtained. Then, as a second step, prediction was performed by fine-tuning the weights using historical data (a process known as "transfer learning").

Here, we suggest that defining the GTV regularizer using covariance information from climate models can increase predictive performance. We term this approach "physics-guided regularization". We rigorously demonstrate the merits of this approach using the SST outputs from the 40 ensemble members of CESM-LENS. Since CESM-LENS simulations are produced on a different spatial grid from that of the SST observations, we interpolated late summer and early fall SSTs from CESM-LENS linearly onto the observation grid. We emphasize that the CESM-LENS outputs are used only to estimate the covariance matrix of the predictors, while the training of our model (estimation of the regression parameters and coefficients) and its performance evaluation (see section 4) are always performed using the observed SST and precipitation series in the training and test periods, respectively.

Letting $\mathbf{X}_{\mathrm{CL}} \in \mathbb{R}^{40n \times p}$ be the detrended and standardized (zero mean and unit variance) matrix of stacked SST variables from all the CESM-LENS members, we define $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ as the sample covariance of $\mathbf{X}_{\mathrm{CL}}$. We also define $\widehat{\mathbf{\Sigma}}_{\mathrm{obs}}$ as the covariance matrix estimated from the observations. These covariance matrices are $p \times p$ matrices in which all considered variables are ordered by longitude, latitude, and month, resulting in repetitive patterns arising from the spatial and temporal dependencies (see Fig. 2a for $\widehat{\mathbf{\Sigma}}_{\mathrm{obs}}$). Visually, there is no striking difference in the dependence structure of SSTs between different months (Fig. 2a, left panel). The highest correlations (both positive and negative) are found in the tropics, with strong SST couplings along the eastern and central tropical Pacific basin (high positive correlation) and between the eastern and western tropical Pacific basin (high negative correlation), features that are a consequence of the ENSO (Wang et al. 2012); for example, see the zoom-in panels in Fig. 2a for the October covariance matrix. Figure 2b shows $\widehat{\mathbf{\Sigma}}_{\mathrm{obs}}$ and $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ for the month of October. Although some differences are observed, the CESM-LENS appears to capture well the spatial structure of the observed SST correlations (i.e., tropics vs extratropics, etc.), and as demonstrated in section 4a, the reduced uncertainty of $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ adds significant predictive skill and robustness to the GTV model.

To finalize the construction of the graph underlying the GTV regularization, we further process the SST covariance matrix $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ estimated from the output of CESM-LENS, using a thresholding procedure with statistical guarantees (see Bickel and Levina 2008; Li et al. 2018). This method simply sets elements of the sample covariance with absolute value under a certain threshold equal to zero, that is, for a threshold θ, the covariance graph $\mathbf{G}$ has edges $\mathbf{E}: = \left\{ (j, k) : |\widehat{\Sigma}_{j,k}| > \theta, \quad j \neq k \right\}$. In addition to the statistical advantage of this thresholding in yielding more consistent covariance estimates, thresholding is also useful from a computational perspective, as it drastically limits the number of edges used in the regularization term [i.e., the number of rows in $\mathbf{\Gamma}$ in Eq. (6)]. We treat the threshold θ as a model parameter and estimate its optimal value in a cross-validation training setting (i.e., similarly to λ_1 and λ_{TV}; see section 3a), which allows us to disregard the smaller, less certain SST correlation values that, if included, would have led to a worse performance (i.e., if $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ were not informative at all, the optimal θ would be close to one).

d. Accounting for nonstationarity in precipitation teleconnections

The last issue that our analysis aims to account for is possible nonstationarities in the strength of the precipitation teleconnections. Traditionally, precipitation in the SWUS has been linked to various large-scale modes of climate variability, and more commonly the state of ENSO (Schonher and Nicholson 1989; Redmond and Koch 1991; Mo and Higgins 1998; McCabe and Dettinger 1999; Cayan et al. 1999). Physically, El Niño (or La Niña) events typically associate with persistent low (or high) atmospheric pressure patterns over the northeastern Pacific (a teleconnection that materializes via quasi-stationary Rossby waves; Trenberth et al. 1998; Castello and Shelton 2004), and thus disturb the location and strength of the wintertime jet stream, which can then bring more (or fewer) winter storms to the SWUS, leading to wet (or dry) conditions over the SWUS and dry (or wet) conditions over the northwestern United States. However, recent research shows that the ENSO effect on the atmospheric pressure and (consequently) on precipitation over the eastern Pacific and North America has been decreasing in strength during the last 3–4 decades, while many studies have highlighted to a greater or lesser extent that the western Pacific climatic state (e.g., SSTs) has been a stronger driver of precipitation variability over North America (Wang et al. 2014; Baxter and Nigam 2015; Teng and Branstator 2017; Seager et al. 2017; Swain et al. 2017; Myoung et al. 2018; Mamalakis et al. 2018). On a similar note, new research (Johnson et al. 2019) shows that during the last 3–4 decades, western Pacific SSTs have been important players in affecting the connection between the tropical atmospheric circulation and the eastern tropical Pacific SSTs, during weak ENSO events, which highlights changes in the tropical Pacific dynamics (see also Mamalakis et al. 2019). Whether these changes in precipitation teleconnections and Pacific dynamics are a result of internal multidecadal climate variability or anthropogenic forcing is still not clear. However, to acknowledge the nonstationary nature of the prediction problem, we herein use a weighted loss function that gives more weight to the more recent years in the training dataset (i.e., the period after the 1970s and 1980s). This is roughly the time that most studies have pinpointed as the start of these changes (Wang et al. 2014; Swain et al. 2016; Mamalakis et al. 2018; Johnson et al. 2019), and it is also the period during which the SWUS precipitation variability (interannual variance) has started to increase (see Fig. 1). As such, with regard to the data-fit term of Eq. (6), we minimize the weighted loss function $\sum_{t=1940}^{1989}\left\{a^{1989-t}\left(y^{(t)}-\mathbf{x}^{(t)}\cdot\boldsymbol{\beta}\right)^2\right\}$, with a being a discount factor set to $a = 0.90$, and $\mathbf{x}^{(t)}\cdot\boldsymbol{\beta}$ indicating the inner product. This simple but effective approach is widely used in the forecasting literature (e.g., Hyndman and Athanasopoulos 2018) and gives preference to the relationship between Pacific SSTs and SWUS precipitation in the more recent decades, while still retaining some information from earlier years.

4. Results

In Fig. 3, we summarize schematically the proposed approach based on the GTV described in the previous sections, for the prediction of winter precipitation totals in SWUS. We inform our prediction using observed Pacific SSTs during the boreal late summer and early fall (July–October) and we form a space–time covariance graph with edge weights corresponding to pairwise correlations (normalized covariances) between SST boxes of 10° × 10°, to constrain our regularization scheme, in addition to the traditionally used LASSO

term. Correlations are obtained from the output of climate models (i.e., using SST outputs from 40 ensemble members of the CESM-LENS, for the period 1940–2005) to decrease estimation uncertainty and improve prediction in the test period.

a. Predictive performance of the GTV model

The GTV model [Eq. (6)] was fitted in the training period (from 1940/41 to 1989/90) and the optimal threshold value θ^* and optimal parameter values $\left(\lambda_1^*, \lambda_{\mathrm{TV}}^*\right)$ were estimated through a fivefold cross-validation procedure (see section 3a). For the case of the areal average precipitation over SWUS, this procedure identified the values $\theta^* = 0.5$ and $\left(\lambda_1^*, \lambda_{\mathrm{TV}}^*\right) = (0.013, 0.0007)$. To test the sensitivity and robustness of the GTV model to this optimal choice of parameters, and also to showcase the advantage of using the CESM-LENS versus the covariance of observations, we start by presenting and discussing results for three different values $\theta = (0.35, 0.5, 0.75)$ and various $(\lambda_1, \lambda_{\mathrm{TV}})$ combinations. Figure 4 shows the October SSTs covariance for those different θ thresholds, highlighting the sparseness of the covariance matrix as the threshold increases. Although not shown in Fig. 4, the dependency graph formed by the thresholded covariances has, as expected, a decreasing number of links (it is sparser) as θ increases. Specifically, the number of links of the GTV graph is 80 503, 32 644, and 5357 for $\theta = 0.35$, 0.5, and 0.75, respectively, highlighting the computational advantages that the reduced-degree graph also offers.

The center-column panels in Fig. 4 show the model performance in the test period, measured by the coefficient of determination R^2, as a function of the different combinations of $(\lambda_1, \lambda_{\mathrm{TV}})$ parameters. On the same panels, the optimal set of parameters $\left(\lambda_1^*, \lambda_{\mathrm{TV}}^*\right)$ obtained from the fivefold cross-validation in the training period, conditional on the three values of θ, is also shown. It is observed that for $\theta^* = 0.5$, the optimal parameters $\left(\lambda_1^*, \lambda_{\mathrm{TV}}^*\right)$ robustly fall within the region for which the model performance in the test period is also optimal. This illustrates that optimally thresholding the covariance to reduce the spatiotemporally correlated predictors used in the regularization avoids overfitting and increases model accuracy. Moreover, our results show that the GTV model explains about $R^2 = 40\%$ of precipitation variance in the test period; even for parameter values other than the optimal $\left(\theta^*, \lambda_1^*, \lambda_{\mathrm{TV}}^*\right)$, R^2 is consistently higher than 30%. This is a significant improvement over the R^2 values obtained in prior work, since commonly used teleconnection indices typically result in a much lower fraction of explained variance, on the order of 10%–20% (see Lee et al. 2018; Deser et al. 2018; see also Fig. 6 herein). The latter indicates that informing the GTV regularizer based on the covariance matrix $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ robustly improves the prediction.

To highlight further the merit of using the climate model covariance $\widehat{\mathbf{\Sigma}}_{\mathrm{CL}}$ versus the covariance of the observations $\widehat{\mathbf{\Sigma}}_{\mathrm{obs}}$, the rightmost column panels in Fig. 4 present the same analysis as the middle column panels, but using $\widehat{\mathbf{\Sigma}}_{\mathrm{obs}}$ instead. It is telling that the performance in the test period in this case is inferior (smaller R^2 values) for all combinations of parameters $(\lambda_1, \lambda_{\mathrm{TV}})$ and threshold values θ.

Having established the robustness of the GTV model using $\widehat{\boldsymbol{\Sigma}}_{\mathrm{CL}}$ with the optimal parameters $\theta^* = 0.5$ and $(\lambda_1^*, \lambda_{\mathrm{TV}}^*) = (0.013, 0.0007)$, we compare in Fig. 5 the predicted and observed precipitation series for the years from 1990/91 to 2018/19. The predicted series explains about 42% of the precipitation variability and captures adequately many of the extreme precipitation years (i.e., the wet years 1992/93, 1994/95, 1997/98, and 2004/05 and the dry years 1998/99, 2001/02, and 2006/07. Also, the probability of a dry (or wet) hit is high (high chance in predicting dry, when actually dry conditions occur, and likewise for wet). Specifically, if we define a wet (dry) year to be a year that falls above (below) the multiyear precipitation average, then our results indicate that our method exhibits a wet hit probability of 64% and dry hit probability of 72%. Moreover, the residuals between the prediction and observations are found to be normally distributed and exhibit insignificant autocorrelation at a 0.05 significance level (see Figs. 5b,c), consistent with the "white noise" assumption in Eqs. (1) and (2).

Finally, the residuals of the prediction do not show statistically significant correlation with the Pacific SSTs (Fig. 5d), indicating that no information in the Pacific SSTs is left unexplored in the predictive model. Only a few small and incoherent SST patterns are found to significantly correlate to precipitation at a 0.05 significance level, probably due the fact that in Fig. 5d we simultaneously test multiple "local" hypotheses, which increases the chances of a type I error (i.e., rejecting a true null hypothesis; Wilks 2016). Thus, we can conclude that our model sufficiently exploits the Pacific SST information, and that any deviation (residual) between our predictions and reality either comes from other forcings not included in our analysis (e.g., SST variability over the Atlantic Ocean; Enfield et al. 2001) or is the result of internal stochastic variability.

b. Benchmarking against other predictive models

In this section, we compare the prediction skill of the GTV, for all climate divisions over the SWUS and the areal average precipitation, to other regularized regression models and models based on commonly used teleconnections (Fig. 6). Specifically, we benchmark our results against the following methods:

- LASSO: standard 'ℓ_1-penalized regression; coefficients are penalized so that the solution is very sparse.

- Fused LASSO: Direct spatial and temporal neighbors are penalized to have similar coefficients.

- GTV (Obs): GTV with the regularization term defined using the covariance matrix estimated from the observations and thresholded at $\theta^* = 0.5$.

- GTV (CESM-LENS): GTV with the regularization term defined using the covariance matrix estimated from CESM-LENS and thresholded at $\theta^* = 0.5$.

- Ordinary least squares using known teleconnection indices.

We highlight that all methods are trained in years from 1940/41 to 1989/90 using the weighting formula described in section 3d to account for nonstationarity and tested in years from 1990/91 to 2018/19. Only observations are used for training and testing; the CESM-

LENS output is used simply to define the regularization term in the GTV (CESM-LENS), but not to actually fit or test the model.

First, we find that the prediction accuracy differs significantly among climate divisions of the studied region. Most notably, prediction of northern climate divisions in California and Utah, specifically, CA(4), CA(5), UT(1), UT(6), and UT(7), is poorer relatively to climate divisions over most of Arizona (AZ). The fact that this holds for all models indicates that the signal of the Pacific SSTs to precipitation is weaker as one moves to northern California, Nevada, and Utah, which is in accordance with other studies (see Schonher and Nicholson 1989; McCabe and Dettinger 1999; Castello and Shelton 2004; Mamalakis et al. 2018; see also our discussion in section 2). With regard to the best performing model, our results show that the proposed GTV model reproduces the highest fraction of precipitation variability over almost all climate divisions, ranging from almost $R^2 = 0.5$ in AZ(7) to $R^2 = 0.1$ in UT(1), and $R^2 = 0.42$ for the areal average precipitation, when using $\widehat{\mathbf{\Sigma}}_{CL}$. When using $\widehat{\mathbf{\Sigma}}_{obs}$, the performance is poorer and similar to the performance of LASSO and fused LASSO, in terms of the areal average precipitation. Fused LASSO performs slightly better than GTV in AZ(3), but it only slightly exceeds $R^2 = 0.1$ for the areal average precipitation.

As a benchmark, we also compare the prediction performance of GTV with known physical teleconnections. Specifically, we train a weighted (see section 3d) linear regression scheme using the averaged July–October Niño-3.4 index as our predictor, which captures ENSO variability and is typically associated with SWUS precipitation. We also use a weighted bivariate regression model combining the Niño-3.4 index and the New Zealand index (NZI) over the same summer months. The NZI has been shown to exhibit high correlation with precipitation over the last four decades (Mamalakis et al. 2018). The latter interhemispheric teleconnection has been suggested to materialize through a western Pacific ocean–atmosphere pathway, whereby SST anomalies in the southwestern Pacific during late boreal summer can modulate time-lagged anomalies of the same sign in the northwestern and central Pacific via perturbation of the regional southern Hadley cell, which in turn affect the jet stream and winter storm tracks to the U.S. west coast. Our results show that ENSO-based predictions explain about 10%–15% of the precipitation variability over most climate divisions. When NZI is added, the prediction performance increases significantly and the explained variance is almost twice as high for the areal average SWUS precipitation. However, in almost all climate divisions, the GTV(CESM-LENS) model outperforms all other models. Similar conclusions are reached also based on the mean square error (see Table 1), where the GTV model is *not* the best performing in only three climate divisions out of the 18 [i.e., in AZ(3), AZ(4), and AZ(7)].

Generally, the results described above, and summarized in Table 1 and Fig. 6, show that GTV(CESM-LENS) robustly outperforms the competing regularized regression schemes and known teleconnections, offering promise for increasing the predictive skill of winter precipitation over the SWUS.

c. Physical interpretation of the predictors

In this section, we seek insight into which SST patterns play an important role in driving winter precipitation variability over the SWUS. In doing so, we seek physical interpretations of the "optimal" solutions of regression coefficients corresponding to each regularization method. We repeat here that we estimate $\hat{\boldsymbol{\beta}}$ by (i) applying a fivefold cross-validation technique to the training data to estimate the regularization parameters of each model, and (ii) minimizing the corresponding loss function using the estimated regularization parameters from (i). Although this analysis is not suitable to draw rigorous causal inferences, it can highlight important sources of predictability for precipitation, which should be physically interpretable.

The optimal weights $\hat{\boldsymbol{\beta}}$ for the LASSO model are presented in Fig. 7a. Keeping in mind that the LASSO regularization promotes sparsity, this method essentially pinpoints the few regions around the Pacific basin, over which late summer and early fall SSTs contained the highest predictive information for November–March precipitation, during the training period from 1940/41 to 1989/90. Specifically, negative regression coefficients on the order of −0.3 are found over the tropical and subtropical western Pacific, and positive coefficients of the same order are found over the Southern Hemisphere midlatitudes. Fused LASSO (Fig. 7b), which promotes the assignment of similar weights to neighboring grid boxes, yields a smoother version of the LASSO solution, in which the majority of Pacific participates in the prediction, but many regions (grid boxes) contribute in a negligible way (many coefficients are on the order of 10^{-3}–10^{-4}). There is some contribution by the northern midlatitude SSTs, but the highest weights are again assigned to the tropical and subtropical western Pacific (negative coefficients), and to the Southern Hemisphere midlatitudes, especially over the southeastern Pacific (positive coefficients). Last, slightly different solutions are obtained by the GTV model when using $\widehat{\boldsymbol{\Sigma}}_{\text{obs}}$ (Fig. 7c) or $\widehat{\boldsymbol{\Sigma}}_{\text{CL}}$ (Fig. 7d). However, in terms of the patterns of the SST predictors (i.e., not in terms of grid by grid comparison), there is some consistency between these two variants over the southwestern Pacific Ocean (where both models exhibit high negative coefficients), and the northeastern and central Pacific basin (positive coefficients).

Although the GTV model (when using $\widehat{\boldsymbol{\Sigma}}_{\text{CL}}$) gives the best prediction performance in the test period, for our physical interpretation we focus on SST patterns that are consistent across all methods. Specifically, in accordance with recent studies (Wang et al. 2014; Seager et al. 2017; Swain et al. 2017; Myoung et al. 2018; Mamalakis et al. 2018), all models highlight to a greater or lesser extent the western Pacific SSTs as important predictors of SWUS precipitation, rather than the eastern Pacific SSTs. Physically, it has been shown that the western tropical Pacific is a region over which anomalous convection can be an important source of Rossby wave energy, which teleconnects through a quasi-stationary Rossby wave train with the atmospheric pressure over the northeastern Pacific, affecting the location of the jet stream, and eventually precipitation totals in North America (Wang et al. 2014). Moreover, the southwestern Pacific (close to New Zealand) has been highlighted in the literature as a special region in leading tropical climate. First, some studies support that climate variability (e.g., SST, sea level pressure, etc.) over the southwestern Pacific leads by a few seasons the ENSO variability (Trenberth and Shea 1987; van Loon and Shea 1987;

Stephens et al. 2007), and specific indices have been suggested to increase predictive skill of ENSO state (Hamlington et al. 2015). Given that ENSO is known to be related with SWUS precipitation during winter (i.e., for zero lead time), the southwestern Pacific SSTs may provide important predictors of precipitation, by leading the ENSO state, and are highlighted by all models in our analysis. By contrast, eastern tropical Pacific SSTs are not shown to be predictive, since our analysis considers nonzero lead times. More recent studies, however, suggest that western Pacific SSTs can also affect precipitation through a western Pacific pathway (i.e., not necessarily through ENSO teleconnections) (Mamalakis et al. 2018). The latter has been suggested to materialize through the seasonal migration of the intertropical convergence zone and the associated expansion of the southern Hadley cell during late summer (Waliser and Gautier 1993; Berry and Reeder 2014; Mamalakis and Foufoula-Georgiou 2018), which allows for persistent SST anomalies to impact the atmospheric circulation and climate variability in the western tropical Pacific, which as noted earlier is a key region of Rossby wave energy. This teleconnection has been increasing in importance during the last 40 years, which is also the time when new, ENSO-independent SST patterns have been emerging and affecting tropical atmospheric circulation (Johnson et al. 2019).

d. Sensitivity of the GTV model to uncertainty in the covariance matrix

Finally, to explore the sensitivity of the GTV model to perturbations of the covariance matrix used to define the regularization term, we perform a bootstrap analysis. Namely, rather than stacking all 40 CESM-LENS trajectories to form the covariance matrix, we resample the 40 trajectories (with replacement) and compute the sample covariance of the new sample. Next, we form our GTV regularization term using this resampled covariance matrix, then fit the GTV scheme in the training period, and finally calculate the coefficient of determination (R^2) in the test period. By repeating this procedure 1000 times, we can quantify how the uncertainty in the covariance matrix propagates to uncertainty in the regression coefficients $\hat{\beta}$ and model performance.

Our results show that the GTV model always captures more than $R^2 = 30\%$ of the variability of the November–March areal-average precipitation, in some cases reaching $R^2 = 45\%$. The bootstrap average is on the order of $R^2 = 40\%$ and the bootstrap standard deviation is about 5% (see Fig. 8a). These results indicate that the GTV model is not particularly sensitive to the uncertainty in the covariance matrix and always outperforms all alternative predictive models.

Regarding the propagation of uncertainty to the regression coefficients, the average vector of $\hat{\beta}$ across the 1000 bootstrap realizations (see Fig. 8b) very closely resembles the results presented in Fig. 7d, indicating the importance of the southwestern Pacific Ocean (high negative coefficients), and the northeastern and central Pacific basin (positive coefficients). Moreover, although in some grid points the standard deviation of $\hat{\beta}$ across the 1000 bootstrap realizations is of the same magnitude as the average value, most of the largest coefficients in Fig. 8b are characterized by small underlying uncertainty in Fig. 8c, which implies that they are not sensitive to covariance perturbations as quantified here. This provides confidence that these SST features mostly located in the western Pacific are indeed important sources of predictability of the November–March SWUS precipitation.

5. Conclusions and future work

In this study, we approached the problem of early prediction of winter precipitation over the SWUS by using machine learning methodologies to increase predictive skill relative to traditional approaches of utilizing dynamical models or relying on empirically established teleconnections. We use late summer and early fall SST information to predict precipitation based on a newly proposed regularized regression scheme, specifically designed to account for high dimensionality and high spatiotemporal dependence structure in the predictor variables, making it well suited to climate applications. The proposed predictive model accounts for high spatiotemporal dependence structures in the predictors expressed as a graph, which is then used to define a graph total variation (GTV) regularizer that promotes similar weights for highly correlated predictors. We also address the short observational record and high dimensionality of the problem by using LASSO terms that promote sparsity, as well as by using large-ensemble outputs from climate models to decrease the structural uncertainty in the estimation of the SST covariance matrix.

Our analysis shows that predictive skill for SWUS precipitation can be increased considerably by using our novel regularization methodology, explaining more than 40% of the average precipitation variability over the SWUS. Our model's performance is higher than any other regularized regression model (LASSO and fused LASSO), and it also outperforms models based on known teleconnection indices. Our results also show that, in accordance with recent literature (DelSole and Banerjee 2017; Ham et al. 2019), climate models can be used in a nonconventional way (e.g., for training rather than predicting and, in our case, for building the graph-based regularizer) toward increasing prediction accuracy. With regard to important regions/sources of precipitation predictability, our analysis highlights the tropical and subtropical western Pacific SSTs as the most consistently important predictors of precipitation, which have increasingly gained attention in the literature (Wang et al. 2014; Swain et al. 2017; Mamalakis et al. 2018). Finally, based on a bootstrap analysis, we show that the proposed model is robust to perturbations in the covariance matrix used to form the GTV regularization term.

The results presented herein suggest some further questions and challenges with regard to the exigent task of seasonal SWUS precipitation prediction. For example, future work should address the intricate nonstationarity of the climate system more explicitly by allowing the regression coefficients to vary with time (Livneh and Badger 2020). This property might be especially important as precipitation variability in California is expected to increase even more under climate change (Swain et al. 2018). It should also address quantification of the underlying uncertainty of the regression coefficients (beyond the uncertainty of the covariance matrix explored herein), which can be translated into confidence intervals of the predicted precipitation. Last, our approach can be extended by using global information from additional climate variables (e.g., ocean heat content, atmospheric pressure, etc.) and using climate model outputs from different projects, like the North American Multimodel Ensemble (Kirtman et al. 2014), phase 6 of the Coupled Model Intercomparison Project (Eyring et al. 2016), or the Decadal Prediction Large Ensemble project (Yeager et al. 2018).

In conclusion, while more complex nonlinear models, such as deep neural networks, have been gaining popularity in modeling of climate data, our work shows that for high-dimensional problems with limited historical records sparse linear models with informative regularization can play an important role in building predictive models for climate variability. This is consistent with a recent review paper that focused on seasonal to subseasonal prediction of climate variables over the entire United States (He et al. 2020) and that highlighted the success of regularized regression models (such as simple LASSO; see also DelSole and Banerjee 2017). In addition, an important advantage of sparse linear models in this context is that they are considerably easier to interpret from a physical perspective, compared to nonlinear models such as deep neural networks. Our results suggest that a promising direction for future research is the development of new models that can incorporate relevant physical knowledge (e.g., from large ensemble simulations of climate models), that can retain the interpretability of sparse linear models, and that have the flexibility to improve the accuracy of current models for seasonal and subseasonal precipitation prediction.

Acknowledgments.

REFERENCES

AghaKouchak A, Feldman D, Hoerling M, Huxman T, and Lund J, 2015: Water and climate: Recognize anthropogenic drought. Nature, 524, 409–411, 10.1038/524409a. [PubMed: 26310750]

Allen R, and Luptowitz R, 2017: El Niño–like teleconnection increases California precipitation in response to warming. Nat. Commun, 8, 16055, 10.1038/ncomms16055. [PubMed: 28681837]

Baxter S, and Nigam S, 2015: Key role of the North Pacific Oscillation–west Pacific pattern in generating the extreme 2013/14 North American winter. J. Climate, 28, 8109–8117, 10.1175/JCLI-D-14-00726.1.

Bellenger H, Guilyardi E, Leloup J, Lengaigne M, and Vialard J, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. Climate Dyn., 42, 1999–2018, 10.1007/s00382-013-1783-z.

Berry G, and Reeder MJ, 2014: Objective identification of the intertropical convergence zone: Climatology and trends from the ERA-Interim. J. Climate, 27, 1894–1909, 10.1175/JCLI-D-13-00339.1.

Bickel PJ, and Levina E, 2008: Covariance regularization by thresholding. Ann. Stat, 36, 2577–2604, 10.1214/08-AOS600.

Bradley RS, Diaz HF, Kiladis GN, and Eischied JK, 1987: ENSO signal in continental temperature and precipitation records. Nature, 327, 497–501, 10.1038/327497a0.

Cai TT, Zhao R, and Zhou HH, 2016: Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. Electron. J. Stat, 10, 1–59, 10.1214/15-EJS1081.

Castello AF, and Shelton ML, 2004: Winter precipitation on the US Pacific coast and El Niño–southern oscillation events. Int. J. Climatol, 24, 481–497, 10.1002/joc.1011.

Cayan DR, Redmond KT, and Riddle LG, 1999: ENSO and hydrologic extremes in the western United States. J. Climate, 12, 2881–2893, 10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2.

Chang Y, Schubert SD, and Suarez MJ, 2000: Boreal winter predictions with the GEOS-2 GCM: The role of boundary forcing and initial conditions. Quart. J. Roy. Meteor. Soc, 126, 2293–2321, 10.1256/smsqj.56714.

Chatterjee S, Steinhaeuser K, Banerjee A, Chatterjee S, and Ganguly A, 2012: Sparse group lasso: Consistency and climate applications. Proc. 2012 SIAM Int. Conf. on Data Mining, SIAM, 47–58, 10.1137/1.9781611972825.5.

Cheng L, and AghaKouchak A, 2015: A methodology for deriving ensemble response from multimodel simulations. J. Hydrol, 522, 49–57, 10.1016/j.jhydrol.2014.12.025.

Coelho CAS, Pezzulli S, Balmaseda M, Doblas-Reyes FJ, and Stephenson DB, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. J. Climate, 17, 1504–1516, 10.1175/1520-0442(2004)017,1504:FCACAS.2.0.CO;2.

Dai A, 2013: The influence of the inter-decadal Pacific oscillation on US precipitation during 1923–2010. Climate Dyn, 41, 633–646, 10.1007/s00382-012-1446-5.

DelSole T, and Banerjee A, 2017: Statistical seasonal prediction based on regularized regression. J. Climate, 30, 1345–1361, 10.1175/JCLI-D-16-0249.1.

Deser C, Alexander MA, Xie SP, and Phillips AS, 2010: Sea surface temperature variability: Patterns and mechanisms. Annu. Rev. Mar. Sci, 2, 115–143, 10.1146/annurev-marine-120408-151453.

——, Simpson IR, Phillips AS, and McKinnon KA, 2018: How well do we know ENSO's climate impacts over North America, and how do we evaluate models accordingly? J. Climate, 31, 4991–5014, 10.1175/JCLI-D-17-0783.1.

de Szoeke SP, and Xie S, 2008: The tropical eastern Pacific seasonal cycle: Assessment of errors and mechanisms in IPCC AR4 coupled ocean–atmosphere general circulation models. J. Climate, 21, 2573–2590, 10.1175/2007JCLI1975.1.

Dettinger M, and Cayan D, 2014: Drought and the California Delta—A matter of extremes. San Franc. Estuary Watershed Sci, 12 (2), 10.15447/SFEWS.2014V12ISS2ART4.

——, Ralph FM, Das T, Neiman PJ, and Cayan D, 2011: Atmospheric rivers, floods, and the water resources of California. Water, 3, 445–478, 10.3390/w3020445.

Ebtehaj AM, and Foufoula-Georgiou E, 2013: On variational downscaling, fusion, and assimilation of hydrometeorological states: A unified framework via regularization. Water Resour. Res, 49, 5944–5963, 10.1002/wrcr.20424.

——, ——, and Lerman G, 2012: Sparse regularization for precipitation downscaling. J. Geophys. Res, 117, D08107, 10.1029/2011JD017057.

Enfield DB, Mestas-Nuñez AM, and Trimble PJ, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. Geophys. Res. Lett, 28, 2077–2080, 10.1029/2000GL012745.

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, and Taylor KE, 2016: An overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev, 9, 1937–1958, 10.5194/gmd-9-1937-2016.

Gershunov A, and Cayan DR, 2003: Heavy daily precipitation frequency over the contiguous United States: Sources of climatic variability and seasonal predictability. J. Climate, 16, 2752–2765, 10.1175/1520-0442(2003)016<2752:HDPFOT>2.0.CO;2.

Goncalves AR, Banerjee A, Sivakumar V, and Chatterjee S, 2017: Structured estimation in high dimensions: Applications in climate. Large-Scale Machine Learning in the Earth Sciences, CRC Press, 13–32.

Ham Y-G, Kim J-H, and Luo J-J, 2019: Deep learning for multi-year ENSO forecasts. Nature, 573, 568–572, 10.1038/s41586-019-1559-7. [PubMed: 31534218]

Hamlington BD, Milliff RF, van Loon H, and Kim K-Y, 2015: A Southern Hemisphere sea level pressure-based precursor for ENSO warm and cold events. J. Geophys. Res. Atmos, 120, 2280–2292, 10.1002/2014JD022674.

Hao Z, Singh VP, and Xia Y, 2018: Seasonal drought prediction: Advances, challenges, and future prospects. Rev. Geophys, 56, 108–141, 10.1002/2016RG000549.

He S, Li X, Sivakumar V, and Banerjee A, 2019: Interpretable predictive modeling for climate variables with weighted lasso. 33rd AAAI Conf. on Artificial Intelligence, New York, NY, AAAI, 1385–1392, 10.1609/aaai.v33i01.33011385.

——, ——, DelSole T, Ravikumar P, Banerjee A 2020: Subseasonal climate forecasting via machine learning: Challenges, analysis, and advances. https://arxiv.org/abs/2006.07972.

Hewitt J, Hoeting JA, Done JM, and Towler E, 2018: Remote effects spatial process models for modeling teleconnections. Environmetrics, 29, e2523, 10.1002/env.2523.

Hirahara S, Ishii M, and Fukuda Y, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. J. Climate, 27, 57–75, 10.1175/JCLI-D-12-00837.1.

Howitt RE, Medellín-Azuara J, MacEwan D, Lund JR, Sumner DA, 2014: Economic analysis of the 2014 drought for California agriculture. Tech. Rep, University of California, 28 pp., https://watershed.ucdavis.edu/2014-drought-report.

——, MacEwan D, Medellín-Azuara J, Lund JR, and Sumner DA, 2015: Economic analysis of the 2015 drought for California agriculture. Tech. Rep, Center for Watershed Sciences, University of California, 16 pp., https://watershed.ucdavis.edu/files/content/news/Economic_Impact_of_the_2014_California_Water_Drought.pdf.

Hyndman RJ, and Athanasopoulos G, 2018: Forecasting: Principles and Practice. 2nd ed. Accessed 14 July 2020, OTexts.com/fpp2.

Johnson NC, L'Heureux ML, Chang C-H, and Hu Z-Z, 2019: On the delayed coupling between ocean and atmosphere in recent weak El Niño episodes. Geophys. Res. Lett, 46, 11 416–11 425, 10.1029/2019GL084021.

Kay JE, and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project. Bull. Amer. Meteor. Soc, 96, 1333–1349, 10.1175/BAMS-D-13-00255.1.

Kim ST, Cai W, Jin F-F, and Yu J-Y, 2014: ENSO stability in coupled climate models and its association with mean state. Climate Dyn, 42, 3313–3321, 10.1007/s00382-013-1833-6.

Kirtman BP, and Coauthors, 2014: The North American Multimodel Ensemble (NMME): Phase-1 seasonal-to-inter-annual prediction; phase-2 toward developing intraseasonal prediction. Bull. Amer. Meteor. Soc, 95, 585–601, 10.1175/BAMS-D-12-00050.1.

Lee S-K, Lopez H, Chung E-S, DiNezio P, Yeh S-W, and Wittenberg AT, 2018: On the fragile relationship between El Niño and California rainfall. Geophys. Res. Lett, 45, 907–915, 10.1002/2017GL076197.

Li G, and Xie S, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. J. Climate, 27, 1765–1780, 10.1175/JCLI-D-13-00337.1.

Li Y, Mark B, Rasutti G, and Willett R, 2018: Graph-based regularization for regression problems with highly-correlated designs. https://arxiv.org/abs/1803.07658.

Liu T, Schmitt RW, and Li L, 2018: Global search for autumnlead sea surface salinity predictors of winter precipitation in southwestern United States. Geophys. Res. Lett, 45, 8445–8454, 10.1029/2018GL079293.

Livneh B, and Badger AM, 2020: Drought less predictable under declining future snowpack. Nat. Climate Change, 10, 452–458, 10.1038/s41558-020-0754-8.

Luo L, Wood EF, and Pan M, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. J. Geophys. Res, 112, D10102, 10.1029/2006JD007655.

Madadgar S, AghaKouchak A, Shukla S, Wood AW, Cheng L, Hsu K-L, and Svoboda M, 2016: A hybrid statistical-dynamical framework for meteorological drought prediction: Application to southwestern United States. Water Resour. Res, 52, 5095–5110, 10.1002/2015WR018547.

Mamalakis A, and Foufoula-Georgiou E, 2018: A multivariate probabilistic framework for tracking the intertropical convergence zone: Analysis of recent climatology and past trends. Geophys. Res. Lett, 45, 13 080–13 089, 10.1029/2018GL079865.

——, Yu J-Y, Randerson JT, AghaKouchak A, and Foufoula-Georgiou E, 2018: A new interhemispheric teleconnection increases predictability of winter precipitation in southwestern US. Nat. Commun, 9, 2332, 10.1038/s41467-018-04722-7. [PubMed: 29899420]

——, ——, ——, ——, and ——, 2019: Reply: A critical examination of a newly proposed interhemispheric teleconnection to southwestern US winter precipitation. Nat. Commun, 10, 2918, 10.1038/s41467-019-10531-3. [PubMed: 31249290]

McCabe GJ, and Dettinger MD, 1999: Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. Int. J. Climatol, 19, 1399–1410, 10.1002/(SICI)1097-0088(19991115)19:13,1399::AID-JOC457.3.0.CO;2-A.

——, Palecki MA, and Betancourt JL, 2004: Pacific and Atlantic Ocean influences on multidecadal drought frequence in the United States. Proc. Natl. Acad. Sci. USA, 101, 4136–4141, 10.1073/pnas.0306738101. [PubMed: 15016919]

Medellín-Azuara JM, MacEwan D, Howitt RE, Sumner DA, and Lund JR, 2016: Economic analysis of the 2016 California drought on agriculture. Tech. Rep Center for Watershed Sciences, University of California, https://watershed.ucdavis.edu/files/DroughtReport20160812.pdf.

Mo KC, and Higgins RW, 1998: Tropical influences on California precipitation. J. Climate, 11, 412–430, 10.1175/1520-0442(1998)011<0412:TIOCP>2.0.CO;2.

Mote PW, Hamlet AF, Clark MP, and Lettenmaier DP, 2005: Declining mountain snowpack in western North America. Bull. Amer. Meteor. Soc, 86, 39–50, 10.1175/BAMS-86-1-39.

Myoung B, Yeh S-W, Kim J, and Kafatos MC, 2018: Impacts of Pacific SSTs on atmospheric circulations leading to California winter precipitation variability: A diagnostic modeling. Atmosphere, 9, 455, 10.3390/atmos9110455.

National Academies of Sciences, Engineering, and Medicine, 2016: Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. The National Academies Press, 336 pp., 10.17226/21873.

Pan B, Hsu K, AghaKouchak A, Sorooshian S, and Higgins W, 2019: Precipitation prediction skill for the west coast United States: From short to extended range. J. Climate, 32, 161–182, 10.1175/JCLI-D-18-0355.1.

Peng Z, Wang QJ, Bennett JC, Pokhrel P, and Wang Z, 2014: Seasonal precipitation forecasts over China using monthly large-scale oceanic–atmospheric indices. J. Hydrol, 519, 792–802, 10.1016/j.jhydrol.2014.08.012.

Quan X, Hoerling M, Whitaker J, Bates G, and Xu T, 2006: Diagnosing sources of U.S. seasonal forecast skill. J. Climate, 19, 3279–3293, 10.1175/JCLI3789.1.

Raftery AE, Gneiting T, Balabdaoui F, and Polakowski M, 2005: Using Bayesian model averaging to calibrate forecast ensembles. Mon. Wea. Rev, 133, 1155–1174, 10.1175/MWR2906.1.

Redmond KT, and Koch RW, 1991: Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices. Water Resour. Res, 27, 2381–2399, 10.1029/91WR00690.

Ropelewski CF, and Halpert MS, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). Mon. Wea. Rev, 114, 2352–2362, 10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.

Schepen A, and Wang QJ, 2013: Towards accurate and reliable forecasts of Australian seasonal rainfall by calibrating and merging multiple coupled GCMs. Mon. Wea. Rev, 141, 4554–4563, 10.1175/MWR-D-12-00253.1.

——, ——, and Robertson D, 2012: Evidence for using lagged climate indices to forecast Australian seasonal rainfall. J. Climate, 25, 1230–1246, 10.1175/JCLI-D-11-00156.1.

——, ——, and ——, 2014: Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. Mon. Wea. Rev, 142, 1758–1770, 10.1175/MWR-D-13-00248.1.

Schonher T, and Nicholson SE, 1989: The relationship between rainfall and ENSO events. J. Climate, 2, 1258–1269, 10.1175/1520-0442(1989)002<1258:TRBCRA>2.0.CO;2.

Schubert S, Chang Y, Wang H, Koster R, and Suarez M, 2016: A modeling study of the causes and predictability of the spring 2011 extreme US weather activity. J. Climate, 29, 7869–7887, 10.1175/JCLI-D-15-0673.1.

Seager R, Henderson N, Cane MA, Liu H, and Nakamura J, 2017: Is there a role for human-induced climate change in the precipitation decline that drove the California drought? J. Climate, 30, 10 237–10 258, 10.1175/JCLI-D-17-0192.1.

Shukla S, Steinemann A, Iacobellis SF, and Cayan DR, 2015: Annual drought in California: Association with monthly precipitation and climate phases. J. Appl. Meteor. Climatol, 54, 2273–2281, 10.1175/JAMC-D-15-0167.1.

Stephens DJ, Meuleners MJ, van Loon H, Lamond MH, and Telcik NP, 2007: Differences in atmospheric circulation between the development of weak and strong warm events in the Southern Oscillation. J. Climate, 20, 2191–2209, 10.1175/JCLI4131.1.

Swain DL, Horton DE, Singh D, and Diffenbaugh NS, 2016: Trends in atmospheric patterns conducive to seasonal precipitation and temperature extremes in California. Sci. Adv, 2, e1501344, 10.1126/sciadv.1501344. [PubMed: 27051876]

——, Singh D, Horton DE, Mankin JS, Ballard TC, Diffenbaugh NS, 2017: Remote linkages to anomalous winter atmospheric ridging over the northeastern Pacific. J. Geophys. Res. Atmos, 122, 12 194–12 209, 10.1002/2017JD026575.

——, Langenbrunner B, Neelin JD, and Hall A, 2018: Increasing precipitation volatility in twenty-first-century California. Nat. Climate Change, 8, 427–433, 10.1038/s41558-018-0140-y.

Tao L, He X, and Wang R, 2017: A hybrid LSSVM model with empirical mode decomposition and differential evolution for forecasting monthly precipitation. J. Hydrometeor, 18, 159–176, 10.1175/JHM-D-16-0109.1.

Teng H, and Branstator G, 2017: Causes of extreme ridges that induce California droughts. J. Climate, 30, 1477–1492, 10.1175/JCLI-D-16-0524.1.

Tibshirani R, 1996: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc,, 58B, 267–288.

——, and Taylor J, 2011: The solution path of the generalized lasso. Ann. Stat, 39, 1335–1371, 10.1214/11-AOS878.

——, Saunders M, Rosset S, Zhu J, and Knight K, 2005: Sparsity and smoothness via the fused lasso. J. Roy. Stat. Soc, 67B, 91–108, 10.1111/j.1467-9868.2005.00490.x.

Trenberth KE, and Shea DJ, 1987: On the evolution of the Southern Oscillation. Mon. Wea. Rev, 115, 3078–3096, 10.1175/1520-0493(1987)115<3078:OTEOTS>2.0.CO;2.

——, Branstator GW, Karoly D, Kumar A, Lau N-C, and Ropelewski C, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. J. Geophys. Res, 103, 14 291–14 324, 10.1029/97JC01444.

van Loon H, and Shea DJ, 1987: The Southern Oscillation. Part VI: Anomalies of sea level pressure on the Southern Hemisphere and of Pacific sea surface temperature during the development of a warm event. Mon. Wea. Rev, 115, 370–379, 10.1175/1520-0493(1987)115<0370:TSOPVA>2.0.CO;2.

Vose RS, and Coauthors, 2014: Improved historical temperature and precipitation time series for U.S. climate divisions. J. Appl. Meteor. Climatol, 53, 1232–1251, 10.1175/JAMC-D-13-0248.1.

Waliser DE, and Gautier C, 1993: A satellite-derived climatology of the ITCZ. J. Climate, 6, 2162–2174, 10.1175/1520-0442(1993)006<2162:ASDCOT>2.0.CO;2.

Wang B, and Coauthors, 2009: Advance and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Climate Dyn, 33, 93–117, 10.1007/s00382-008-0460-0.

Wang C, Deser C, Yu J-Y, DiNezio P, and Clement A, 2012: El Niño–Southern Oscillation (ENSO): A review. Coral Reefs of the Eastern Pacific, Glymn P, Manzello D, and Enochs I Eds., Springer, 85–106.

Wang SY, Hipps L, Gillies RR, and Yoon JH, 2014: Probable causes of the abnormal ridge accompanying the 2013–2014 California drought: ENSO precursor and anthropogenic warming footprint. Geophys. Res. Lett, 41, 3220–3226, 10.1002/2014GL059748.

Wang T, and Miao J-P, 2018: Twentieth-century Pacific decadal oscillation simulated by CMIP5 coupled models. Atmos. Ocean. Sci. Lett, 11, 94–101, 10.1080/16742834.2017.1381548.

Wilks DS, 2016: "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. Bull. Amer. Meteor. Soc, 97, 2263–2273, 10.1175/BAMS-D-15-00267.1.

Willard JD, Jia X, Xu S, Steinbach M, and Kumar V, 2020: Integrating physics-based modeling with machine learning: A survey. https://arxiv.org/abs/2003.04919.

Wu Z, Wang B, Li J, and Jin F-F, 2009: An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO. J. Geophys. Res, 114, D18120, 10.1029/2009JD011733.

Yeager SG, and Coauthors, 2018: Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. Bull. Amer. Meteor. Soc, 99, 1867–1886, 10.1175/BAMS-D-17-0098.1.

Yu J-Y, Zou Y, Kim ST, and Lee T, 2012: The changing impact of El Niño on US winter temperatures. Geophys. Res. Lett, 39, L1570, 10.1029/2012GL052483.

Zhang T, and Coauthors, 2018: Predictability and prediction of Southern California rains during strong El Niño events: A focus on the failed 2016 winter rains. J. Climate, 31, 555–574, 10.1175/JCLI-D-17-0396.1.
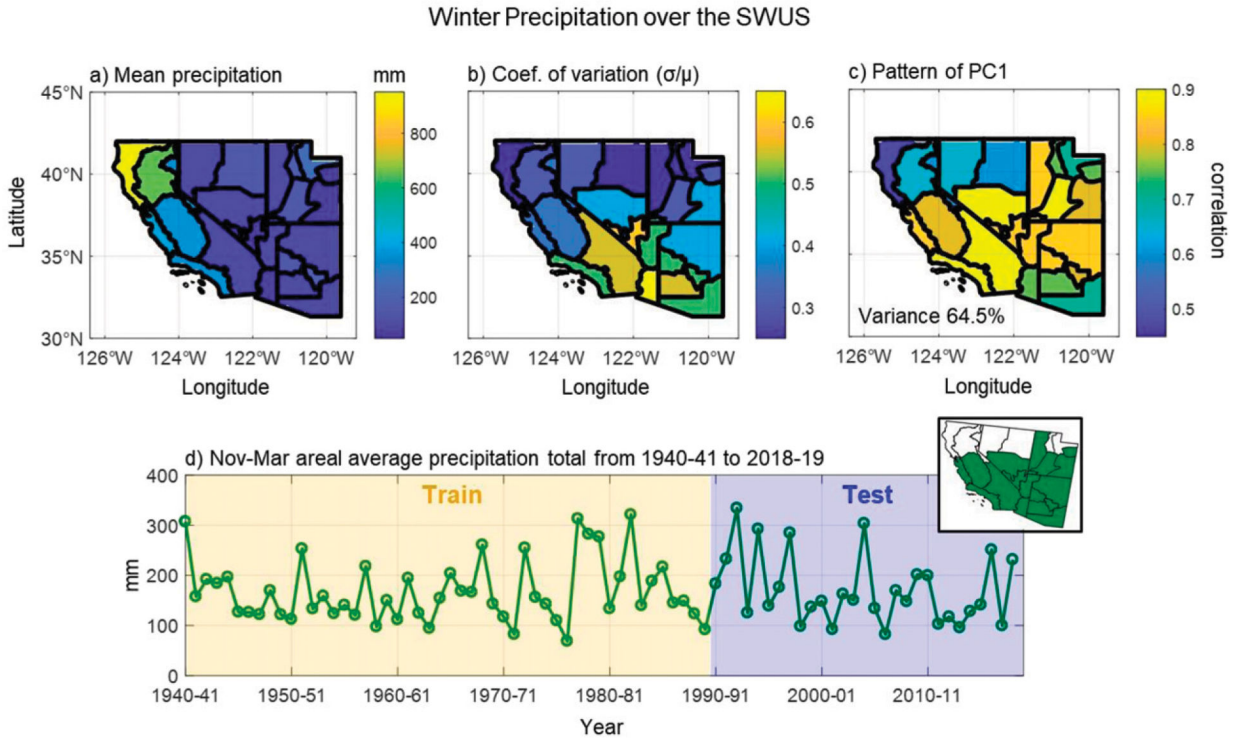
Fig. 1.
Spatial patterns of winter precipitation statistics over the southwestern United States. (a) Multiyear mean of November–March precipitation over SWUS, for the period from 1940/41 to 2018/19. (b) Coefficient of variation (sample standard deviation divided by sample mean) of November–March precipitation for the period from 1940/41 to 2018/19. (c) Correlation of the first area-weighted principal component (PC1) of the November–March precipitation over the SWUS and precipitation in each climate division. (d) Series of the area-weighted average precipitation over the climate divisions considered in this study (see panel to the right). In our study, we use years 1940/41–1989/90 as a training period (for model fitting), and years 1990/91–2018/19 as a test period (for model evaluation).

a) Covariance Matrix of Observed SSTs ($\widehat{\Sigma}_{obs}$)
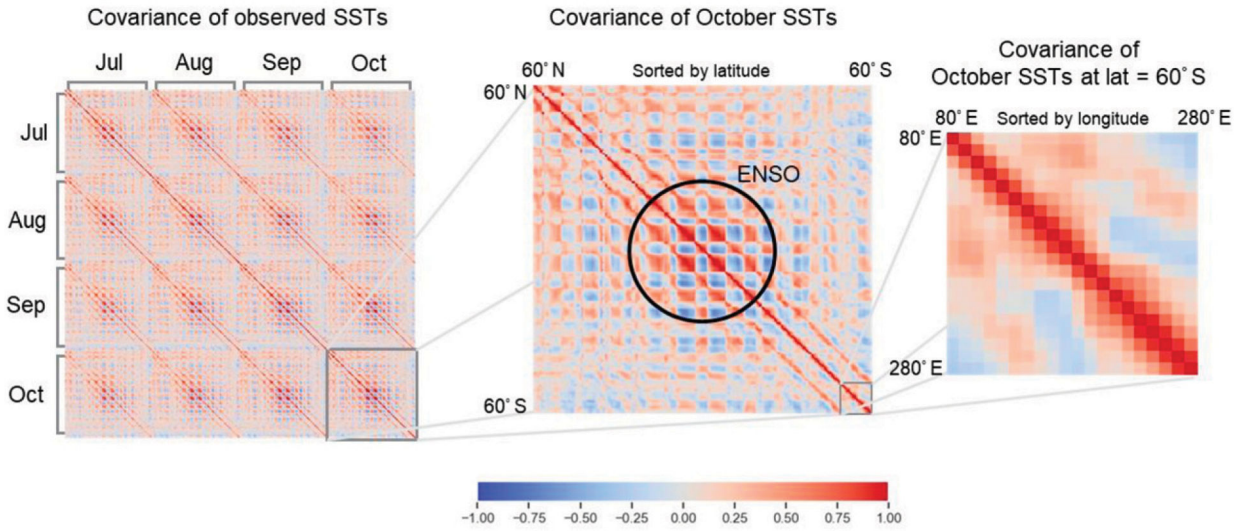
b) Covariance Matrix of Observed ($\widehat{\Sigma}_{obs}$) and Simulated SSTs ($\widehat{\Sigma}_{CL}$)
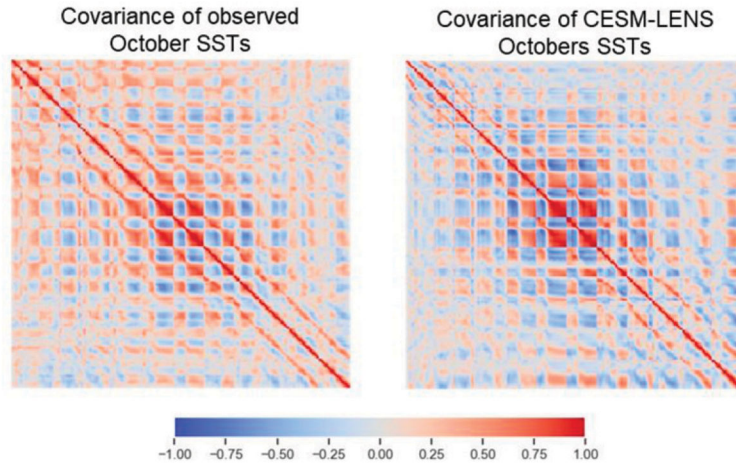
Fig. 2.
Space–time covariance of sea surface temperatures (SSTs) in the Pacific Ocean. (a) Sample covariance matrix of the observed Pacific SSTs over longitude–latitude and for four months. Zoom-in covariance in October highlights the spatial extent of the tropical ENSO signal and a further zoom-in at the fixed latitude of 60°S shows the spatial longitudinal dependence. (b) Comparison of the sample covariance of October Pacific SSTs as estimated from the observations and the output of CESM-LENS.
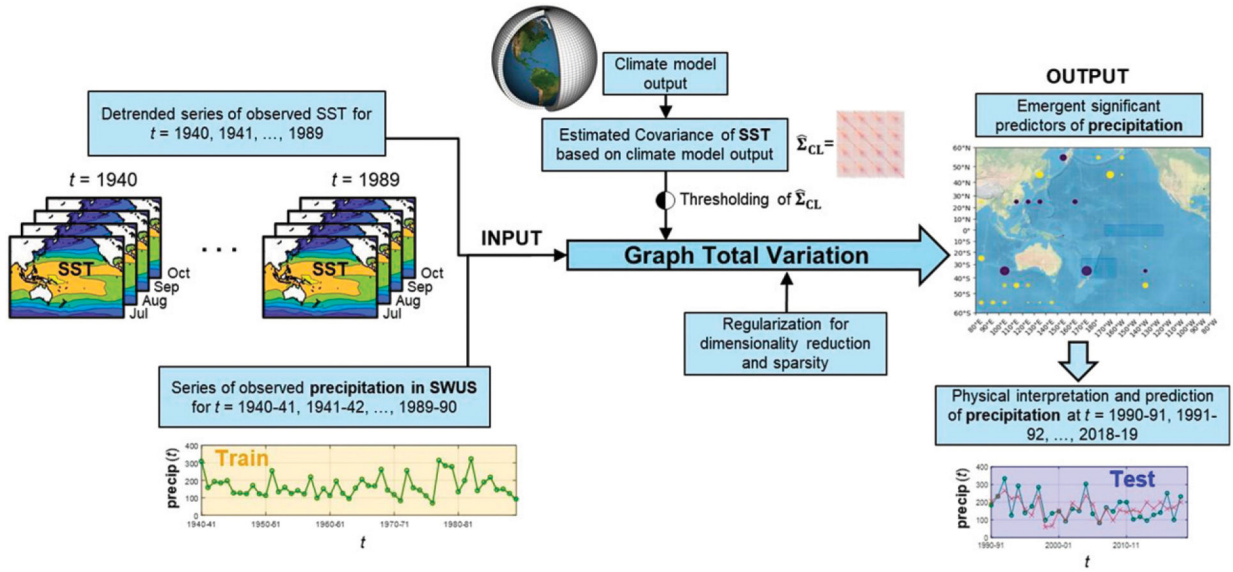
Fig. 3.

Schematic for the graph-guided regularization [graph total variation (GTV)] for predicting winter precipitation over SWUS. We use observed Pacific SSTs as input to the predictive model and we form a space–time covariance graph with edge weights corresponding to pairwise SST covariances to constrain the model via a GTV regularization term. The covariance matrix is estimated based on the output of climate models and subjected to a hard thresholding (see section 3c) to increase the consistency of the dependency among predictors for improved performance of the GTV algorithm.
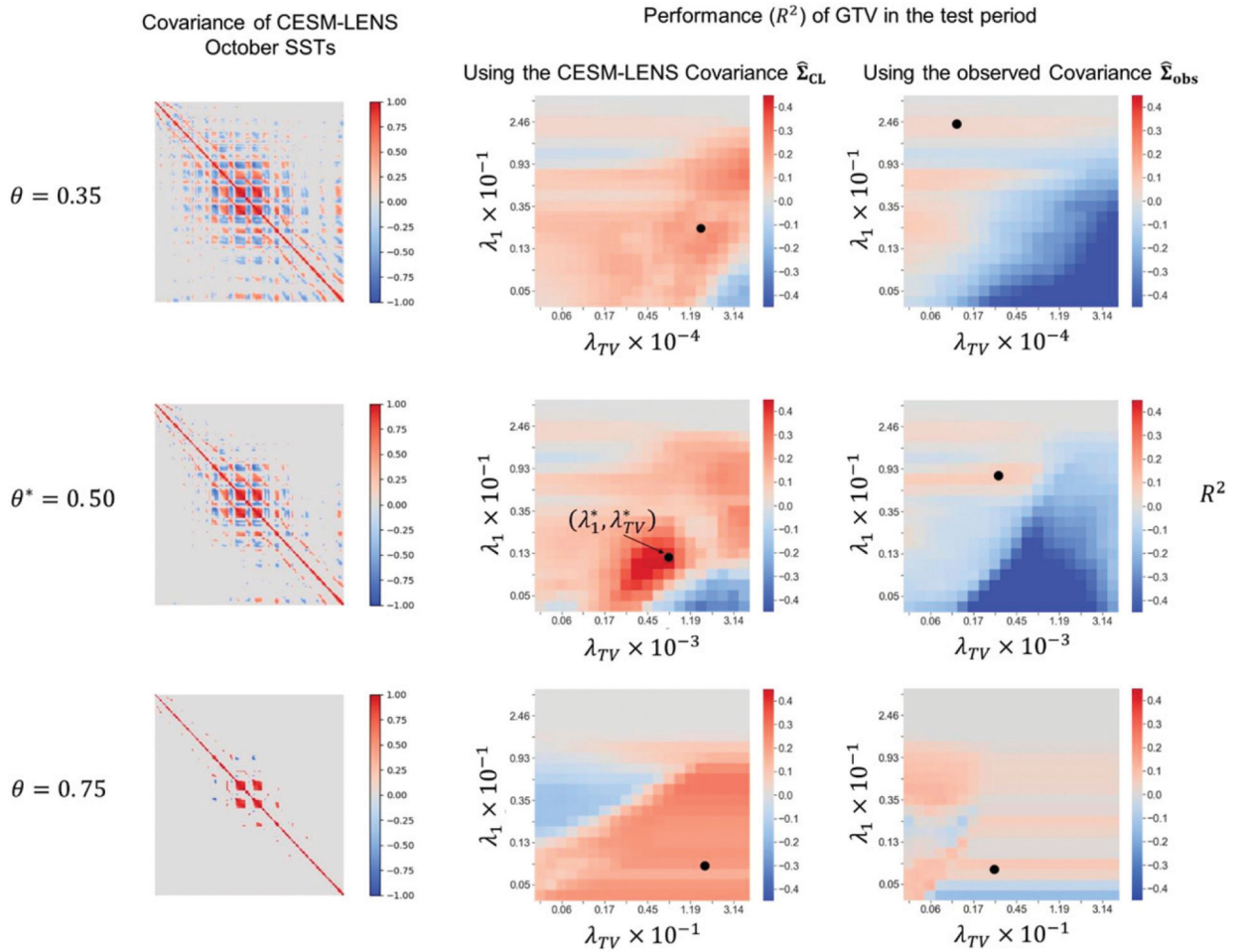
Fig. 4.

Sensitivity analysis of the GTV model performance for a range of covariance thresholds θ and regularization parameters λ_1 and λ_{TV}. (left) The covariance of the October SSTs (as estimated based on the CESM-LENS) for three different thresholds ($\theta = 0.35$, 0.5, and 0.75). (center), The coefficient of determination (R^2) between the areal average observed and model predicted precipitation in the test period (1990/91–2018/19), when the CESM-LENS covariance is used to define the GTV regularizer. (right) As in the center panels, but when the observed SST covariance is used. In all panels, and conditional on the corresponding values of θ, the optimal $(\lambda_1^*, \lambda_{TV}^*)$ pair for each model (obtained by using a fivefold cross-validation in the training period 1940/41–1989/90) is shown (black dots). These results highlight that 1) the use of the CESM-LENS covariance, instead of the observational covariance, to inform the GTV regularizer leads to a highly robust and improved predictive performance, as judged by the larger domain of regularization parameters with high R^2 values (see center column plots, as compared to their right counterparts), and 2) our choice to use a threshold of $\theta^* = 0.5$, which was based on a fivefold cross-validation in the training period (see section 4a), shows to yield the most robust and highest predictive performance.
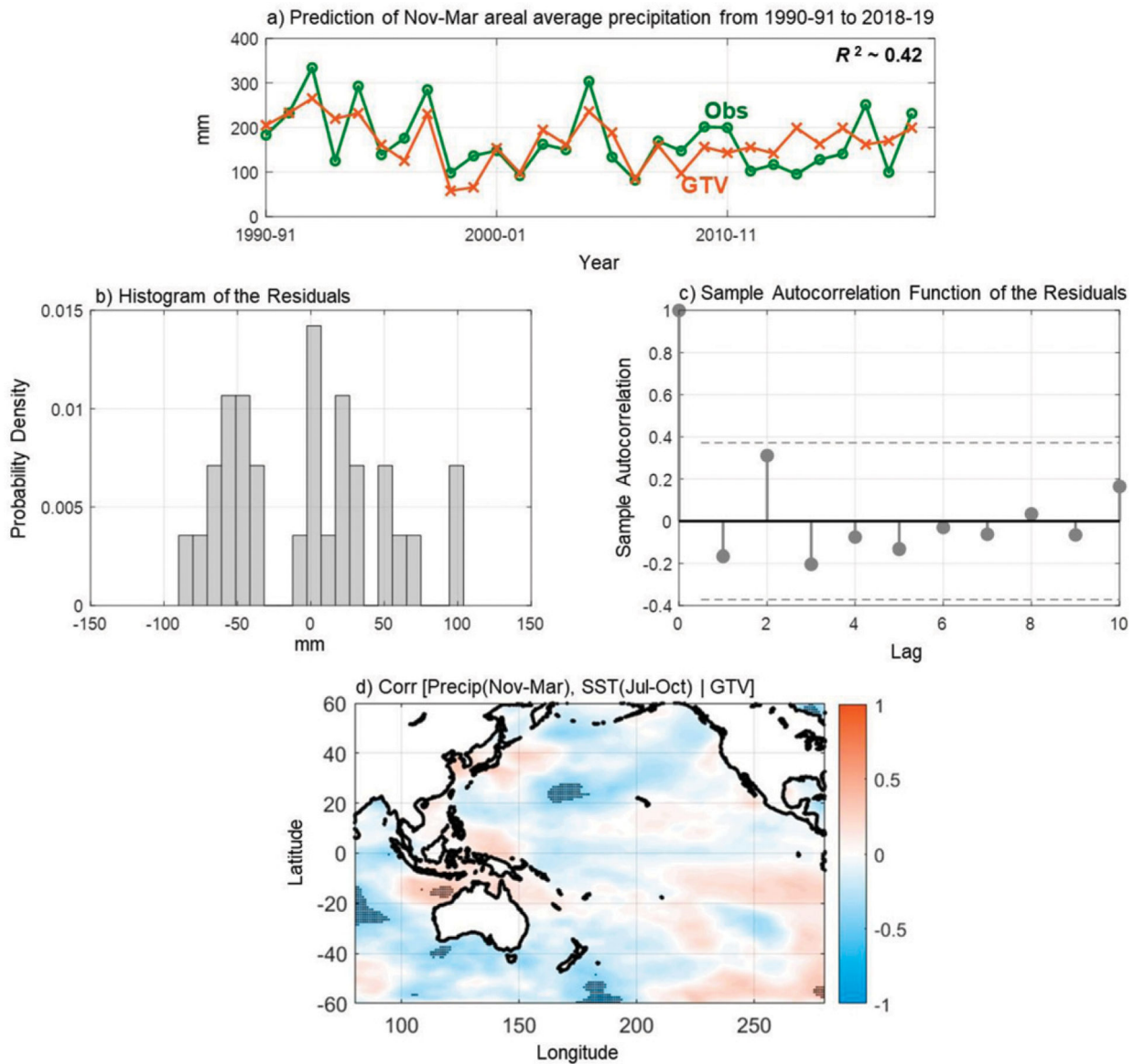
Fig. 5.
Evaluation of the prediction of winter (November–March) precipitation. (a) Series of observed (green) and predicted (brown) November–March areal average SWUS precipitation during the test period from 1990/91 to 2018/19. Prediction is made using the sample covariance from the CESM-LENS output. (b),(c) Histogram and autocorrelation function of the residual time series during the test period [residuals are between GTV predictions and observations shown in (a)]. The null hypothesis that the residuals are normally distributed is not rejected at a 0.05 significance level. Also, the null hypothesis that there is no year to year linear dependence (autocorrelation) in the residual time series is not rejected at the 0.05 significance level. (d) Partial correlation between SWUS precipitation in November–March and linear detrended grid point SSTs in July–October, after accounting for the GTV prediction. Stippling indicates locally significant correlations. The absence of

significant correlation patterns indicates that no more predictive information can be extracted from the Pacific basin SSTs, providing confidence for the fitted model.
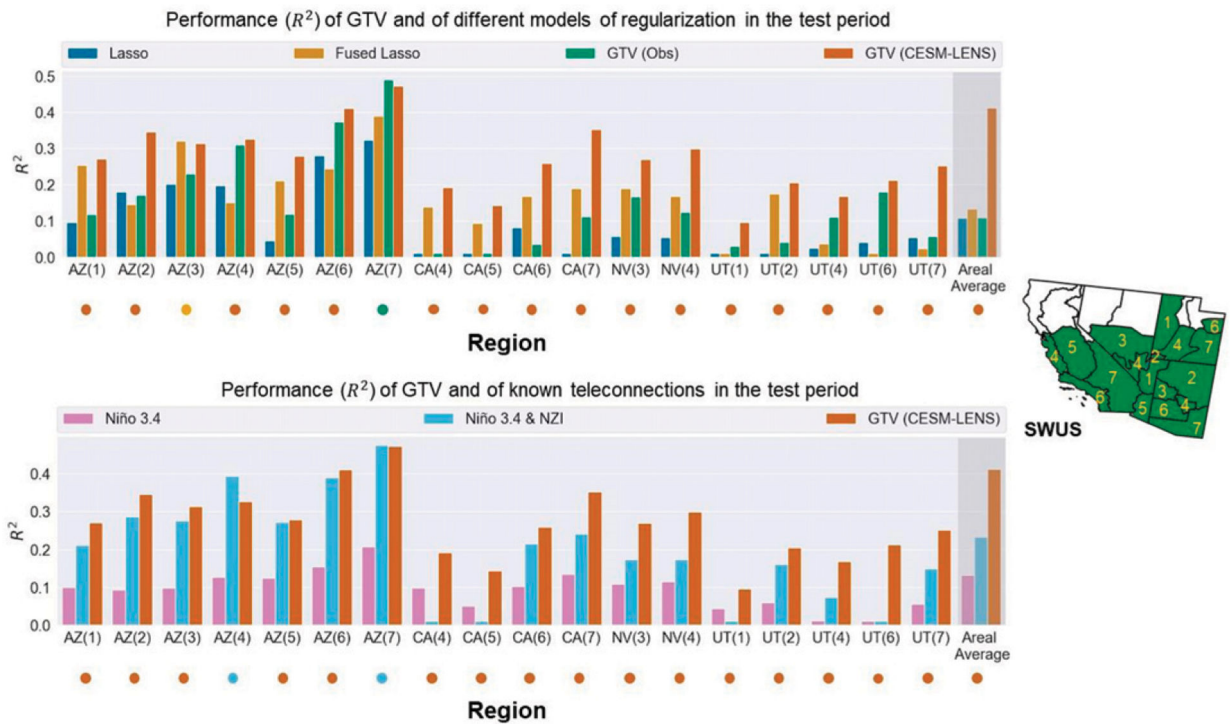
Fig. 6.
Performance of (top) GTV and different methods of regularization and (bottom) known teleconnections in predicting precipitation totals over different SWUS divisions in the test period from 1990/91 to 2018/19. The coefficient of determination (R^2) is presented. It is shown that GTV with model-estimated covariance of SSTs outperforms all other regularization methods (top) as well as statistical regression on two known teleconnections (bottom).
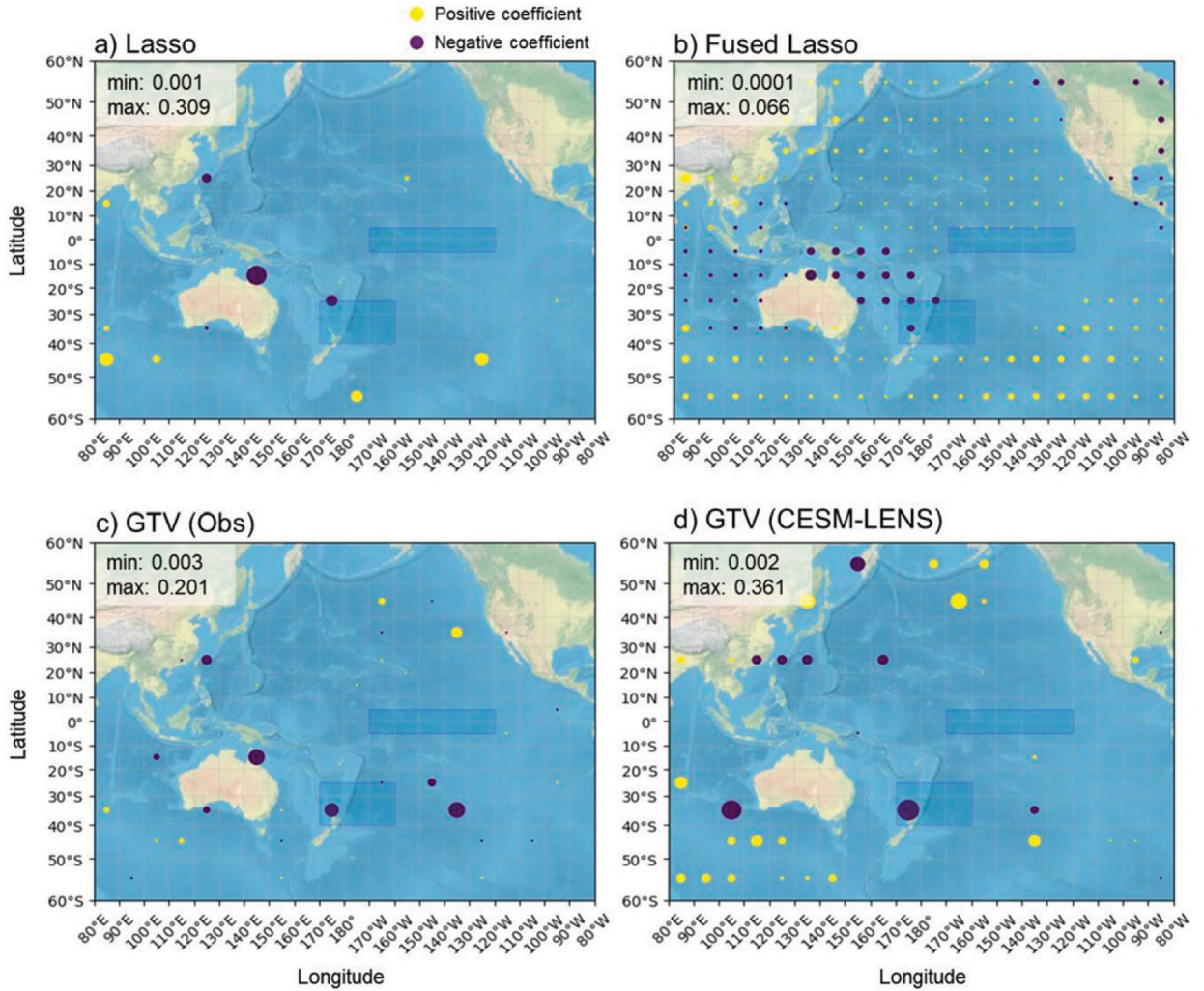
Fig. 7.
The emergent predictors of the areal average SWUS winter precipitation for different models of regularization: (a) LASSO, (b) fused LASSO, (c) GTV using the sample covariance of the observed SSTs, and (d) GTV using the sample covariance from the CESM-LENS output. The $\widehat{\beta}$ values are presented (colored circles) after training each method in the training period from 1940/41 to 1989/90, using a fivefold cross-validation technique. The color of the circles indicates the sign of the $\widehat{\beta}$ values (yellow for positive and purple for negative), while the size of circles is proportional the their magnitude; for each method, the minimum and maximum $\widehat{\beta}$ values (in absolute terms) are also given. Niño-3.4 and NZI boxes are also shown. All models highlight to a greater or lesser extent the western and southwestern Pacific SSTs as important predictors of SWUS precipitation.
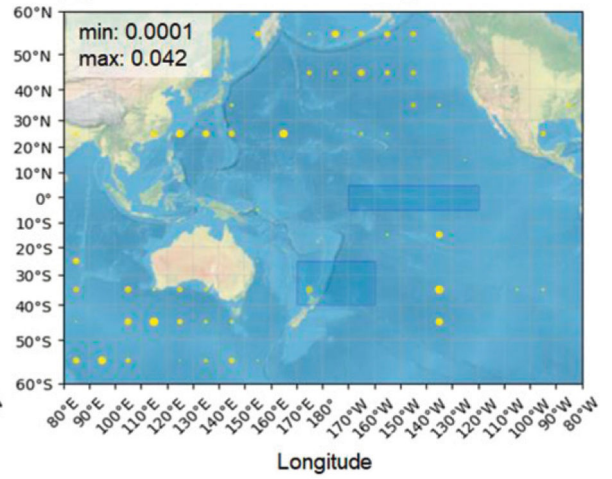
a) Performance of GTV (R^2) for 1000 bootstrap realizations
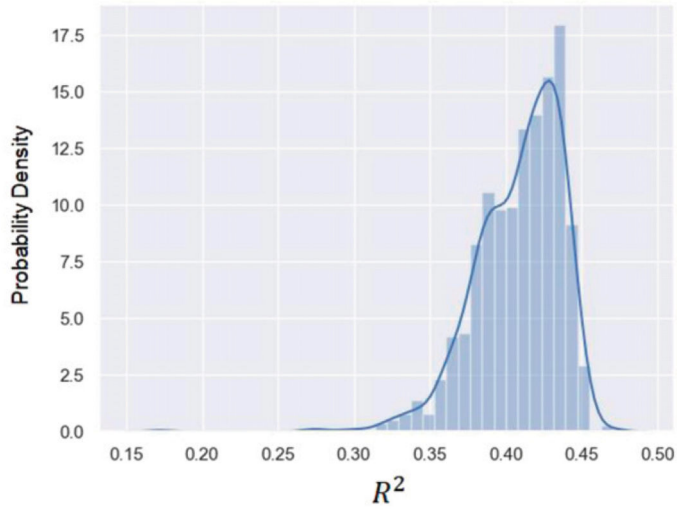
b) $\widehat{\beta}$ values of GTV: **bootstrap average**

min: 0.0001
max: 0.358
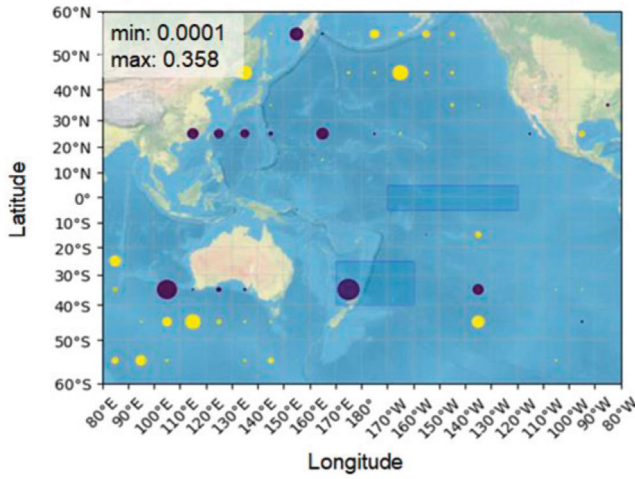
c) $\widehat{\beta}$ values of GTV: **bootstrap st. deviation**

min: 0.0001
max: 0.042

Fig. 8.

Bootstrap investigation of the sensitivity of the GTV to the uncertainty of the covariance estimated from CESM-LENS. (a) The histogram of the coefficient of determination (R^2) between the observed November–March SWUS precipitation and the GTV prediction across all 1000 bootstrap realizations. (b) The vector average of the 1000 $\widehat{\beta}$ vectors from the 1000 bootstrap realizations. For each realization, training is performed in the period from 1940/41 to 1989/90, using a fivefold cross-validation technique. (c) As in (b), but the standard deviation of the 1000 $\widehat{\beta}$ vectors is presented. The small uncertainty of the most important predictors (grids with the largest $\widehat{\beta}$ values) is noteworthy.

Table 1.

Mean square error (MSE) of different methods of regularization and teleconnections in predicting precipitation totals over different SWUS divisions in the test period from 1990/91 to 2018/19. Precipitation series has been standardized (zero mean and unit variance). For the GTV model the covariance threshold of $\Theta^* = 0.5$ has been used. Bold font indicates the method with the lowest MSE for each climate division.

Region	Niño-3.4	Niño-3.4 and NZI	LASSO	Fused LASSO	GTV (Obs)	GTV (CESM-LENS)
Arizona (1)	1.1515	1.0091	1.1556	0.9549	1.1279	**0.9321**
Arizona (2)	1.0219	0.8037	0.9234	0.9631	0.9349	**0.7372**
Arizona (3)	1.1344	0.9102	1.0036	**0.8540**	0.9691	0.8625
Arizona (4)	0.9812	**0.6815**	0.9017	0.9536	0.7758	0.7571
Arizona (5)	1.1927	0.9918	1.3016	1.0754	1.2016	**0.9833**
Arizona (6)	0.9684	0.6985	0.8250	0.8656	0.7178	**0.6744**
Arizona (7)	0.9239	0.6104	0.7885	0.7105	**0.5936**	0.6135
California (4)	0.8560	0.9695	0.9469	0.8174	0.9549	**0.7672**
California (5)	0.9169	1.1322	0.9663	0.8756	0.9663	**0.8270**
California (6)	1.0367	0.9062	1.0598	0.9606	1.1127	**0.8552**
California (7)	1.1252	0.9870	1.3009	1.0544	1.1543	**0.8409**
Nevada (3)	0.9632	0.8927	1.0182	0.8761	0.9006	**0.7889**
Nevada (4)	1.1680	1.0907	1.2489	1.0980	1.1567	**0.9255**
Utah (1)	1.1706	1.2818	1.2238	1.2196	1.1871	**1.1074**
Utah (2)	1.1295	1.0078	1.1973	0.9915	1.1512	**0.9545**
Utah (4)	0.9243	0.8656	0.9108	0.9000	0.8312	**0.7773**
Utah (6)	0.8958	0.9667	0.8537	0.8849	0.7305	**0.7012**
Utah (7)	0.8462	0.7622	0.8470	0.8740	0.8450	**0.6701**
Area-weighted average	0.9818	0.8671	1.0087	0.9803	1.0074	**0.6652**