

# Uncovering the Forced Climate Response from a Single Ensemble Member Using Statistical Learning

SEBASTIAN SIPPEL

*Institute for Atmospheric and Climate Science, and Seminar for Statistics, ETH Zurich, Zurich, Switzerland*

NICOLAI MEINSHAUSEN

*Seminar for Statistics, ETH Zurich, Zurich, Switzerland*

ANNA MERRIFIELD

*Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland*

FLAVIO LEHNER AND ANGELINE G. PENDERGRASS

*National Center for Atmospheric Research, Boulder, Colorado*

ERICH FISCHER AND RETO KNUTTI

*Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland*

(Manuscript received 22 December 2018, in final form 13 May 2019)

## ABSTRACT

Internal atmospheric variability fundamentally limits predictability of climate and obscures evidence of anthropogenic climate change regionally and on time scales of up to a few decades. Dynamical adjustment techniques estimate and subsequently remove the influence of atmospheric circulation variability on temperature or precipitation. The residual component is expected to contain the thermodynamical signal of the externally forced response but with less circulation-induced noise. Existing techniques have led to important insights into recent trends in regional (hydro-) climate and their drivers, but the variance explained by circulation is often low. Here, we develop a novel dynamical adjustment technique by implementing principles from statistical learning. We demonstrate in an ensemble of Community Earth System Model (CESM) simulations that statistical learning methods, such as regularized linear models, establish a clearer relationship between circulation variability and atmospheric target variables, and need relatively short periods of record for training (around 30 years). The method accounts for, on average, 83% and 78% of European monthly winter temperature and precipitation variability at gridcell level, and around 80% of global mean temperature and hemispheric precipitation variability. We show that the residuals retain forced thermodynamical contributions to temperature and precipitation variability. Accurate estimates of the total forced response can thus be recovered assuming that forced circulation changes are gradual over time. Overall, forced climate response estimates can be extracted at regional or global scales from approximately 3–5 times fewer ensemble members, or even a single realization, using statistical learning techniques. We anticipate the technique will contribute to reducing uncertainties around internal variability and facilitating climate change detection and attribution.

---

Denotes content that is immediately available upon publication as open access.

---

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-18-0882.s1>.

---

*Corresponding author:* Sebastian Sippel, [sebastian.sippel@env.ethz.ch](mailto:sebastian.sippel@env.ethz.ch)

---

## 1. Introduction

Climate resilience hinges on anticipating externally forced changes in the climate system, which are often challenging to distinguish from internal variability or “climate noise” on local to regional scales or on time scales up to a few decades (Madden 1976; Schneider and Kinter 1994; Deser et al. 2012a,b). Internal variability is a dominant source of uncertainty in near-term

DOI: 10.1175/JCLI-D-18-0882.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

regional climate predictions (Hawkins and Sutton 2009), and hinders detection and attribution studies by reducing signal-to-noise ratios and thus potential predictability of externally driven changes in the climate system (Stott et al. 2000; Lambert et al. 2004).

Sometimes referred to as the irreducible uncertainty of prediction, internal variability can be quantified by repeatedly running the same climate model, subject to the same external forcing, from slightly different initial atmospheric states (i.e., in an initial condition large ensemble; Deser et al. 2014; Kay et al. 2015). In a large ensemble, it is common practice to estimate a climate variable's total forced response, defined as the variable's total response to external forcing, by averaging across the full set of realizations. However, with only a single, externally forced realization of the climate system (i.e., the real world), disentangling forced responses from internal variability requires a different approach.

“Dynamical adjustment” techniques have been developed with the ultimate goal of removing the estimated influence of internal atmospheric circulation variability on target climate variables in the observational record or in models (Wallace et al. 2012; Deser et al. 2016; Smoliak et al. 2015; Saffioti et al. 2016). This research followed upon a large suite of earlier studies aiming to understand the relationship between modes of internal atmospheric variability and temperature or precipitation (Wallace et al. 1995; Thompson et al. 2000; Compo and Sardeshmukh 2010; Foster and Rahmstorf 2011; Iles and Hegerl 2017). Conceptually, variability in a target climate variable might be seen as a realization of internal variability that is superimposed upon the response to external forcing (Wallace et al. 1995; Deser et al. 2016). External forcing could either affect the target variable thermodynamically (“forced thermodynamical component,” which for our purposes means due to the warming itself, i.e., related to energy balance changes), or it could impose changes upon the atmospheric circulation that affect the target variable dynamically [“forced dynamical component”; see, e.g., Wallace et al. (1995), and illustrated in Fig. 1]. To explain projected changes in the global hydrologic cycle, both thermodynamical and dynamical changes (such as a poleward shift of midlatitude storm tracks; Yin 2005; Perlwitz et al. 2017) are typically invoked (Emori and Brown 2005; Seager et al. 2010; He and Soden 2017). Evidence for forced changes in key climate variables such as temperature, and some precipitation characteristics, is unequivocal in the historical record. In contrast, historical and projected changes induced by atmospheric circulation are typically more uncertain in both observations and models (Shepherd 2014; Xie et al. 2015; Collins et al. 2018).

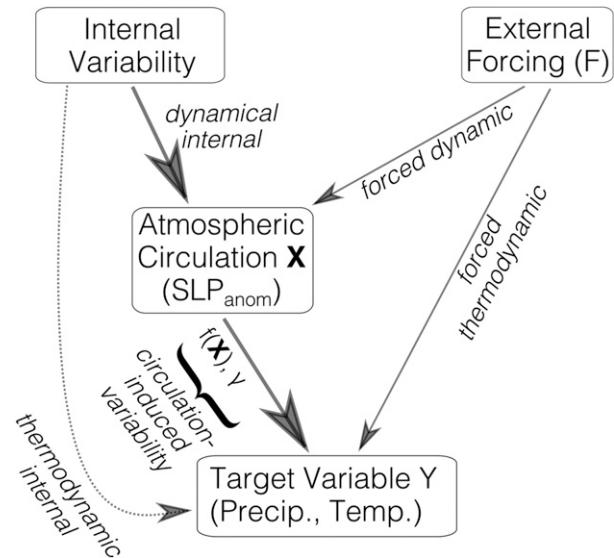


FIG. 1. Conceptual illustration of causal relationships assumed by dynamical adjustment techniques. External forcing causes a thermodynamical response in the target variable  $Y$ , assumed to be distinct from random internal variability that affects  $Y$  through circulation anomaly patterns (e.g., sea level pressure anomalies). The residuals of a regression of  $Y$  on sea level pressure anomalies reflect a “noise-filtered” response to external forcing. The total response to external forcing (forced thermodynamical and dynamical) may be revealed through the residuals of a regression of  $Y$  on  $SLP_{anom}$  with the external influence on  $SLP_{anom}$  removed.

The fundamental idea of dynamical adjustment is to derive an estimate of the circulation-induced contribution to variability in a target climate variable, such as temperature or precipitation (Fig. 1). Subsequently, the estimate is used to obtain a residual time series, that is, the component of the target climate variable's variability that cannot be explained by circulation. The residuals are expected to contain forced thermodynamical components (Deser et al. 2016). An estimate of the total forced response (i.e., including forced circulation changes) could be obtained if one assumes that forced circulation components are reasonably smooth in time (cf. section 2b), or by relying on simulated forced circulation changes (e.g., Deser et al. 2016). Yet, the key step in dynamical adjustment is to estimate the relationship between atmospheric circulation and a target climate variable. Accordingly, an improved statistical estimate of the circulation-induced contribution would translate into a higher signal-to-noise ratio in the residual (Smoliak et al. 2015). Different dynamical adjustment methods, among them circulation analog techniques (Deser et al. 2016; Merrifield et al. 2017; Lehner et al. 2017) or regression-based approaches (Smoliak et al. 2015; Saffioti et al. 2016), are different ways to approach the statistical estimation step.

The application of dynamical adjustment techniques has led to important insights into the origins and drivers of trends in hydroclimatic variables. For example, Wallace et al. (1995) identified that nearly half of the observed Northern Hemisphere winter surface air temperature (SAT) variance in the twentieth century is attributable to a “cold ocean, warm land” spatial pattern that occurs randomly over time and was therefore not necessarily associated with external forcing. Foster and Rahmstorf (2011) showed that a reevaluation of global temperature trends by adjusting for the influence of known factors such as ENSO, volcanic aerosols, and solar variability yields a more precise estimate of the global warming signal in the residual time series. Deser et al. (2016) found that after accounting for circulation-induced variability, SAT trends in a large ensemble of model simulations are closer to the model’s total forced response. Similarly, observed cooling trends in SAT at the local scale, for instance, in Switzerland, can be reconciled with the simulated total forced response when atmospheric circulation is accounted for (Saffioti et al. 2015, 2016). Furthermore, dynamical adjustment advances the time of emergence of externally forced signals (Deser et al. 2016; Lehner et al. 2017), and consequently could facilitate detection and attribution. Nonetheless, residual time series after accounting for circulation effects cannot be seen as a pure estimate of the imprint of external forcing (Smoliak et al. 2015). This is because residual variability might still contain other sources of variability such as land–atmosphere feedbacks (Merrifield et al. 2017) or effects related to land surface initial conditions (Lehner et al. 2017).

Dynamical adjustment of precipitation is more challenging than for temperature (Saffioti et al. 2016) because precipitation more spatially heterogeneous and thus less predictable from circulation. In addition, precipitation depends on location and orography (Houze 2012), and in some regions, forced precipitation changes are driven by atmospheric circulation (Yin 2005; He and Soden 2017; Siler et al. 2019). Nonetheless, a large fraction of precipitation variability in Europe and North America is related to internal variability in atmospheric circulation (Saffioti et al. 2016; Fereday et al. 2018; Lehner et al. 2018; Guo et al. 2019). Fereday et al. (2018) show that the majority of projection uncertainty for European winter precipitation is associated with the models’ atmospheric circulation response to external forcing.

The objective of the present study is to incorporate statistical learning principles into the framework of dynamical adjustment. We evaluate whether statistical learning can improve upon existing techniques. The manuscript is set up as follows:

We frame dynamical adjustment and its assumptions in the context of statistical learning (sections 2a and 2b)

and illustrate a candidate statistical learning technique for dynamical adjustment (“regularized linear models,” section 2c). Sections 3a–3c detail the study’s data and experimental setup. In section 4a, we evaluate the performance of statistical techniques in estimating circulation-induced components of temperature and precipitation variability. Next, we evaluate the consistency of the residual, “dynamically adjusted” trends in forced model simulations in Europe against the model’s forced response (section 4b) and illustrate a simple partitioning of the underlying thermodynamical and circulation-induced components of the forced response. Last, we apply and evaluate our methodology at the global scale (section 4c).

## 2. Dynamical adjustment as a statistical learning problem

In this section, we revisit dynamical adjustment in order to conceptualize it as a statistical learning problem (sections 2a and 2b, represented schematically in Fig. 1), and then introduce a specific set of statistical learning techniques to apply (section 2c).

### a. Isolating the thermodynamical component of the forced response

Variations in atmospheric circulation and associated variations [e.g., in sea level pressure (SLP)] drive weather variability, for instance, through different advective flow regimes, extratropical cyclones (McMurdie and Houze 2006), or anticyclones (Colucci 2015). Therefore, dynamical adjustment studies (Deser et al. 2016; Smoliak et al. 2015) rely on a proxy for circulation (i.e., in this study  $\mathbf{X} := \text{SLP}$ , Fig. 1).

Estimating the circulation-induced component ( $\hat{\mathbf{Y}}_{\mathbf{X}}$ ) of  $\mathbf{Y}$  yields

$$\hat{\mathbf{Y}}_{\mathbf{X}} = \hat{f}(\mathbf{X}) \approx \underbrace{\mathbf{X}\hat{\boldsymbol{\gamma}}}_{\text{linear approximation}}, \quad (1)$$

where  $\mathbf{X}$  represents a  $N \times (p + 1)$  matrix, with each column an input feature (e.g., SLP anomalies located on a grid across the spatial domain of interest) with  $N$  observations, and  $\mathbf{Y}$  is the target vector of length  $N$  and  $\hat{\boldsymbol{\gamma}}$  is a  $(p + 1)$  vector of regression coefficients including the intercept. We denote  $\hat{\mathbf{Y}}_{\mathbf{X}}$  as an estimate of  $\hat{\mathbf{Y}} = \hat{f}(\mathbf{X})$  based on the predictor matrix  $\mathbf{X}$  and a suitable statistical model  $f$ ;  $\hat{\mathbf{R}}$  denotes the residual component that remains unexplained, that is,  $\hat{\mathbf{R}} = \mathbf{Y} - \hat{\mathbf{Y}}_{\mathbf{X}}$ .

By removing estimated circulation components  $\hat{\mathbf{Y}}_{\mathbf{X}}$  from the target variable  $\mathbf{Y}$ , the residual component ( $\hat{\mathbf{R}}$ ) is expected to maintain the thermodynamical component of the forced response,

$$\underbrace{\hat{\mathbf{R}}}_{\text{thermodynamical}} = \mathbf{Y} - \hat{\mathbf{Y}}_{\mathbf{X}}. \quad (2)$$

In other words, the function  $\hat{f}(\mathbf{X})$  or its linear approximation  $\mathbf{X}\hat{\gamma}$  encapsulate the physical relationship between the prevailing atmospheric circulation pattern (on a specific day or month) and the target variable  $\mathbf{Y}$ . The residual component ( $\hat{\mathbf{R}}$ ) represents a time series with estimated circulation-induced variability removed. Forced circulation-induced changes would be part of the estimated circulation-induced component ( $\hat{\mathbf{Y}}_{\mathbf{X}}$ ) if these circulation changes project onto SLP (Smoliak et al. 2015). The statistical estimation step of  $\hat{f}$  (or  $\hat{\gamma}$ ) is where statistical learning techniques and principles can be applied. A consistent statistical prediction of  $\hat{\mathbf{Y}}_{\mathbf{X}}$  is directly relevant to the signal-to-noise ratio in the residual component. Previous dynamical adjustment studies (e.g., Smoliak et al. 2015; Deser et al. 2016) estimated circulation-induced components with various regression methods (see, e.g., section 3c), but yielded largely consistent results (Deser et al. 2016).

One potential complication is that the effects of external forcing  $\mathbf{F}$  may have effects beyond the thermodynamical response of the target variable; it may also affect the atmospheric circulation  $\mathbf{X}$  (a forced dynamic response; Fig. 1). Hence, a spurious correlation between  $\mathbf{X}$  and  $\mathbf{Y}$  induced by  $\mathbf{F}$  would potentially bias the estimation of  $\hat{f}$  (Peters et al. 2017, chapter 9.3). In other words, a forced thermodynamical change could be erroneously explained by some SLP trend if forced circulation trends are large and not accounted for in the training step. This biased estimation of  $\hat{f}$  can be avoided, for instance, if 1) model training is based on a control run without external forcing, or 2) if forced trends in circulation can be assumed small for other reasons (e.g., training on daily circulation with minor low-frequency variability, section 2c), or 3) via high-pass filtering the SLP field *only for the training step* assuming a smooth forced response in SLP (Smoliak et al. 2015; Lehner et al. 2018).

### b. Recovering the total forced response

Dynamical adjustment methods have been designed from the outset with the explicit goal of reducing uncertainties around the total forced response (Deser et al. 2016). While the forced dynamical component can be estimated in a large ensemble through averaging over all ensemble member's estimated circulation-induced components (Deser et al. 2016; Saffioti et al. 2017), this approach would fall short if only one realization is available (e.g., in observations).

An estimate of the total forced response within this framework would require an estimated circulation-induced component ( $\hat{\mathbf{Y}}_{\mathbf{X}}$ ), with forced effects on circulation ( $\mathbf{X}$ )

removed. Both dynamical and thermodynamical components of the total forced response would be maintained in the residuals. Under the confounding influence of an external variable  $\mathbf{F}$  that affects both  $\mathbf{X}$  and  $\mathbf{Y}$  (Peters et al. 2017, chapter 9.3), “instrumental variable estimation” is typically used to infer unconfounded relationships between  $\mathbf{X}$  and  $\mathbf{Y}$ . That is, an “instrumental” variable  $\mathbf{Z}$  that affects  $\mathbf{X}$  but is unaffected by  $\mathbf{F}$  is used to determine the unconfounded relationship between  $\mathbf{X}$  and  $\mathbf{Y}$ .

Here, we propose a simple approach along those lines that is applicable to a single realization of a model simulation or in observations. In the observable climate system, a proxy for internal variability that is entirely unaffected by external forcing ( $\approx \mathbf{Z}$ ) is likely not available. In the absence of such a proxy, one might assume that, first, external forcing  $\mathbf{F}$  is rather smooth and therefore causes low- but not high-frequency variability in atmospheric circulation (see, e.g., Wallace et al. 1995), whereas the physical relationship between  $\mathbf{X}$  and  $\mathbf{Y}$  is dominated by day-to-day weather (i.e., high-frequency) variability. Second, the system is assumed additive in its components (Fig. 1). Then, one might construct a proxy variable for internal circulation variability. Here, we use a high-pass filter on each predictor column in  $\mathbf{X}$ , such that

$$\hat{\mathbf{X}}_{\text{high-freq.}} = \mathbf{X} - \underbrace{\hat{\mathbf{X}}}_{\text{low-freq.}}, \quad (3)$$

where  $\hat{\mathbf{X}}_{\text{low-freq.}}$  represents a trend estimate for each column  $p$ . Then, we obtain

$$\hat{\mathbf{Y}}_{\mathbf{X}_{\text{high-freq.}}} = \hat{\mathbf{X}}_{\text{high-freq.}} \hat{\gamma}. \quad (4)$$

The residual component is expected to maintain forced thermodynamical and circulation-induced components,

$$\underbrace{\hat{\mathbf{R}}}_{\text{thermodynamical+dynamical}} = \mathbf{Y} - \hat{\mathbf{Y}}_{\mathbf{X}_{\text{high-freq.}}}. \quad (5)$$

Despite the assumption that external forcing on circulation is smooth, any short-term external forcing that would affect the target climate variable through thermodynamical changes (e.g., cooling in response to a volcanic eruption) would still be contained in the residuals (see, e.g., section 4c for a discussion). High-pass filtering the circulation field reduces low-frequency variability in the predicted circulation-induced component  $\hat{\mathbf{Y}}_{\mathbf{X}_{\text{high-freq.}}}$ . Hence, internal variability at very low frequencies (e.g., at centennial time scales depending on the properties of the high-pass filter; see section 3b) would remain in the residual. Hence, we recommend that SLP trends are subtracted such that (multi-) decadal

variability in SLP remains intact, since these represent an important aspect of internal climate system variability (Trenberth 1995; DelSole et al. 2011; Sutton et al. 2018).

*c. Applying statistical learning techniques in dynamical adjustment*

In this subsection, we describe a specific set of statistical learning tools, regularized linear models and a nonlinear extension (single-index models), as a means to incorporate statistical learning tools and principles into dynamical adjustment.

1) REGULARIZED LINEAR MODELS

Regularized linear models (RLMs) estimate relationship such as those discussed in section 2a by incorporating a constraint to penalize model complexity, called “regularization” (Hastie et al. 2001). The constraint is used to shrink regression coefficients in order to find a parsimonious approximation of the system. The main advantage of RLMs is to address high-variance (overfitting) issues with the least squares fit when the number of correlated predictors  $p$  is large relative to the number of samples  $n$  (i.e., if  $n \gg p$  does not hold), while maintaining advantages of linear methods such as prediction accuracy and interpretability. The reduction in variance of the fit comes at the cost of a bias, which is a necessity to address the bias–variance trade-off (Hastie et al. 2001), that is, the compromise between choosing a flexible model (i.e., small bias), but that is also parsimonious to avoid overfitting (i.e., low variance). The number of predictors  $p$  in dynamical adjustment might lie in the order of  $10^2$ – $10^3$  (e.g.,  $p = 400$  in a  $40^\circ \times 40^\circ$  SLP field at  $2^\circ$  resolution). The number of samples  $n$  depends on the available data, but would unlikely exceed  $n \approx 300$  for monthly observational data for a given season. Therefore, RLMs represent a natural choice for this purpose.

Consider a linear model of the form

$$Y = \gamma_0 + \sum_{j=1}^p X_j \gamma_j + \varepsilon, \quad (6)$$

where the ordinary least squares solution is achieved by minimizing the residual sums of squares (RSS), that is,

$$\begin{aligned} \hat{\gamma}^{\text{OLS}} &= \underset{\gamma}{\operatorname{argmin}} \{ \text{RSS} \} \\ &= \underset{\gamma}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left( y_i - \gamma_0 - \sum_{j=1}^p x_{ij} \gamma_j \right)^2 \right\}. \quad (7) \end{aligned}$$

An important class of penalties for regularized linear regression is the elastic-net family of regularization methods (Zou and Hastie 2005). The idea is to shrink the vector of regression coefficients ( $\gamma$ ) based on a shrinkage parameter

$\lambda$  and a parameter  $\alpha$  that balances different forms of shrinkage. The elastic-net solution is obtained by

$$\hat{\gamma}^{\text{elastic-net}} = \underset{\gamma}{\operatorname{argmin}} \left\{ \text{RSS} + \lambda \sum_{j=1}^p [(1 - \alpha) \gamma_j^2 + \alpha |\gamma_j|] \right\}. \quad (8)$$

In other words, the parameter  $\alpha$  balances between penalizing the sum of absolute values of regression coefficients [the “lasso” penalty, based on the  $L^1$  norm for  $\alpha = 1$  (i.e.,  $\lambda \sum_{j=1}^p |\gamma_j|$ )] and the sum of squared regression coefficients [“ridge regression,” based on the  $L^2$  penalty for  $\alpha = 0$  (i.e.,  $\lambda \sum_{j=1}^p \gamma_j^2$ )]. The regularization parameter  $\lambda \geq 0$  determines the magnitude of shrinkage. The apparently small difference in shrinkage type between the lasso and ridge regression (or intermediate types if  $0 < \alpha < 1$ , the elastic-net penalty) leads to an important practical difference in the resulting vector of regression coefficients ( $\gamma_\alpha$ ) and interpretation of the statistical model:

As  $\lambda$  increases, ridge regression shrinks regression coefficients, that is, reduces the magnitude of regression coefficients. In the case of orthonormal columns in  $\mathbf{X}$ , ridge regression shrinks coefficients proportional to  $\lambda$  (Hastie et al. 2001). In the case of correlated predictors in  $\mathbf{X}$ , ridge regression shrinks inverse proportionally to the variance of the principal components in  $\mathbf{X}$  and thus distributes regression coefficients between all predictors  $p$  (Hastie et al. 2001). In contrast, the lasso shrinks regression coefficients *and* performs feature selection (i.e., truncates below a given threshold). Intermediate values of  $\alpha$  ( $0 < \alpha < 1$ ) in the elastic net can be viewed as a combination of the lasso and ridge penalties (see also Hastie et al. 2001 for more details on the shrinkage). The bias obtained because of shrinkage is small if a good sparse linear approximation of the target variable exists (in the sense of a small  $L^1$  or  $L^2$  norm).

Different choices of  $\alpha$  account for possibly different assumptions around the data structure, data-generating processes, and interpretation of the regression model in the context of predicting a target climate variable based on a spatial field of predictors:

Ridge regression yields a smooth spatial map of nonzero regression coefficients. Ridge regression (or the elastic net with low  $\alpha$ ) would likely be more appropriate if all gridcell predictors are relevant for predicting the target variable, or if the circulation field would be contaminated with Gaussian noise for the prediction (e.g., in observations). Conversely, the lasso penalty would impose sparsity by choosing only a small subset of predictors (i.e., individual locations) and discarding the rest. The lasso penalty would yield a better prediction if only a few predictors would be relevant for the target

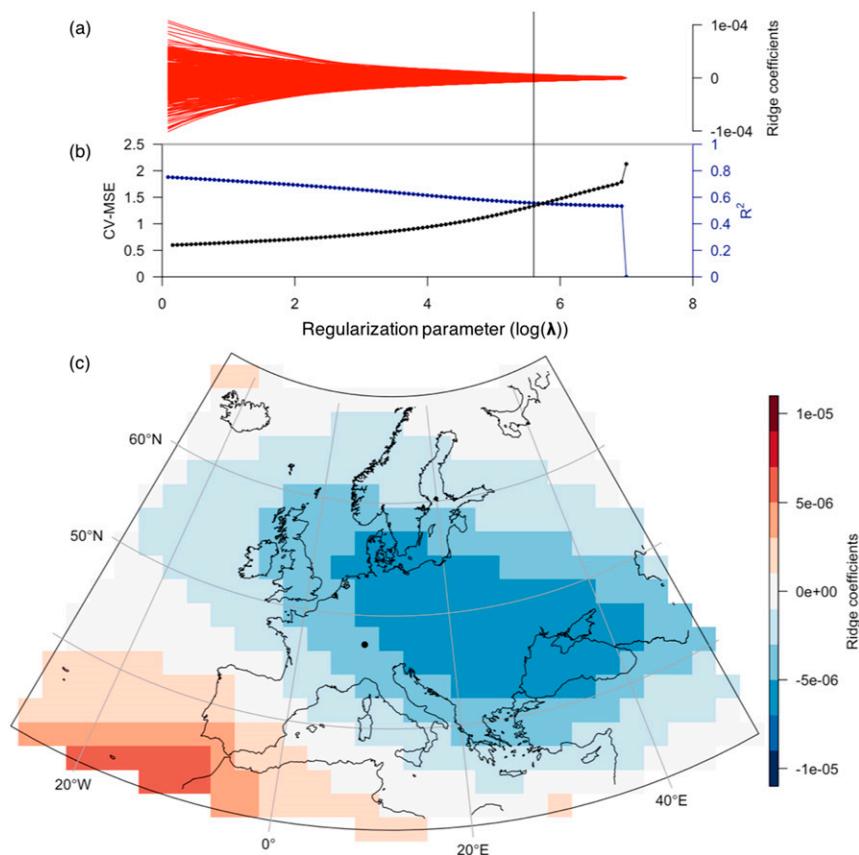


FIG. 2. Illustrative example of the ridge regression regularization path for estimating the effect of atmospheric circulation anomaly patterns on winter precipitation over northern Switzerland (black dot). (a) Individual grid cells' regression coefficients illustrate the nature of coefficient shrinkage; and (b) cross-validated MSE (and associated  $R^2$  values) along the regularization path indicated by  $\log(\lambda)$  values. (c) Map of ridge regression coefficients at one snapshot along the regularization path indicated by the thin black line above.

variable. The elastic net offers a compromise: coefficient shrinkage for a set of correlated predictors (e.g., a region of high or low SLP) is similar to ridge regression, that is, yielding smooth coefficients in space, while irrelevant predictors are eliminated. In this study, the parameter  $\alpha$  is set to three different values (ridge regression:  $\alpha = 0$ , the elastic net:  $\alpha = 0.5$ , and the lasso:  $\alpha = 1$ ).

The regularization or flexibility parameter  $\lambda$  is typically determined by cross validation (in this study: 10-fold cross validation). That is, the training dataset is partitioned into 10 folds that contain each 10% of the data. Next, for each fold, a sequence along 100 candidate  $\lambda$  values (regularization path) of RLMs is estimated based on the remaining 90% of the data. The regularization parameter  $\lambda$  is selected to minimize mean squared prediction error over the holdout data (i.e., the “unseen” 10% not used for model fitting) and all folds. An illustrative example of a regularization path and coefficient map for ridge regression is shown in Fig. 2 in the context of estimating the

influence of the winter sea level pressure on precipitation over a location in northern Switzerland (see an animation of the full regularization path in the online supplemental material). High values of  $\lambda$  yield very small coefficients and thus low predictive performance [high mean-square error (MSE), low  $R^2$ ], while low  $\lambda$  values allow for good performance but eventually begin to overfit. The example shows that the regression coefficients are 1) spatially coherent, 2) negative in a region located to the northeast, and 3) largely positive to the southwest of the location in Switzerland. Hence, these patterns capture a key feature of midlatitude weather variability in a data-driven way: a midlatitude synoptic-scale cyclone, centered over east-central Europe, and an associated high pressure system over southwest Europe and North Africa. Such a pressure pattern would be predicted to drive precipitation over the target grid cell, and could be interpreted physically as advection of moist Atlantic air into central Europe toward the rearward flank of the cyclone.

## 2) A NONLINEAR EXTENSION: REGULARIZED SINGLE-INDEX MODELS

RLMs cannot model a nonlinear relationship between a field of atmospheric circulation variables and a target climate variable. We therefore test a simple nonlinear extension of RLMs, following the idea of a “sparse single-index model” (Alquier and Biau 2013): The idea is to use the prediction from a RLM ( $\mathbf{X}\hat{\boldsymbol{\gamma}}$ ) and allow nonlinearity of the form

$$\mathbf{Y} = h(\mathbf{X}\hat{\boldsymbol{\gamma}}) + \varepsilon = h(\hat{\mathbf{Y}}_{\text{linear pred.}}) + \varepsilon. \quad (9)$$

Here, the nonlinear function  $h(\cdot)$  is determined by using a LOWESS smoother: fitting a locally weighted second-order polynomial (Cleveland et al. 1991) that is determined using the univariate predictor derived from the fitted RLM ( $\mathbf{X}\hat{\boldsymbol{\gamma}}$ ) and predictand  $\mathbf{Y}$ .

### 3. Data, methods, and experimental setup

#### a. Data

To evaluate statistical learning methods for dynamical adjustment and compare them against existing techniques, we apply them to a set of fully coupled climate simulations with the Community Earth System Model (CESM), version 1.2.2. These simulations consist of 1) a 4700-yr control run with constant preindustrial forcing, and 2) a 21-member ensemble simulation, with individual ensemble members that span 1850–2100 driven by historical CMIP5 forcing during 1850–2005, and RCP8.5 during 2006–2100 (Meinshausen et al. 2011). Each ensemble member initially branches from the control with different atmosphere and ocean initial conditions separated by 20-yr intervals. All simulations are run with an atmospheric resolution of  $1.9^\circ \times 2.5^\circ$  in space and 30 min in time, with 30 vertical levels using the Community Atmosphere Model, version 5 (CAM5.3; Neale et al. 2010). A  $1^\circ$  ocean grid with 60 levels in the vertical with the Parallel Ocean Program 2 (POP2; Smith et al. 2010) is used. CESM1 includes fully coupled atmosphere, ocean, sea ice, and land surface components (Hurrell et al. 2013; Meehl et al. 2013). The model setup has been used and evaluated in other studies (e.g., Stolpe et al. 2019).

#### b. Experimental setup

##### 1) PERFORMANCE EVALUATION IN A CLIMATE MODEL CONTROL RUN WITH FIXED FORCING

A performance evaluation for estimating temperature and precipitation variability from atmospheric circulation (see section 4a) is set up as follows:

- 1) 70 grid cells are selected randomly (weighted by area) as predictands from the European domain (see Table S1 in the online supplemental material), the SLP field around each predictand location is used as predictor.
- 2) Performance evaluation is conducted with respect to 1) temperature and precipitation as respective target variables, 2) different statistical methods (described in sections 2c and 3c), 3) the length of the available training sample (10–1000 yr), 4) the spatial domain size of the SLP field used for prediction ( $12^\circ \times 12^\circ$  up to  $60^\circ \times 60^\circ$  centered on the specific grid cell), for 5) daily versus monthly training data.
- 3) Anomalies are derived for each predictor and target time series relative to its respective monthly mean seasonal cycle. In addition to the instantaneously prevailing circulation anomalies on the specific day/month, lagged circulation values are included in  $\mathbf{X}$  for the antecedent month: That is, for training in 1) monthly and 2) daily resolution, the first 10 principal components of the 1) previous month and 2) averaged separately for 1–2, 3–7, and 8–30 lag days are included as predictors in  $\mathbf{X}$ . However, including lags changes the performance of the methods only very marginally.
- 4) Models are trained and cross validated on a period within the control run (years 500 to 1500), and evaluated in a different period (years 1600 to 2000).
- 5) Performance is evaluated against the “fraction of variance explained” by the prediction [ $R^2$ ; denoted here as squared sample correlation between the out-of-sample prediction ( $\hat{\mathbf{Y}}$ ) and the target variable ( $\mathbf{Y}$ )] at monthly resolution at each location. The measure is used in order to provide a standardized measure for performance that is comparable across regions and seasons.

In addition to these method tests over Europe, prediction performance of temperature and precipitation is illustrated at the global scale for the EOF benchmark and the regularized linear models using a 50-yr training period and a  $20^\circ \times 20^\circ$  training domain.

##### 2) DYNAMICAL ADJUSTMENT APPLICATION TO AN ENSEMBLE OF MODEL SIMULATIONS WITH TIME-VARYING FORCING

The configuration described above also forms the basis for a regional-scale evaluation of dynamically adjusted trend slopes in model simulations with time-varying forcing (section 4b) with only minor modifications:

- An estimate for the circulation-induced component of temperature and precipitation is obtained using RLMs at daily resolution in all grid cells within the European

domain [Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) regions (IPCC 2012), north Europe (NEU), central Europe (CEU), and Mediterranean (MED)]. Subsequently predictions are spatially and temporally aggregated to regional and seasonal averages.

- Two configurations are used to disentangle 1) the forced thermodynamic component and 2) the total forced response (cf. sections 2a and 2b). In 1), training and prediction variables are not detrended. In 2), to account for potentially confounding trends in atmospheric circulation, a high-pass locally weighted scatterplot smoothing (LOWESS) filter based on regression that is local in time (Cleveland et al. 1991) is applied prior to model training to remove low-variance components from each predictor time series (and predictand time series for training). LOWESS parameters are fixed at a window size of 75% of the time series length ( $n = 250$ yr) with a second-order polynomial fit. The detrending procedure removes variability at centennial time scales, but decadal variability remains largely unaffected.
- For the dynamical adjustment of one ensemble member  $m$ , all model training is conducted in an independent ensemble member  $m + 1$  and in a 50-yr period with historical time-varying forcing (1969–2018), to mimic an application to real-world observations.

The dynamically adjusted total forced response of temperature and precipitation at the global scale (section 4c) is configured similar to the regional evaluation. Predictions are derived at monthly resolution for each grid cell and subsequently aggregated to global annual averages. Because the globally aggregated predictions tend to underestimate interannual variability, the variance of the global mean prediction is adjusted in a final step by linearly regressing the detrended global mean target time series on the aggregated prediction:

$$\mathbf{Y}_{\text{global mean}} = \beta \underbrace{\hat{\mathbf{Y}}_{\text{global mean}}}_{\text{area-weighted sum over grid cell predictions}} + \varepsilon. \quad (10)$$

Because the prediction of global mean precipitation (in contrast to temperature) achieved through aggregation is poor (Fig. 9), the latter is improved by using ridge regression instead of simple linear regression in Eq. (10) within the training dataset.

Statistical learning is performed using the R packages “glmnet” (Friedman et al. 2010) for RLMS, and “pls” (Wehrens and Mevik 2007) for partial least squares (PLS) regression in R, version 3.4.3 (R Core Team 2018).

### c. Benchmark regression methods for dynamical adjustment

To benchmark the performance of the regularized linear models and the sparse single-index models (described in section 2c), we implement two regression methods used in earlier dynamical adjustment studies (Saffioti et al. 2016; Smoliak et al. 2015). In addition, an illustrative comparison to the performance of the circulation analog method is provided (Deser et al. 2016, described in the supplemental material).

#### 1) PRINCIPAL COMPONENT REGRESSION

Empirical orthogonal function (EOF) analysis is widely used in climate science (see, e.g., von Storch and Zwiers 2002). For example, Saffioti et al. (2016) determines the relationship in Eq. (1) by regressing the target climate variable onto the principal components that are associated with the first five EOFs in the circulation field  $\mathbf{X}$ . This procedure is almost identical to principal component regression described in Hastie et al. (2001) except that the number of principal components that are retained for regression is fixed instead of determined as a tuning parameter by cross validation. In this study, EOF-based dynamical adjustment with a fixed number of five EOFs retained (as in Saffioti et al. 2016) serves as the standard benchmark for comparison to the other methods.

EOF regression methods have the advantage of being related to modes of climate variability (e.g., the NAO; Hurrell 1995), and are thus interpretable in this context. However, methodological choices (such as domain size and number of modes used in the model) affect the performance skill of EOF regression methods. This is, in part, because higher-order SLP EOFs tend to capture small-scale features that are not representative of the overall circulation field, which can introduce spurious structure into the derived circulation-target variable relationship.

#### 2) PARTIAL LEAST SQUARES

Smoliak et al. (2015) used partial least squares regression to estimate the effect of circulation on target climate variables. Similarly to principal component regression, a regression step is performed based on a set of  $M$  derived orthogonal input directions (latent vectors). Each latent vector  $\mathbf{k}$  is constructed by successively weighting each normalized input variable by its correlation with the target variable, and subsequently the target variable is regressed on the respective latent vector  $\mathbf{k}$ ; followed by an orthogonalization step of the input variables with respect to the latent vector  $\mathbf{k}$ , from which the latent vector  $\mathbf{k} + 1$  is constructed (for details, see, e.g., Hastie et al. 2001; Smoliak et al. 2015).

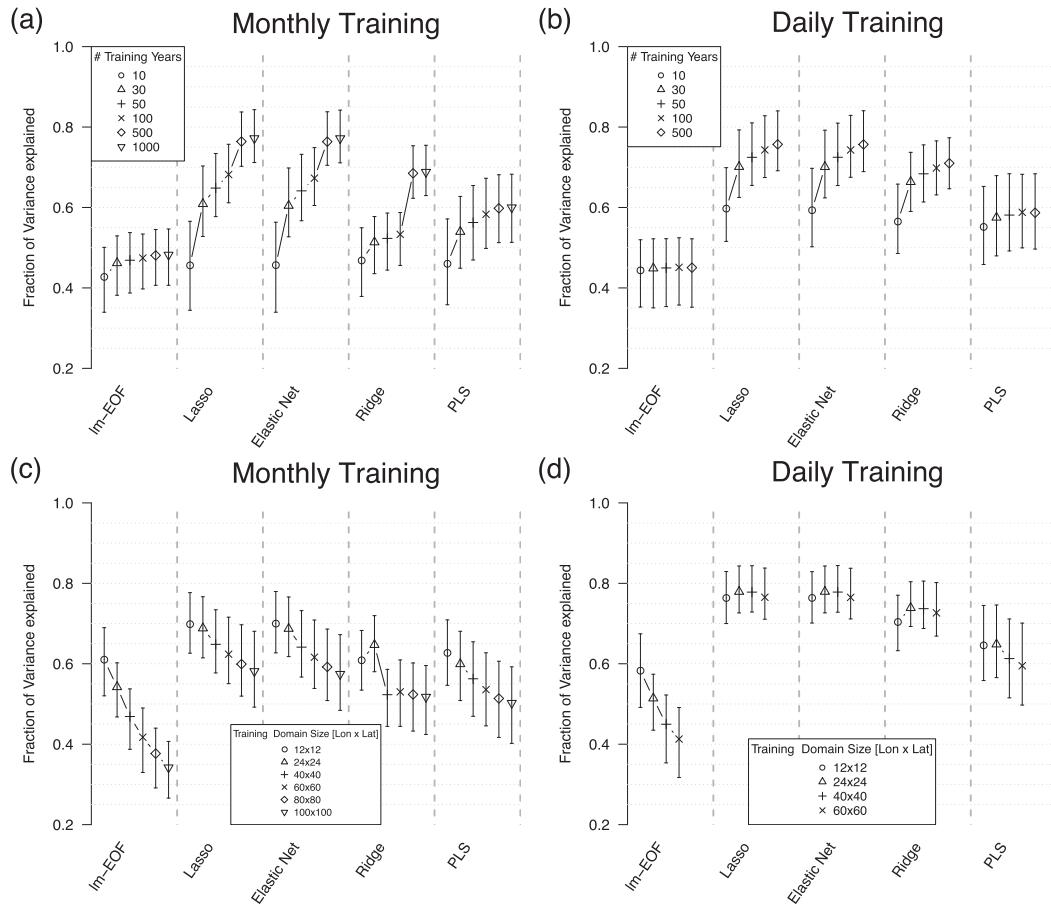


FIG. 3. Performance of statistical prediction for monthly precipitation over the European domain. Plots show the fraction of variance explained ( $R^2$ ) in an independent test period for different statistical methods used for prediction, showing effects of (a),(b) training sample length and (c),(d) spatial extent of training sample (50-yr training), and trained on (a),(c) monthly vs (b),(d) daily training data.  $R^2$  is calculated for comparability in both cases for monthly data; diamonds show averaged  $R^2$  values and error bars represent the 25th and 75th quantile across European locations.

PLS regression methods have the advantage of being more flexible than EOF methods, because their derived inputs are designed based on their relationship with the target variable. However, similar to EOF regression, PLS is based on a fixed cutoff  $k_{\max}$  (of input directions), and is potentially susceptible to overfitting as more latent vectors are incorporated (Smoliak et al. 2015).

#### 4. Results and discussion

In section 4a, we evaluate the performance of statistical techniques to estimate precipitation and temperature variability from atmospheric circulation in a climate model control simulation with fixed forcing (the key step in dynamical adjustment). The associated question of whether the dynamically adjusted residual trends of such predictions contain accurate estimates of the total

or thermodynamical forced components in model simulations with time-varying forcing is evaluated in section 4b and illustrated at the global scale in section 4c.

##### a. Performance evaluation of statistical learning methods for dynamical adjustment

###### 1) COMPARISON OF STATISTICAL METHODS AND INFLUENCE OF TRAINING CHARACTERISTICS OVER EUROPE

Across Europe, an average of 48% and 60% of the variance in monthly precipitation and temperature, respectively, is explained by SLP with the benchmark EOF method (“lm-EOF” in Fig. 3 and Fig. S1) at grid-cell scale in winter. These estimates are largely unaffected by training sample length and whether monthly or daily data are used for training (cf. Figs. 3a and 3b for

precipitation, Fig. S1 for temperature). For monthly training data and very short training samples (about 10 yr), RLMs show a performance comparable to the EOF benchmark (Figs. 3a, 4). However, increasing the sample length beyond 10 yr, RLMs show a remarkable performance increase, explaining on average approximately 78% and 83% of variance in monthly precipitation and temperature, respectively (Fig. 3a and Fig. S1). The performance of PLS lies in between the RLMs and the EOF-based benchmark. The circulation analog method performs slightly better than the EOF-based benchmark (Fig. 4). Moreover, if the daily training resolution is exploited, RLMs clearly outperform the EOF benchmark even for very short training sample lengths (Fig. 3b). For example, the performance obtained with a daily training dataset of only 30 yr in length, with around 70% variance explained, is only achieved with 100–500 yr of monthly training data for precipitation (Fig. 3). Hence, for training, daily data are much more efficient than monthly data, indicating a clear benefit of capturing circulation dynamics at daily time scales (rather than using “monthly mean weather”) for dynamical adjustment.

A nonlinear sparse single-index model improves prediction performance for precipitation slightly further, while this is not the case for temperature (“lasso+LOWESS” in Figs. 4a and 4b). This finding indicates that nonlinearities in daily circulation dynamics are more relevant for estimating the circulation-induced component in precipitation compared to temperature, which is likely related to the skewed precipitation distribution.

A relatively small domain appears preferable if only a short training record is available (here: 50 yr and monthly training resolution), as prediction performance decreases with increasing domain size (Fig. 3c) for all regression methods. However, the dependence on domain size is weak for RLMs if a long training sample or daily data are available (Fig. 3d), while the EOF-based method with fixed number of EOFs shows decreasing performance with larger domains. Overall, good performance for predicting precipitation can be achieved for relatively modest domain sizes (around  $24^\circ \times 24^\circ$ , centered on the respective grid cell), whereas slightly larger domains are preferable for temperature dynamical adjustment (around  $40^\circ \times 40^\circ$  in Europe; Fig. S1). Note that these domains are comparable in size to those in a “weather type” classification scheme for regional downscaling (Boé et al. 2006), but smaller than domain sizes previously used for dynamical adjustment (Wallace et al. 2012; Smoliak et al. 2015; Deser et al. 2016).

Overall, our results indicate that RLMs in combination with daily data for model training and prediction

improve estimates of circulation-induced temperature or precipitation variability. The main advantage of RLMs (over the EOF benchmark, for example) lies in their flexibility to capture locally relevant information about the circulation field that may not be captured by the first few EOFs (or latent vectors in PLS). For dynamical adjustment based on training data of 30–50 yr (comparable to observations), estimates from RLMs benefit considerably from exploiting daily circulation data for training. Accordingly, using RLMs with daily data might open up the possibility for dynamical adjustment at daily time scales, that is, targeting actual weather situations rather than only monthly mean weather.

Furthermore, the results confirm that there is a relatively weak dependence of prediction performance on domain size for the flexible regression methods if enough training data, especially at the daily time scale, are available. In contrast, the EOF benchmark suffers severely from a large domain size, irrespective of training sample length, as the locally relevant information encapsulated in the first few EOFs decreases with increasing domain size.

## 2) ESTIMATING CIRCULATION EFFECTS AT THE GLOBAL SCALE

Next, we illustrate the ability to explain temperature and precipitation variability based on atmospheric circulation at the global scale, which is a prerequisite for the application of dynamical adjustment beyond the midlatitudes. The global area-weighted average of monthly precipitation variance explained at gridcell level is about  $R^2 \approx 0.62$  for the lasso and elastic-net regression (and slightly lower for ridge regression) for both DJF and JJA (Figs. 5b,d). Using a single-index model based on lasso regression (lasso+LOWESS) yields a slightly higher fraction of variance explained  $R^2 \approx 0.67$ . The standard EOF-based benchmark yields on average  $R^2$  values around only 0.22 (DJF) and 0.3 (JJA). Circulation-induced “predictability” obtained in Northern Hemisphere (NH) winter (Fig. 5a) and summer (Fig. 5c) shows distinct but climatologically consistent differences: In boreal winter, atmospheric circulation explains a large fraction of precipitation variability in NH mid- and high-latitude regions, especially along continental west coasts and adjacent oceanic areas. In contrast, precipitation in dry regions such as the Sahara or the Arabian Peninsula is not explained well by atmospheric circulation, and low predictability extends to the Southern Ocean and southeastern Australia. In boreal summer in mid- and high-latitude regions, precipitation is generally explained better in the SH than in the NH. Predictability is low in the SH over the Namib

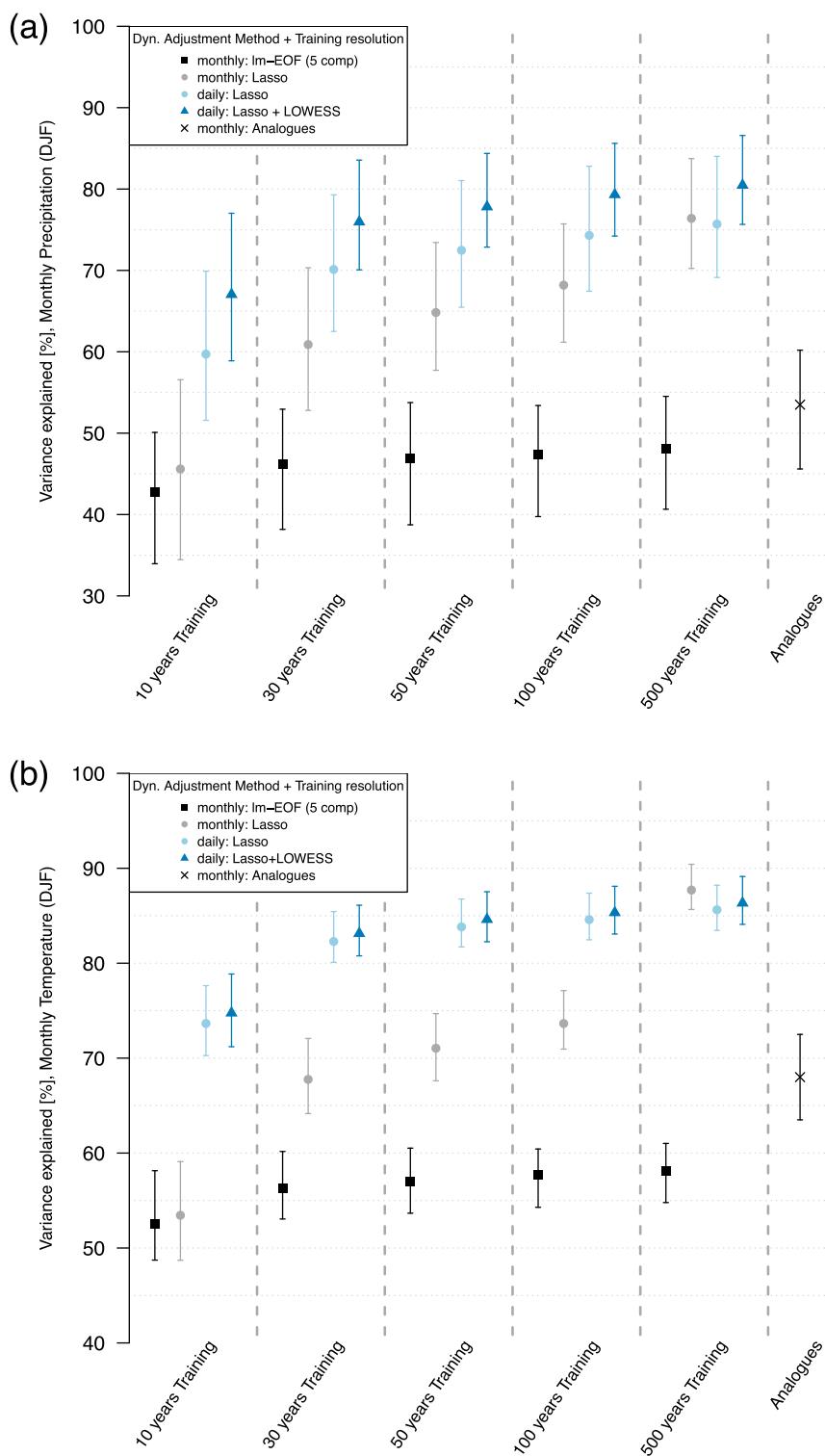


FIG. 4. Prediction performance for (a) monthly precipitation and (b) monthly temperatures in winter over the European domain. The performance of regularized linear models is compared to the EOF-based benchmark regression method (“lm-EOF, 5 comp.”) for training on monthly and daily data and different lengths of the training sample. The  $R^2$  values based on atmospheric circulation analogs (Deser et al. 2016) are shown for comparison.

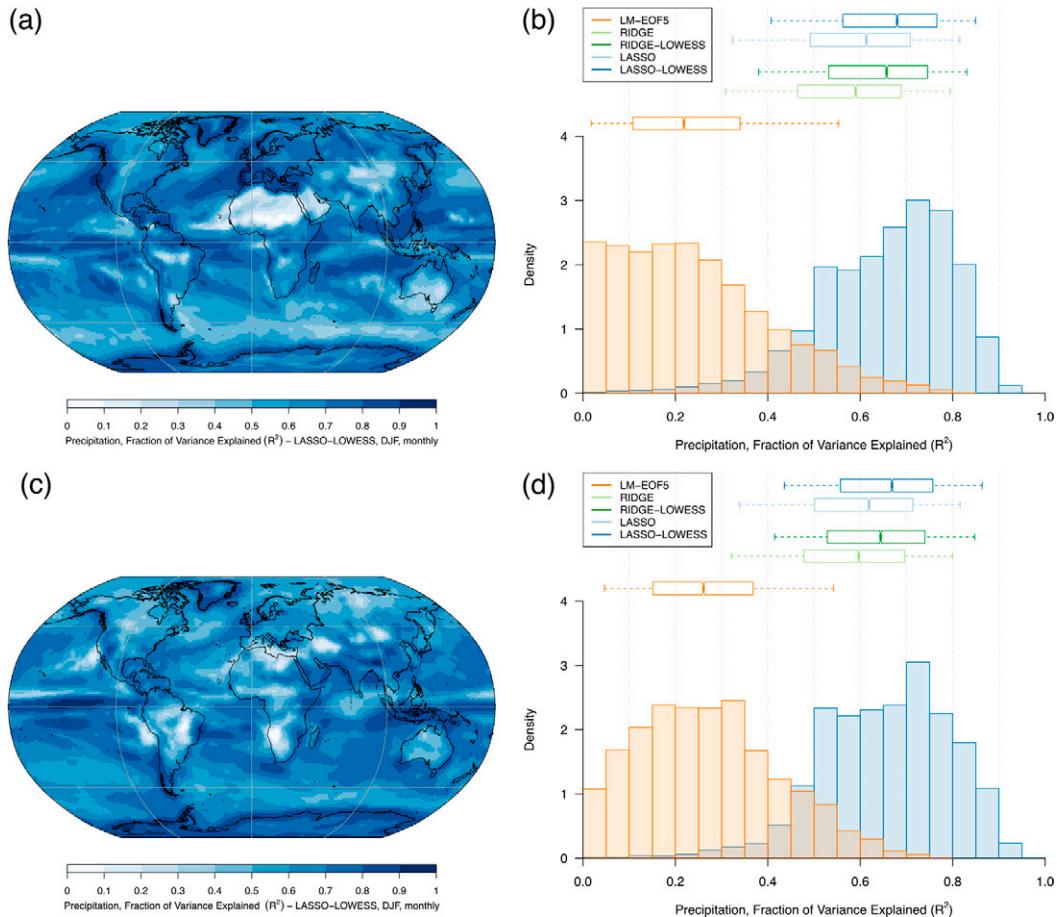


FIG. 5. (a),(c) Performance of monthly precipitation predictions based on regularized linear models in (a) Northern Hemisphere winter (DJF) and (c) Northern Hemisphere summer (JJA). (b),(d) Comparison of prediction performance using a standard EOF-based method (“lm-EOF,” 5 comp.), and the performance of regularized linear models, all based on daily training data, in (b) Northern Hemisphere winter (DJF) and (d) Northern Hemisphere summer (JJA).

Desert and over central South America. In the NH, dry regions such as the Mediterranean and inside continental interiors (central North America, central Eurasia) are not very well explained from atmospheric circulation. Overall, circulation-induced components are more important in winter. In dry regions, the vertical structure of the atmosphere is likely important, hence including vertical information would potentially benefit the ability of circulation to explain precipitation variability (but is beyond the scope of the present study).

For temperature anomalies, monthly circulation-induced predictability is slightly higher than for precipitation ( $R_{DJF}^2 \approx 0.73$ ,  $R_{JJA}^2 \approx 0.67$ ; Fig. 6). The EOF-based benchmark method explains only around 30% of the variance in monthly temperatures. The climatological patterns of predictability are noticeably different: Land regions appear generally more driven by atmospheric

circulation than oceanic regions with the exception of temperatures in the equatorial ENSO region (which are also well predicted). Low predictability of temperatures over the ocean is likely associated with memory effects due to ocean energy storage and lagged ocean–atmosphere interaction (Deser and Timlin 1997), which is not captured by instantaneous circulation anomalies (or the first few lagged EOFs).

In summary, atmospheric circulation patterns can be used to estimate temperature or precipitation variability consistently at the global scale using RLMs, although performance varies depending on background climate. Accordingly, dynamical adjustment applications could be extended in the future beyond midlatitude land areas, for instance, toward ocean regions where including lagged circulation information would become a key requirement (Deser and Timlin 1997).

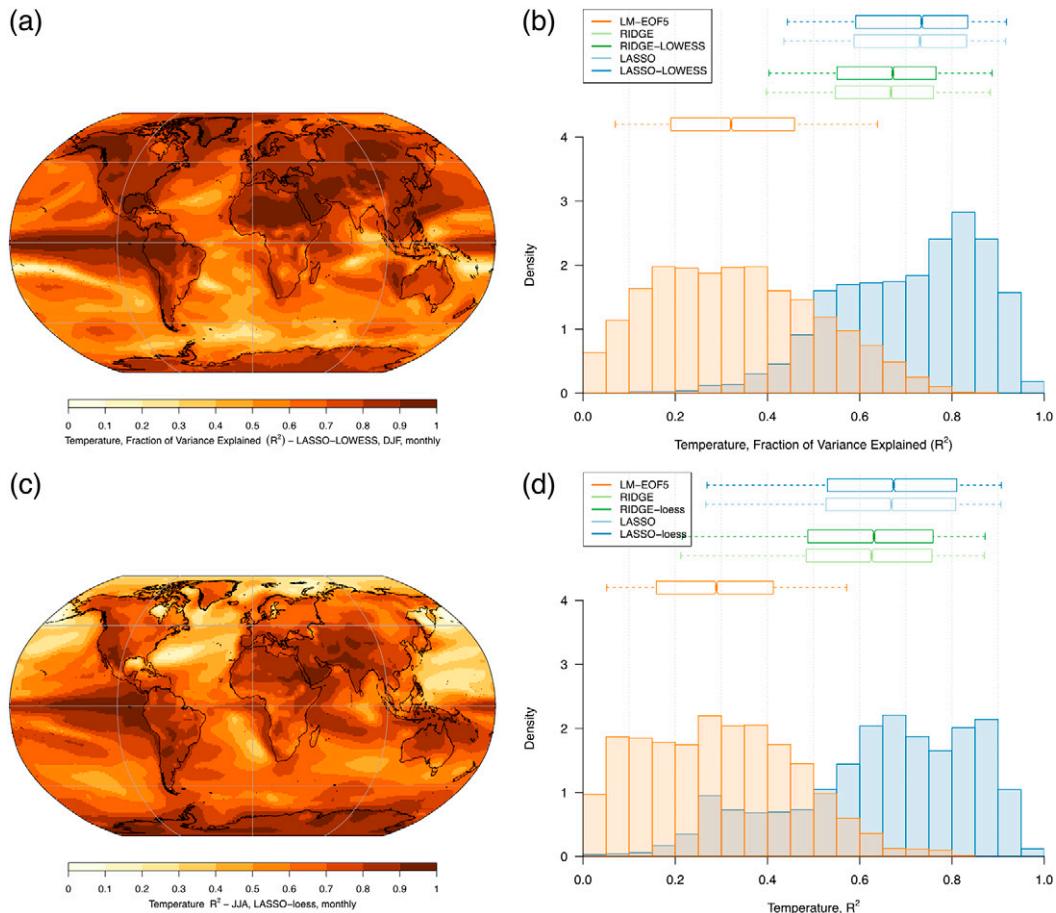


FIG. 6. As in Fig. 5, but performance evaluation for temperature instead of precipitation.

*b. Dynamical adjustment at the regional scale over Europe: Uncovering the total forced response and dynamic versus thermodynamic components*

In this subsection, we evaluate the consistency of the residual trend slopes obtained by dynamically adjusting all 21 ensemble members over Europe to uncover 1) the total forced response (evaluated against the 21-member ensemble mean “total forced” response), and 2) thermodynamically and dynamically driven components of the total forced response.

1) UNCOVERING THE TOTAL FORCED RESPONSE WITH IMPROVED SIGNAL-TO-NOISE RATIO

Winter precipitation over both NEU and MED is highly variable from year to year (Figs. 7a and 8a, illustrated for the first three ensemble members, and for all 21 members for NEU and MED, respectively). In addition, NEU shows an externally forced increasing trend in the twenty-first century, while winter precipitation over the MED region is projected to decrease in

the ensemble mean forced signal (Figs. 7b, 8b). Nonetheless, individual ensemble members deviate from the ensemble mean, which is evident in the spread of 30-yr trend slopes in individual ensemble members in both regions (bottom panel in Figs. 7 and 8).

The prediction of winter precipitation based on detrended circulation anomalies explains the bulk of year-to-year variability in both regions (NEU:  $R^2 = 0.92$ , MED:  $R^2 = 0.90$ , evaluated against detrended time series, Table 1). However, long-term trends (i.e., spanning several decades) remain unexplained by detrended circulation anomalies in both regions (light blue lines in Figs. 7a,c,e). Hence, precipitation trends must result from a forcing external to year-to-year circulation variability. The dynamically adjusted residual time series contain a smooth trend toward increasing precipitation in NEU, and a smooth drying trend in MED, which both match the ensemble mean estimate of the total forced response well (blue lines in Figs. 7b and 8b).

To evaluate the consistency of the residual trend slopes for RLMs (shown in Figs. 7 and 8),  $t$  tests are

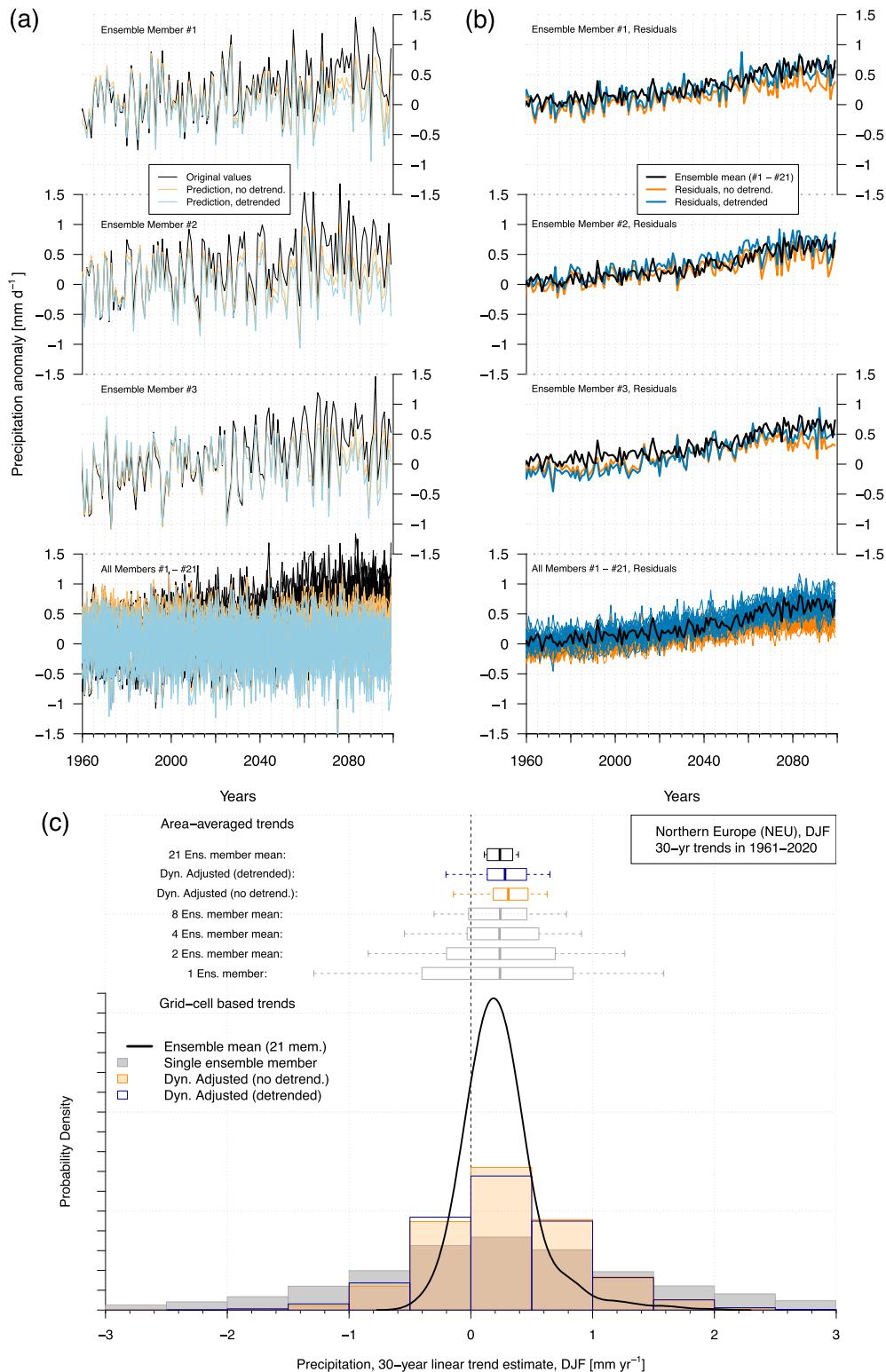


FIG. 7. Dynamical adjustment of winter precipitation in northern Europe in forced runs in the 21-member CESM ensemble. (a) Original time series and predictions  $\hat{Y}_x$  (“no detrend.”) and  $\hat{Y}_{x_{\text{high-freq.}}}$  (“detrended”) for three individual ensemble members (1–3) and all members. (b) Residual, dynamically adjusted time series for individual ensemble members (1–3), and for all ensemble members in comparison to the ensemble mean forced response (black). (c) Northern Europe 30-yr trend slopes for dynamically adjusted time series of winter precipitation in the 1961–2020 forced period, compared to the ensemble mean.

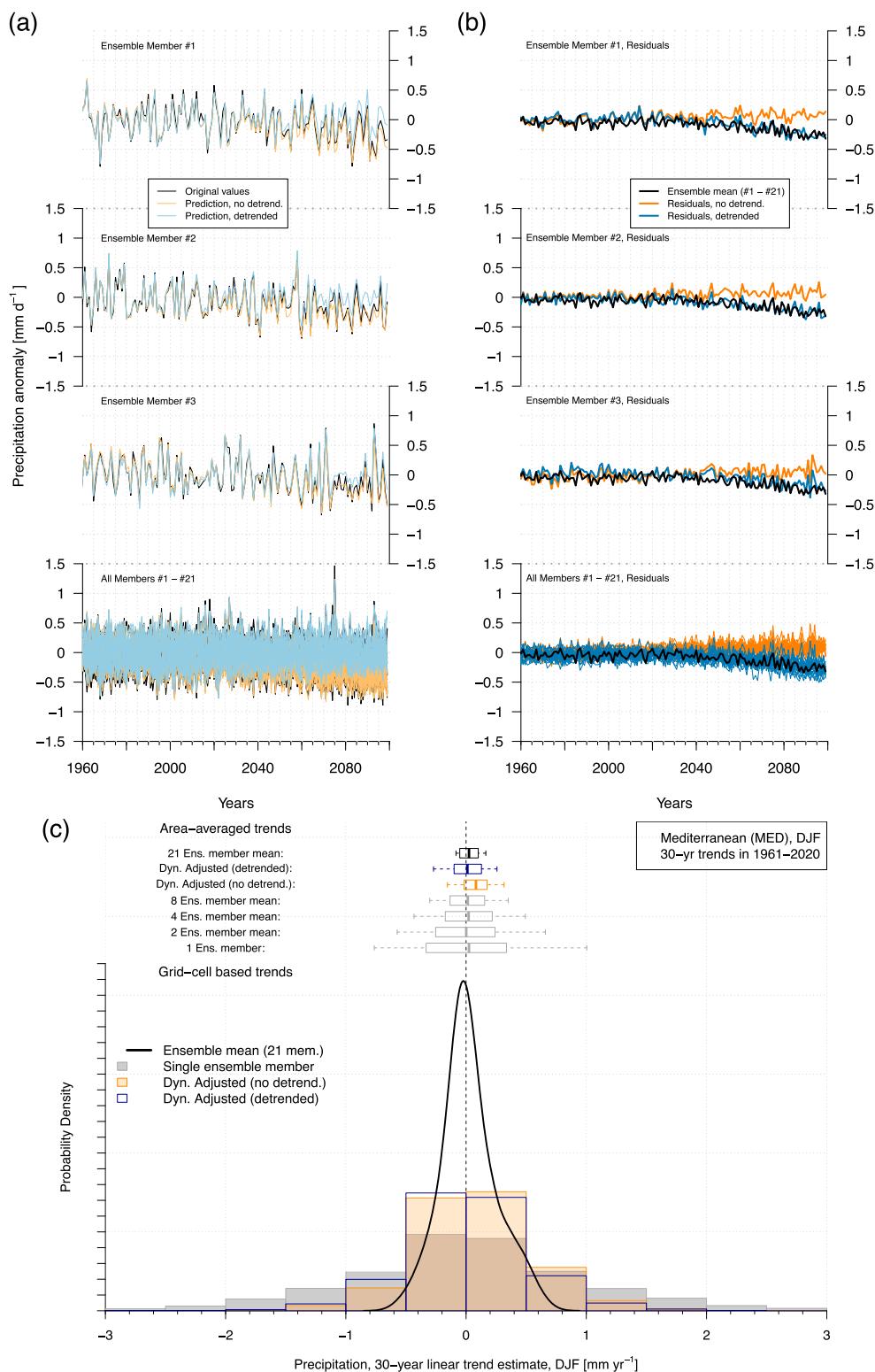


FIG. 8. As in Fig. 7, but dynamical adjustment for Mediterranean winter precipitation in the 21-member CESM ensemble.

TABLE 1. Regional-scale dynamical adjustment over Europe.

Variable	Season	Region	$R^2$ (against detrended series)	TC <sup>a</sup> (1961–2020)	TC <sup>a</sup> (2001–60)	TC <sup>a</sup> (2031–90)	PP <sup>b</sup> (original, 1961–2020)	PP <sup>b</sup> (dynamically adjusted, 1961–2020)	ToE <sup>c</sup> (original, one std dev)	ToE <sup>c</sup> (adjusted, one std dev)
Precipitation	DJF	NEU	0.92	0.73	0.26	0.60	0.15	0.61	2062	2011
Precipitation	DJF	CEU	0.87	0.72	0.71	0.76	0.09	0.46	2058	2020
Precipitation	DJF	MED	0.90	0.84	0.91	0.90	0.03	0.21	—	—
Precipitation	DJF	EU	0.74	0.75	0.19	0.54	0.29	0.52	2056	2019
Precipitation	JJA	NEU	0.86	0.33	0.27	0.29	0.41	0.75	—	2005
Precipitation	JJA	CEU	0.75	0.85	0.26	0.83	0.10	0.32	—	—
Precipitation	JJA	MED	0.85	0.93	0.26	0.17	0.04	0.26	—	—
Precipitation	JJA	EU	0.64	0.46	0.46	0.04*	0.43	0.57	—	2017
Temperature	DJF	NEU	0.81	0.58	0.40	0.69	0.42	0.75	2032	1997
Temperature	DJF	CEU	0.83	0.67	0.37	0.46	0.38	0.75	2031	2004
Temperature	DJF	MED	0.83	0.94	0.88	0.49	0.79	0.93	2010	1998
Temperature	DJF	EU	0.79	0.63	0.34	0.42	0.66	0.87	2012	1997
Temperature	JJA	NEU	0.59	0.87	0.26	0.60	0.83	0.89	1998	1993
Temperature	JJA	CEU	0.66	0.61	0.68	0.87	0.87	0.92	1999	1994
Temperature	JJA	MED	0.77	0.25	0.45	0.05	0.91	0.95	1999	1991
Temperature	JJA	EU	0.58	0.59	0.81	0.21	0.93	0.95	1994	1987

<sup>a</sup>TC: Trend consistency of residual trend slopes. Columns give  $p$  values of  $t$  test of 30-yr residual trend slopes in 21-member ensemble against the original collection of trend slopes; significant deviations at  $\alpha = 0.05$  are marked with an asterisk.

<sup>b</sup>PP: Potential predictability is a metric that determines the fraction of predictability of decadal mean climate variable in a single climate system realization from the forced response alone [see Stott et al. (2000)], defined as  $\{\text{var}(\text{ens.mean}) - \text{var}(\text{control})/n\} / \{\text{var}(\text{ens.mean}) + \text{var}(\text{control})\}$ ;  $\text{var}(\text{ens.mean})$  is the variance of an ensemble mean of size  $n$ , and  $\text{var}(\text{control})$  is the variance in a control run.

<sup>c</sup>ToE: Time of emergence is given as a time when the long-term trend exceeds one standard deviation of variability in a preindustrial control period.

performed between 30-yr linear trend slopes calculated from the original precipitation time series from each ensemble member and the residual trend slopes obtained after dynamical adjustment for three different periods (present, near future, end of century). The null hypothesis of statistically indistinguishable trend slopes cannot be rejected on a  $\alpha = 0.05$  confidence level for any time period in NEU or MED. In fact, the null hypothesis is rejected in only one out of the 48 tests (for summer/winter, temperature/precipitation, the three European regions and three time periods; Table 1). Hence, we conclude that dynamical adjustment using RLMs based on detrended circulation anomalies yields an accurate estimate of the total forced response. However, unlike RLMs, the residual trend slopes for precipitation obtained from the nonlinear estimates (using the sparse single-index model) deviate slightly from the ensemble mean toward the end of the twenty-first century (not shown). This could be due to some higher-order changes in the sea level pressure patterns that are not removed by the detrending procedure, and we therefore focus on the regularized linear models for dynamical adjustment for the remainder of the manuscript.

Signal-to-noise ratios increase substantially in the dynamically adjusted residual trend slopes (Figs. 7b, 8b) because a large fraction of circulation-induced variability is accounted for. Consequently, the forced trend emerges more clearly, and uncertainties related to 30-yr trend slopes decrease considerably (lower panels in Figs. 7 and 8). For instance, the spread around dynamically adjusted residual 30-yr trends from a single ensemble member is reduced to approximately the spread of a hypothetical ensemble mean of 4–8 members (lower panels in Figs. 7 and 8). Climatological metrics that are linked to signal-to-noise ratios such as the “time of emergence” (cf. Hawkins and Sutton 2012; Lehner et al. 2017) shift earlier. For instance, a “ $1\sigma$  emergence” in NEU precipitation occurs in approximately 2011 after dynamical adjustment instead of 2062, while the MED forced trend is evident in the residual time series but unlikely to emerge before 2100 (Table 1). Among the regions documented here, precipitation signals are likely to emerge before 2020 (at a  $1\sigma$  level) for NEU and CEU in winter, and for NEU in summer; whereas without dynamical adjustment these forced signals would only emerge in the second half of the twenty-first century. The forced temperature trend emerges in all of the seasons and regions before 2005. Last, “potential predictability,” the predictability of a decadal mean climate variable in one ensemble member from the forced response alone (Stott et al. 2000; Lambert et al. 2004, see Table 1 herein for detailed definition), increases across all regions and seasons (Table 1) with temperature

and precipitation predictability in the ranges of 75%–95% and 21%–61%, respectively, in the 1960–2020 period.

Overall, our analysis shows that the residual trend slopes after dynamical adjustment based on linear statistical learning techniques reveal 1) an accurate estimate of the total forced response for all tested regions, seasons and variables that we test, and 2) considerably reduced uncertainties around the forced trend signal. This example uses a training step within a single ensemble member and a short training period (50 yr). Hence, applying the technique to observations would be straightforward.

## 2) DATA-DRIVEN PARTITIONING OF THE TOTAL FORCED RESPONSE INTO THERMODYNAMIC AND DYNAMIC COMPONENTS

Dynamical adjustment was designed to disentangle thermodynamical and dynamical components of the forced response (Deser et al. 2016). Here, we illustrate thermodynamical and circulation-induced contributions to forced trends in simulated European precipitation using the data-driven approach (described in sections 2a and 2b). That is, 1) thermodynamical components and 2) the total forced response are estimated, by subtracting from precipitation time series (i) the predicted total (i.e., forced and internal) circulation-induced component and (ii) the predicted circulation-induced component with low-variance circulation components removed.

For NEU winter precipitation, estimates of the forced thermodynamical component (orange lines in Fig. 7b) and the total forced response (blue lines) in the residuals produce virtually identical trend slopes. This indicates that forced trends in seasonal winter precipitation over NEU are driven by thermodynamical changes with only a minor contribution of forced circulation-induced changes. This result is consistent with a thermodynamical increase in the atmospheric moisture holding capacity that drives increases in NEU precipitation over the twenty-first century (see, e.g., Bindoff et al. 2013). In contrast, winter precipitation over the MED region throughout the twenty-first century shows a slightly negative forced trend (Fig. 8b). However, contrary to NEU, nondetrended circulation anomalies capture Mediterranean drying remarkably well, whereas circulation anomalies with low-variance components removed predict no trend. Hence, the forced thermodynamical component in the residuals reveals no change or only a weak change (orange lines in Fig. 8b), whereas the estimated total forced response matches the ensemble mean very well (Fig. 8b). This indicates that projected decreases in Mediterranean precipitation in the twenty-first century result from

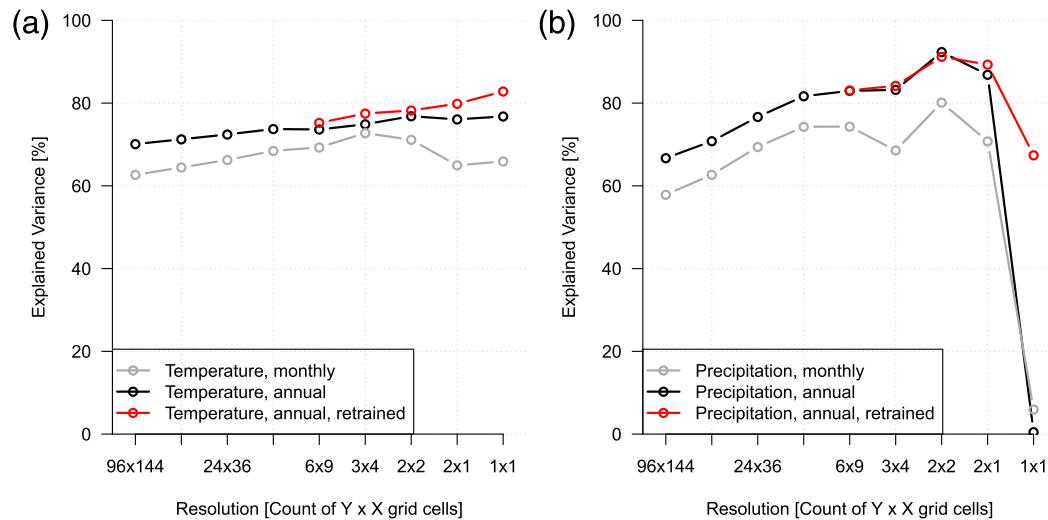


FIG. 9. Aggregation of gridscale predictions (in original  $96^{\circ} \times 144^{\circ}$  resolution) of (a) temperature and (b) precipitation to regional, continental, and global-scale predictions, and from monthly to annual, increases the fraction of variance that can be explained by atmospheric circulation. For continental-scale and global-scale predictions, a second linear training step based on the continentally/globally aggregated predictions improves the predictions slightly further.

externally forced changes upon atmospheric circulation, consistent with theoretical expectations and earlier literature (Yin 2005; Seager et al. 2014; He and Soden 2017; Kröner et al. 2017).

In summary, the simple approach illustrated here successfully partitions precipitation change over Europe into thermodynamical and dynamical components. The partitioning indicates that projected Mediterranean drying in the twenty-first century arises from forced circulation changes, whereas increased winter precipitation in northern Europe is driven by forced thermodynamical changes associated with an increase in atmospheric moisture in a warmer climate. However, a few caveats should be noted: First, uncovering the total forced response through high-pass filtering circulation patterns relies on an assumption that forced circulation trends are smooth, and that internal and forced components are additive (section 2b). Depending on the properties of the high-pass filter, multidecadal or centennial variability in atmospheric circulation could be removed, and hence would be interpreted as a part of the residual time series. Second, it remains to be tested whether such an approach could be extended to other regions or quantities such as daily extremes. Third, potential future changes in the *physical relationship* between circulation anomalies and precipitation, for example, through shifts in mean versus extreme precipitation (Allen and Ingram 2002) or increased precipitation variability (Pendergrass et al. 2017) would not be captured as a circulation component, and any such

changes would remain in the residuals (i.e., a thermodynamical component).

### c. Can dynamical adjustment reveal the total forced response at the global scale?

In a final step, we evaluate to what extent the total forced response in global annual mean precipitation (GMP) and temperature (GMT) can be revealed by dynamical adjustment from an individual ensemble member. GMP is indicative of changes in the global hydrological cycle as driven, for instance, by aerosol or greenhouse gas forcing (Allen and Ingram 2002; Salzmann 2016), and therefore comprises a relevant target and test case for dynamical adjustment.

Average performance increases through spatial and temporal aggregation (Fig. 9) up to almost 80% variance explained for GMT and almost 90% variance explained for precipitation at the hemispheric scale (Fig. 9). GMP cannot be predicted from an aggregation of gridcell predictions (Fig. 9) because the magnitude of GMP variation is much smaller than hemispheric precipitation variability as Northern and Southern Hemisphere precipitation anomalies compensate each other (not shown). A second training step (instead of simple aggregation; section 2c) improves the prediction of global mean precipitation (i.e., about 70% variance explained, red line in Fig. 9).

GMP, along with its estimated circulation-induced components, is illustrated for the first two ensemble members during 1950–2020 in Fig. 10a. Both ensemble

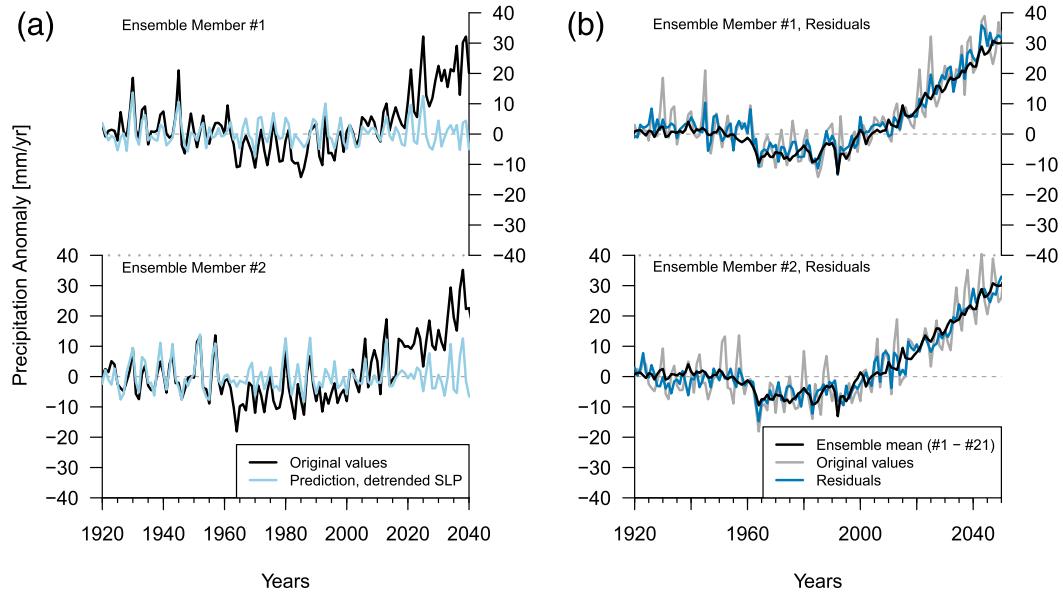


FIG. 10. Global dynamical adjustment of precipitation (a) accounts for interannual variability in an individual simulation, and (b) reveals an estimate of the total forced response from an individual ensemble member that closely resembles the ensemble mean.

members show considerable interannual variability and forced trends. The GMP total forced response, illustrated by the 21-member ensemble mean (black lines in Fig. 10b), shows a small reduction during 1960–90. This is consistent with increased aerosol forcing that counteracted GHG forcing (Salzmann 2016), and radiatively induced effects that dominated over changes induced by warming (Myhre et al. 2018). This period is followed by a relatively rapid forced GMP increase. The estimated circulation-induced component matches interannual precipitation variability to a reasonable degree (blue lines in Fig. 10a). The residual, dynamically adjusted time series for both ensemble members (blue lines in Fig. 10b) matches the ensemble mean estimate of the forced response closely (Table 2 for comparison of trend slopes), which indicates that the total forced GMP response can

indeed be uncovered as a residual from dynamical adjustment. Moreover, the GMP response to short-term external forcing, for instance, induced by major volcanic eruptions (e.g., Mount Agung in 1963, Mount Saint Helens and El Chichón in 1980 and 1982, and Mount Pinatubo in 1992; Robock 2000), appears to be well captured by a single, dynamically adjusted ensemble member (Fig. 10b).

GMT in individual ensemble members is not as dominated by interannual variability as GMP, but variability around the forced trend is nonetheless considerable (Fig. 11a). The estimated circulation-induced component (Fig. 11a) captures most interannual GMT variation. The dynamically adjusted time series (Fig. 11b) aligns closely with the ensemble mean, and thus appears to accurately reflect the total forced response. Similarly to GMP, short-term cooling in response to volcanic eruptions

TABLE 2. Global-scale dynamical adjustment.

Variable	Season	Region	$R^2$	TC <sup>a</sup> (present)	TC <sup>a</sup> (near future)	TC <sup>a</sup> (future)	PP <sup>b</sup> (original)	PP <sup>b</sup> (adjusted)	ToE <sup>c</sup> (original, one std dev)	ToE <sup>c</sup> (adjusted, one std dev)
Temperature	Annual	Global	0.78	0.11	0.61	0.03	0.83	0.92	1989	1982
Temperature	Annual	NH	0.82	0.35	0.90	0.09	0.88	0.98	1984	1981
Temperature	Annual	SH	0.82	0.36	0.89	0.09	0.73	0.95	1990	1982
Precipitation	Annual	Global	0.65	0.73	0.88	0.85	0.40	0.69	2005	2000
Precipitation	Annual	NH	0.88	0.44	0.92	0.27	0.10	0.61	2018	2005
Precipitation	Annual	SH	0.94	0.48	1.00	0.43	0.02	0.54	2048	2003

<sup>a</sup> TC: Trend consistency of residual trend slopes (see Table 1 for additional explanations).

<sup>b</sup> PP: Potential predictability.

<sup>c</sup> ToE: Time of emergence.

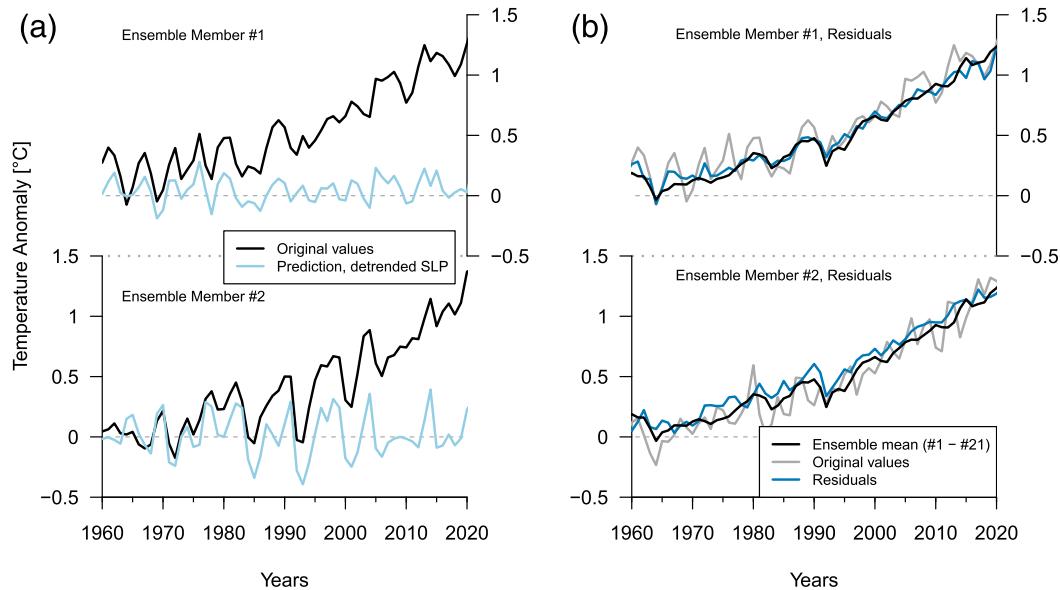


FIG. 11. Global dynamical adjustment of temperature (a) accounts for interannual variability in an individual simulation, and (b) reveals an estimate of the total forced response from an individual ensemble member that closely resembles the ensemble mean.

appears to be captured in the dynamically adjusted members, while in unadjusted members such external cooling signals are dominated by internal variability.

Overall, we conclude that dynamical adjustment using statistical learning methods successfully uncovers accurate estimates of the GMP and GMT total forced response from a single ensemble member. This implies that the ensemble size required to extract the total forced climate response can be reduced substantially. For an arbitrary ensemble member, the time of emergence of global or hemispheric climate signals is advanced and potential predictability increases (Table 2). Furthermore, dynamical adjustment at the global scale reveals thermodynamical responses to short-term external forcing such as induced by volcanoes from a single ensemble member (Figs. 10, 11). However, an important caveat is that forced circulation trends are assumed to be smooth in time and also additive, and are removed from the circulation field prior to dynamical adjustment. This assumption implies that, first, circulation variability at centennial time scales is reduced. Second, unlike short-term forced thermodynamical components, any short-term forced circulation component would likely be included as part of internal circulation variability.

## 5. Conclusions

The overall objective of this study was to incorporate statistical learning principles into dynamical adjustment. Specifically, our goals were 1) to evaluate statistical

estimates of circulation-induced components of precipitation and temperature variability at local, regional, and global scales; and 2) to evaluate whether the externally forced component of these responses can be isolated.

First, we framed dynamical adjustment conceptually in the context of statistical learning and introduced a specific set of statistical learning techniques to apply regularized linear models and their nonlinear extension. Then, we evaluated the performance of regularized linear models in estimating the relationship between atmospheric circulation and temperature/precipitation in a long climate model control run. RLMs, along with optimizing temporal and spatial characteristics of the training domain, provide a parsimonious representation and considerably improved performance. For instance, an average of 83% and 78% variance is explained by atmospheric circulation in monthly winter temperature and precipitation, respectively, at gridcell scale in Europe. Similarly, around 80% of the variance can be explained by atmospheric circulation for annual mean temperature and large-scale precipitation. By reducing training resolution from monthly to daily data, the available training sample is used much more efficiently. For instance, 30–50 yr of training on daily European winter precipitation data (i.e., comparable to the length of observational records) achieve similar performance to a sample of around 100–500 yr if only monthly time resolution is exploited for training.

Second, we evaluated the consistency of the “dynamically adjusted” residual trend slopes in a set of

simulations where forcing varies over time, with a 21-member ensemble. Our evaluation focused both regionally over Europe and at the global scale, comparing the total forced response from dynamical adjustment against an estimate from the ensemble mean. We show that residual trend slopes of temperature and precipitation from dynamical adjustment represent an accurate estimate of the total forced climate response both regionally and globally. More nuanced thermodynamical changes related to, for example, short-term volcanic forcing, also appear to be captured from a single dynamically adjusted ensemble member. Signal-to-noise ratios and related potential predictability metrics improve considerably (consistent with earlier studies; Deser et al. 2016; Lehner et al. 2017). For instance, uncertainties in forced 30-yr precipitation trends in Europe are reduced through dynamical adjustment and approach the uncertainty of trend slopes obtained from a 4–5-member ensemble mean. A similar reduction of minimum-required ensemble size can be achieved at the global scale.

Last, we illustrated for regional precipitation over Europe that a data-driven partitioning of the total forced response into its dynamic and thermodynamic components might be achieved via estimating 1) internal, high-pass-filtered circulation components (a total forced response estimate is maintained as a residual), and 2) the total contribution of circulation (forced thermodynamic components are maintained as a residual). Our results illustrate that increases in northern European winter precipitation are driven by thermodynamical changes (i.e., a warming-induced increase in the moisture-holding capacity of the atmosphere). In contrast, projected twenty-first-century decreases in Mediterranean winter precipitation are largely driven by forced circulation changes, consistent with theoretical expectations and earlier studies (Seager et al. 2010; He and Soden 2017; Kröner et al. 2017). The partitioning method relies on the assumption that 1) the high-pass filtering of the SLP field removes forced circulation components, and 2) future changes in circulation patterns project onto present-day patterns.

Overall, our study has shown that incorporating statistical learning techniques into dynamical adjustment helps to better constrain the externally forced response at local, regional, and global scale. We anticipate that the methodology evaluated in this study might enable a number of climate science applications.

An improved understanding and interpretation of local or regional (hydro-) climatic variability, trends, and their drivers could be achieved through application to the observed record. Increased signal to noise in observations (and model simulations) will further reduce

uncertainties around internal climate variability and benefit climate change detection and attribution. Our results point at possible extensions of the dynamical adjustment framework toward continental or global scales, oceanic regions, and the possibility of adjusting daily climate variability.

Furthermore, the methodology could enable an evaluation of forced versus internal components of climate model simulations against observations, for instance, in the context of CMIP6 (Eyring et al. 2016). This could include evaluation of short-term responses such as those induced by volcanic eruptions, or data-driven partitioning of the total forced response into thermodynamical and dynamical components.

*Acknowledgments.* We thank Eniko Székely, Clara Deser, Iselin Medhaug, Andrea Dittus, and three anonymous reviewers for useful comments and discussion. Funding and expertise provided by the Swiss Data Science Centre within the project “Data Science-informed attribution of changes in the Hydrological cycle” (DASH) is gratefully acknowledged (ID C17-01). A.G.P. was supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy’s Office of Biological and Environmental Research (BER) Cooperative Agreement DE-FC02-97ER62402, and the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement 1852977.

## REFERENCES

- Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419**, 224, <https://doi.org/10.1038/nature01092>.
- Alquier, P., and G. Biau, 2013: Sparse single-index model. *J. Mach. Learn. Res.*, **14**, 243–280.
- Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952, <https://doi.org/10.1017/CBO9781107415324.022>.
- Boé, J., L. Terray, F. Habets, and E. Martin, 2006: A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling. *J. Geophys. Res.*, **111**, D23106, <https://doi.org/10.1029/2005JD006889>.
- Cleveland, W., E. Grosse, and W. Shyu, 1991: Local regression models. *Statistical Models in S*, J. Chambers, and T. Hastie, Eds., Chapman & Hall, 309–376.
- Collins, M., and Coauthors, 2018: Challenges and opportunities for improved understanding of regional climate dynamics. *Nat. Climate Change*, **8**, 101, <https://doi.org/10.1038/s41558-017-0059-8>.
- Colucci, S., 2015: Synoptic meteorology—Anticyclones. *Encyclopedia of Atmospheric Sciences*, 2nd ed. Elsevier, 273–279.

- Compo, G. P., and P. D. Sardeshmukh, 2010: Removing ENSO-related variations from the climate record. *J. Climate*, **23**, 1957–1978, <https://doi.org/10.1175/2009JCLI2735.1>.
- DelSole, T., M. K. Tippett, and J. Shukla, 2011: A significant component of enforced multidecadal variability in the recent acceleration of global warming. *J. Climate*, **24**, 909–926, <https://doi.org/10.1175/2010JCLI3659.1>.
- Deser, C., and M. S. Timlin, 1997: Atmosphere–ocean interaction on weekly timescales in the North Atlantic and Pacific. *J. Climate*, **10**, 393–408, [https://doi.org/10.1175/1520-0442\(1997\)010<0393:AOIOWT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0393:AOIOWT>2.0.CO;2).
- , R. Knutti, S. Solomon, and A. S. Phillips, 2012a: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775, <https://doi.org/10.1038/nclimate1562>.
- , A. Phillips, V. Bourdette, and H. Teng, 2012b: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, <https://doi.org/10.1007/s00382-010-0977-x>.
- , —, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, <https://doi.org/10.1175/JCLI-D-13-00451.1>.
- , L. Terray, and A. S. Phillips, 2016: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *J. Climate*, **29**, 2237–2258, <https://doi.org/10.1175/JCLI-D-15-0304.1>.
- Emori, S., and S. Brown, 2005: Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophys. Res. Lett.*, **32**, L17706, <https://doi.org/10.1029/2005GL023272>.
- Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>.
- Fereday, D., R. Chadwick, J. Knight, and A. A. Scaife, 2018: Atmospheric dynamics is the largest source of uncertainty in future winter European rainfall. *J. Climate*, **31**, 963–977, <https://doi.org/10.1175/JCLI-D-17-0048.1>.
- Foster, G., and S. Rahmstorf, 2011: Global temperature evolution 1979–2010. *Environ. Res. Lett.*, **6**, 044022, <https://doi.org/10.1088/1748-9326/6/4/044022>.
- Friedman, J., T. Hastie, and R. Tibshirani, 2010: Regularization paths for generalized linear models via coordinate descent. *J. Stat. Software*, **33**, <https://doi.org/10.18637/jss.v033.i01>.
- Guo, R., C. Deser, L. Terray, and F. Lehner, 2019: Human influence on winter precipitation trends (1921–2015) over North America and Eurasia revealed by dynamical adjustment. *Geophys. Res. Lett.*, **46**, 3426–3434, <https://doi.org/10.1029/2018gl081316>.
- Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, Vol. 1, Springer, 533 pp.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, <https://doi.org/10.1175/2009BAMS2607.1>.
- , and —, 2012: Time of emergence of climate signals. *Geophys. Res. Lett.*, **39**, L01702, <https://doi.org/10.1029/2011GL050087>.
- He, J., and B. J. Soden, 2017: A re-examination of the projected subtropical precipitation decline. *Nat. Climate Change*, **7**, 53–57, <https://doi.org/10.1038/nclimate3157>.
- Houze, R. A., 2012: Orographic effects on precipitating clouds. *Rev. Geophys.*, **50**, RG1001, <https://doi.org/10.1029/2011RG000365>.
- Hurrell, J. W., 1995: Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, <https://doi.org/10.1126/science.269.5224.676>.
- , and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, <https://doi.org/10.1175/BAMS-D-12-00121.1>.
- Iles, C., and G. Hegerl, 2017: Role of the North Atlantic Oscillation in decadal temperature trends. *Environ. Res. Lett.*, **12**, 114010, <https://doi.org/10.1088/1748-9326/aa9152>.
- IPCC, 2012: Summary for policymakers. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. Field et al., Eds., Cambridge University Press, 19 pp., [https://www.ipcc.ch/site/assets/uploads/2018/03/SREX\\_FD\\_SPM\\_final-2.pdf](https://www.ipcc.ch/site/assets/uploads/2018/03/SREX_FD_SPM_final-2.pdf).
- Kay, J., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Kröner, N., S. Kotlarski, E. Fischer, D. Lüthi, E. Zuber, and C. Schär, 2017: Separating climate change signals into thermodynamic, lapse-rate and circulation effects: Theory and application to the European summer climate. *Climate Dyn.*, **48**, 3425–3440, <https://doi.org/10.1007/s00382-016-3276-3>.
- Lambert, F. H., P. A. Stott, M. R. Allen, and M. A. Palmer, 2004: Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.*, **31**, L10203, <https://doi.org/10.1029/2004GL019545>.
- Lehner, F., C. Deser, and L. Terray, 2017: Toward a new estimate of “time of emergence” of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *J. Climate*, **30**, 7739–7756, <https://doi.org/10.1175/JCLI-D-16-0792.1>.
- , —, I. R. Simpson, and L. Terray, 2018: Attributing the U.S. Southwest’s recent shift into drier conditions. *Geophys. Res. Lett.*, **45**, 6251–6261, <https://doi.org/10.1029/2018GL078312>.
- Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.*, **104**, 942–952, [https://doi.org/10.1175/1520-0493\(1976\)104<0942:EOTNVO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<0942:EOTNVO>2.0.CO;2).
- McMurdie, W., and R. Houze, 2006: Weather systems. *Atmospheric Science: An Introductory Survey*, J. M. Wallace and P. V. Hobbs, Eds., Elsevier, 313–373.
- Meehl, G. A., and Coauthors, 2013: Climate change projections in CESM1 (CAM5) compared to CCSM4. *J. Climate*, **26**, 6287–6308, <https://doi.org/10.1175/JCLI-D-12-00572.1>.
- Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213, <https://doi.org/10.1007/s10584-011-0156-z>.
- Merrifield, A., F. Lehner, S.-P. Xie, and C. Deser, 2017: Removing circulation effects to assess central US land–atmosphere interactions in the CESM large ensemble. *Geophys. Res. Lett.*, **44**, 9938–9946, <https://doi.org/10.1002/2017GL074831>.
- Myhre, G., and Coauthors, 2018: Sensible heat has significantly affected the global hydrological cycle over the historical period. *Nat. Commun.*, **9**, 1922, <https://doi.org/10.1038/s41467-018-04307-4>.
- Neale, R. B., and Coauthors, 2010: Description of the NCAR Community Atmosphere Model (CAM5.0). NCAR Tech. Note NCAR/TN-486+STR, 268 pp., [www.cesm.ucar.edu/models/cesm1.1/cam/docs/description/cam5\\_desc.pdf](http://www.cesm.ucar.edu/models/cesm1.1/cam/docs/description/cam5_desc.pdf).

- Pendergrass, A. G., R. Knutti, F. Lehner, C. Deser, and B. M. Sanderson, 2017: Precipitation variability increases in a warmer climate. *Sci. Rep.*, **7**, 17966, <https://doi.org/10.1038/s41598-017-17966-y>.
- Perlwitz, J., T. Knutson, J. Kossin, and A. N. LeGrande, 2017: Large-scale circulation and climate variability. *Climate Science Special Report: Fourth National Climate Assessment*, Vol. I, D. J. Wuebbles et al., Eds., U.S. Global Change Research Program, 161–184, <https://doi.org/10.7930/J0RV0KVQ>.
- Peters, J., D. Janzing, and B. Schölkopf, 2017: Elements of causal inference: Foundations and learning algorithms. MIT Press, 288 pp.
- R Core Team, 2018: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>.
- Robock, A., 2000: Volcanic eruptions and climate. *Rev. Geophys.*, **38**, 191–219, <https://doi.org/10.1029/1998RG000054>.
- Saffioti, C., E. M. Fischer, and R. Knutti, 2015: Contributions of atmospheric circulation variability and data coverage bias to the warming hiatus. *Geophys. Res. Lett.*, **42**, 2385–2391, <https://doi.org/10.1002/2015GL063091>.
- , —, S. C. Scherrer, and R. Knutti, 2016: Reconciling observed and modeled temperature and precipitation trends over Europe by adjusting for circulation variability. *Geophys. Res. Lett.*, **43**, 8189–8198, <https://doi.org/10.1002/2016GL069802>.
- , —, and R. Knutti, 2017: Improved consistency of climate projections over Europe after accounting for atmospheric circulation variability. *J. Climate*, **30**, 7271–7291, <https://doi.org/10.1175/JCLI-D-16-0695.1>.
- Salzmann, M., 2016: Global warming without global mean precipitation increase? *Sci. Adv.*, **2**, e1501572, <https://doi.org/10.1126/sciadv.1501572>.
- Schneider, E., and J. Kinter, 1994: An examination of internally generated variability in long climate simulations. *Climate Dyn.*, **10**, 181–204, <https://doi.org/10.1007/BF00208987>.
- Seager, R., N. Naik, and G. A. Vecchi, 2010: Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming. *J. Climate*, **23**, 4651–4668, <https://doi.org/10.1175/2010JCLI3655.1>.
- , H. Liu, N. Henderson, I. Simpson, C. Kelley, T. Shaw, Y. Kushnir, and M. Ting, 2014: Causes of increasing aridification of the Mediterranean region in response to rising greenhouse gases. *J. Climate*, **27**, 4655–4676, <https://doi.org/10.1175/JCLI-D-13-00446.1>.
- Shepherd, T. G., 2014: Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.*, **7**, 703, <https://doi.org/10.1038/ngeo2253>.
- Siler, N., C. Proistosescu, and S. Po-Chedley, 2019: Natural variability has slowed the decline in western U.S. snowpack since the 1980s. *Geophys. Res. Lett.*, **46**, 346–355, <https://doi.org/10.1029/2018GL081080>.
- Smith, R., and Coauthors, 2010: The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM). Rep. LAUR-01853, 141, 140 pp., <http://www.cesm.ucar.edu/models/cesm1.0/pop2/doc/sci/POPRefManual.pdf>.
- Smoliak, B. V., J. M. Wallace, P. Lin, and Q. Fu, 2015: Dynamical adjustment of the Northern Hemisphere surface air temperature field: Methodology and application to observations. *J. Climate*, **28**, 1613–1629, <https://doi.org/10.1175/JCLI-D-14-00111.1>.
- Stolpe, M. B., I. Medhaug, U. Beyerle, and R. Knutti, 2019: Weak dependence of future global mean warming on the background climate state. *Climate Dyn.*, in press.
- Stott, P. A., S. Tett, G. Jones, M. Allen, J. Mitchell, and G. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137, <https://doi.org/10.1126/science.290.5499.2133>.
- Sutton, R., G. McCarthy, J. Robson, B. Sinha, A. Archibald, and L. Gray, 2018: Atlantic multidecadal variability and the U.K. ACSIS program. *Bull. Amer. Meteor. Soc.*, **99**, 415–425, <https://doi.org/10.1175/BAMS-D-16-0266.1>.
- Thompson, D. W., J. M. Wallace, and G. C. Hegerl, 2000: Annular modes in the extratropical circulation. Part II: Trends. *J. Climate*, **13**, 1018–1036, [https://doi.org/10.1175/1520-0442\(2000\)013<1018:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1018:AMITEC>2.0.CO;2).
- Trenberth, K. E., 1995: Atmospheric circulation climate changes. *Climatic Change*, **31**, 427–453, <https://doi.org/10.1007/BF01095156>.
- von Storch, H., and F. W. Zwiers, 2002: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp., <https://doi.org/10.1017/CBO9780511612336>.
- Wallace, J. M., Y. Zhang, and J. A. Renwick, 1995: Dynamic contribution to hemispheric mean temperature trends. *Science*, **270**, 780–783, <https://doi.org/10.1126/science.270.5237.780>.
- , Q. Fu, B. V. Smoliak, P. Lin, and C. M. Johanson, 2012: Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. *Proc. Natl. Acad. Sci. USA*, **109**, 14 337–14 342, <https://doi.org/10.1073/pnas.1204875109>.
- Wehrens, R., and B.-H. Mevik, 2007: The pls package: Principal component and partial least squares regression in R. *J. Stat. Software*, **18**, <https://doi.org/10.18637/jss.v018.i02>.
- Xie, S.-P., and Coauthors, 2015: Towards predictive understanding of regional climate change. *Nat. Climate Change*, **5**, 921–930, <https://doi.org/10.1038/nclimate2689>.
- Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.*, **32**, L18701, <https://doi.org/10.1029/2005GL023684>.
- Zou, H., and T. Hastie, 2005: Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc.*, **67B**, 301–320, <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.