

A Weather-Type-Based Cross-Time-Scale Diagnostic Framework for Coupled Circulation Models

ÁNGEL G. MUÑOZ

Atmospheric and Oceanic Sciences, and NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey, and International Research Institute for Climate and Society, Earth Institute, Columbia University, Palisades, New York

XIAOSONG YANG

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey, and Cooperative Programs for the Advancement of Earth System Science, University Corporation for Atmospheric Research, Boulder, Colorado

GABRIEL A. VECCHI

Department of Geosciences, Princeton Environmental Institute, and Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey

ANDREW W. ROBERTSON

International Research Institute for Climate and Society, Earth Institute, Columbia University, Palisades, New York


WILLIAM F. COOKE

Cooperative Programs for the Advancement of Earth System Science, University Corporation for Atmospheric Research, Boulder, Colorado, and NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey

(Manuscript received 23 February 2017, in final form 23 July 2017)

ABSTRACT

This study proposes an integrated diagnostic framework based on atmospheric circulation regime spatial patterns and frequencies of occurrence to facilitate the identification of model systematic errors across multiple time scales. To illustrate the approach, three sets of 32-yr-long simulations are analyzed for northeastern North America and for the March–May season using the Geophysical Fluid Dynamics Laboratory’s Low Ocean–Atmosphere Resolution (LOAR) and Forecast-Oriented Low Ocean Resolution (FLOR) coupled models; the main difference between these two models is the horizontal resolution of the atmospheric model used. Regime-dependent biases are explored in the light of different atmospheric horizontal resolutions and under different nudging approaches. It is found that both models exhibit a fair representation of the observed circulation regime spatial patterns and frequencies of occurrence, although some biases are present independently of the horizontal resolution or the nudging approach and are associated with a misrepresentation of troughs centered north of the Great Lakes and deep coastal troughs. Moreover, the intraseasonal occurrence of certain model regimes is delayed with respect to observations. On the other hand, inter-experiment differences in the mean frequencies of occurrence of the simulated weather types, and their variability across multiple time scales, tend to be negligible. This result suggests that low-resolution models could be of potential use to diagnose and predict physical variables via their simulated weather type characteristics.

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

Corresponding author: Á. G. Muñoz, agms@princeton.edu

DOI: 10.1175/JCLI-D-17-0115.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

1. Introduction

Global circulation models (GCMs) are essential tools for both scientific research (e.g., [Delworth et al. 2012](#); [Taylor et al. 2012](#)) and the development of climate services ([Vaughan and Dessai 2014](#)) at different time scales. Since, by definition,¹ models are not a perfect representation of reality, different types of errors must be accounted for before GCMs can be used as a fair depiction of the real world.

Common approaches to diagnose systematic errors involve the computation of metrics aimed at providing an overall summary of the performance of the model in reproducing the particular variables of interest in the study, normally tied to specific spatial and temporal scales. For example, [Gleckler et al. \(2008\)](#) used different metrics to assess the climatological behavior of GCMs, [Covey et al. \(2016\)](#) focused on model evaluation of the rainfall diurnal cycle, and several studies have analyzed the present-time rainfall characteristics in state-of-the-art climate change simulations (see, e.g., [Reichler and Kim 2008](#); [Delworth et al. 2012](#); [Ryu and Hayhoe 2014](#); [van der Wiel et al. 2016](#)). Moreover, the number of open source packages to evaluate the fidelity of GCMs compared to observations (e.g., [Phillips et al. 2014](#); [Mason and Tippet 2016](#); [Gleckler et al. 2016](#)) is rapidly increasing. Nonetheless, the evaluation of the goodness of a model is not always tied to the understanding of the physical processes that are correctly represented, distorted, or even absent in the model world. As the physical mechanisms are more often than not related to interactions taking place at multiple time and spatial scales (e.g., [Nakamura et al. 2013](#); [Robertson et al. 2015](#); [Muñoz et al. 2015, 2016](#)), cross-scale model diagnostic tools are not only desirable but required.

A complementary alternative to the common diagnostic approach mentioned above can be achieved via a nonlinear dynamical system perspective ([Lorenz 1963](#); [Palmer 1999](#)). Thus, typical questions like “How well does this model represent the observed seasonal rainfall patterns?” are framed as part of the more general question “Can the model adequately represent the available states of the system?” (in a suitably coarse-grained phase space; see [Ghil and Robertson 2002](#)). Since only certain physical states can be accessed by the real-world system under study, as in statistical or quantum mechanics, the rationale is that models that cannot faithfully reproduce those states should not be expected

to provide, for example, the right rainfall or temperature patterns for the correct reasons.

How to identify the available states? Theoretically, they are related to the concept of multiple flow equilibria ([Lorenz 1969](#); [Charney and DeVore 1979](#); [Reinhold and Pierrehumbert 1982](#)) and quasi-stationary regimes or metastable fixed points capable of attracting the chaotic trajectory of the system; they can be recognized in a state (or phase) space as regions that are visited more frequently—or, equivalently, regions surrounding a local density maximum. The identification of these clusters in the state space poses a nontrivial statistical problem ([Stephenson et al. 2004](#); [Christensen et al. 2015](#)), and in general they correspond to proxies of the available states of the system, and not the states themselves. From a practical perspective, in the atmosphere those clusters are typically associated with recurrent daily circulation types or “weather types” (WTs), which have been widely studied in the literature ([Lorenz 1969](#); [Charney and DeVore 1979](#); [Reinhold and Pierrehumbert 1982](#); [Lorenz 2006](#); [Vautard 1990](#); [Michelangeli et al. 1995](#); [Robertson and Ghil 1999](#); [Moron et al. 2008a,b](#); [Johnson and Feldstein 2010](#); [Riddle et al. 2013](#)), as they can lead to both important positive and negative socioeconomic impacts.

Formally speaking, weather types are statistical constructs associated with recurrent circulation configurations that can be used to study weather regimes, a more physical concept that often requires additional conditions [e.g., the average time spent on each regime must be long compared to other oscillations in the system, as in [Lorenz \(2006\)](#)]. Having stated that formal difference, for the purposes of this work both concepts are considered largely interchangeable. If correctly defined, these weather types can be understood as “building blocks” or some sort of “alphabet” that can be used to describe all the physically acceptable events of the system ([Muñoz et al. 2015, 2016](#)). This is shown schematically in [Fig. 1](#); the events (e.g., rainy days) can be explained by the occurrence of individual weather types (letters) or particular WT sequences/transitions (words).

The underlying hypothesis used here is that climate variability across time scales can be described in terms of the frequency of occurrence of weather types at the different scales, with external (or internal) forcings taking the form of shifts in the residence time of the system in the different basins of attractions of the state space. This idea is related to the fluctuation–dissipation theorem in climate ([Leith 1975](#)), which relates the mean response to small perturbations of a nonlinear dynamical system to fluctuations in the unforced system. As indicated by [Leith \(1975\)](#), the nonlinear transfer of energy and enstrophy in the climate system implies sources

¹ The word *model* is derived from Latin *modulus*, a “measure” or “standard” of something, which evolved to today’s idea of “likeness made to scale,” for example in clay or wax, or via mathematical or numerical expressions.

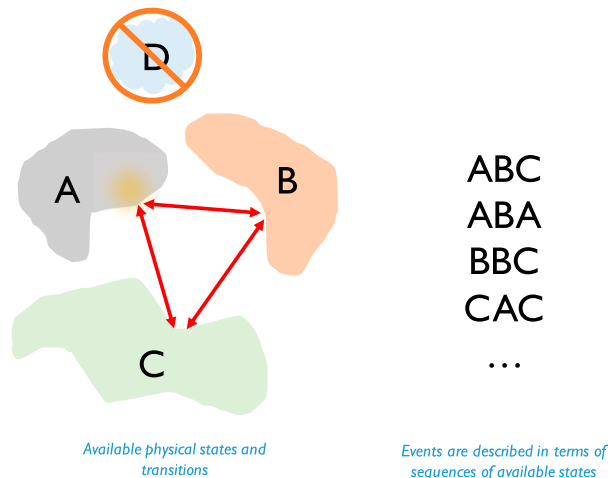


FIG. 1. Schematic showing (left) three available states (A, B, C) and a forbidden or inaccessible state (D). The system transitions between the available states (represented by the arrows). Those transitions define sequences of states (right) that describe events; e.g., ABC could be related to extreme rainfall events and BBC to heat waves.

of these variables in some scales of motion and dissipation in others; hence, considering the present cross-time scale diagnostics approach can help to understand the problem better.

Beyond the theoretical interest of associating weather types with physically acceptable states of the system, the present approach is useful in those cases in which it is possible to identify circulation-dependent systematic errors in climate models; a weather-type rectification (correction) has the potential to improve the related physical fields and events in the simulations.

A regime approach has been explored to study general biases and uncertainty in climate models at specific time scales (e.g., Palmer and Weisheimer 2011; Perez et al. 2014; Christensen et al. 2015). Since it is possible to analyze the behavior of the circulation regimes over a wide range of different time scales—for example, by aggregating their frequency of occurrence at sub-seasonal, seasonal, interannual and longer time scales—and since they can be associated with physical processes occurring across time scales (Moron et al. 2015; Muñoz et al. 2015, 2016), it is proposed here that the weather-typing approach provides a natural and unified framework for cross-time-scale diagnostics of GCMs. Although the scheme is deemed especially useful when considering seamless prediction systems (Hoskins 2013), it can be also employed to analyze causes of bias at any particular time scale. Furthermore, the same set of weather types can be used to diagnose a wide range of variables in a physically consistent way, thus shedding light on the causes of biases rather than focusing on the bias itself.

The goal of the present study is to illustrate the proposed approach using simulations produced by two coupled GCMs developed by the Geophysical Fluid Dynamics Laboratory (GFDL), with physical configurations designed to reproduce key aspects of the observed climate variability. For the sake of the illustration, the study is only focused on rainfall during the March–May (MAM) season for northeastern North America (NENA). The MAM atmospheric circulation regimes in this region are relatively well understood, especially in relation to flood events in the Ohio and Mississippi river basins (Nakamura et al. 2013; Robertson et al. 2015).

The rest of the paper is organized as follows. After introducing the datasets and summarizing the methods in the next section, observed and modeled rainfall climatologies are analyzed in section 4. The weather-type diagnostics are presented in section 5 for daily, sub-seasonal, interannual, and “decadal” time scales, while section 6 deals with possible sources of bias and an overall discussion of the results. Finally, the concluding remarks are presented in section 7.

2. Data

Both observations and model outputs were used in the present study. These are described in the following subsections. The time period considered in all datasets is MAM 1981–2012.

a. Observations

Observed rainfall fields were obtained from the gauge-based NOAA–NCEP–CPC Unified Precipitation gridded dataset (Chen et al. 2008). This product has daily temporal resolution and a spatial resolution of $1^\circ \times 1^\circ$.

The analysis of the observed daily circulation over NENA was derived from the NCEP–NCAR Reanalysis Project (version 2, or NNRv2) 500-hPa geopotential data on a $2.5^\circ \times 2.5^\circ$ grid (Kalnay et al. 1996; Kistler et al. 2001), and also from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) for the same variable on a $0.5^\circ \times 0.5^\circ$ grid (Rienecker et al. 2011). These two datasets were used to investigate the impact of horizontal resolution and reanalysis methods in the identification of the weather types.

b. Model data and experimental design

This study explores the impact of (a) horizontal resolution and (b) Newtonian relaxation toward observed fields on model biases associated with the variability of circulation regimes at multiple time scales.

The three sets of climate simulations used are 32-yr long, with 5 members each, produced by two kindred coupled GCMs developed by the Geophysical Fluid Dynamics

Laboratory. They share the same ocean, land, and ice model components inherited from the GFDL coupled models version 2.1, CM2.1 (Delworth et al. 2006), and version 2.5, CM2.5 (Delworth et al. 2012).

The models used in this study are the Low Ocean–Atmosphere Resolution (LOAR; van der Wiel et al. 2016), and the Forecast-Oriented Low Ocean Resolution (FLOR; Vecchi et al. 2014). Unlike CM2.1 and CM2.5, LOAR and FLOR use essentially the same atmospheric model component, based on a finite-volume dynamical core on a cubed sphere (Putman and Lin 2007), integrating along 32 vertical levels and with horizontal resolutions of $2^\circ \times 2^\circ$ (C48) for LOAR and $0.5^\circ \times 0.5^\circ$ (C180) for FLOR. The dynamical time step is modified to match the individual model’s atmospheric resolution, and the atmospheric physics for both models is similar to that in CM2.5 (Delworth et al. 2012; Vecchi et al. 2014). The ocean model component is the Modular Ocean Model (MOM) version 5 (Griffies 2012), at $1^\circ \times 1^\circ$ and configured as reported in Vecchi et al. (2014). The land surface model is the Land Model version 3 (LM3; Milly et al. 2014), with the same horizontal resolution as the atmospheric model component. The sea ice model component is the Sea Ice Simulator (SIS), having three vertical layers, one snow and two ice, and five thickness categories, as reported in Delworth et al. (2006) and references therein. The time resolution for all model output is daily.

A typical way to explore how well uncoupled GCMs reproduce the observed natural variability is to force them with observed SSTs [e.g., à la the Atmospheric Model Intercomparison Project (AMIP) (Gates et al. 1999)]. Although this approach is, by definition, not possible for coupled GCMs, a common alternative is to use a Newtonian nudging to relax certain model fields, such as SSTs, toward observations (e.g., Rosati et al. 1997). This method has been shown useful to reproduce key aspects of the natural variability of the climate system and to decrease model biases [e.g., Luo et al. (2008), and references therein]. Thus, all three experiments used in the study were nudged to observed fields as follows.

In two sets of numerical experiments, LOAR_{sst} and FLOR_{sst}, the models’ SSTs were relaxed to the daily-interpolated observed HadISST product (Rayner et al. 2003), SST_O, using a restoring time scale $\tau = 5$ days, such that

$$\partial_t \text{SST} = \widehat{\text{SST}} + \frac{1}{\tau} (\text{SST}_O - \text{SST}), \quad (1)$$

where ∂_t represents the partial time derivative of the field, and $\widehat{\square}$ the coupled model tendency of the field.

These two experiments were used to analyze the impact of horizontal resolution on model biases.

Since nudging only SSTs is often inadequate (Fujii et al. 2009), as the variability in the climate system is not only controlled by sea surface temperatures, for the third set of numerical experiments, called FLOR_{sst+strat}, MERRA’s 6-hourly stratospheric temperatures and winds above 100 hPa were nudged in addition to the SST nudging described above. The same ansatz presented in Eq. (1) was used but with a relaxation time $\tau = 6$ hours, and a tapering factor γ such that it is one for model layers above 50 hPa (i.e., 1, 4, 8, 14, 21, 30, and 41 hPa), and it is zero for layers below 100 hPa; from 50 to 100 hPa, γ varies linearly from one to zero, thus transitioning in a smooth way from the stratosphere to the troposphere. This same stratospheric nudging approach has been used recently by Jia et al. (2017) to show the role of the extratropical stratosphere as an important source of predictability at interannual scale. Both FLOR_{sst} and FLOR_{sst+strat} were used in this study to explore the role of nudging in model biases.

3. Methods

Here the general methodology is described. The diagnostic approach per se is described in section 5.

a. Cluster analysis

To identify the proxies of the available states of the system, daily circulation regimes were determined performing a k -means analysis (e.g., Robertson and Ghil 1999) on observed and modeled 500-hPa geopotential height anomalies, each at its own spatial resolution to avoid the possibility of spurious results induced by the spatial interpolation process.

The k -means technique is a partitioning method that classifies all days into a predefined number of clusters, minimizing the intracluster sum of variance, \mathcal{W} , for a partition \mathcal{P} (the Voronoi diagram generated by the classification process):

$$\mathcal{W}(\mathcal{P}) = \sum_i^k \sum_{X \in p_i} ds^2(X, \bar{p}_i), \quad (2)$$

where $i = 1 \dots k$, k is the total number of clusters in the solution, p_i denotes the different clusters in the partition, and $ds^2(X, \bar{p}_i)$ represents the Euclidean line element (distance) between the cluster’s maps X and centroid \bar{p}_i . In this work, no areal weighting was necessary, as the selected domain is relatively small and it does not extend into high latitudes (30° – 50° N; more details below).

This procedure assigns only one cluster to each day on record. Daily geopotential data were first projected onto its six leading empirical orthogonal functions (same number for all datasets), accounting for at least 95% of the total observed variance. No additional time filtering was applied to the data before clustering, thus retaining the annual seasonal cycle and interannual, subseasonal, and synoptic weather time scales for diagnosis. No projection between the modeled and observed leading empirical orthogonal functions was performed (i.e., the modeled WTs were directly obtained using exactly the same procedure as for the observed ones). As with the avoidance of any kind of spatial interpolation before the k -means analysis, this was done in order to evaluate each model's weather types in an unbiased way.

Two approaches were tested to define the 500-hPa geopotential anomalies (departures from the long-term mean) that characterize each weather type. In one, the anomalies of all members were concatenated along the time coordinate before the k -means process was performed, as in Muñoz et al. (2016). In the second approach, the classification of the 500-hPa geopotential anomalies was performed for each member independently, then computing the ensemble mean of the same weather type across members. The latter method was used in the present study as it respects the chronology and transition probabilities of the weather types in the presence of nonstationary data.

The classifiability index (CI; Michelangeli et al. 1995) measures the similarity of partitions of the data, and permits the identification of the minimum number of clusters required to achieve a good classification. Classification performance is measured via the pattern (or anomaly) correlation coefficient ρ_{ij} between pairs of cluster centroids \bar{p}_i and \bar{q}_j in partitions \mathcal{P} and \mathcal{Q} , respectively, which allows the matrix $\rho[\mathcal{P}(k), \mathcal{Q}(k)]$ to be computed, with elements given by

$$\rho(\bar{p}_i, \bar{q}_j) \equiv \frac{\sum_m p'_{im} q'_{jm}}{\sqrt{\sum_m p'^2_{im} \sum_m q'^2_{jm}}} \quad \forall i, j = 1 \dots k, \quad (3)$$

where the sums consider all the M grid points in the domain, and the grid point deviations are computed as usual:

$$p'_{im} = \bar{p}_{im} - \frac{1}{k} \sum_r \bar{p}_{rm}, \quad (4)$$

$$q'_{jm} = \bar{q}_{jm} - \frac{1}{k} \sum_r \bar{q}_{rm}, \quad (5)$$

where $m = 1 \dots M$. The maximal value of the i th row of the $k \times k$ matrix ρ in Eq. (3) represents the cluster in \mathcal{Q}

that best fits the i th cluster in \mathcal{P} . If the two partitions are identical, of course, $\rho = 1$.

In this study, a total of $N = 100$ partitions are used, obtained from 100 different initial random seeds of the algorithm; the single partition that matches most closely the remaining ones is then selected. This similarity is computed by averaging the values of $\rho(\mathcal{P}_n, \mathcal{P}_{n'}) \forall n, n' = 1 \dots N$, and, following Michelangeli et al. (1995),

$$\text{CI}(k) \equiv \frac{1}{N(N-1)} \sum_{1 \leq n \neq n' \leq N} \rho[\mathcal{P}(k)_n, \mathcal{P}(k)_{n'}]. \quad (6)$$

Clustering solutions within the range of $k = 2$ –10 were explored for the NENA geographical domain, bounded by 30° – 50°N and 105° – 69°W . A k -means seven-cluster solution was found to yield a statistically significant value of the classifiability index ($p < 0.1$), compared to a red-noise background modeled as a first-order Markov process having the same covariance at lag 0 and 1 days as the atmospheric data (Michelangeli et al. 1995); therefore the solution, which is the same one reported by Robertson et al. (2015), was selected for further analysis.

This set of clusters, or weather types, can be interpreted as a set of geopotential regimes that typify the daily variability, and are considered to be the so-called building blocks or letters mentioned in the introduction. Other cluster solutions were explored, verifying that the general features of the circulation regimes presented below are robust and sufficient for the purposes of this study.

b. Similarity metrics

There are multiple metrics to evaluate how well a model reproduces observations (e.g., Mason and Stephenson 2008; Jolliffe and Stephenson 2012); selection among them depends on the attribute or type of question to be addressed. For the sake of simplicity, this work focuses on the similarity of spatial patterns and temporal behavior between the simulated and observed weather types. Two metrics are chosen for that purpose: pattern correlation and the scatter index. As indicated before, the k -means analysis is performed on the native grid of the models and the observations; however, when comparing results, all calculations were always carried out in a common low resolution grid of $2.5^\circ \times 2.5^\circ$ (the one of the NNRPV2 dataset).

The pattern correlation ρ is computed following Eq. (3), but considering the fact that the partitions now correspond to those modeled and observed.

The scatter index ξ is defined here as the root-mean-square error operator of the simulated and observed frequencies of occurrence of the weather types, normalized

by the weather type mean frequency (e.g., Perez et al. 2014). The frequency of occurrence is defined as the proportion of time that each weather type occurs at each time scale window of interest, averaged across ensemble members for the case of model simulations (see an example below). Two versions of the scatter index are used in this work, one to evaluate the mean frequency state, $\bar{\xi}_{|t_s}$, and another for the temporal variability at different time scale windows t_s , $\xi_{\sigma|t_s}$:

$$\bar{\xi}_{|t_s} \equiv \frac{\sqrt{k \sum_i^k (f_i^S - f_i^O)^2}}{\sum_i^k f_i^O} \Big|_{t_s}, \quad (7)$$

$$\xi_{\sigma|t_s} \equiv \frac{\sqrt{k \sum_i^k [\sigma(f_i^S) - \sigma(f_i^O)]^2}}{\sum_i^k \sigma(f_i^O)} \Big|_{t_s}, \quad (8)$$

where f_i^S and f_i^O denote the simulated and observed frequencies of the $i = 1 \dots k$ weather types, and σ denotes the corresponding sample standard deviation. The different time scales (or time scale windows) are presented in section 5. A model with a perfect representation of the observed frequency of occurrence satisfies $\bar{\xi}_{|t_s} = 0$, and in consequence its representation of the variability in the frequency of occurrence is also perfect ($\xi_{\sigma|t_s} = 0$). In general that is not the case, and each index provides information on these two statistics.

To evaluate the scatter index at a particular time scale implies performing the corresponding calculations on the time window defined for that purpose. To illustrate this idea, consider as an example the overall evaluation of a model representation of the observed weather type frequencies during the window defined as 16–31 May, for all seasons in the period 1981–2012. The scatter indices are then computed considering the observed and simulated frequencies of occurrence of weather types for those thirty-one 16-day-long windows, using Eqs. (7) and (8), where the vertical bars indicate that the indices are to be evaluated for the particular time scale (window) t_s .

Once the scatter indices have been computed for the different time scales of interest, they can be presented using boxplots to also convey information about their uncertainties in the models; for example, if all members in a model perfectly agree on the value of a scatter index for a particular time scale, the corresponding box in the boxplot will collapse to an horizontal line. More details about these boxplots are discussed in section 5.

To analyze persistence of the weather types, the empirical probability density functions were computed for each dataset and regime. To compare between the different numerical experiments and the reanalysis, the Euclidean distance between each pair of the corresponding Weibull distribution parameters, α (scale) and β (shape), was used to assess similarity. The Weibull distribution

$$f(x|\alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right] \quad (9)$$

is commonly used in survival or duration analysis, and provided better fit than other less flexible distributions, like the exponential one (which corresponds to the special case of $\beta = 1$).

4. Northeastern North America's rainfall climatology

It is a common practice to evaluate the fidelity of a model compared to observations through the analysis of the climatological behavior of the variable of interest. This section uses the MAM rainfall climatology of the region under study (30°–50°N and 105°–69°W; see Fig. 2) to illustrate some key ideas of the present diagnostic approach.

The observed rainfall climatology (Fig. 2a) exhibits a clear precipitation gradient with higher amounts ($\approx 3 - 4 \text{ mm day}^{-1}$) in a wide region covering approximately 30°–40°N and 97°–84°W (most parts of Alabama, Mississippi, Louisiana, Arkansas, Tennessee, southern Missouri, and southwestern Kentucky), with basically no precipitation along the western border of NENA (Texas, Oklahoma, Kansas, Nebraska, the Dakotas, and southern Ontario), 0.5–1 mm day^{-1} to the north (southern Ontario and Quebec) and 2–3.5 mm day^{-1} along the eastern coast (from northern Florida to Maine).

A visual inspection of the models' climatology (Figs. 2b–d) clearly shows some important biases in both magnitudes and spatial distribution of rainfall, with simulations at higher resolution (Figs. 2c,d) exhibiting some improvement with respect to the low-resolution one (Fig. 2b).

What are the causes of such biases? Typically, issues involving model resolution or physical parameterizations (specially for rainfall) are frequently identified as the sources of bias—and indeed higher horizontal and vertical resolution, as well as better physics, tends to improve the representation of the observed fields. Nonetheless, it is also possible that the resolution and parameterizations are fit for the purpose, and something else is failing. Whatever the causes, a complementary

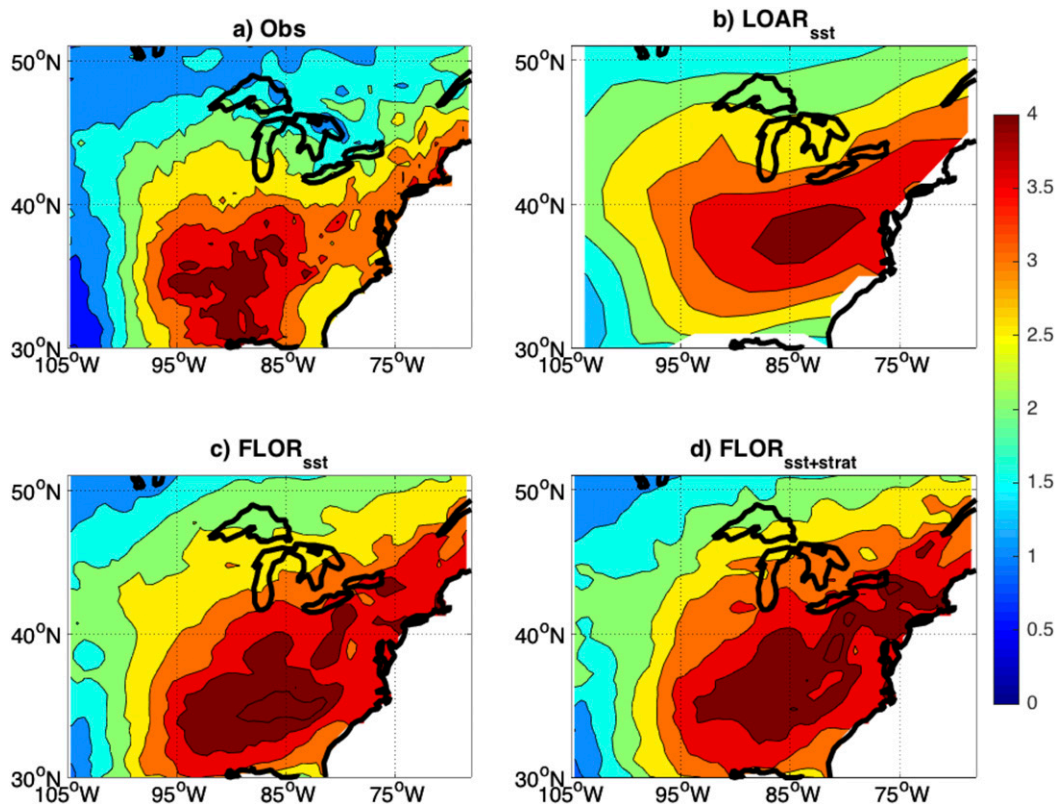


FIG. 2. Rainfall climatology (MAM 1981–2012) for northeastern North America: (a) observations, (b) $LOAR_{sst}$, (c) $FLOR_{sst}$, and (d) $FLOR_{sst+strat}$. Units are in $mm\ day^{-1}$. Ocean has been masked in (b)–(d). Each dataset is plotted using its native horizontal resolution.

approach consisting of the diagnosis of the physical mechanisms conducive to precipitation—rather than how well the precipitation itself is represented in the model—is also useful and most often required.

Rainfall variability in NENA is dominated by synoptic-scale circulation patterns (Archambault et al. 2008; Nakamura et al. 2013; Robertson et al. 2015; Roller et al. 2016) that are, in turn, modulated by well-known climate modes at multiple time scales like El Niño–Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), the Pacific–North American pattern (PNA), or the Madden–Julian oscillation (MJO; e.g., Ropelewski and Halpert 1986, 1987; Barnston and Livezey 1987; Wheeler and Hendon 2004).

Although some studies have found an inconclusive link between ENSO and NENA’s climate (e.g., Ropelewski and Halpert 1986), others indicate that this mode can modulate the frequency of low pressure systems and storms in the region (Trenberth and Caron 2000; Frankoski and DeGaetano 2011). Similarly, pressure and geopotential height anomalies associated with the NAO modify the frequency of occurrence of nor’easters, as well as storm track locations over the

North Atlantic basin via changes to the position and orientation of the North Atlantic jet stream (Jones and Davis 1995; Archambault et al. 2008). There is also evidence of links between PNA and NENA’s rainfall anomalies (Leathers et al. 1991; Notaro et al. 2006), and between MJO’s phases 5–7 and precipitation rate in the region (Becker et al. 2011; Zhou et al. 2012).

The effect of these and other climate modes conducive to rainfall in NENA can be studied through their influence in the occurrence, persistence, and evolution of daily weather types. For example, Roller et al. (2016) analyzed NENA’s wintertime mean and extreme precipitation, storm tracks, and teleconnections using a set of five WTs defined in terms of 850-hPa winds; similar studies have been conducted in other parts of the world for different seasons (Moron et al. 2008a,b, 2012, 2015; Muñoz et al. 2015, 2016).

The rainfall climatology (Fig. 2) discussed in this section can thus be understood in terms of the (non) linear contribution of the different mechanisms represented by persistence and transitions of the region’s daily circulation regimes (WTs, or atmospheric states). A misrepresentation of the physical interactions involved,

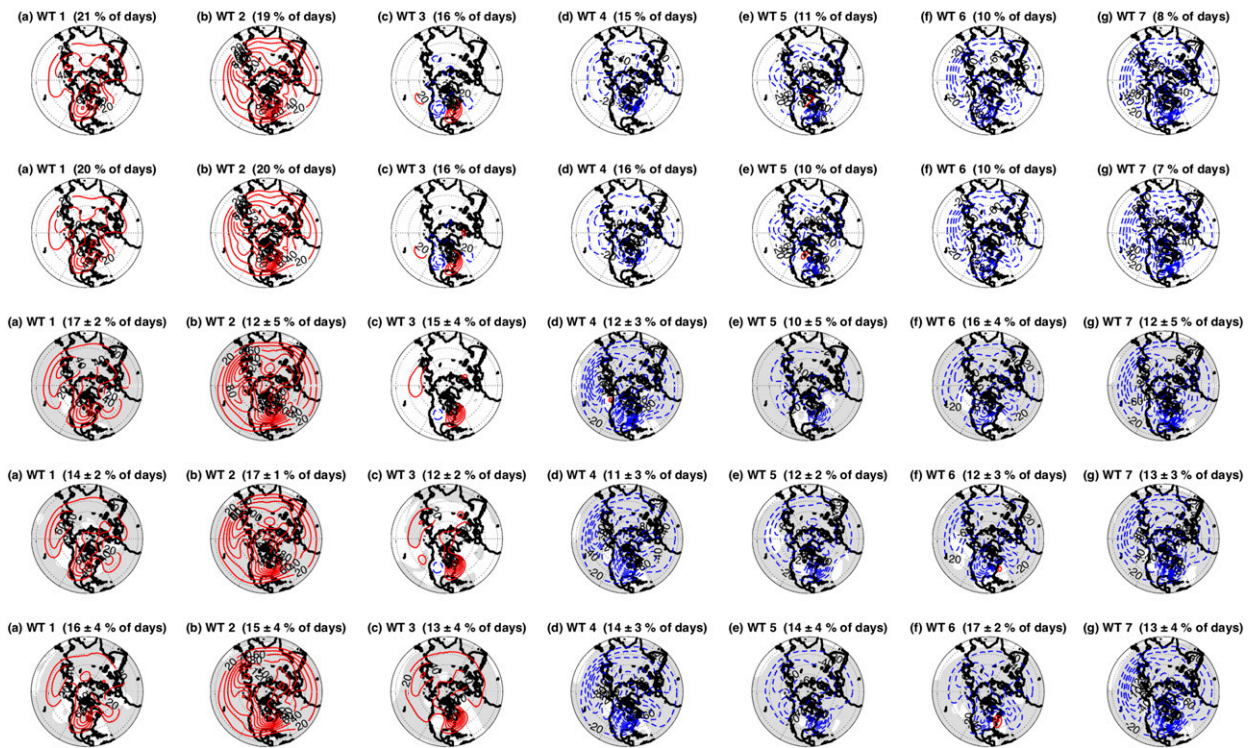


FIG. 3. Hemispheric view of observed (NNRPv2 and MERRA) and modeled ($LOAR_{sst}$, $FLOR_{sst}$, $FLOR_{sst+strat}$) weather types (WTs), using geopotential height anomalies at 500 hPa (contour interval is 20 gpm). Red solid (blue dashed) lines indicate positive (negative) anomalies. In the model experiments, regions showing statistically significant ($p < 0.05$) ensemble-mean anomalies are shaded in gray. Relative frequency of occurrence for the entire 32-yr period is indicated in parentheses, with model experiments showing the ensemble mean \pm one standard deviation. Each dataset is plotted using its native horizontal resolution.

even with perfect rainfall parameterizations, can generate bias not only in the climatological precipitation field but also in higher-order statistics of the rainfall variability at different time scales. The same can be said about other variables physically associated with the behavior of those WTs.

Hence, the next section focuses on diagnosing how well the models represent the observed behavior of the WTs for different time scales.

5. Cross-time-scale diagnostics

This section first discusses the circulation regimes obtained for the observations and experiments, before turning to different tools to diagnose the weather types evolution at daily to decadal scales. Although the method is general and can be used for longer time scales, this study is constrained by the availability of longer observed rainfall records.

The computed set of seven NENA WTs found to best typify the daily circulation regimes for MAM is presented in Fig. 3 plotted over the entire hemisphere to better identify wavelike patterns, and in Fig. 4 for a more

regional analysis of the weather types. The regimes with the highest observed frequency of occurrence are located to the left and the less frequently occurring to the right (see top row in those figures; the model regimes were ordered to follow each observed pattern). Equation (3) was used to identify the “model-equivalent WT” and physical interpretation of the hemispheric patterns (Fig. 3) was used in those cases in which the pattern correlation coefficients provided too similar values for two given regimes.

Two different reanalysis products were used to analyze the robustness of the method identifying the observed WTs, and the dependence on horizontal resolution (NNRPv2 at $\sim 2.5^\circ$ vs MERRA at $\sim 0.5^\circ$). Tables 1 and 2 show a comparison of the anomaly correlation coefficient between WTs using NNRPv2 and MERRA as a reference, respectively. The results are indeed consistent across reanalyses and resolutions.

The observed weather types (top two rows of Figs. 3 and 4) exhibit synoptic-scale meridionally elongated dipolar wave patterns (WT1, WT3, WT5, WT6), and monopole/dipole patterns that are zonally elongated (WT2, WT4, WT7). Weather regimes that are predominantly

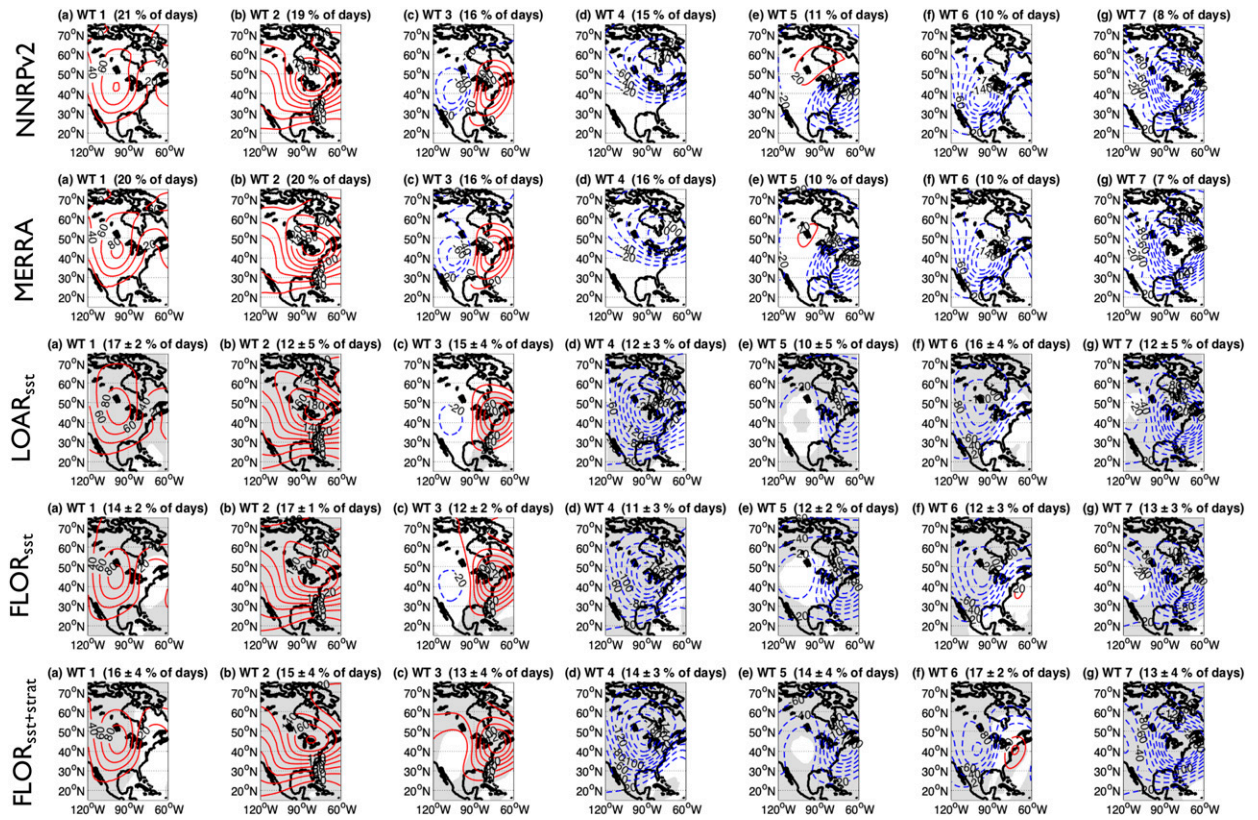


FIG. 4. As in Fig. 3, but for most of North America.

associated with positive geopotential anomalies (WT1 and WT2) tend to be more frequent—and more persistent, as will be shown later—than the more transient regimes (WT4–WT7), which show spatial structures with dominant negative geopotential anomalies (see Figs. 3 and 4). Although this general separation between more persistent and more transient circulation regimes is present in the simulations, the ordering in those cases—which reflects their mean frequencies—is not exactly the same as in the observations.

As it will be shown in section 5b, there are preferred sequences of states whose pattern evolution and typical time scales suggest eastward propagation of baroclinic

waves (e.g., state transitions 3 → 1 → 3; top rows in Fig. 3). Generally speaking, in NENA the MAM daily circulation regimes tend to be associated with continental ridges west of the Great Lakes (WT1; top rows in Fig. 4), continental ridges over the Great Lakes (WT2), northeastern seaboard ridges (WT3), troughs centered north of the Great Lakes (WT4), deep coastal land troughs (WT5), southeastern seaboard ridges (WT6), and deep northeastern troughs (WT7). The seaboard ridges (WT3 and WT6) have been shown to be associated with extreme rainfall events in the Ohio River basin; these two circulation types tend to occur more often during La Niña years, and phase 5 of the MJO also

TABLE 1. Anomaly correlation coefficient (Spearman) between the model-equivalent ensemble-mean weather types (WTs) in each experiment and the corresponding circulation regime in NINRPv2, computed on the reanalysis grid for the domain sketched in Fig. 4. Numbers in bold indicate the highest correlations found in the model experiments, all values being statistically significant (Student’s *t* test; $p < 0.05$). No statistically significant difference was found between the average anomaly correlation coefficients of the experiments.

Dataset vs NINRPv2	WT1	WT2	WT3	WT4	WT5	WT6	WT7	Average
MERRA	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.98
LOAR _{sst}	0.93	0.93	0.37	0.54	0.66	0.87	0.91	0.74
FLOR _{sst}	0.89	0.96	0.20	0.66	0.72	0.89	0.94	0.75
FLOR _{sst+strat}	0.90	0.97	−0.11	0.78	0.66	0.84	0.95	0.71

TABLE 2. Anomaly correlation coefficient (Spearman) between the model-equivalent ensemble-mean WTs in each experiment and the corresponding circulation regime in MERRA, computed on the reanalysis grid for the domain sketched in Fig. 4. Numbers in bold indicate highest correlations found in the model experiments, all values being statistically significant (Student's t test; $p < 0.05$). No statistically significant difference was found between the average anomaly correlation coefficients of the experiments.

Dataset vs MERRA	WT1	WT2	WT3	WT4	WT5	WT6	WT7	Average
NNRPv2	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.98
LOAR _{sst}	0.92	0.93	0.36	0.58	0.73	0.86	0.91	0.75
FLOR _{sst}	0.90	0.96	0.18	0.70	0.72	0.89	0.94	0.75
FLOR _{sst+strat}	0.91	0.96	-0.10	0.82	0.74	0.85	0.96	0.73

modulates their seasonal frequency, with lower occurrence for WT3 and higher presence for WT6 [for additional details see Robertson et al. (2015)].

There is no simple answer to the question of what set of experiments best represents the observed weather types, as the answer depends on the basis for the comparison. The shapes, magnitudes, locations, tilts, and frequencies of the weather types are characteristics to consider. On average, both LOAR and FLOR models do a good job reproducing the observed circulation regimes, although some important biases are present (Figs. 3 and 4; Tables 1 and 2). For example, the spatial features of WT1, the most frequently observed, are well represented in all experiments, but the ensemble-mean frequency of occurrence is about 30% too low in FLOR_{sst}. The observed dipolar configuration in WT3 is basically absent in the FLOR_{sst+strat} ensemble mean for that regime, and for WT4 all experiments exhibit southwardly shifted and widely enhanced negative geopotential anomalies with respect to the observations. WT5 appears severely tilted (about +45°) in the simulations, arguably due to a misrepresentation and location of the weak positive geopotential anomaly observed north of 50°N near and over the Great Lakes (Fig. 4). FLOR tends to show positive geopotential anomalies along the eastern coast of the United States in WT6, especially in the FLOR_{sst+strat} experiment, which also exhibits important frequency biases for that regime. Finally, although the simulated WT7 spatial patterns are very well represented (see high correlations in Tables 1 and 2), all experiments overestimate its frequency of occurrence.

Since there are no striking differences in the weather type representations in terms of the resolution of the reanalyses, the NNRPv2 product is selected as the reference in all following discussions about circulation regimes characteristics, unless otherwise indicated.

The persistence of these circulation patterns and their transitions control, both in the real world and in the simulations, aspects like the moisture that is advected into NENA, and thus this weather-within-climate approach is normally used to better understand the

physical mechanisms behind the occurrence or not of (extremely) rainy days in a region (Robertson and Ghil 1999; Moron et al. 2008a, 2013, 2015; Muñoz et al. 2015, 2016). As discussed in section 4, the systematic errors in the simulated weather regimes can help to explain the biases in other variables, as for example in the rainfall field and its climatological behavior. To illustrate this idea, Fig. 5 shows the average daily rainfall regimes (RR), defined by compositing rainfall values associated with each weather type.

As expected, FLOR tends to simulate better the rainfall field than LOAR at a local level; however, at a regional scale, there are not statistically significant differences ($p < 0.05$; two-sample Student's t test) in the spatial correlation coefficients between the three experiments. Table 3 shows the anomaly correlation coefficients between the rainfall regimes in the models and observations. Overall, the most significant biases appear in the daily rainfall regimes associated with the weather types showing greater bias over NENA (Fig. 5), that is, WT4 and WT5. Errors in magnitude and spatial distribution of the rainfall anomalies are related to misrepresented atmospheric circulations and moisture fluxes in WT4 in all experiments (not shown), exhibiting positive rainfall anomalies along the coast as a result of the deformed and displaced trough discussed above (Fig. 4). In spite of the tilt that WT5 exhibits in the simulations, the associated daily rainfall patterns for NENA are not too biased, except along the northeastern coast where the anomalies should have opposite sign; this is attributed in part to the fact that the circulation and moisture fluxes over NENA are similar to the observed one, even when that is not the case at a larger scale. Likewise, although WT3 is the circulation regime with the worst large-scale anomaly correlation coefficients (e.g., Table 1), its representation of the associated physical mechanisms controlling rainfall over NENA provides a very good depiction of the observed rainfall regime RR3. Having this regime well represented in the simulations is important to adequately represent extreme rainfall events over the Ohio River basin (Robertson et al. 2015).

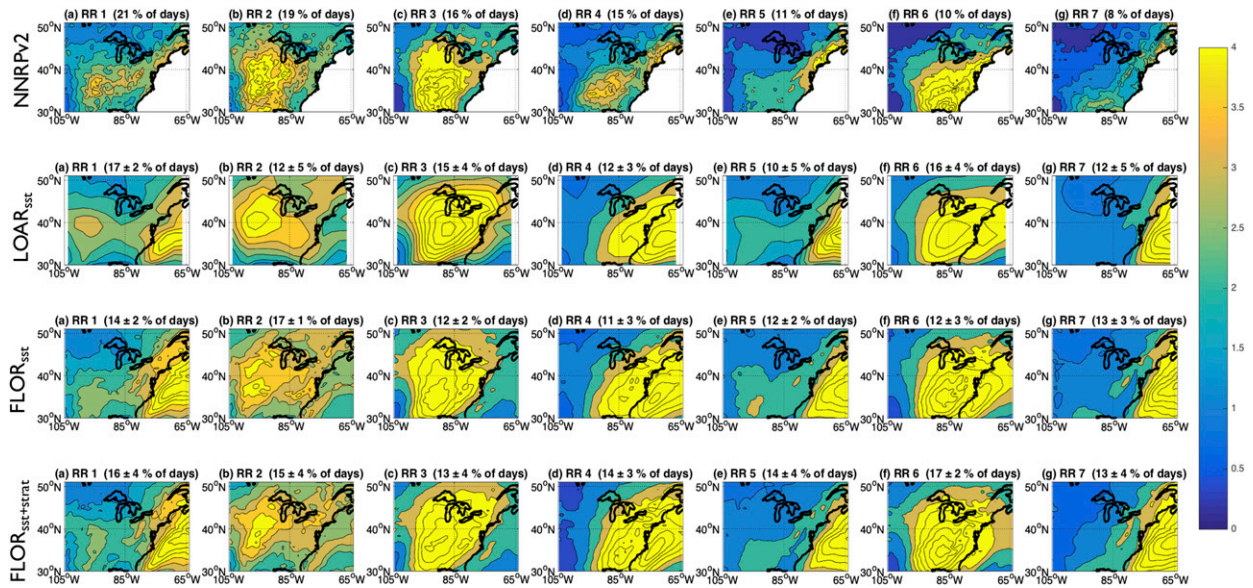


FIG. 5. Observed and modeled ($LOAR_{sst}$, $FLOR_{sst}$, $FLOR_{sst+strat}$) rainfall regimes (RR) associated with each weather type ($mm\ day^{-1}$). Relative frequency of occurrence for the entire 32-yr period is indicated in parentheses, with model experiments showing the ensemble mean \pm one standard deviation. Each dataset is plotted using its native horizontal resolution.

Biases both in the spatial patterns (involving shape, magnitudes, tilt, location) and the frequency of occurrence of the simulated weather types contribute to biases in rainfall and other variables at different time scales. In studies involving a large number of years (e.g., 100 years), two different k -means solutions should be computed to analyze, for example, whether the climate change signal is modifying the weather types' spatial patterns or the dimensionality of the phase space. For studies like the present one, however, which involves just three decades, the spatial patterns are normally assumed to be constant. This allows for the analysis of the weather types' temporal variability at different time scales in terms of changes in their frequency of occurrence at those time scales.

A possible way to summarize how well the simulations reproduce both the mean frequency of occurrence and its variability across multiple time scales is through the use of the corresponding scatter indices [see Eqs. (7) and

(8)], presented in Fig. 6 for two particular subseasonal windows (20–30 March and 4–14 May; further discussed in section 5c), for the interannual variability considering all MAM seasons in the 1981–2012 period (section 5d), and for the first and last decades in the same period of years (section 5e). As with the case of the spatial patterns (e.g., Table 1), the analysis of the frequencies of occurrence indicates that no particular model or experiment can be overall considered the best one across all the time scales in study. Certainly, there are differences in terms of the median and dispersion values at particular scales, with higher errors and uncertainties in the subseasonal case; nonetheless, within-scale differences in the medians are normally negligible. Furthermore, compared to $LOAR_{sst}$, $FLOR_{sst}$ tends to have the same or a lower dispersion for both scatter indices (Fig. 6).

The following subsections provide more details on daily transition statistics using Klee diagrams (sections

TABLE 3. Anomaly correlation coefficient (Spearman) between the model-equivalent ensemble-mean rainfall regimes (RRs) in each experiment and the corresponding rainfall pattern in the observations, computed on the observations grid for the domain sketched in Fig. 5. Numbers in bold indicate highest correlations found in the model experiments, all values being statistically significant (Student's t test; $p < 0.05$). No statistically significant difference was found between the average anomaly correlation coefficients of the experiments.

Experiment	RR1	RR2	RR3	RR4	RR5	RR6	RR7	Average
$LOAR_{sst}$	0.80	0.83	0.87	0.51	0.57	0.80	0.87	0.75
$FLOR_{sst}$	0.78	0.81	0.90	0.48	0.64	0.79	0.88	0.75
$FLOR_{sst+strat}$	0.83	0.82	0.88	0.38	0.69	0.62	0.90	0.73

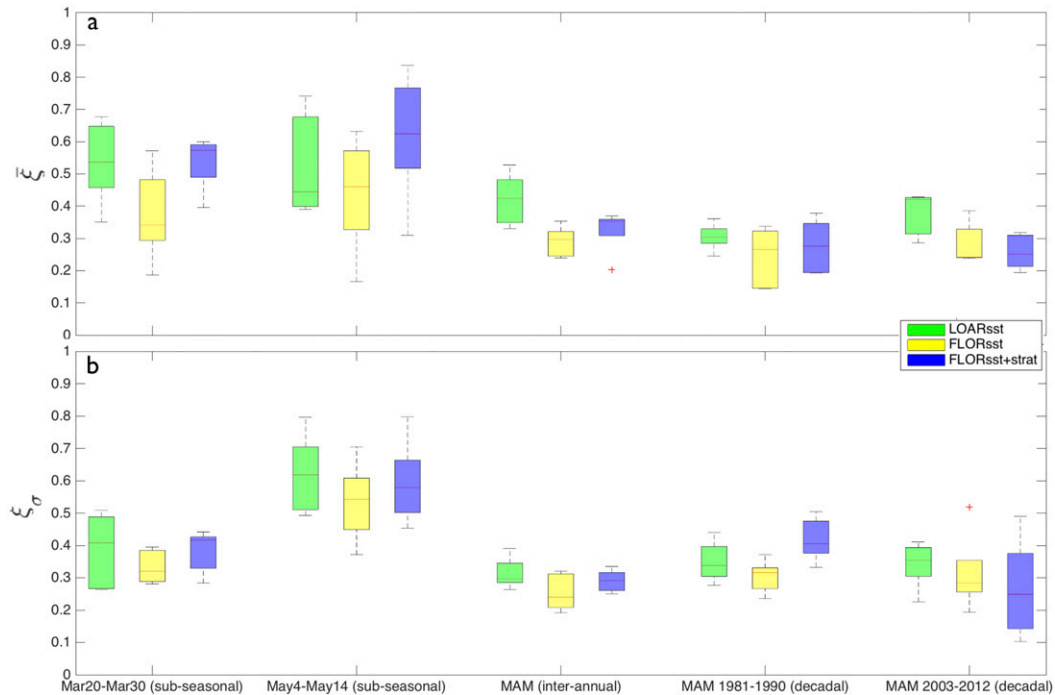


FIG. 6. Boxplots of scatter indices for (a) the mean and (b) standard deviation of the frequency of occurrence of weather types across multiple time scales. Perfect coincidence between simulations and observations corresponds to a scatter index value of zero. For each boxplot, the central mark indicates the ensemble median, and the bottom and top edges of the box indicate the ensemble 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers (values >2.7 standard deviations) are plotted individually using the “+” symbol. For additional details see sections 5c, 5d, and 5e.

5a and 5b) and further analysis pertinent to the temporal variability of the simulated circulation regimes and to understanding Fig. 6 (sections 5c, 5d, and 5e).

a. Klee diagrams

Klee diagrams, shown in Fig. 7, are a way to represent the temporal evolution of the available states of the system (Muñoz et al. 2015). These diagrams consist of a simple matrix plot sketching the daily evolution of weather types for the entire period under study. They are equivalent to the representation of the Viterbi state sequences, except for the fact that no Viterbi algorithm or hidden Markov model (e.g., Robertson et al. 2006) is involved in the process. A Klee diagram is the basis for analyzing daily transitions and temporal variability at subseasonal, seasonal, decadal, and longer time scales. Moreover, it has been used to define subseasonal-to-seasonal states for forecast purposes, using hybrid dynamical–statistical models (Muñoz et al. 2016).

If a simulation has exactly the same Klee diagram as the one computed from the observations, then it has a perfect representation of the observed weather types’ evolution across time scales, independently of how well the model represents their spatial patterns; of course,

even if the models were perfect, the Klee diagrams would not be exactly the same due to atmospheric chaos. Nonetheless, the idea is conceptually useful, and helps to define several ways to diagnose similarity in the temporal evolution and associated statistics. Moreover, having simulated Klee diagrams that exhibit similar characteristics to observations is also of practical use because it is then possible to use a combination of modeled frequencies of occurrence and observed spatial patterns to represent the evolution of the circulation regimes at a given time scale. In such cases, this approach has important implications for prediction, an idea that is briefly discussed in section 6, and that will be explored in more detail elsewhere.

Overall, the observed and simulated Klee diagrams for NENA show high similarity (Fig. 7). Since it is not possible to have an ensemble mean of Klee diagrams because they represent categories and not real numbers, the analysis must be performed on a member-per-member basis. Visual inspection suggests that all experiments exhibit a dominance of WT1 and WT2 toward the end of the season, and of WT6 and WT7 during the first half, all consistent with the observations and consistent with the predominance of negative height

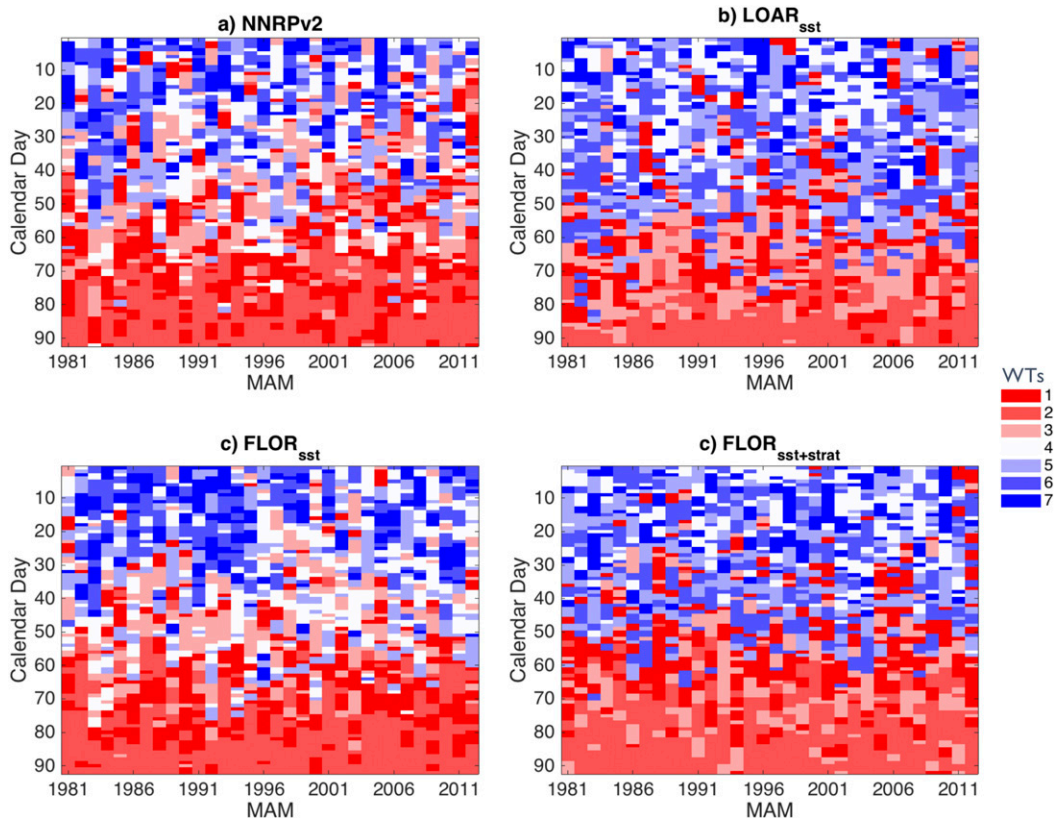


FIG. 7. Klee diagrams for (a) NNRPv2 and numerical experiments (b) $LOAR_{sst}$, (c) $FLOR_{sst}$, and (d) $FLOR_{sst+strat}$. Each tile corresponds to a particular day, and the colors represent different weather types (see color bar). Only one member per experiment is shown (others are similar).

anomalies in March and positive ones in May within the MAM season (Fig. 4). However, some biases are apparent, like the presence in $LOAR_{sst}$ and $FLOR_{sst+strat}$ of too many days with WT3 at the end of the season, which, for example, could provide false alarms in those cases in which WT3 is related to floods in the Ohio River basin.

Different statistics can easily be computed from the Klee diagrams to help diagnose and compare dynamical models. Some examples—like daily transition probabilities, mean durations (persistence) for each weather type, and other summaries to analyze the evolution of the circulation regimes at different time scales—are discussed in the following subsections.

b. Daily transitions and typical durations

Daily transition matrices are commonly used to characterize persistence and preferred state transitions. Their diagonal sketches the persistence probabilities for each weather type, and the off-diagonal elements represent the conditional probabilities of transition to a particular regime (along the horizontal axes of the matrix; “posterior WT”) given that a different weather

type (along the vertical axes; “prior WT”) occurred on the previous day.

Daily transitions in NENA, presented in Fig. 8, are dominated by persistence, with continental ridges over the Great Lakes (WT2) being the most probable regime to persist, especially during May (Fig. 7). The most frequently observed statistically significant transitions ($p < 0.1$; see Vautard et al. 1990) suggest circuits like $1 \rightarrow (2, 3) \rightarrow 1$ or $4 \rightarrow 3 \rightarrow 1$, that can be associated with the propagation of baroclinic waves over NENA, but also transitions leading to highly persistent states, like $6 \rightarrow 7 \rightarrow 4$, associated with anomalously high moisture fluxes from the Gulf of Mexico and the Caribbean conducive to rainy days along the eastern coast, followed by days with a high pressure system stationed in the northern part of the region, typically conducive to low-level divergence and no rain (see Figs. 4, 5, and 7).

All simulations for NENA adequately represent the fact that the persistence probabilities are significantly higher than nonself transitions. As expected, some biases exist; for example, the persistence of WT4 in the $LOAR_{sst}$ and $FLOR_{sst}$ experiments (see Fig. 8) is overestimated, and there are consistent biases in all

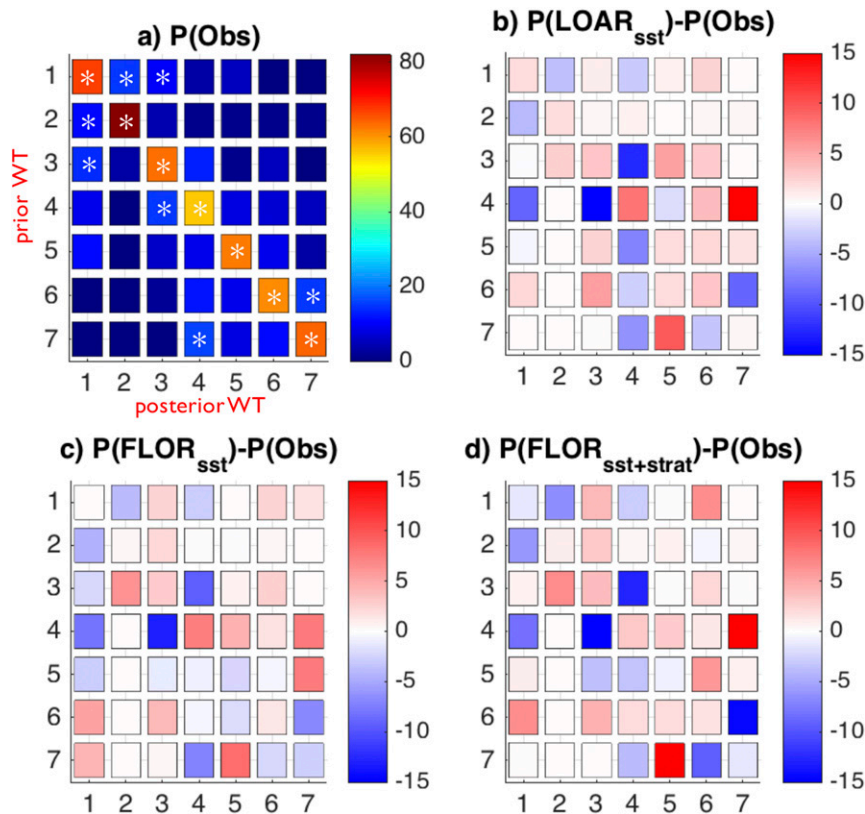


FIG. 8. (a) Daily transition probabilities P for observed (NRRPv2) weather types (see label bar). A star indicates statistically significant transitions ($p < 0.1$; see Vautard et al. 1990). Vertical and horizontal axes correspond to the prior and posterior WTs, respectively. (b)–(d) Differences between the $LOAR_{sst}$, $FLOR_{sst}$, and $FLOR_{sst+strat}$ ensemble-mean transition probabilities and those presented in (a), respectively. Average differences among (b)–(d) are not statistically significant ($p < 0.05$, using a two-sample Student's t test).

experiments for transitions like $3 \rightarrow 4$, $4 \rightarrow 3$, $4 \rightarrow 1$, $7 \rightarrow 5$, and $4 \rightarrow 7$. But in general, all models tend to capture most of the observed transition probabilities, with differences always less than $\pm 15\%$; moreover, no statistically significant differences ($p < 0.05$; two-sample Student's t test) exist between the average transition probabilities of the experiments.

It is also important to evaluate if the simulations fairly represent the typical persistence of the weather types. Figure 9 shows relative frequency histograms and their corresponding Weibull distribution fit for the most common durations; for comparison, Table 4 presents the values of the Weibull distribution parameters α and β , which measure the scale (or characteristic life) and the shape (or slope) of the distribution, respectively.

This analysis indicates that WT1–WT3 (predominant in May) tend to persist more than the other weather types (as expected from the visual inspection of the observations' Klee diagram; Fig. 7), WT4 tends to transition faster than the other regimes, and WT5–WT7 exhibit similar duration

probability density functions (PDFs), all these facts in agreement with the observations.

As with the spatial patterns of the weather types between models and observations, there is not a unique set of experiments that is consistently better than the others (i.e., for which all the regimes adequately represent the observed persistence distributions); nonetheless, $FLOR_{sst}$ tends to have better self-transition statistics for WT2, and WT4–WT6 (Table 4; bold numbers indicate experiments with the best representation of self-transitions, or persistence, for each WT). Overall, although the main characteristics of the duration PDFs are captured well by the experiments (e.g., the fact that can be well modeled by a Weibull distribution, higher persistence in WT1–WT3), certain features are not. It is possible to identify WT5 and WT6 as the regimes with the worst representation in terms of persistence in all simulations (Table 4); for example, WT5 overestimates the number of “early” transitions by at least a factor of 1.5. As complementary information, Table 5 summarizes the average persistence of each WT

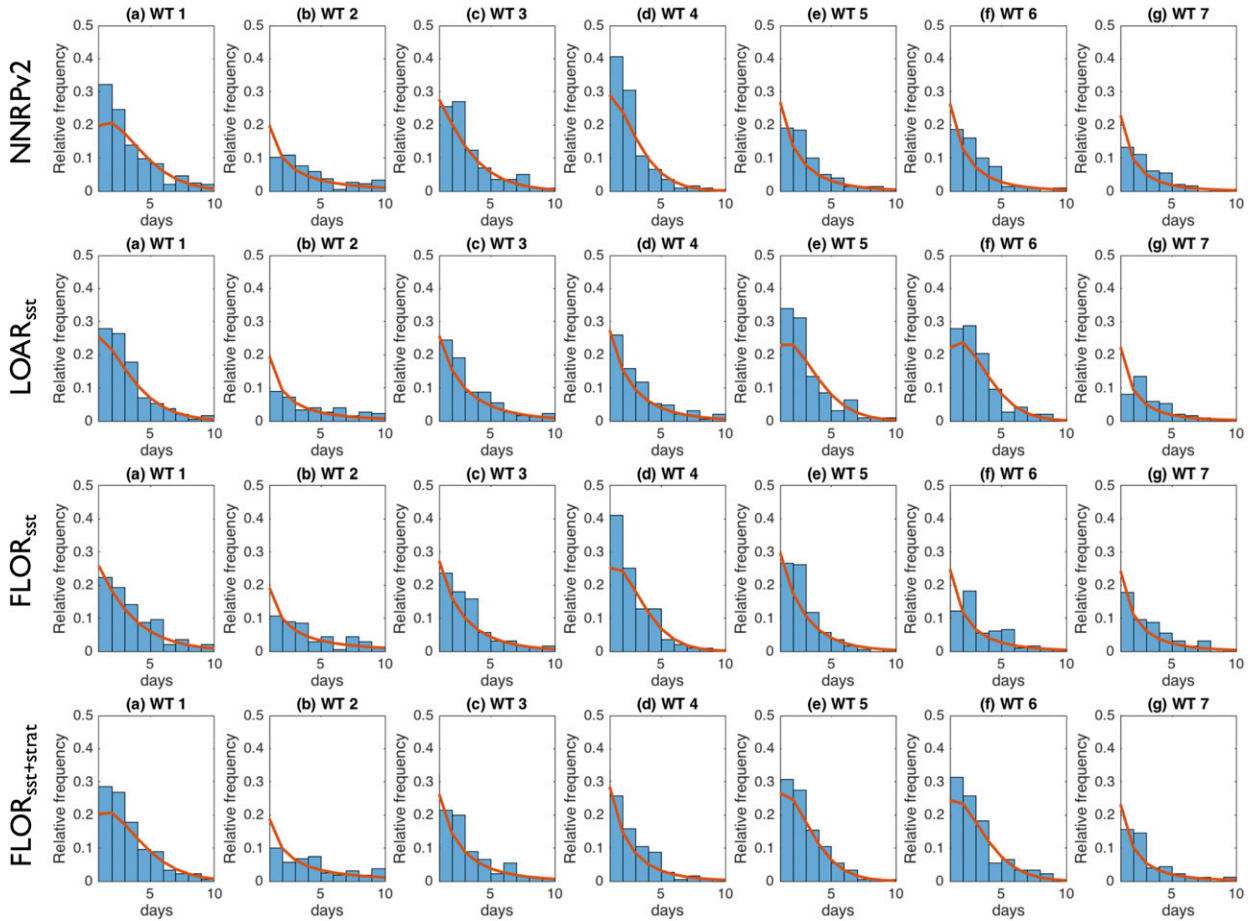


FIG. 9. Average durations (in days) for each weather type (see label bar) in NNRpv2, LOAR_{sst}, FLOR_{sst}, and FLOR_{sst+strat}. Red curves sketch the corresponding Weibull fit, whose parameters α and β are presented in Table 4.

in the NNRpv2 and the numerical experiments; results are consistent with the discussion presented above and Table 4.

c. Subseasonal evolution

The typical subseasonal evolution of the weather types can be analyzed using their “climatological” frequency of occurrence as computed by the regime’s appearance for each calendar day, after an 11-day moving average is applied to the frequency of occurrence time

series in order to filter out the shortest time scales. This metric, hereafter referred to as “subseasonality” to avoid the cacophonous phrase “subseasonal seasonality,” is shown in Fig. 10. Clearly, MAM is a transition season between boreal winter and summer that could be characterized by low occurrence of ridge configurations (WT1–WT3) during March, and their dominance during May.

Generally speaking, the sketched subseasonal evolution of the weather types in the numerical experiments is

TABLE 4. Best fit of Weibull parameters (α, β) for the model-equivalent ensemble-mean weather type durations sketched by the red curves in Fig. 9. Bold pairs indicate minimum Euclidean distance $ds = \sqrt{(\Delta\alpha^2 + \Delta\beta^2)}$, where Δ denotes differences between simulated and observed parameters.

	WT1	WT2	WT3	WT4	WT5	WT6	WT7
NNRPv2	(3.52,1.46)	(1.71,0.55)	(2.56,1.13)	(2.53,1.36)	(1.47,0.76)	(1.30,0.73)	(0.83,0.62)
LOAR _{sst}	(2.90,1.29)	(1.13,0.53)	(2.22,0.83)	(1.89,0.84)	(3.09,1.48)	(3.12,1.57)	(0.82,0.61)
FLOR _{sst}	(2.73,1.04)	(1.71,0.54)	(2.05,0.87)	(2.86,1.48)	(1.94,0.95)	(1.21,0.68)	(1.10,0.66)
FLOR _{sst+strat}	(3.47,1.43)	(1.87,0.53)	(1.82,0.78)	(1.56,0.82)	(2.74,1.47)	(2.95,1.45)	(0.93,0.63)

TABLE 5. Average persistence (in days) for each WT in the NNRPv2 dataset and the numerical experiments. Values close to reanalysis are presented in bold.

	WT1	WT2	WT3	WT4	WT5	WT6	WT7
NNRPv2	4.06	3.47	3.36	3.21	2.62	2.48	2.14
LOAR _{sst}	3.59	2.94	3.31	2.96	3.66	3.69	2.16
FLOR _{sst}	3.59	3.56	3.09	3.46	2.88	2.47	2.40
FLOR _{sst+strat}	3.98	3.89	2.95	2.62	3.39	3.57	2.26

similar to the observations, although there is an overall delayed subseasonality in several weather types (cf. the slopes of the curves in Figs. 10b–d with the ones in Fig. 10a). Biases are present both in the subseasonal frequency of occurrence of some weather types and in their actual evolution. For example, the frequency of WT4 is underestimated in all experiments, especially after calendar day 50 (Fig. 10); the model-equivalent WT5–WT7 tend to have more or less the same relative frequency of occurrence in April than in March, which is

inconsistent with observations; WT1 has its peak of occurrence typically between calendar days 65 and 75 (4–14 May), but this maximum appears about 10 days earlier in FLOR_{sst+strat}.

Because of the clear differences between the beginning and the end of the season, two periods were considered for further analysis: 20–30 March and 4–14 May (see black boxes in Fig. 10).

Errors in the median values of the simulated frequencies of occurrence of weather types, as measured by the scatter index (Fig. 6a), tend to be similar between the two periods under consideration, although with higher dispersion during the second half of the season, which is mostly due to the misrepresentation of the subseasonality of WT4–WT6. On the other hand, errors in the standard deviations of the occurrences are more clearly discriminated (Fig. 6b), being similar in all three experiments but with higher values for the 4–14 May period. For the end-of-March section under study, the median values for the standard deviations of the

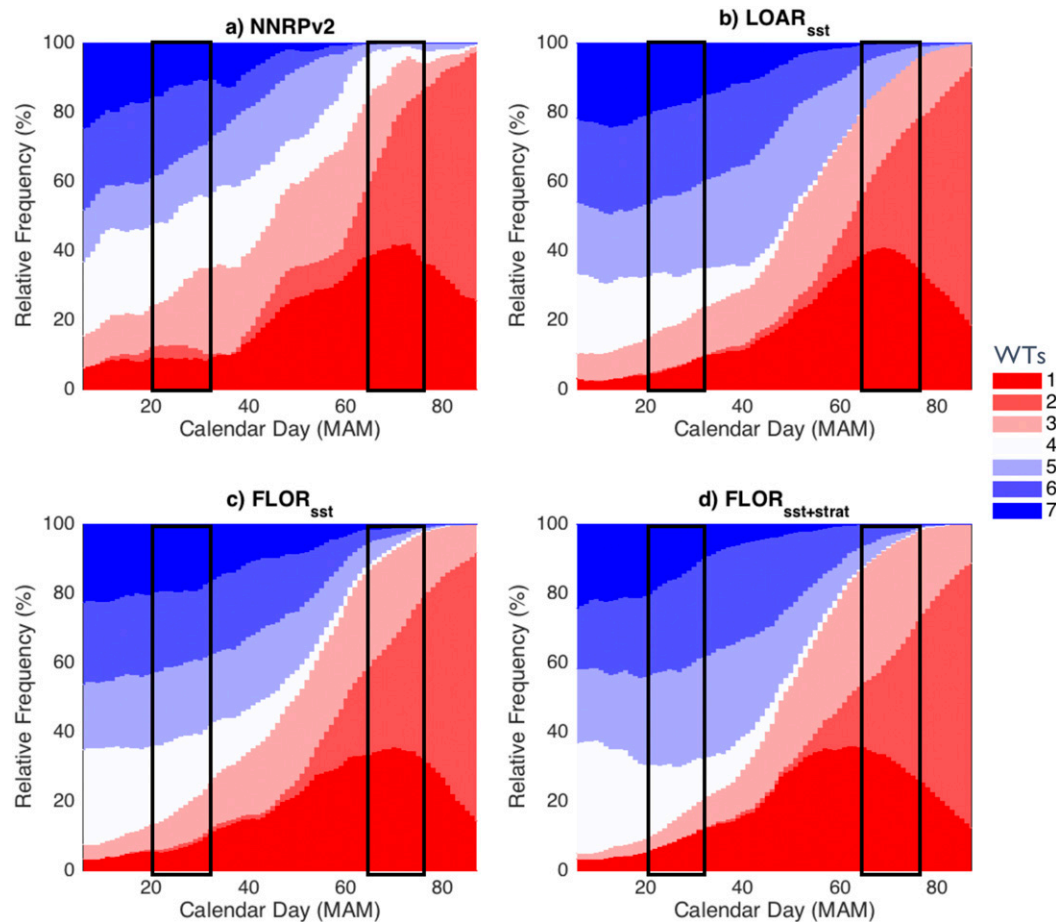


FIG. 10. Ensemble-mean subseasonal frequency of occurrence for each weather type, smoothed with an 11-day moving average. Periods in the black boxes were selected for further analysis (see section 5c).

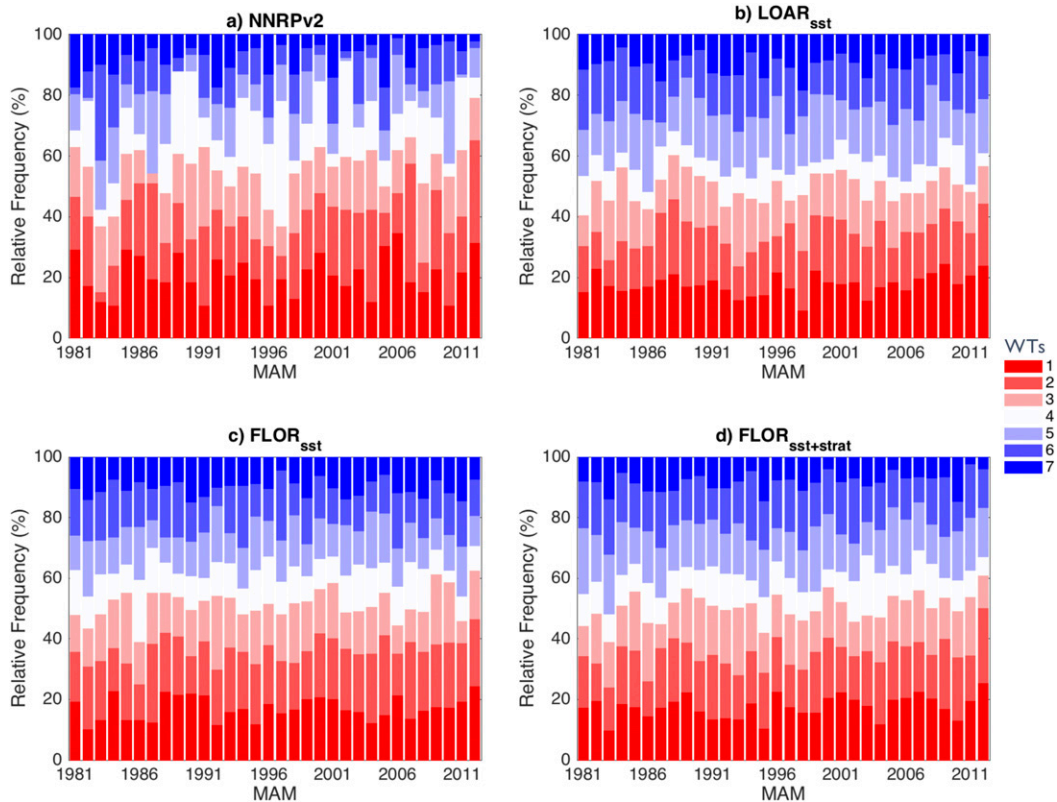


FIG. 11. (a) Observed and (b)–(d) ensemble-mean interannual frequency of occurrence for each weather type (see label bar), for all MAM seasons.

simulated frequencies are similar to the other time scales analyzed in this work (interannual and decadal), although $LOAR_{sst}$ exhibits higher and similar dispersions at subseasonal scale than at the other scales.

d. Interannual variability

Analysis of the ensemble-mean interannual evolution of the frequency of weather types, shown in Fig. 11, indicates that the highest root-mean-square errors, presented in Table 6, occur for patterns associated with troughs north of the Great Lakes (WT4), especially for $LOAR_{sst}$ (~11.7 days per season, compared to ~10.0 and ~9.8 days for $FLOR_{sst}$ and $FLOR_{sst+strat}$, respectively). On the other hand, WT3 and WT7—northeastern seaboard ridges and deep troughs—have the best average representation of the interannual variability in all experiments, with slightly lower errors for $LOAR_{sst}$ and $FLOR_{sst+strat}$ (~5.1 days for both WT3 and WT7 in both experiments, compared to ~6 days for the same weather types in $FLOR_{sst}$). Nonetheless, on average, there are no statistically significant differences ($p < 0.05$, using a two-sample Student’s *t* test) between the experiments, with an overall median error of 7.3 days, which is lower than

10% of the total days in the season. This result is consistent with the scatter index values for the mean frequency of occurrence at interannual scales (Fig. 6a).

The variability in the frequencies—measured by the associated standard deviation—is, again, very similar between the experiments. Moreover, the fact that there is relatively low dispersion in the variability scatter index for almost all time scales (Fig. 6b) implies agreement between the different members in each numerical experiment, suggesting low model uncertainty in the reported values of this parameter. Nonetheless, overall Fig. 11 suggests that the experiments do not seem to capture well the observed interannual variability.

TABLE 6. Ensemble-mean RMSE (in days per season) for the interannual evolution of the frequency of occurrence of the weather types in each experiment, with respect to NNRpv2.

Experiment	WT1	WT2	WT3	WT4	WT5	WT6	WT7
$LOAR_{sst}$	7.4	7.3	5.1	11.7	10.2	7.9	5.2
$FLOR_{sst}$	7.4	6.4	6.2	10.0	7.7	7.1	5.7
$FLOR_{sst+strat}$	6.6	6.9	5.0	9.8	7.1	7.8	5.1

e. Decadal differences

There are not enough years to formally study interdecadal variability, and thus only the differences in the frequency of occurrence of weather types were analyzed for the first and last decades, after smoothing the frequency of occurrence time series using an 11-yr moving average, to pick up the long-time-scale signals of interest in the present analysis.

Differences in medians and dispersions of the mean frequency of occurrence are negligible for the two decades (Fig. 6a), although dispersion in the simulated mean frequencies of occurrence are definitively higher in the FLOR experiments than in LOAR_{sst} for the 1981–90 decade. FLOR_{sst} and FLOR_{sst+strat} show slightly lower medians for the scatter index of the standard deviations for the 2003–12 decade, although with higher dispersions than LOAR_{sst}. Overall, both decades show the same ranges and values for the statistics considered.

6. Discussion

The approach presented here offers tailored diagnostics to understand possible sources of model biases at multiple time scales. The basic idea is that the inherent biases at the synoptic or low-frequency variability scale in models could be rectified at larger scales, according to how climate drivers excite the observed weather types differentially on longer subseasonal-to-seasonal and seasonal-to-decadal scales. A variety of diagnostic metrics were explored to identify model errors at several time scales. Nonetheless, putting together the big picture provided by the different metrics is in general not a trivial task.

The analysis performed indicates that, at large scale, the simulations have trouble representing the observed spatial pattern associated with WT3 (Table 2), although at regional scale—over NENA—the spatial configuration of the northeastern seaboard ridges is actually good enough to provide a fair representation of the observed rainfall regimes (Table 3), suggesting that the observed physical mechanisms that control rainfall in that case (e.g., moisture and heat transport from the Gulf of Mexico; Fig. 4) are present in the simulations. This has important implications for flood prediction in the Ohio River basin (Nakamura et al. 2013; Robertson et al. 2015).

In contrast, the lowest rainfall pattern correlations are obtained for the regime associated with troughs north of the Great Lakes (RR4, WT4; see Table 3), attributed earlier to the southward displacement of the geopotential anomaly with respect to the observed pattern, and its enhancement over most of North America. This is truly the worst circulation regime simulated in the experiments.

Not only is the spatial pattern poorly simulated, but also the depiction of the observed temporal variability across all time scales is the worst of all WTs. These issues are hypothesized here to be part of the same pathology, and they could be related to problems in the models' rendition of tropical–extratropical interactions, as the seasonal frequency of occurrence of WT4 is significantly correlated with the Niño-3.4 SST index (not shown), a link that will be treated—along with other teleconnection indices and circulation regimes—in a future paper, following the methodology discussed in Muñoz et al. (2015).

The other weather regimes have a fair representation of the geopotential anomalies and rainfall patterns over NENA, although this is not necessarily true at the continental or hemispheric scale.

Altogether, there are not significant differences in the average performance of LOAR and FLOR in terms of the reproduction of the spatial patterns and temporal variability of the observed weather types, although a certain experiment can be better than the others when considering particular characteristics (e.g., mean frequency at interannual scale, or the persistence of WT4; see section 5). In this work, the question of which model is better is really a question of what nudging approach performs better and what is the impact of horizontal resolution.

Nudging both SST and stratospheric fields did not consistently improve the representation of the weather types in the model, and indeed in some cases provided worse results than the SST-only nudging experiment (FLOR_{sst}). It is possible that some stratosphere–troposphere and ocean–atmosphere interactions are not being well simulated, and that some improvement can be achieved if the vertical resolution in the model is increased to adequately account for the stratospheric processes. This is a matter of future research.

As indicated earlier, no significant improvement was found when increasing horizontal resolution, either in the reanalysis products or in the model experiments (LOAR_{sst} and FLOR_{sst}). This is attributed to the fact that synoptic-scale 500-hPa geopotential height anomalies do not really require high resolution in order to reproduce the key physical mechanisms associated with, for example, propagation of Rossby waves that perturbs circulation patterns, or the moisture advection conducive to rainfall; in addition, the topography in NENA is not tall or complex enough as to negatively impact the low-resolution model. Yet, high horizontal resolution could be important in other regions of the world.

Although high spatial resolution does not seem to be necessary to satisfactorily reproduce observed weather types, high-resolution models like FLOR have the advantage of providing physically related variables like

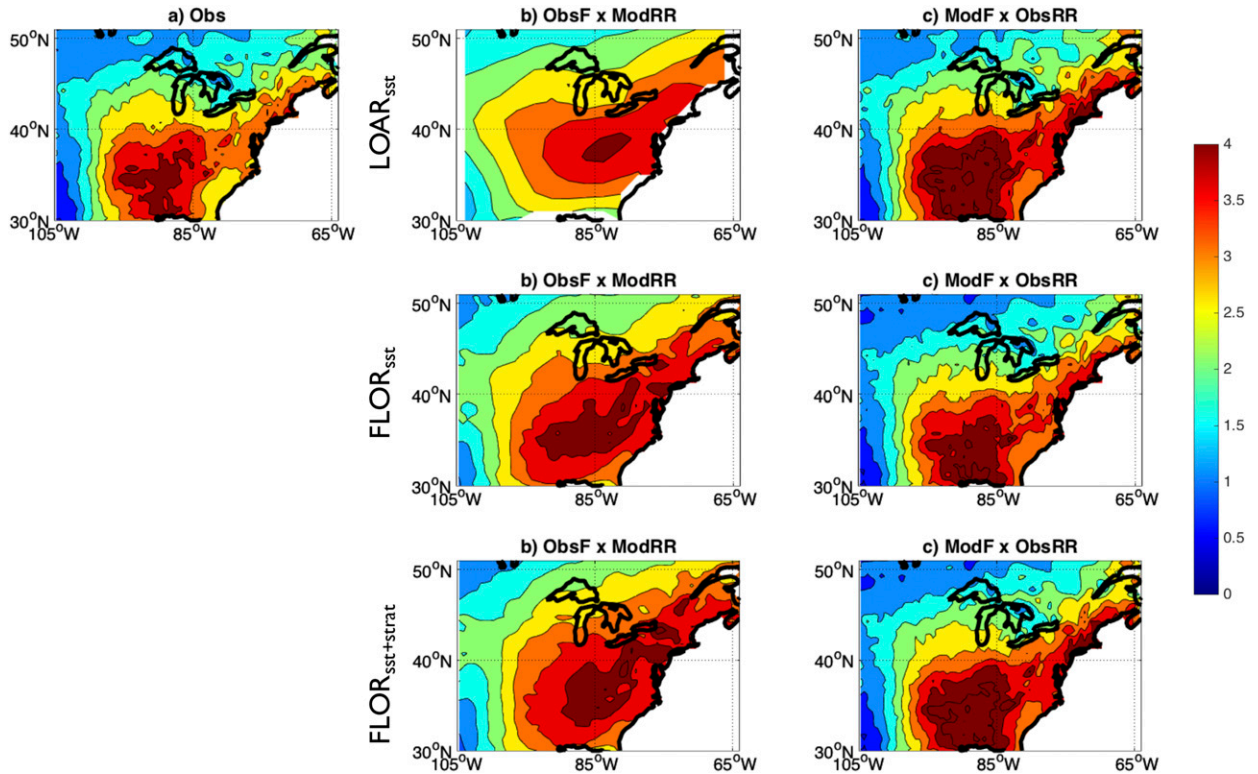


FIG. 12. (a) Observed rainfall climatology (MAM 1981–2012), and reconstruction of the rainfall climatology using, for each numerical experiment, and linear combinations of (b) the observed frequencies of occurrence of weather types (ObsF) and the modeled rainfall regimes (ModRR) and (c) the modeled frequencies of occurrence of weather types (ModF) and the observed rainfall regimes (ObsRR). Units in mm day^{-1} . Ocean has been masked in (b). Each dataset is plotted using its native horizontal resolution.

rainfall or surface temperature at a resolution preferred by decision-makers and for the provision of climate services (Vaughan and Dessai 2014). Nonetheless, it is possible to exploit the fair representation of weather type characteristics by faster coupled models like LOAR to “reconstruct” fields of interest (e.g., the rainfall climatology in NENA).

Since the rainfall climatology $\langle p \rangle$ can be computed in terms of a linear combination of rainfall regimes RR_i and the total frequency of occurrence of those regimes F_i ,

$$\langle p \rangle = \sum_i^k F_i RR_i \quad (10)$$

for $i = 1 \dots k$ regimes, if the model has an adequate representation of only one of these fields, then observations could be used to compute the other factor in the linear combination. For example, the rainfall regimes of the experiments analyzed in this work (ModRR) could be used in conjunction with the observed frequencies of occurrences (ObsF) to reconstruct the observed rainfall climatology, as shown in Figs. 12a and 12b. For other

approaches that can be used for prediction purposes, see Moron et al. (2008b).

As expected, the experiments with FLOR provide better spatial rainfall patterns than $LOAR_{sst}$ because of their higher resolution. Although there are biases, the use of the observed frequency of occurrences has improved the original simulated rainfall climatologies (Figs. 2b–d). Further improvement is obtained if the observed rainfall regimes (ObsRR) are used in conjunction with the modeled frequencies (ModF; Fig. 12c). With a satisfactory bias-correction method, this approach has the potential to provide relatively economic—at least from a computational point of view—diagnostic products and forecasts.

7. Concluding remarks

This work discussed a new diagnostic framework to evaluate the performance of models across multiple time scales, based on their representation of the observed spatial and temporal variability of weather types. Under this nonlinear system dynamics perspective, “good” models are those that correctly reproduce the

observed characteristics of the weather types at multiple time scales.

The framework takes advantage of the weather-typing decomposition, in terms of the spatial patterns of the circulation regimes and their temporal evolution, to analyze model performance at multiple time scales focusing on the evaluation of tailored statistics like daily transition probabilities, weather type mean durations, and subseasonal, interannual, and longer-term frequencies of occurrence. Furthermore, since the circulation regimes are normally linked to concrete climate modes, they can also be used to diagnose model biases from a physical perspective, like deformations or displacements of particular geopotential height configurations that control the occurrence of rainfall in a region of the world.

To illustrate how the diagnostic approach works, it was applied to three different sets of numerical experiments using Geophysical Fluid Dynamics Laboratory coupled circulation models. The simulations tend to represent well the location, shape, and magnitude of daily circulation regimes and associated rainfall patterns, although some important biases were reported and discussed. Further research is being conducted to perform an in-depth analysis of possible tropical–extratropical interactions that might not be well represented by the models.

Finally, the present framework can also be used for model intercomparison, and can be applied to uncoupled and regional models.

Acknowledgments. The authors are grateful to Tony Barnston and Simon Mason (IRI) and Nathaniel Johnson and Ángel Adames (GFDL) for discussions about different aspects of the paper. ÁGM was supported by National Oceanic and Atmospheric Administration's Oceanic and Atmospheric Research, under the auspices of the National Earth System Prediction Capability. AWR was supported by NOAA Next Generation Global Prediction System (NGGPS) project Grant NA16NWS4680014. This paper is dedicated to Eneko Muñoz and Cathy Vaughan (ÁGM's nena).

REFERENCES

- Archambault, H. M., L. F. Bosart, D. Keyser, and A. R. Aiyyer, 2008: Influence of large-scale flow regimes on cool-season precipitation in the northeastern United States. *Mon. Wea. Rev.*, **136**, 2945–2963, doi:10.1175/2007MWR2308.1.
- Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.
- Becker, E. J., E. H. Berbery, and R. W. Higgins, 2011: Modulation of cold-season U.S. daily precipitation by the Madden–Julian oscillation. *J. Climate*, **24**, 5157–5166, doi:10.1175/2011JCLI4018.1.
- Charney, J. G., and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205–1216, doi:10.1175/1520-0469(1979)036<1205:MFEITA>2.0.CO;2.
- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, doi:10.1029/2007JD009132.
- Christensen, H. M., I. M. Moroz, and T. N. Palmer, 2015: Simulating weather regimes: Impact of stochastic and perturbed parameter schemes in a simple atmospheric model. *Climate Dyn.*, **44**, 2195–2214, doi:10.1007/s00382-014-2239-9.
- Covey, C., P. J. Gleckler, C. Doutriaux, D. N. Williams, A. Dai, J. Fasullo, K. Trenberth, and A. Berg, 2016: Metrics for the diurnal cycle of precipitation: Toward routine benchmarks for climate models. *J. Climate*, **29**, 4461–4471, doi:10.1175/JCLI-D-15-0664.1.
- Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, doi:10.1175/JCLI3629.1.
- , and Coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25**, 2755–2781, doi:10.1175/JCLI-D-11-00316.1.
- Frankoski, N. J., and A. T. DeGaetano, 2011: An East Coast winter storm precipitation climatology. *Int. J. Climatol.*, **31**, 802–814, doi:10.1002/joc.2121.
- Fujii, Y., T. Nakaegawa, S. Matsumoto, T. Yasuda, G. Yamanaka, and M. Kamachi, 2009: Coupled climate simulation by constraining ocean fields in a coupled model with ocean data. *J. Climate*, **22**, 5541–5557, doi:10.1175/2009JCLI2814.1.
- Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55, [https://doi.org/10.1175/1520-0477\(1999\)080<0029:AOOTRO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2).
- Ghil, M., and A. W. Robertson, 2002: “Waves” vs. “particles” in the atmosphere's phase space: A pathway to long-range forecasting? *Proc. Natl. Acad. Sci. USA*, **99** (Suppl. 1), 2493–2500, doi:10.1073/pnas.012580899.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007JD008972.
- , C. Doutriaux, P. Durack, K. Taylor, Y. Zhang, D. Williams, E. Mason, and J. Servonnat, 2016: A more powerful reality test for climate models. *Eos, Trans. Amer. Geophys. Union*, **97**, doi:10.1029/2016EO051663.
- Griffies, S., 2012: Elements of the Modular Ocean Model (MOM). GFDL Tech. Rep. 7, 614 pp.
- Hoskins, B. J., 2013: The potential for skill across the range of the seamless weather-climate prediction problem: A stimulus for our science. *Quart. J. Roy. Meteor. Soc.*, **139**, 573–584, doi:10.1002/qj.1991.
- Jia, L., and Coauthors, 2017: Seasonal prediction skill of northern extratropical surface temperature driven by the stratosphere. *J. Climate*, **30**, 4463–4475, <https://doi.org/10.1175/JCLI-D-16-0475.1>.
- Johnson, N. C., and S. B. Feldstein, 2010: The continuum of North Pacific sea level pressure patterns: Intraseasonal, interannual, and interdecadal variability. *J. Climate*, **23**, 851–867, doi:10.1175/2009JCLI3099.1.
- Jolliffe, I., and D. Stephenson, Eds., 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. 2nd ed. Wiley and Sons, 292 pp.
- Jones, G. V., and R. E. Davis, 1995: Climatology of nor'easters and the 30 kPa jet. *J. Coast. Res.*, **11**, 1210–1220, <http://journals.fcla.edu/jcr/article/view/79985>.

- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267, doi:[10.1175/1520-0477\(2001\)082<0247:TNNYRM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2).
- Leathers, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific–North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate*, **4**, 517–528, doi:[10.1175/1520-0442\(1991\)004<0517:TPATPA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1991)004<0517:TPATPA>2.0.CO;2).
- Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.*, **32**, 2022–2026, doi:[10.1175/1520-0469\(1975\)032<2022:CRAFD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1975)032<2022:CRAFD>2.0.CO;2).
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, doi:[10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
- , 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, doi:[10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- , 2006: Regimes in simple systems. *J. Atmos. Sci.*, **63**, 2056–2073, doi:[10.1175/JAS3727.1](https://doi.org/10.1175/JAS3727.1).
- Luo, J.-J., S. Masson, S. K. Behera, and T. Yamagata, 2008: Extended ENSO predictions using a fully coupled ocean–atmosphere model. *J. Climate*, **21**, 84–93, doi:[10.1175/2007JCLI1412.1](https://doi.org/10.1175/2007JCLI1412.1).
- Mason, S. J., and D. Stephenson, 2008: How do we know whether seasonal climate forecasts are any good? *Seasonal Climate: Forecasting and Managing Risk*, A. Troccoli et al., Eds., Springer, 259–289.
- , and M. K. Tippet, 2016: Climate Predictability Tool version 15.3. Columbia University Academic Commons, doi:[10.7916/D8NS0TQ6](https://doi.org/10.7916/D8NS0TQ6).
- Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.*, **52**, 1237–1256, doi:[10.1175/1520-0469\(1995\)052<1237:WRRSQS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1237:WRRSQS>2.0.CO;2).
- Milly, P. C. D., and Coauthors, 2014: An enhanced model of land water and energy for global hydrologic and Earth-system studies. *J. Hydrometeorol.*, **15**, 1739–1761, doi:[10.1175/JHM-D-13-0162.1](https://doi.org/10.1175/JHM-D-13-0162.1).
- Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye, 2008a: Weather types and rainfall over Senegal. Part I: Observational analysis. *J. Climate*, **21**, 266–287, doi:[10.1175/2007JCLI1601.1](https://doi.org/10.1175/2007JCLI1601.1).
- , —, —, and —, 2008b: Weather types and rainfall over Senegal. Part II: Downscaling of GCM simulations. *J. Climate*, **21**, 288–307, doi:[10.1175/2007JCLI1624.1](https://doi.org/10.1175/2007JCLI1624.1).
- , —, and M. Ghil, 2012: Impact of the modulated annual cycle and intraseasonal oscillation on daily-to-interannual rainfall variability across monsoonal India. *Climate Dyn.*, **38**, 2409–2435, doi:[10.1007/s00382-011-1253-4](https://doi.org/10.1007/s00382-011-1253-4).
- , P. Camberlin, and A. W. Robertson, 2013: Extracting sub-seasonal scenarios: An alternative method to analyze seasonal predictability of regional-scale tropical rainfall. *J. Climate*, **26**, 2580–2600, doi:[10.1175/JCLI-D-12-00357.1](https://doi.org/10.1175/JCLI-D-12-00357.1).
- , A. W. Robertson, J.-H. Qian, and M. Ghil, 2015: Weather types across the Maritime Continent: From the diurnal cycle to interannual variations. *Front. Environ. Sci.*, **2**, 65, doi:[10.3389/fenvs.2014.00065](https://doi.org/10.3389/fenvs.2014.00065).
- Muñoz, Á. G., L. Goddard, A. W. Robertson, Y. Kushnir, and W. Baethgen, 2015: Cross–time scale interactions and rainfall extreme events in southeastern South America for the austral summer. Part I: Potential predictors. *J. Climate*, **28**, 7894–7913, doi:[10.1175/JCLI-D-14-00693.1](https://doi.org/10.1175/JCLI-D-14-00693.1).
- , —, S. J. Mason, and A. W. Robertson, 2016: Cross–time scale interactions and rainfall extreme events in southeastern South America for the austral summer. Part II: Predictive skill. *J. Climate*, **29**, 5915–5934, doi:[10.1175/JCLI-D-15-0699.1](https://doi.org/10.1175/JCLI-D-15-0699.1).
- Nakamura, J., U. Lall, Y. Kushnir, A. W. Robertson, and R. Seager, 2013: Dynamical structure of extreme floods in the U.S. Midwest and the United Kingdom. *J. Hydrometeorol.*, **14**, 485–504, doi:[10.1175/JHM-D-12-059.1](https://doi.org/10.1175/JHM-D-12-059.1).
- Notaro, M., W.-C. Wang, and W. Gong, 2006: Model and observational analysis of the northeast U.S. regional climate and its relationship to the PNA and NAO patterns during early winter. *Mon. Wea. Rev.*, **134**, 3479–3505, doi:[10.1175/MWR3234.1](https://doi.org/10.1175/MWR3234.1).
- Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. *J. Climate*, **12**, 575–591, doi:[10.1175/1520-0442\(1999\)012<0575:ANDPOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<0575:ANDPOC>2.0.CO;2).
- , and A. Weisheimer, 2011: Diagnosing the causes of bias in climate models—Why is it so hard? *Geophys. Astrophys. Fluid Dyn.*, **105**, 351–365, doi:[10.1080/03091929.2010.547194](https://doi.org/10.1080/03091929.2010.547194).
- Perez, J., M. Menendez, F. J. Mendez, and I. J. Losada, 2014: Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Climate Dyn.*, **43**, 2663–2680, doi:[10.1007/s00382-014-2078-8](https://doi.org/10.1007/s00382-014-2078-8).
- Phillips, A. S., C. Deser, and J. Fasullo, 2014: Evaluating modes of variability in climate models. *Eos, Trans. Amer. Geophys. Union*, **95**, 453–455, doi:[10.1002/2014EO490002](https://doi.org/10.1002/2014EO490002).
- Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, doi:[10.1016/j.jcp.2007.07.022](https://doi.org/10.1016/j.jcp.2007.07.022).
- Rayner, N. A., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:[10.1029/2002JD002670](https://doi.org/10.1029/2002JD002670).
- Reichler, T., and J. Kim, 2008: How well do coupled models simulate today’s climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311, doi:[10.1175/BAMS-89-3-303](https://doi.org/10.1175/BAMS-89-3-303).
- Reinhold, B. B., and R. T. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.*, **110**, 1105–1145, doi:[10.1175/1520-0493\(1982\)110<1105:DOWRQS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2).
- Riddle, E. E., M. B. Stoner, N. C. Johnson, M. L. L’Heureux, D. C. Collins, and S. B. Feldstein, 2013: The impact of the MJO on clusters of wintertime circulation anomalies over the North American region. *Climate Dyn.*, **40**, 1749–1766, doi:[10.1007/s00382-012-1493-y](https://doi.org/10.1007/s00382-012-1493-y).
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1).
- Robertson, A. W., and M. Ghil, 1999: Large-scale weather regimes and local climate over the western United States. *J. Climate*, **12**, 1796–1813, doi:[10.1175/1520-0442\(1999\)012<1796:LSWRAL>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1796:LSWRAL>2.0.CO;2).
- , S. Kirshner, P. Smyth, S. P. Charles, and B. C. Bates, 2006: Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland. *Quart. J. Roy. Meteor. Soc.*, **132**, 519–542, doi:[10.1256/qj.05.75](https://doi.org/10.1256/qj.05.75).
- , Y. Kushnir, U. Lall, and J. Nakamura, 2015: Weather and climatic drivers of extreme flooding events over the Midwest of the United States. *Extreme Events: Observations, Modeling, and Economics*, *Geophys. Monogr.*, Vol. 214, Amer. Geophys. Union, 113–124.

- Roller, C. D., J.-H. Qian, L. Agel, M. Barlow, and V. Moron, 2016: Winter weather regimes in the northeast United States. *J. Climate*, **29**, 2963–2980, doi:10.1175/JCLI-D-15-0274.1.
- Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, doi:10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.
- , and —, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626, doi:10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.
- Rosati, A., K. Miyakoda, and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.*, **125**, 754–772, doi:10.1175/1520-0493(1997)125<0754:TIOOIC>2.0.CO;2.
- Ryu, J. H., and K. Hayhoe, 2014: Understanding the sources of Caribbean precipitation biases in CMIP3 and CMIP5 simulations. *Climate Dyn.*, **42**, 3233–3252, doi:10.1007/s00382-013-1801-1.
- Stephenson, D. B., A. Hannachi, and A. O'Neill, 2004: On the existence of multiple climate regimes. *Quart. J. Roy. Meteor. Soc.*, **130**, 583–605, doi:10.1256/qj.02.146.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Trenberth, K. E., and J. M. Caron, 2000: The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation. *J. Climate*, **13**, 4358–4365, doi:10.1175/1520-0442(2000)013<4358:TSORSL>2.0.CO;2.
- van der Wiel, K., and Coauthors, 2016: The resolution dependence of contiguous U.S. precipitation extremes in response to CO₂ forcing. *J. Climate*, **29**, 7991–8012, doi:10.1175/JCLI-D-16-0307.1.
- Vaughan, C., and S. Dessai, 2014: Climate services for society: Origins, institutional arrangements, and design elements for an evaluation framework. *Wiley Interdiscip. Rev.: Climate Change*, **5**, 587–603, doi:10.1002/wcc.290.
- Vautard, R., 1990: Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Mon. Wea. Rev.*, **118**, 2056–2081, doi:10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2.
- , K. C. Mo, and M. Ghil, 1990: Statistical significance test for transition matrices of atmospheric Markov chains. *J. Atmos. Sci.*, **47**, 1926–1931, doi:10.1175/1520-0469(1990)047<1926:SSTFTM>2.0.CO;2.
- Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, doi:10.1175/JCLI-D-14-00158.1.
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.
- Zhou, S., M. L'Heureux, S. Weaver, and A. Kumar, 2012: A composite study of the MJO influence on the surface air temperature and precipitation over the continental United States. *Climate Dyn.*, **38**, 1459–1471, doi:10.1007/s00382-011-1001-9.