Empirical Modeling and Stochastic Simulation of Sea Level Pressure Variability

SERGEY KRAVTSOV
Atmospheric Sciences Group, Department of Mathematical Sciences, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin

NATALIA TILININA AND YULIA ZYULYAEVA
P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

SERGEY K. GULEV
P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, and Moscow State University, Moscow, Russia

(Manuscript received 9 July 2015, in final form 9 September 2015)

ABSTRACT

The scope of this work is stochastic emulation of sea level pressure (SLP) for use in error estimation and statistical prediction studies. The input SLP dataset whose statistics are to be emulated was taken from the 1979–2013 ERA-Interim dataset at full 6-hourly temporal and 0.75° spatial resolutions over the Northern Hemisphere. Upon subtracting the monthly climatological mean value and mean diurnal cycle, the SLP anomalies (SLPA) were projected onto the subspace of 1000 leading empirical orthogonal functions of the daily-mean SLPA, which account for the vast majority (>99%) of the full 6-hourly fields’ variance for each season. The main step of this method is the estimation of a linear autoregressive moving-average empirical model for the daily SLPA principal components (PCs) via regularized multiple linear regression; this model was driven, at the stage of simulation, by state-dependent (multiplicative) noise. Last, a diagnostic statistical scheme has been developed and implemented for accurate interpolation of simulated daily SLPA to 6-hourly temporal resolution. Upon transforming the simulated 6-hourly SLPA PCs into the physical space and adding a seasonal climatological mean and mean diurnal cycle, the resulting SLP variability was compared with the actual variability in the ERA-Interim dataset. It is shown that this empirical model produces independent realizations of SLP variability that are nearly indistinguishable from the observed variability over a wide range of statistical measures; these measures include, among others, spatial patterns of bandpass- and low-pass-filtered variability, as well as diverse characteristics of midlatitude cyclone tracks.

1. Introduction

Empirical data modeling in climate science has been previously used as an observational diagnostic (Penland and Sardeshmukh 1995; Kravtsov et al. 2005), model output analysis (Kondrashov et al. 2006, 2011), and statistical prediction tool (Penland 1989, 1996; Winkler et al. 2001; Kondrashov et al. 2005, 2013; Chekroun et al. 2011). The emphasis in all of these areas has traditionally been placed on identifying an optimal low-order model that would be able to faithfully mimic certain features of the observed low-frequency variability, such as weather regimes and low-frequency oscillations in the midlatitude or tropical climate subsystems [see Kravtsov et al. (2010) for a review].

There were reasons for such an emphasis. First of all, the large-scale low-frequency structures, presumably associated with low-order intrinsic dynamics of the climate system, are likely to be most predictable (Kleeman 2007, 2008). Furthermore, the robust estimation of a model that is purely data based has natural limitations related to the available (and short) length of the data record. It is intuitive that the number of independent observations in the data record being modeled has to be much larger than the number of model coefficients to be...
estimated. Yet, because the coefficients in a separate equation composing in part a multidimensional empirical model can in principle be estimated in isolation from the coefficients for other equations and the optimal number of effective predictors in each equation is a priori unknown, there still remains hope for at least hypothetical feasibility of constructing an empirical model that is able to capture the full spectrum of observed variability in select atmospheric fields. For example, the intermediately sized empirical model of Kravtsov et al. (2011) demonstrated capability of reproducing various aspects of high-dimensional air–sea interaction over the Southern Ocean.

In this respect, stochastic modeling of sea level pressure (SLP) variability across the entire range of spatial and time scales may have a number of important applications. SLP is a key thermodynamic characteristic of the atmosphere. Snapshots of SLP in midlatitudes are dominated by the synoptic eddies that largely control day-to-day weather variations (Gulev et al. 2002). On longer time scales, the SLP anomalies are organized into larger-scale structures, such as the North Atlantic Oscillation (NAO; Hurrell 1995; Benedict et al. 2004; Franzke et al. 2004; Vallis and Gerber 2008). Löptien and Ruprecht (2005), Kravtsov and Gulev (2013), and Kravtsov et al. (2015) argued for a strong kinematic connection between synoptic eddies and low-frequency SLP variability. Much of the SLP variability may thus be stochastic, albeit with signatures of large-scale, possibly nonlinear dynamics (Stephenson et al. 2000). Stochastic modeling of the type undertaken in the study that is presented here may help to better isolate and study the atmospheric weather regimes, their variability, and their predictability (Hertig and Jacobet 2014). Furthermore, stochastic emulation provides an avenue for modeling SLP extremes, such as out-of-the-ordinary deepening rates and pressure falls associated with intense storms (Tan et al. 2008; Nakajo et al. 2014). Combining empirical modeling with dynamical simulations of global and regional climate via various downscaling strategies extends the utility of these models to addressing extreme climate events (Maraun et al. 2015). Also, modeling long-term climatic responses rooted in ocean dynamics requires accurate representation of the stochastic forcing functions used to drive the ocean models (Aiken and England 2005; Vanem 2011; Legatt et al. 2012; Oddo et al. 2014; Turki et al. 2015); these functions are strongly dependent on SLP.

With these applications in mind, we developed in this paper an empirical stochastic model of Northern Hemisphere SLP and used it to produce surrogate SLP time series that mimic the statistics of their observed counterpart. Our main goals are 1) to document a method for constructing a prototype high-dimensional model that would mimic the detailed structure of the observed SLP variability and 2) to conduct an evaluation of this model’s performance using both Eulerian and Lagrangian frameworks for characterizing SLP variability. Note that the said model is not expected to produce climate realizations that are pathwise similar to the observed climate; on the contrary, the climate simulated by such a model would be, by construction, statistically independent of the actual observed climate realization. Hence, we should judge the success of the model’s performance by comparing not the pathwise convergence but rather the long-term statistical properties of the observed and simulated SLP variability, such as the spatiotemporal SLP spectra or the composite characteristics of the individual cyclones within storm tracks.

Further presentation is built around the three main tasks of the paper—empirical SLP model construction (top row in Fig. 1), simulation (bottom row in Fig. 1), and validation (statistical comparisons of the input and output SLP fields; left column of Fig. 1)—and is organized as follows: In section 2, we introduce the SLP dataset to be analyzed and modeled (input and step 1 in Fig. 1), as well as the data-compression (appendix A; step 2 in Fig. 1) and interpolation (appendix B; step 3 in Fig. 1) procedures. Section 3 (steps 3–5 in Fig. 1) concerns the construction of our empirical model and the details of numerical simulations (appendixes C and D; steps 4 and 5 in Fig. 1). In section 4, we present an account of comparisons between the observed and simulated SLP variability in the phase space, as well as in the physical space (step 6 in Fig. 1) over a wide range of statistical measures (left column of Fig. 1). These results are summarized and further discussed in section 5, which also outlines a suite of major applications for the proposed empirical climate emulators.

2. Dataset and preprocessing

We utilized the SLP dataset from the 1979–2013 European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim; Dee et al. 2011), with 6-hourly output restricted to the Northern Hemisphere at 0.75° × 0.75° spatial resolution (input box in Fig. 1). This dataset has a total length of \( N = 12,784 \) days and contains \( 12,784 \times 4 = 51,136 \) six-hourly SLP snapshots; each snapshot is represented by \( M = 480 \times 121 = 58,080 \) spatial points.

a. Data compression and manipulation

We first formed SLP anomalies (SLPA) by subtracting the long-term monthly SLP “climatology” (the
1979–2013 SLP mean for each month) from the full SLP data (step 1 in Fig. 1). Next, we computed the empirical orthogonal function (EOF; Monahan et al. 2009) decomposition of the daily-mean SLPA time series (step 2 in Fig. 1). EOF-based data compression has been routinely used in empirical weather/climate modeling (Kravtsov et al. 2005, among others), as well as in reduced-space optimal interpolation of historical datasets (Kaplan et al. 1997, 2000). We will also apply it here in the context of emulating the full spectrum of the synoptic SLP variability, from hourly to weekly and longer time scales.

The columns (time series) composing the original $N \times M$ daily-mean SLPA data matrix $P$ were traditionally weighted by the square root of the cosine of their corresponding latitudes prior to computing $L = 1000$ leading EOF patterns $V$ (an $M \times L$ matrix) and the same number of corresponding principal component (PC) time series $x$ (an $N \times L$ matrix) associated with these patterns. We will see below that these EOFs account for the vast majority of the total SLPA variance, which was used as our sole criterion for EOF truncation.

We normalized all daily PCs so obtained by the standard deviation of the leading PC (PC-1). Hereinafter, we will use the symbol $x$ to describe the matrix of these normalized PCs; the procedures described in section 3 will seek to find the best empirical model for producing independent surrogate realizations of $x$. We refer the reader to appendix A for the conversion formulas between the fields in the EOF phase space and physical space.

The daily-mean SLPA field reconstructed from the phase space of $L = 1000$ EOFs using Eq. (A1) accounts for over 99.9% of the total variance contained in the full SLPA field for each of its seasonal (December–February, March–May, June–August, and September–November) subsets. For illustration, Fig. 2 shows 12 leading nonweighted dimensional EOF patterns computed using Eq. (A3); the panel captions list the fraction of the total variance accounted for by each of these EOFs. The EOF spectrum is relatively flat, with the leading NAO/Arctic Oscillation (AO)-type EOF accounting for only 9% of the total variance. In general, the EOFs turn out to be ranked in consistency with their spatial and time scales: the larger-scale lower-frequency EOF/PC modes tend to lead the spectrum while the higher-ranked EOF/PC modes exhibit features with progressively smaller spatial and time scales. This arrangement for spatial scales is already apparent from Fig. 2; we will present a more explicit demonstration of the correspondence between EOF rank and time scale later in section 4 when discussing the performance of our empirical SLPA models.

b. Interpolation to 6-hourly time grid

Consider now 6-hourly deviations from the daily mean SLPA, referred to here as SLPA (step 1 in Fig. 1). Similar to simply removing and setting aside the mean monthly seasonal cycle, we computed and removed the mean diurnal cycle from the monthly (all Januarys, all Februarys, etc.) subsets of the full
SLPAA data. The standard deviations associated with the mean diurnal cycle (left column of Fig. 3) are larger over mountain regions but are generally small elsewhere. We will not attempt to model these variations, but we still have an option to add the fixed observed diurnal cycle to our simulated SLP surrogates, in the same way we will be adding the mean seasonal cycle (section 3).
The variability of the SLPAA anomalies is spatially distributed in a way that is similar to that of the classical storm-track pattern (right column of Fig. 3) and is thus likely related to the propagating synoptic features apparent in the daily-mean fields (not shown here). Indeed, projecting the 6-hourly SLPAA anomalies into the phase space of the daily SLPA EOFs \( \mathbf{V} \) (step 2 in Fig. 1) using Eq. (A4) with the input SLPAA time series \( \mathbf{P} \) on the right-hand side (and the 6-hourly left-hand side denoted by \( \mathbf{x}' \)) and then projecting back into physical space using Eq. (A1) (modified, once again, to include primed variables) produces an SLPAA reconstruction that captures nearly 100% of the total SLPAA variance; the reconstruction residual has standard deviations with maximum values in the mountain regions of about 0.6 hPa (not shown) and much smaller values elsewhere.

Hence, it appears reasonable to diagnostically model the 6-hourly SLPAA using the daily-mean SLPA time series. This diagnostic modeling (step 3 in Fig. 1) was set up in the phase space of the daily SLPA EOFs and employed multiple linear regression to connect the extended predictor vector \( \mathbf{x} \) and response variables represented by \( \mathbf{x}' \) (see appendix B for details). The results are summarized in Fig. 4. The skill of regression turns

![Fig. 3. Standard deviation (hPa) of SLP variability associated (left) with the mean diurnal cycle and (right) with the residual 6-hourly SLP deviations from the daily-mean values for (top) December and (bottom) June.](image-url)
out to depend on the EOF rank. The regression captures nearly 100% of the total 6-hourly SLPAA variance for leading EOFs (note near-zero values of the unexplained variance for these EOFs in the top row of Fig. 4); for higher-ranked EOFs, the regression skill monotonically decreases and saturates at about 50%. Recall that these fractional accuracies refer to the deviations (SLPAA) from the daily-mean SLPA, which are by themselves small relative to the latter daily-mean anomalies. The actual values of the standard deviation for the regression residual $R$ (Fig. 4, bottom-left panel) are also small, as compared with the standard deviation of the daily PCs (recall also that these PCs were normalized by the standard deviation of the leading PC, and therefore the typical amplitude scale of the dominant variability is set to unity). In the physical space (Fig. 4, bottom-right panel), the all-year standard deviations of the 6-hourly regression residual peak at 1.2 hPa so that the residual variance is, once again, a smaller than the raw SLPAA variance by a factor of 4–5 (see right column of Fig. 3).

In summary, the regression modeling of the 6-hourly SLPAA anomalies using daily SLPA as predictors is fairly successful and captures most of the actual 6-hourly SLPAA variance. The 6-hourly residual variability, although small, is not negligible (Fig. 4) and will be modeled separately from the daily-mean variability (see section 3).

3. Empirical model construction and simulation

The main computational tasks addressed in this section are 1) estimation of the propagator for the empirical model designed to emulate the statistics of daily SLPA (step 4 in Fig. 1), 2) use of this model to produce surrogate daily SLPA (step 4 in Fig. 1), and 3) empirical interpolation of the simulated daily anomalies from task 2 onto a 6-hourly time grid (step 5 in Fig. 1). All of these operations are performed in the phase space of the daily SLPA's EOFs (see section 2). Task 1 employs the training $N \times L$ data matrix of daily SLPA PCs $x$ (section 3a).
The model simulation in task 2 involves stochastic forcing, whose generation can be approached in several ways, as discussed in detail in section 3b and appendix C. Upon producing the empirical model-based surrogate realization of the daily PCs $\mathbf{x}$ (step 4 in Fig. 1), its corresponding surrogate 6-hourly deviations $\mathbf{x}'$ from the daily means are obtained as per task 3 (step 5 in Fig. 1) using the diagnostic relationships described in section 2b and given in Eqs. (B1)–(B3) (see appendix B); furthermore, the residual 6-hourly variability unexplained by Eqs. (B1)–(B3) is also modeled stochastically and added to form the total surrogate 6-hourly anomalies $\mathbf{x}'$ (section 2b and appendix D). Last, the simulated daily-mean PCs and their simulated 6-hourly anomalies are combined and these total surrogate 6-hourly PCs are projected into the physical space using Eq. (A1) (step 6 in Fig. 1); adding to these surrogate SLPA fields the observed mean monthly seasonal cycle and the observed mean diurnal cycle completes the emulation procedure (step 7 in Fig. 1).

a. Empirical model for daily SLP

We followed the general empirical model reduction (EMR) method of Kravtsov et al. (2005, 2010, 2011) to build our empirical model for the daily SLPA. The EMR model has a three-level structure of the following form:

$$
\begin{align*}
\mathbf{d}x &= \mathbf{xA}^{(1)} + \mathbf{r}^{(1)}, \\
\mathbf{d}r^{(1)} &= [\mathbf{r}^{(1)} \mathbf{x}]\mathbf{A}^{(2)} + \mathbf{r}^{(2)}, \quad \text{and} \\
\mathbf{d}r^{(2)} &= [\mathbf{r}^{(2)} \mathbf{r}^{(1)}} \mathbf{x}]\mathbf{A}^{(3)} + \mathbf{r}^{(3)},
\end{align*}
$$

(1)

where the differentials on the left-hand side denote the daily increments of the corresponding variables. All model levels in Eq. (1) are linear. The first level in isolation, with the residual $\mathbf{r}^{(1)}$ represented at the simulation stage by the spatially correlated white noise, would make up a classical linear inverse model (Penland 1989, 1996; Penland and Sardeshmukh 1995; Winkler et al. 2001). Instead, however, the daily increments of the first-level residual $\mathbf{d}r^{(1)}$ are in turn modeled as a linear function of the extended predictor vector $[\mathbf{r}^{(1)} \mathbf{x}]$ to form the second level of our multilevel regression model. In the same way, the third level connects the daily increments of the second-level residual $\mathbf{d}r^{(2)}$ and the extended predictor vector $[\mathbf{r}^{(2)} \mathbf{r}^{(1)} \mathbf{x}]$ involving the variables from the previous two model levels.

The matrices of the model coefficients $\mathbf{A}$ and the level residuals were found by regularized multiple linear regression (MLR). Whereas the residuals of the first and second level involve some serial correlations, the third level’s residual $\mathbf{r}^{(3)}$ is essentially white in time, with zero lag-1 autocorrelation (see Fig. 5), thus justifying cutting off the number of model levels at three; stochastic modeling of this residual at the model’s simulation stage will be addressed in section 3b (see also appendix C).

Note that, although the model construction procedure is sequential from the first level down to the third level, the system given in Eq. (1), when rewritten as one equation containing the time-lagged variables, is formally equivalent to the autoregressive moving-average model (Box et al. 1994).

In performing the MLR, each column vector of regression coefficients in matrices $\mathbf{A}$ on the right-hand side of Eq. (1) was estimated separately. For example, at the main level, the time series of the $i$th PC’s increments $\mathbf{dx}_i$ was used as a response variable regressed onto the matrix of predictor variables $\mathbf{x}$ to find the $i$th column of $\mathbf{A}^{(1)}$. Furthermore, at all model levels, we employed the same grouping of the effective regression variables as in estimating the linear models of 6-hourly SLPA residuals in section 2b (see appendix B). Namely, the $i$th equation at all levels used only a subset $\mathbf{S}_i$ of the right-hand side predictors, as given by Eq. (B2). This restriction is technically necessary to achieve robust estimation of regression coefficients by reducing the number of regression variables and is formally justified by the fact that the neighboring EOFs share the time and spatial scales and are thus most likely to be dynamically related.

On top of this manual model selection, we used the partial least squares regularization (see Kravtsov et al. 2005, 2010, 2011) to optimize the model estimation procedure; the number of latent variables for each level was set to 40 by trial and error. Our general logic for choosing the value of $W$ in Eq. (B2) and the number of latent variables in the partial least squares regression was to...
intentionally choose a larger-than-optimal number of variables [as estimated, e.g., by cross-validation procedures described in Kravtsov et al. (2011)] but still small enough to achieve a ratio of at least 20 between the number of independent observations and the number of regression coefficients to be estimated.

To capture the seasonal changes in SLPA variance and other statistical characteristics, we estimated 12 separate models of the type in Eq. (1) for the 12 monthly subsets of our SLPA dataset; in other words, the propagators $A$ on the right-hand side of Eq. (1) were seasonally dependent.

b. Simulation procedures

Once the propagator of the model depicted in Eq. (1) has been estimated, one needs to assume a parameterization for the residual forcing $r^{(3)}$ to run the model from an arbitrary initial condition and to achieve our end goal of producing a synthetic climate—surrogate independent realizations of the SLPA PCs. The generation of the stochastic forcing for the prognostic daily model in Eq. (1) is discussed in appendix C, and the application of the diagnostic model in Eq. (B3) for the interpolation of simulated daily PCs to 6-hourly resolution is covered in appendix D.

The simulated daily-mean SLPA PCs and the simulated 6-hourly deviations from the daily mean PCs were added together to form a surrogate PC time series at 6-hourly resolution, which was then transformed back into the physical space and compared, upon adding the mean seasonal and diurnal cycles (output box in Fig. 1), with the observed SLP variability (input box in Fig. 1).

The model that is described by Eq. (1) is a very numerically efficient emulator of the observed SLP variability: one 1979–2013 simulation using this model takes about 5 min of wall-clock time on the standard 2.7-GHz Intel Core i7 Apple MacBook Pro laptop computer. Let us reemphasize here that our model is not meant to reproduce the observed SLP trajectory but rather is designed to emulate the statistics of the observed SLP variations as closely as possible.

4. Comparisons between actual and surrogate climates

We first generated 100 surrogate SLP realizations and compared the observed and simulated SLP statistics, first in the phase space (section 4a) and then in the physical space, in the latter case using one of the 100 available realizations. In particular, we compared the distributions of the raw and bandpass-filtered SLP variance (section 4b) and analyzed the patterns and variability of cyclone tracks in the observed and simulated SLP data (section 4c).

a. Comparisons in the phase space

The empirical model in Eq. (1) produces surrogate realizations of the observed SLPA PCs with the EOF spectrum of variance, as well as the probability density function and autocorrelation function (ACF) of individual PCs, that are statistically indistinguishable from their observed counterparts (not shown). As an illustration, we compare here the observed and simulated integral correlation time scales (ICTs) of the observed and simulated EOFs (Fig. 6); these ICTs were defined, for each PC, as the sum of the absolute values of this PC’s ACF over the range of lags from 0 to 50 days. Note also that Fig. 6 provides a confirmation of our previously postulated property of the EOF decomposition ranked by the variance to also tend to arrange the EOFs roughly in order of decreasing time scales. The entire observed ICT spectrum is captured very well by the model-generated data, with perhaps a very slight enhanced persistence bias in the simulated time scales of the leading 100 EOF modes.

b. Comparisons in physical space: Spectral analysis and anomaly propagation

1) SPECTRAL ANALYSIS AND ANOMALY PROPAGATION

We then transformed the simulated phase-space anomalies to the SLPA physical space (see appendix A) and looked first at the seasonal distributions of the standard deviations of the data filtered within different frequency bands. This diagnostic corresponds to a widely adopted Eulerian framework for studying storm tracks [see, e.g., Hoskins et al. (1983), Blackmon et al. (1984), Hoskins and Sardeshmukh (1987), Wallace et al. (1988), Christoph et al. (1997), Gulev et al. (2002), and Hoskins and Hodges (2002), among others]. Following the method of Gulev et al. (2002), we analyzed statistics of bandpass-filtered data for the four ranges corresponding to atmospheric ultrahigh-frequency variability (periods of 0–2 days), synoptic-scale variability (2–6 days), slow synoptic-process variability (6–12 days), and low-frequency variability (>12 days); similar subdivisions were used by Ayrault et al. (1995), Rogers (1997), and others. As in Hoskins and Sardeshmukh (1987), Gulev et al. (2002), and many other studies of synoptic variability, we adopted a Lanczos filter (Duchon 1979) for SLP bandpass filtering. This filter is characterized by a very sharp cutoff at the ends of the selected frequency range, which is important for distinguishing the variability associated with neighboring frequency ranges. We reduced Gibbs oscillations produced by the Lanczos filtering by the smoothing applied in the frequency domain.
FIG. 6. ICT of the observed and simulated PCs (see text for details) for (top) all PCs and (bottom) the leading 100 PCs. The observed ACFs are shown by the blue lines, and the 95% spreads on the basis of 100 surrogate SLPA realizations are shown by the red dashed lines.
Figures 7 and 8 show standard deviations of the wintertime [January–March (JFM)] and summer-season [July–September (JAS)] bandpass-filtered SLP anomalies. For each season, the model captures well both the locations and magnitude of the observed SLP variability in each of the four frequency ranges considered. The differences in the SLP standard deviation between the simulation and observations are really minor, rarely exceeding 10% of the observed standard deviations; these differences are not statistically significant according to the Student's t test.

To complement the diagnosis of the bandpass-filtered variance above, we also computed the SLPA spectra for two regions that are representative of conditions over the midlatitude Atlantic and Pacific Oceans, respectively (Fig. 9). The spectra of the observed and simulated data are, once again, extremely close and are statistically indistinguishable from each other throughout the entire frequency range (the uncertainty associated with the average spectrum is approximately given by the spread of individual annual spectra shown in Fig. 10 divided by the square root of the number of the independent spectral estimates: $35^{1/2} \approx 6$).

Another essential aspect of the observed SLP variability that one would want the empirical model simulation to reproduce concerns the propagation of synoptic SLPA along the storm tracks. One way to diagnose such propagation is to construct one-point lagged correlation maps (see Fig. 10). In this particular case—for the year and the location chosen to be in the middle of the Pacific Ocean—there turned out to be a near-perfect correspondence in the pattern and speed of the observed and simulated SLPA propagation in both the 0–2- and 2–6-day spectral bands (recall, once again, that we do not expect to match the exact history of the simulated "1980" to that of the observed 1980 we only attempt to reproduce the overall statistics of the SLPA variability). The patterns and speeds of anomaly propagation do exhibit sampling variability in both the observations and model simulations, however (not shown). Hence, to meaningfully compare the observed and simulated anomaly propagation speeds, we derived the 35 individual-year estimates of this quantity on the basis of the slopes of propagation lines such as those shown in Fig. 10 and then estimated statistical significance of the difference between the observed and simulated ensemble-averaged propagation speeds using the t test. This analysis was also applied to diagnose the anomaly propagation in the Atlantic. The results are summarized in Table 1 and demonstrate that the observed and simulated propagation speeds are statistically indistinguishable except for the 0–2-day bandpass-filtered SLPA over the Atlantic Ocean, which tend to propagate faster in model simulations. The reasons behind this discrepancy are not clear but may have something to do with phase locking of the observed SLPA with the diurnal cycle; we did not attempt to represent the dynamics of such phase locking in the current empirical model.

As an alternative, propagation characteristics of SLP synoptic patterns in the observed and simulated data can be diagnosed by computing the leading EOF pair of the bandpass-filtered regional SLPA and the associated lagged correlation function between the corresponding PCs. Figure 11 shows an example that is based on the analysis of the 2–6-day bandpass-filtered anomalies in the Atlantic region. Note that the absolute phase of the leading EOF pair describing a propagating wave is arbitrary, and therefore the phase shift between the observed and simulated EOF pair in Fig. 11 does not represent a mismatch. More to the point, the observed wave patterns and relative phasing of EOF-1 and EOF-2 as well as the PC-1/PC-2 cross-correlation function are all captured very well in the structure of the simulated SLPA, implying practically the same phase velocities of propagation in the original and simulated data (see Table 1). This result holds for all individual years.

In summary, Eulerian analysis of storm tracks demonstrated a striking similarity between the observed and simulated spatiotemporal patterns of SLPA variability across the whole spectral range, including an excellent match of storm-track locations and intensities as well as characteristics of SLPA propagation along the storm tracks.

2) CYCLONE TRACKING

To further study the details of reproducing the observed spatiotemporal SLP characteristics in the model simulations, we performed a cyclone-tracking analysis of both datasets using the algorithm described in Rudeva and Gulev (2007) and postprocessing procedures by Tilinina et al. (2013). The feature tracking is a widely used Lagrangian diagnostic technique that is largely independent from the Eulerian approaches to describing weather and climate variations (Zolina and Gulev 2002; Rudeva and Gulev 2007, 2011; Raible et al. 2008; Neu et al. 2013). Yet, Kravtsov and Gulev (2013) and Kravtsov et al. (2015) recently showed that tracking and recording even the basic life-cycle characteristics of cyclones and anticyclones—such as the path, effective radius, and intensity—are sufficient for describing a major fraction of the total and the dominant fraction of the low-frequency variability in the SLP field. The ability of our model to capture the observed geographical distribution and life-cycle properties of the cyclone tracks would thus be a very powerful indication...
FIG. 7. Standard deviations (hPa) of the wintertime (JFM) (left) observed and (center) simulated bandpass-filtered SLPA in physical space and (right) the relative difference (observed − simulated; %) for the (top) 2-day high-pass-filtered, (top middle) 2–6-day bandpass-filtered, (bottom middle) 6–12-day bandpass-filtered, and (bottom) 12-day low-pass-filtered anomalies.
FIG. 8. As in Fig. 7, but for the summer season (JAS).
of the model’s skill in reproducing the spatiotemporal structure of the observed SLP variability.

Following the method of Neu et al. (2013), among others, we only analyzed in this section the tracks for the cyclones that traveled a distance of 1000 km or more during their lifetime. This filtering removes from consideration artificial cyclone noise that result from potential uncertainties in both the tracking input data and the tracking algorithm. The patterns of cyclone-track densities from the output of our empirical model (Figs. 12a,d) match the observed densities very well, for both the winter and summer seasons (observations not shown). There is a substantial (~15%–20%) negative bias in the number of cyclones the model generates relative to the full-resolution ERA-Interim cyclone-tracking results (Figs. 12b,e). This apparent bias is dominated by the truncation of the simulated 6-hourly SLPAA anomalies to the 1000 EOFs of the daily SLP variability (section 2). Indeed, truncating the original ERA-Interim SLP in the same way and tracking the cyclones in this truncated observed dataset produces a much better match with the cyclone densities derived from the simulated SLP field (Figs. 12c,f), essentially removing a major part of the original bias in cyclone numbers, especially over the continents. In technical terms, this bias arises because the differences between the original and truncated SLP fields, although small in terms of the total variance for which they account, are locally large enough to split the trajectories in the original, full-resolution SLP data, thereby increasing the total number of tracks counted there. This situation is most typical for the cyclone tracks that undergo splitting over mountain ranges, whereas the EOF truncation effectively “sews” them back together in longer continuous tracks. In the following presentation, we will be comparing the tracking results for the simulated and truncated ERA-Interim fields.
Note that there are still some lingering 5%–10% mismatches detected in the simulated cyclone numbers even when compared with the cyclone tracking of the truncated ERA-Interim SLP (Fig. 13). On top of this general bias, the simulated cyclone numbers exhibit interannual variability and longer-term trends that are both consistent, in magnitude, with the observed variability. Since there is no external forcing in this particular

**Fig. 10.** SLPA propagation of the (a) 2-day high-pass-filtered anomalies and (b) 2–6-day bandpass-filtered anomalies in observations and stochastic simulations. Shown are one-point lagged correlation maps for the observed (red contours) and simulated (black contours) JFM data for 1980 relative to the point 40°N, 170°E in the Pacific region. The lag values are given in each panel label. The slope of the line tracking the propagating anomaly is related to the propagation speed (see Table 1).
rendition of the empirical model, the simulated long-term variations of the number of simulated cyclones are generated by the model internally and may be sufficient, by themselves, to rationalize the observed variations. Uncertainty estimation and attribution of the observed climate variability are two important applications of the climate emulators (see section 5 for further discussion).

Having established that our empirical model generates reasonable distribution of cyclone tracks, we turn now to discussing the statistics of the observed and simulated cyclone life cycles. Figure 14 uses the so-called Q–Q plots (Wilk and Gnanadesikan 1968) to document a striking similarity between the observed and simulated probability distributions of the fundamental cyclone life-cycle characteristics: in particular, minimum pressure, lifetime, maximum deepening rate, and maximum propagation speed. We colored red the points on these diagrams that fall within the 95% spread of the synthetic diagrams obtained via bootstrap resampling of the simulated cyclone data; the points outside this confidence range are shown in blue. Fairly minor differences between the observed and simulated life cycles are more pronounced in winter (Figs. 14a–d) and indicate somewhat shallower, shorter-lived, and more slowly deepening cyclones. These biases are dominated by the cyclones that achieve their minimum pressure over the ocean (not shown, but see below).

To investigate further the biases in the simulated cyclone life-cycle characteristics, we computed the two-dimensional cyclone-speed/deepening-rate histograms conditioned on the track being located either over ocean or over land (Fig. 15). Following the method of Tilinina et al. (2013), cyclone trajectories were attributed to the ocean or to the land regions on the basis of whether the

<table>
<thead>
<tr>
<th></th>
<th>Simulated τ and (c_{min}–c_{max}) (km h ^{-1})</th>
<th>ERA τ and (c_{min}–c_{max}) (km h ^{-1})</th>
<th>Δc (simulated – ERA)</th>
<th>t statistic (t^* = 1.99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-day high-pass filtered, ATL</td>
<td>73.1 (61.7–83.6)</td>
<td>67.4 (54.2–78.8)</td>
<td>5.7</td>
<td>2.97</td>
</tr>
<tr>
<td>2–6-day bandpass filtered, ATL</td>
<td>48.5 (39.5–58.0)</td>
<td>48.8 (36.9–62.2)</td>
<td>–0.3</td>
<td>0.29</td>
</tr>
<tr>
<td>2-day high-pass filtered, PAC</td>
<td>69.1 (60.7–82.0)</td>
<td>67.6 (57.2–80.5)</td>
<td>1.6</td>
<td>1.3</td>
</tr>
<tr>
<td>2–6-day bandpass filtered, PAC</td>
<td>50.1 (34.6–62.0)</td>
<td>50.8 (43.7–63.2)</td>
<td>–0.7</td>
<td>0.53</td>
</tr>
</tbody>
</table>

FIG. 11. SLPA propagation diagnosed by EOF analysis: the leading EOF pair of (a) simulated and (b) observed 2–6-day bandpass-filtered JFM SLPA over the Atlantic region for 1982. (c) Lagged correlations between the PCs corresponding to the leading EOF pair for 1982 (red line is observed; black line is simulated); both the observed and simulated anomalies exhibit eastward propagation (see also Fig. 9). (d) The range of the PC1–PC2 lagged correlations computed for individual years (red lines show observed; black lines show simulated).
cyclone achieved its minimum central pressure over these respective regions. From these diagrams, it does not appear that the biases in the simulated propagation speeds and deepening rates are easily attributable to the tracks passing either over ocean or over land. In fact, the differences between the observation-based and simulation-based histograms are subtle. All in all, the empirical model simulation shows, once again, a great skill in capturing the detailed spatiotemporal characteristics of cyclones and their life cycles over the Northern Hemisphere.

5. Summary and discussion

a. Summary

We constructed an empirical model that is based on the ERA-Interim SLP dataset (Dee et al. 2011). This model produces independent (uncorrelated with one another and with the observed data) realizations of the SLP evolution—synthetic climates—that replicate various statistical properties of the observed sea level pressure variability very well. The model construction method was based on the work of Kravtsov et al. (2005, 2010) but was adapted to process a much higher-dimensional input data and included a somewhat different treatment of the seasonal cycle and state-dependent forcing used to drive the model. The model dimension was dictated by the number of spatial degrees of freedom in the observed SLP dataset, which we showed to be less than $L = 1000$, as based on data compression through the standard empirical orthogonal function analysis (Monahan et al. 2009). We also showed that much of the 6-hourly SLP variability can be diagnostically related to the daily-mean SLP fields, thereby providing the basis for statistical interpolation.
of the daily-mean SLPs simulated by our empirical model to 6-hourly time resolution; the residual 6-hourly variability was modeled separately using an empirical closure scheme. The resulting empirical model captures well the basic statistical characteristics of the observed data—means, variances, and autocorrelation functions—in the phase space, the patterns of the full and bandpass-filtered variances in the physical space, as well as the pathways and life cycles of the midlatitude cyclones. This indicates a great success of this model in simulating the whole spatiotemporal structure of the observed SLP variability.

b. Discussion

Earth’s climate is extremely complex and involves dynamical interactions across a wide range of spatial and time scales. Given a relatively short duration of the observed instrumental records, climatic-data modeling thus far has concentrated on reproducing relatively low-dimensional subsets of the observed climate evolution. A daring—and to a large extent successful—attempt of this study was to try to describe, statistically, the observed variability in the SLP, which is one of the most important climatic fields. Our ability to construct such a skillful empirical SLP model demonstrates that the number of effective (spatial) degrees of freedom in the observed climate variability is not too large and that the observed data record is not too short for comprehensive statistical data modeling to be feasible. An excellent statistical performance of this model combined with its predictive potential may come handy in historical SLP reconstruction applications (Kaplan et al. 2000; Ansell et al. 2006).

The model constructed in this paper is also a proof-of-concept, prototype model and, as such, provides a method for simulating other climatic fields. For example, we have already done some work on applying this method to modeling observed daily vector winds in the Northern Hemisphere and have demonstrated comparable skill of such a model in reproducing the observed
vector-wind statistics. These results will be reported on elsewhere. We also plan to apply our current method to sea surface temperature (SST) and, later on, combined SST–SLP modeling on monthly and longer time scales. Another possibility for extending and utilizing these results is in adding to the model framework the dependence of the model operator on external variables that are associated, for example, with interannual-to-multidecadal intrinsic SST variability. These variables are external because the empirical model would not simulate them but rather would use them as fixed time-dependent inputs during the simulations. Such a procedure will potentially allow one to address SST impact on SLP variability and predictability on different time scales (Latif et al. 2000).

The latter proposed extension to the SST-dependent empirical models of atmospheric circulation provides an example of these models’ application in climate prediction and attribution studies. Indeed, given future projections of possible SST evolution derived from comprehensive dynamical climate models and assuming approximate stationarity of the SST effect on the circulation dynamics in the near future, one can use the SST-dependent empirical models to address long-term variability in future weather statistics and circulation regimes (e.g., expected changes in the frequency of blocking); alternatively, such empirical models can serve as an “atmospheric” component of a hybrid coupled ocean–atmosphere model. This line of research can potentially be extended further by using surface heat-flux data over the ocean in conjunction with or instead of SST, thus explicitly accounting for the effects of ocean diabatic forcing on low-level atmospheric baroclinicity and, hence, on cyclone generation and development (Branscome et al. 1989, among others).

Another area of the empirical models’ application on which we are currently working is their coupling with comprehensive dynamical models for regional downscaling and risk assessment (see, e.g., Maraun et al. 2015). The empirical models are constructed to match well the observed atmospheric climatological behavior and statistics in a small number of select important fields. Dynamical models, on the other hand, are more detailed and complete but may also exhibit more-pronounced biases in the isolated atmospheric characteristics that are well simulated by empirical models. Such biases would necessarily translate into those in regional dynamical downscaling using high-resolution models forced on lateral boundaries by the output of a coarser-resolution global dynamical model. A possible

Fig. 14. The Q–Q plots comparing the probability distributions of the observed (vertical axis) and simulated (horizontal axis) cyclone life-cycle characteristics by plotting their quantiles against each other; the quantile interval used is 1%. The cyclone subsets included in the computation were the same as in Fig. 13. The red dots show locations in the diagram plot at which the simulated distributions are statistically indistinguishable from the observed distributions at the 5% significance level according to the bootstrap test (see text). Shown are (a),(e) minimum pressure (hPa), (b),(f) lifetime (h), (c),(g) maximum deepening rate [hPa (6 h)$^{-1}$], and (d),(h) maximum propagation speed (km h$^{-1}$) for the (top) winter and (bottom) summer seasons.
strategy for correcting these biases is to nudge a global dynamical model using the output of the empirical model and then to proceed with the regional downscaling in the usual way.

We used ERA-Interim to get our proof-of-concept results, but it is a legitimate question as to whether our proposed emulation technique would perform equally well with other available reanalysis products. In fact, one of the possible further applications of our method can be exactly for evaluation and intercomparison of these reanalyses. Hodges et al. (2011) and Tilinina et al. (2013) reported high consistency in characteristics of cyclone activity across different modern-era reanalyses, including ERA-Interim. Statistical reconstruction skills may depend, however, on the number of observations that go into the assimilation procedure in different reanalyses (presumably because of dynamical model biases that would be emphasized in the products that assimilate fewer data). In a similar way, the products with higher spatial resolution may be more difficult to emulate, especially on mesoscales. Applying our emulation technique to a suite of available reanalysis products (as well as model-generated climates, e.g., within the CMIP5 ensemble) could add another, potentially useful assessment line for their evaluation.

The excellent statistical performance and extreme numerical efficiency of the empirical models will come in handy in applications that would require an extension of the observed time series to a much longer period in search of rare or extreme events. The most

FIG. 15. Two-dimensional cyclone-speed/deepening-rate histograms for the observed and simulated data. The cyclone subsets included in the computation were the same as in Figs. 13 and 14. The histograms were computed separately for the cyclone tracks over ocean and over land. See panel labels and legends for details.
APPENDIX A

Manipulation of EOFs and PCs

The rule for projecting simulated SLPA variability from the EOF space back to physical space is as follows:

\[
P = xV^T. \tag{A1}
\]

Here (and hereinafter) \(V\) defines the nonweighted (and also dimensional) EOF patterns. Using Eq. (A1), one can write

\[
P^T x/(N-1) = V (x^T x)/(N-1) = V \Sigma, \tag{A2}
\]

where \(\Sigma = (x^T x)/(N-1)\) is the diagonal covariance matrix computed in the phase space of the weighted SLPA EOFs. Rearranging Eq. (A2), we get the formula for computing the nonweighted EOFs \(\hat{V}\) from the original SLPA data matrix \(P\) and its PCs \(x\) (which—as we recall—were based on the weighted covariance matrix):

\[
\hat{V} = [P^T x/(N-1)] \Sigma^{-1}. \tag{A3}
\]

Last, given the nonweighted EOF patterns \(\hat{V}\) and the original SLPA \(P\), the original PCs \(x\) can be recovered by

\[
x = P \times \text{pinv}(V^T), \tag{A4}
\]

where \(\text{pinv}(A)\) is the pseudoinverse of a nonsquare matrix \(A\).

APPENDIX B

Interpolation to 6-Hourly Resolution

Let us denote the response-variable subsets of 0000, 0600, 1200, and 1800 UTC 6-hourly SLPA PCs as \(x_i(t, 0), x_i(t, 6), x_i(t, 12),\) and \(x_i(t, 18)\), respectively; here \(i = 1, L\) is the PC-component index and \(t = 1, N\) is the time index. Let us also define the extended predictor vector for the \(i\)th component of a given response variable (viz., one of the four variables introduced in the previous sentence) as

\[
X_{(i)}(t) = [x_i(t-2) x_i(t-1) x_i(t) x_i(t+1) x_i(t+2)]. \tag{B1}
\]

The individual entries in \(X_{(i)}\) are lagged copies of daily SLPA PCs, with lags ranging from \(-2\) to \(+2\) days. The subscript \(S_i\) denotes the subset of SLPA PCs used in the regression describing the \(i\)th component of the SLPA response variable that we are modeling [e.g., \(x_i(t, 0)\)].

We used a half-window size \(W = 50\) to define subsets \(S_i\) as follows:

\[
S_i = \begin{cases} 
[i - W, i + W], & \text{if } i > W \text{ and } i \leq L - W \\
[1, 2W + 1], & \text{if } i \leq W \\
[L - 2W, L], & \text{if } i > L - W
\end{cases} \tag{B2}
\]

Thus, a given time slice (0000, 0600, 1200, or 1800 UTC) of the 6-hourly PC component is linearly modeled using the extended vector \(X_{(i)}\), composed, per Eqs. (B1) and (B2), of the lagged copies of the nearest \(2W + 1 = 101\) daily-mean PCs as follows:

\[
x_i(t, T) = X_{(i)}(t) B_i(T) + R_i(t, T). \tag{B3}
\]

where \(B_i(T)\) is the \(i\)th column [of length \(5(2W + 1) = 505\)] of the linear-regression coefficient matrix \(B(T)\), the variable \(R_i(t, T)\) is the \(i\)th component of the regression residual, and the tag \(T\) takes one of the four values 0, 6, 12, or 18. The vectors \(B_i(T)\) and residuals \(R_i(t, T)\) were estimated using partial least squares regression (see, e.g., Kravtsov et al. 2011) with the number of latent variables set to 100. Using a larger number of latent variables does not affect the goodness of fit. As per usual in this paper, the regression fits in Eq. (B3) were performed separately for the 12 monthly subsets of data obtained by concatenating the data points for all Januaries, all Feburaries, and so forth.

The setup of the parametric regression in Eqs. (B1)–(B3) involves two central features: the lagged daily
predictors in Eq. (B1) and the grouping of the predictors in the phase-space neighborhood of a given response variable in Eq. (B2). The former property was introduced to reflect the fact that the typical lifetimes of the synoptic disturbances are on the order of a few days. Therefore, to interpolate accurately the evolution of these disturbances, we have to take into account local (in time) variations over a comparable time period; using a smaller maximum lag of 1 day produces a somewhat poorer regression fit of the observed 6-hourly anomalies, whereas using larger lags introduces essentially no further improvements in our fit. Grouping the predictors in the neighborhood of the individual EOF whose time series is being modeled is justified by the dependence of the spatial and time scales of the EOF modes on the EOF rank (see the last paragraph of section 2a). Because of this dependence, the time series of neighboring EOFs, which have spatial and time scales similar to those of a given response variable, are most likely to be the ones that represent skillful predictors of the response. From a technical perspective—setting aside the above dynamical arguments—such a choice of our parametric model is dictated by the necessity to minimize the number of effective predictor variables to get robust estimates of the regression coefficients and avoid overfitting.

APPENDIX C

Stochastic Forcing Options

A widely adopted approach is to model \( r^{(3)} \) (or its monthly subsets) as the spatially correlated Gaussian white noise using the Cholesky decomposition of its covariance matrix (Press et al. 1994). As an alternative, one can bypass the assumption of Gaussianity and simply randomly sample the \( r^{(3)} \) from the actual library of the regression residuals. It turns out that, perhaps as somewhat of a surprise, both of these procedures lead to a substantial inflation of the persistence of simulated PCs (not shown), which clearly has to be due to our omitting nonlinear effects by imposing the linear structure of the regression model in Eq. (1).

The original EMR formulation of Kravtsov et al. (2005) did include nonlinear (quadratic) terms in the main model level to simulate nonlinear effects (see also Kondrashov et al. 2006, 2011; Strouline et al. 2010). This was possible because of a low dimensionality of their EMRs, on the order of 10–15. In contrast, our current SLP model’s main level has 1000 variables to be simulated, which makes the estimation of coefficients involved in a general quadratic form based on so many variables statistically infeasible. An alternative parameterization of the nonlinear interactions not explicitly represented in our linear model in Eq. (1) could be developed by estimating the dependence of the amplitude of the third-level stochastic forcing on the “large scale” model state. Sura et al. (2005) argued that such state-dependent, or multiplicative, noise is a viable paradigm for the description of nonlinear effects in the atmospheric low-frequency variability.

We developed a recipe for incorporating a state-dependent noise forcing in our model by loosely following the ideas of D’Andrea and Vautard (2001), Chekrour et al. (2011), and Kondrashov et al. (2013). The motivation for this new procedure comes from an observation that while the individual-equation forcing residuals are all white in time, there are also nonzero simultaneous and—more important—lagged correlations between the forcing residuals of different equations. These lagged correlations apparently introduce a low-frequency modulation in the structure of the forcing covariance matrix, which our new procedure for forcing resampling is designed to capture. In particular, the forcing at the third level of the model in Eq. (1) was applied in the form of the 15-day-long snippets sampled from the library of the observed third-level residuals \( r^{(3)} \) available after the regression procedure; the model performance is insensitive to the choice of the snippet length in the range of 7–25 days. We chose the starting point of each forcing snippet to be the time [in the library of the observed residuals \( r^{(3)} \)] at which the observed main-level state \( x \) was closest to the current simulated state in terms of the Euclidean distance in the phase space of the leading 100 EOFs of SLPA (this phase space contains essentially all of the large-scale low-frequency SLPA modes). To avoid synchronization of the simulated variability with the observed SLP history, we did not allow the model to choose consecutive snippets from the forcing library used; namely, the starting point of the current snippet had to be separated by more than 7 days from the end point of the previous snippet (5–10-day delays produced similar results). To account for seasonal effects, the above search for the closest state was performed within a subset of the original data corresponding to the current month of model simulation. For example, if the current simulated state belonged to January, the Euclidean distance between this state and the observed state was computed over the all-January subset of the observed daily states, and so forth for other months. The whole procedure described above turned out to improve the model performance dramatically, including the near-correction of the persistence bias that was inherent in other stochastic forcing schemes that we tried.
APPENDIX D

Application of the 6-Hourly Diagnostic Interpolation

The 6-hourly interpolation of the simulated daily SLPA PCs used two steps. In the first step, we computed the part of the 6-hourly variability that is diagnostically captured by the regression-based model in Eq. (B3) of appendix B, using the simulated daily SLPA PCs as the input. Next, the residual variability, which is not captured by Eq. (B3), was separately computed and then added to the 6-hourly variability estimated during step 1 (this step is optional: it does not introduce any changes in the “Eulerian” diagnosis of section 4b, but the cyclone tracking of section 4c turns out to be a bit more robust without the addition of the residual variability).

To compute this residual variability, we used a procedure analogous to the choice of the forcing for the daily-mean model Eq. (1) that is described in appendix C. In particular, for each simulated day, we searched (within a given month’s subset of the observed data) for the day in the observed dataset in which the observed daily state was the closest to the current simulated daily state; this was done, once again, in the phase space of the 100 leading EOFs of SLPA. Then we picked the 24-h snippet (four 6-hourly points) of the observed residual on the day [see Eq. (B3)] as a current contribution to the simulated 6-hourly residual variability.

REFERENCES


