# Forecasting Eastern Mediterranean Droughts

GIDON ESHEL

*Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois*

MARK A. CANE

*Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*

BRIAN F. FARRELL

*Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts*

(Manuscript received 16 October 1999, in final form 3 February 2000)

ABSTRACT

A dynamically motivated statistical forecasting scheme for eastern Mediterranean winter rainfall is presented. The scheme is based on North Atlantic sea level pressure precursors. The resulting forecasts are robust and statistically significant at ~13 months lead time, and improve at ~7 months lead. It is suggested that these forecasts form a foundation for an operational early-warning system for eastern Mediterranean droughts.

## 1. Introduction

The eastern Mediterranean's (EM) hot and rainless summers render highly capricious winter rainfall the region's sole source of renewable water. Because rain-based agriculture plays a primary role in the still developing local economies, EM winter rainfall is socioeconomically crucial. The principal objective of this paper is to devise a skillful, long lead time forecasting scheme for EM winter-mean rainfall anomalies.

The attribute of EM rainfall that facilitates its forecasting is its strong teleconnection with the North Atlantic (NA). While substantial evidence for NA influence on Mediterranean climate variability has emerged (e.g., Hurrell 1995; Hurrell and van Loon 1997; Rodwell et al. 1999; R. T. Sutton 1999, personal communication), it has mostly been peripheral to primarily NA-centered analyses. Two exceptions to this pattern are particularly important for the current work. Lamb and Peppler (1987) advanced some evidence suggesting that the intensity of Moroccan winter rainfall is affected by the North Atlantic oscillation (NAO). They further conjectured about possible predictability of Moroccan rainfall driven by the NA teleconnection, but left the substantiation of this conjecture to future work. Another exception to the above pattern is Rodwell and Hoskins's (1996) study of EM summer aridity. The conventional explanation of Mediterranean aridity is in terms of zonally symmetric Hadley subsidence (weakest during summer). Rodwell and Hoskins (1996) challenge this argument, shifting the focus instead to advective heat flux divergence due to large thermal contrasts between the NA and the monsoon region. Building on the above results, Eshel and Farrell (2000, hereafter EF0) focus on EM rainfall, almost exclusively a winter phenomenon. They advance a simple mechanism linking winter NA climate anomalies to EM rainfall variability, thus combining and extending the scope of the NA–Mediterranean teleconnections proposed by Cullen and deMenocal (2000) and Lamb and Peppler. The winter NA–EM teleconnection forms the physical foundation of the presented forecasting scheme.

Put in a broader context, EF0's work joins a vast body of literature on extratropical stationary waves and their downstream climatic effects. Because major human population centers are affected by stationary wave variability [e.g., downstream of the Aleutian low and the NAO pattern (Kushnir and Wallace 1989; Kushnir 1994; Hurrell 1995)], such variability is of great societal relevance (e.g., Kushnir 1999; Rodwell at al. 1999). The dynamics of quasi-stationary extratropical planetary waves is reasonably well understood theoretically (Held 1983). On the other hand, the scope of their role in observed mid- to high-latitude climate anomalies is still not fully appreciated, and is one of the most intensely

*Corresponding author address:* Dr. Gidon Eshel, Department of the Geophysical Sciences, University of Chicago, 5734 S. Ellis Ave., Chicago, IL 60637.
E-mail: geshel@uchicago.edu

studied issues in current climate research.[1] Among the numerous outstanding questions, interannual to multidecadal predictability of extratropical climate stands out as perhaps the most societally urgent and potentially beneficial. With superior observational data coverage and some indications of predictability (see below), the NA appears to be particularly promising for an exploratory investigation into potential predictability. This is the general context in which this study is embedded.

Encouraging tentative evidence suggesting NA predictability has recently emerged. Hansen and Bezdek (1996) and Sutton and Allen (1997) presented coherently evolving NA sea surface temperature (SST) anomaly patterns, and suggested that their propagation can yield inter- to multiannual downstream predictability. Rodwell et al. (1999) simulated the atmospheric response to observed SST variability and proposed that the roots of much of the relatively low-frequency NAO variance can be traced back to NA SST fluctuations. Because SST evolution is slow relative to characteristic atmospheric timescales (e.g., Kushnir 1994; Hansen and Bezdek 1996), a corollary of this proposition is that the NA presents considerable potential for predictability of European climate more than a year in advance (Rodwell et al. 1999; Kushnir 1999). This notion is strongly supported by the results reported by Johansson et al. (1998), who show that European temperatures can be predicted with considerable skill up to a year in advance using NA predictors. Further, complementing the atmospheric and surface analyses is mounting evidence for large-scale, persistent subsurface ocean anomalies accompanying the atmospheric variability. Curry et al. (1998) track ocean thermal and thickness anomalies as they propagate from the subpolar gyre, where they form in response to atmospheric anomalies, to the subtropics, over approximately six years. Joyce et al. (2000) show slowly evolving anomalies in the Gulf Stream position in the northwestern subtropical gyre, and accompanying large-scale variability of the subtropical mode water. These results are observationally consistent with those reported by Molinari et al. (1997), Taylor (1996), and Taylor and Stephens (1998), and are also conceptually consistent with at least some elements of the coupled mechanism argued for by Latif and Barnett (1994, 1996).

Thus fundamental climatic importance of stationary wave and rainfall variability (and covariability), their potential predictability, and the extreme socioeconomical importance of rainfall in the EM motivate the current study. With EF0's proposed NA–EM teleconnection in mind, we set out to test the feasibility and usefulness of an NA-based statistical forecasting of winter EM rainfall anomalies.

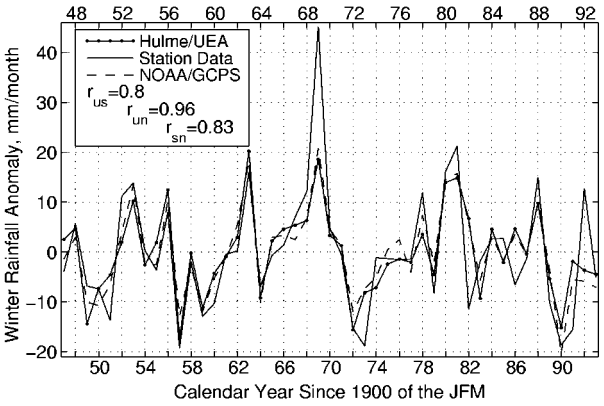


FIG. 1. Three records of winter (Oct–Mar) eastern Mediterranean–mean precipitation anomalies [$\mathbf{P}_{EM}$(yr) in the text]. The records are derived from UEA (Hulme 1992), a subset of high quality, long-coverage station data (EF0), and the NOAA/GCPS gridded dataset (Baker et al. 1995). Because they are used for forecasting, the records are detrended (the small trends represent mostly unresolved low frequencies). The linear correlations between the various records are given in the legend with *r*'s subscripts denoting the correlated records: *u*, UEA; *s*, EF0's station data; and *n*, NOAA/GCPS.

## 2. Data

As in EF0, we define winter as October–March (the rainy months), and the EM as 22°–37°E, 35°–45°N, encompassing parts of Lebanon, Syria, Turkey, and Greece. The somewhat peculiar spatial definition follows the dynamics of the NA–EM teleconnection described by EF0. That is, the choice of boundaries reflects the geographical extent of the area whose teleconnection to the NA is most pronounced. Thus the backbone of this work—the annual-resolution time series $\mathbf{P}_{EM}$(yr) of rainfall anomalies (Fig. 1)—is constructed by averaging monthly mean rainfall anomalies over the EM from October to March.

Figure 1 shows three highly, but imperfectly, correlated rain records (see EF0 for a detailed discussion of the rain records, abbreviated below). Their inconsistencies are not unexpected given the extreme geographical diversity within the EM and the fact that rainfall is one of the hardest geophysical fields to observe. Hulme's (1992; Fig. 2) indicates that the EM is well covered by reliable stations in the University of East Anglia (UEA) dataset. Further, in the 5° × 5° version of the UEA analysis used in Fig. 1, all EM grid points possess at least a 30-yr time series (using the data treatment described by Hulme; see Hulme's Fig. 3). Hence the UEA precipitation record appears representative of the true EM precipitation, and the approximate agreement among Fig. 1's records suggests that any one of them is adequate. In the current study we use the UEA rain

[1] For example, Battisti et al. (1995), Griffies and Bryan (1997), Hurrell (1995), Hurrell and van Loon (1997), Kushnir (1994), Kushnir and Held (1996), Kushnir and Wallace (1989), Latif and Barnett (1994), (1996), Molteni and Palmer (1993), Rodwell et al. (1999), Sutton and Allen (1997), Taylor and Stephens (1998), Wallace et al. (1992), and Weng and Neelin (1998), among others.

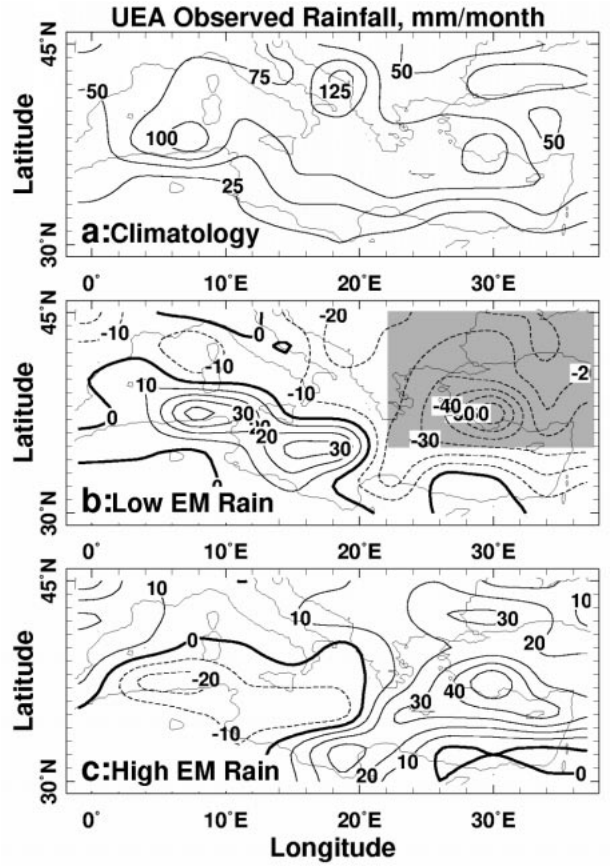


FIG. 2. Oct–Mar mean Mediterranean UEA precipitation (Hulme 1992): (a) climatological field, (b) anomalies during low EM rain winters, and (c) anomalies during high EM rain winters. The spatial extent of the EM box is given by the shading in (b).

record; all the results were reproducible with the other rain datasets.

To characterize dynamical states during $\mathbf{P}_{EM}$ extrema, we form canonical ''rainy'' and ''dry'' states by averaging October–March atmospheric anomalies of the five extreme years: with the year indicating the season's January–February–March (JFM), 1963, 1969, 1980, 1981, and 1988 are wet years, while 1963, 1969, 1980, 1981, and 1988 are dry years. [Extreme seasons of both signs are common to all three of Fig. 1's datasets, with the minor exception of 1987/88, the fifth rainiest year in the station and UEA's records, but seventh in the National Oceanic and Atmospheric Administration/Global Climate Perspectives System (NOAA/GCPS) record.] Thus anomalies characterizing EM precipitation extremes represent averages of the 30 monthly mean anomalies comprising the five most extreme winters. For a given variable, ''low–high differences'' correspond to the difference between averages of that variable during the five driest and the five rainiest EM winters, thus incorporating 60 monthly mean anomalies. Figure 2 shows the climatological winter-mean rainfall (panel a), as well as the averaged 30 monthly mean anomalies during the two extreme periods. Figure 2b also indicates the spatial extent of the EM as defined above. The anomalies are very large, of the same order of magnitude as the climatological means, and are antisymmetrical in the east–west direction, for reasons discussed below.

Tropospheric variables (winds, humidity, temperature) are taken from the National Centers for Environmental Prediction–National Center for Atmospheric Research reanalysis project (Kalnay et al. 1996). The reanalysis is strongly constrained by data in the data-dense Mediterranean and appear satisfactory over the Mediterranean [Trenberth and Guillemot (1998): e.g., the reanalysis EM-mean winter 1000-mb temperature time series correlates at 0.92 with the analogous time series derived from ship-observed surface temperatures (daSilva et al. 1994)]. Whenever possible, the validity of the reanalysis data is examined by comparisons with other independent datasets.

Sea surface variables (both observed and calculated) are taken from daSilva et al. (1994), which derives from ship observation (Comprehensive Ocean–Atmosphere Data Set, Slutz et al. 1985). Significance tests we use in the analysis are described in some detail in the appendix.

## 3. Statistical $\mathbf{P}_{EM}$ hindcasting

### *a. Dynamical motivation*

This paper is motivated by EF0, who show that EM rainfall anomalies are local manifestations of large-scale perturbations of NA planetary waves. Near Greenland, where these perturbations are particularly pronounced, they are characterized by persistent $O(50$ m) geopotential height anomalies. Accompanying Greenland anomalies is a reversed-polarity, Mediterranean-wide node centered over the northern Adriatic [Fig. 3, see also Figs. 2 and 3 in Lamb and Peppler (1987)]. Consequently, elevated (reduced) Greenland pressure is consistent with anomalous cyclone (anticyclone) over the Mediterranean. In the EM, this results in anomalous southerlies during Greenland highs, and northerlies during Greenland lows (Fig. 4). As (mean, as well as instantaneous) temperatures rise toward the south, EM southerlies (northerlies) warm (cool) the EM. Because heat advection by horizontal and vertical motions dominate the EM heat budget (Rodwell and Hoskins 1996), cooling (warming) by horizontal winds is dynamically consistent with enhanced (reduced) subsidence (cf. the Mediterranean patterns shown in Figs. 3c and 3d, and in Figs. 4a and 4c). This is related to variability with height and latitude of EM potential temperature and specific humidity. The southward temperature increase means that EM winter isentropes slope down toward the south. Consequently, to the extent that the meridional winds are nearly adiabatic (isentrope following), air subsides when flowing north to south during periods of low Greenland pressure and EM rainfall, and ascends when

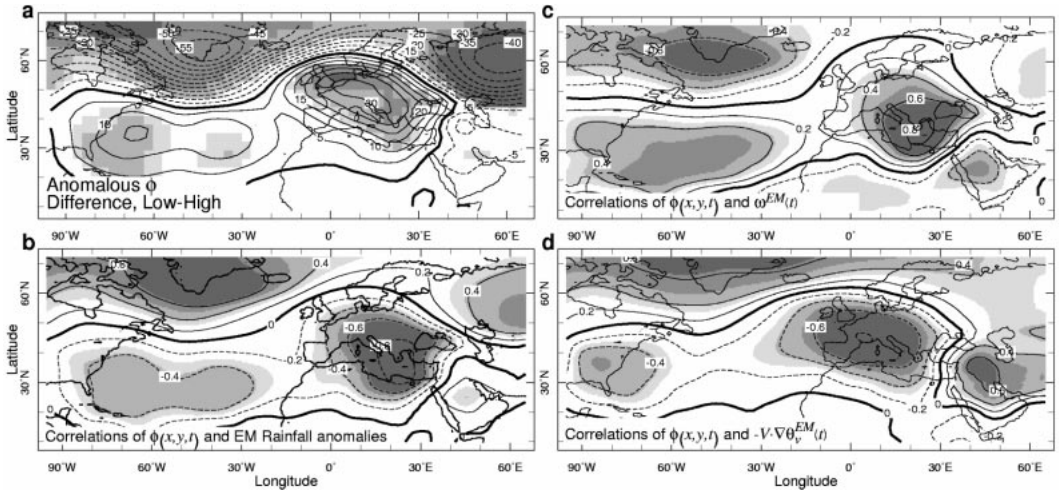


FIG. 3. Observed 850-mb anomalies associated with winter EM rainfall variability over the NA–Mediterranean sector. (a) Low–high EM rainfall difference of anomalous geopotential height (m; i.e., the difference between the averages of the five driest and five rainiest winters). (b)–(d) Correlations of height anomalies with EM-mean rainfall (Hulme 1992), subsidence, and horizontal advective heat flux divergence anomalies (Kalnay et al. 1996), respectively. Shadings indicate significance of 0.1, 0.05, 0.01, and 0.005, allowing for persistence. Time series are detrended prior to computing correlations.

flowing south to north during periods of enhanced Greenland pressure and EM rainfall. Because water vapor mixing ratio decrease rapidly with height, the vertical motions profoundly affect the moisture budget of the lower troposphere. During northward upgliding the lower troposphere is downstream of the moist boundary layer and is, thus, being advectively moistened. Conversely, subsidence dries the lower troposphere by advecting into it dry air from aloft. By modifying large-scale condensation and static stability with respect to moist ascent, these moisture advection patterns directly affect the intensity of deep convection and rain generation, causing the observed EM rainfall anomalies. This mechanism is entirely consistent with the opposite NAO dependencies of EM (EF0) and Moroccan (Lamb and Peppler 1987) rains; as western Mediterranean meridi-

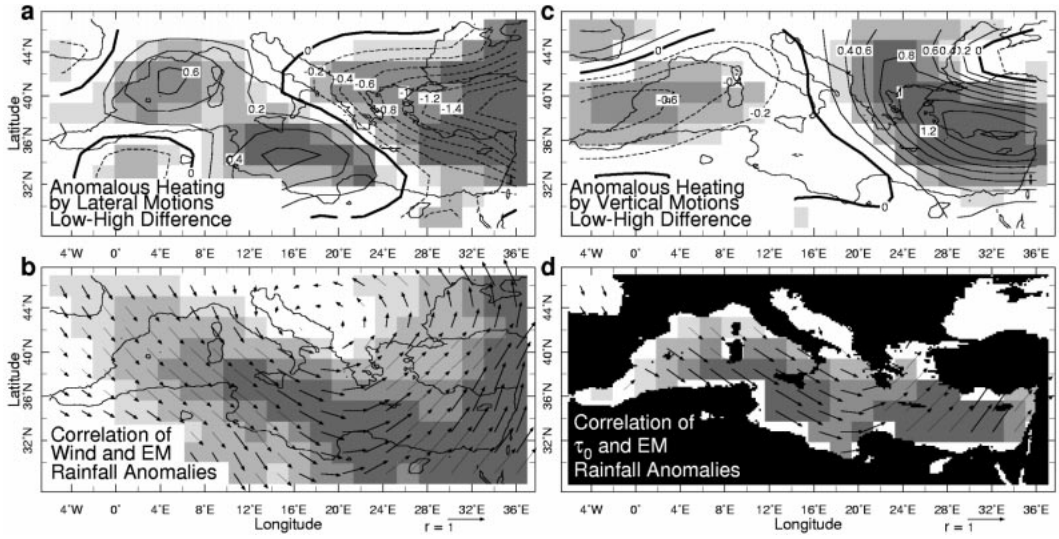


FIG. 4. Observed Mediterranean anomalies associated with winter EM rainfall variability, at (a)–(c) 850 mb and (d) the sea surface. (a), (c) Differences (between the averages of the driest and rainiest winters) of anomalous (a) horizontal and (c) vertical advective heat flux divergence, K day$^{-1}$ (Kalnay et al. 1996). Vector correlations between the anomalous EM-mean rainfall time series (Hulme 1992), and (b) 850-mb winds (Kalnay et al. 1996) and (d) sea surface wind stress calculated from ship observations (daSilva et al. 1994). Shadings indicate significance of 0.1, 0.05, 0.01, and 0.005, allowing for persistence. In (b) and (d), the significance corresponds to correlation magnitude, and reference $r = 1$ arrows are shown near bottom-right corners. Time series are detrended prior to computing correlations.

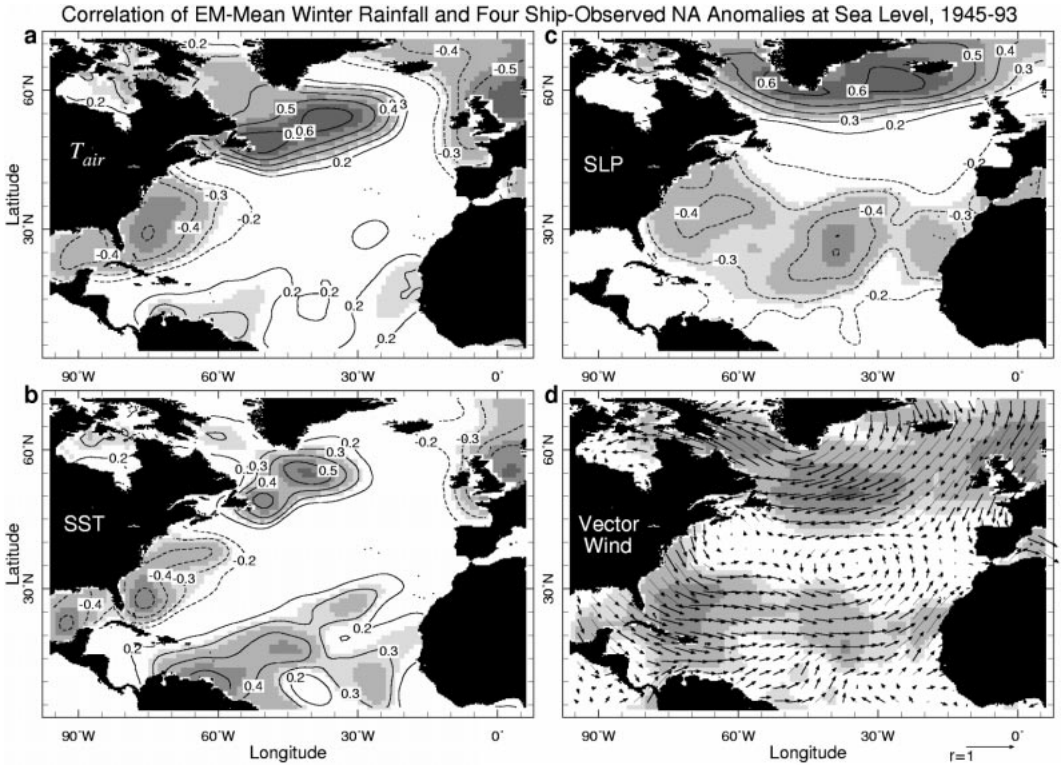


FIG. 5. Correlations of the winter- and EM-mean rainfall anomaly time series with four ship-observed NA anomaly fields [air temperature, sea surface temperature, sea level pressure, and vector winds, indicated on the American landmass in each panel, from daSilva et al. (1994)]. Only $|r| \geq 0.2$ contours are shown, at 0.1 increments. The four shadings denote local significance of 0.1, 0.05, 0.01, and 0.005. For the vector wind correlations in (d), a reference $r = 1.0$ arrow is shown in the lower right, and the significance corresponds to the correlation magnitude.

onal wind anomalies are opposite to those in the EM, the above considerations, with opposite signs, hold in the two locals, yielding the observed opposite response.

The NA–EM teleconnection, along with the fast response time of the atmosphere and the large-scale disparity between the EM and the NA, suggest that the state of the (much smaller) EM can be approximated as instantaneously adjusting to the evolving state of the NA. If NA evolution tends to favor certain phase-space trajectories, the EM state, which is determined by the NA anomalies that the trajectories correspond to, is somewhat predictable. The degree to which this approximation holds is established by simultaneous and lagged analyses, and is tested below.

Consistent with the EF0 teleconnection, EM rainfall anomalies display high, large-scale, simultaneous correlations with various NA surface variables (Fig. 5). Most important for prediction, the correlations also hold with *past* anomalies (Fig. 6), suggesting that NA evolution does indeed favor certain phase-space trajectories that contribute to EM predictability. Figure 6 shows that NA sea level pressure anomalies (SLPAs) contain information about future $\mathbf{P}_{\mathrm{EM}}$. [Very similar correlation patterns emerge using other EM rain records (EF0).] This is the basis of our forecasting scheme. The procedure for forecasting $\mathbf{P}_{\mathrm{EM}}(t + \text{lead})$ using NA anomalies at time $t$ has two stages. In the first, time-independent spatial patterns of NA anomalies most useful for forecasting $\mathbf{P}_{\mathrm{EM}}(t + \text{lead})$ are calculated for each of the considered leads. In the second stage, predictor data for each lead are projected on the lead's pattern. The resulting univariate annual resolution time series are then used, in a multiple regression framework, to hind-/forecast $\mathbf{P}_{\mathrm{EM}}$.

### *b. Pattern calculation and projection*

To lower the risk of overfitting to noise, we form overlapping 3-month averages of SLPA data (DJF, JFM, etc.). We consider leads of up to 18 months, beyond which we do not anticipate useful predictability. [Lead time is measured from midpredictor period to midrainfall period (e.g., lead = 8.5 month corresponds to inferring Oct–Mar $\mathbf{P}_{\mathrm{EM}}$ from previous Mar–Apr–May-mean predictor data).] The $\mathbf{P}_{\mathrm{EM}}$ time series is linearly best-fit detrended, as are the annually resolved, lead-specific, SLPA time series at each grid point. For each of the 18 considered leads, we compute local SLPA–$\mathbf{P}_{\mathrm{EM}}$ correlations, and $t$ test their significance, allowing for copersistence in the time series involved as described in section b of the appendix. (Fisher $\zeta$ transform yielded essentially identical patterns, not shown.) Local corre-

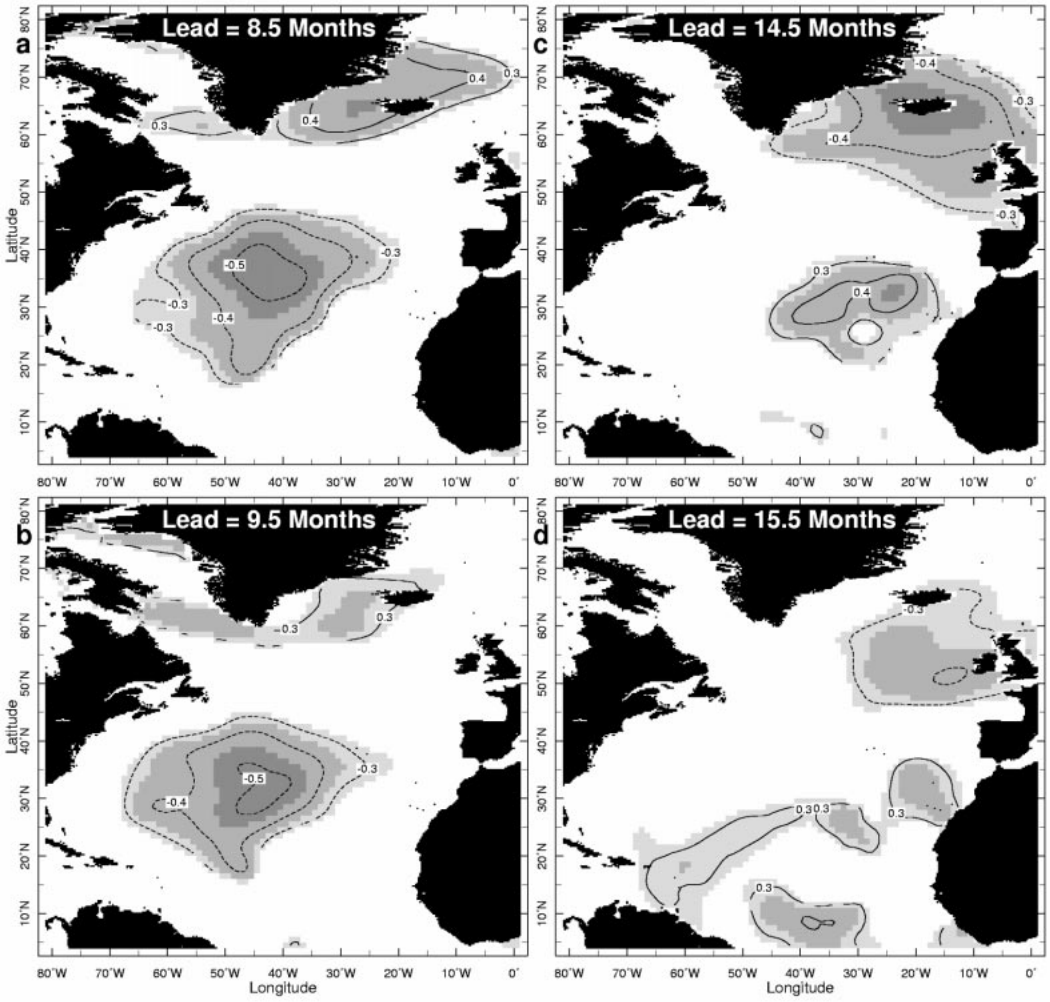


FIG. 6. Correlation of EM winter rainfall anomalies with *past* SLPAs (from daSilva et al. 1994). Each panel's SLPA lead is indicated near the top. Shadings indicate local significance as before; all panels are field significant at $p < 0.05$.

lations that fail to achieve $p < 0.05$ significance are set to zero, and $p < 0.05$ ones are retained. This yields a correlation map for each lead considered, whose nonzero elements are $p < 0.05$ significant gridpoint (local) correlations with $\mathbf{P}_{\mathrm{EM}}$. We then compute the area-weighted average absolute correlation at lead $l$, denoted $f_l$. This is our field-significance statistic. The field as a whole is deemed significant at $p < 0.05$ when $f_l > f_l^t$, where $f_l^t$ is the threshold areal fraction for the considered predictor at lead $l$ (found following the Monte Carlo method described in the appendix).

Figure 7a shows that SLPA correlations proved field significant in 9 out of the 17 leads considered, far in excess of what can be expected by chance. The associated correlation patterns form various subtropical–subpolar dipoles, four examples of which are shown in Fig. 6. The long-lead correlation maps fall into two general categories. In the first, in the late winter/spring ($7 <$ lead $< 11$ months, e.g., Figs. 6a and 6b), patterns are similar to the simultaneous one (Fig. 5c), with negative correlation in the subtropics, and positive ones around Greenland. The second category of correlation patterns occurs for the longer leads that are still significant (late summer/fall, e.g., Figs. 6c and 6d). These patterns tend to be less spatially coherent, but clearly display sign reversal of both (subtropical and subpolar) nodes. The two categories are separated by three insignificant leads (11–13 months). Together, Figs. 5c, 6a,b, and 6c,d suggest that the dependence of EM rainfall on NA SLPAs oscillates while being damped. Figure 7b gives this notion more precision. The panel shows that while EM rainfall is significantly simultaneously correlated with more than 60% of the domain, this percentage drops to essentially zero in the peak of the preceding summer. The percentage rises again farther back in time, achieving high significance in the early spring preceding the insignificant summer (at 8–11-month lead). After deteriorating back to insignificance by mid-

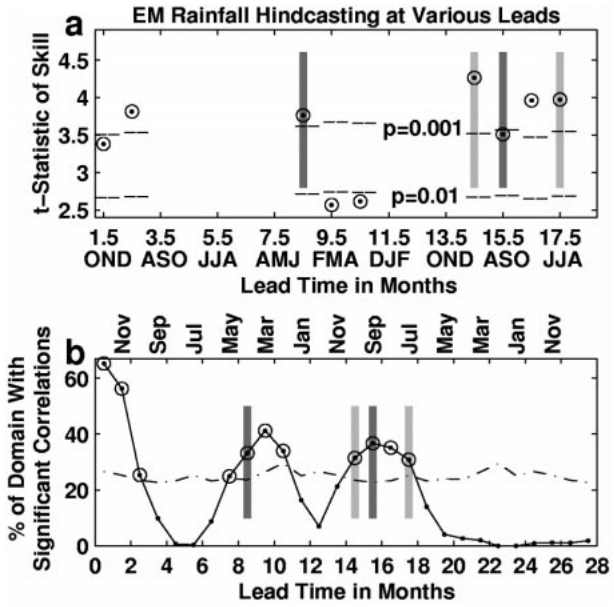


FIG. 7. (a) Hindcasting $\mathbf{P}_{EM}$ using NA SLPAs at various leads. Leads not shown failed the field-significance test and did not permit hindcasting. For the significant leads, the annual data are projected on the calculated ''most predictable pattern'' (see text), and the resultant time series is used, in a univariate regression analysis, to hindcast EM rainfall anomalies. The hindcast correlation is converted to a $t$ value that is plotted, along with the 0.01 and 0.001 significance levels. (b) A demonstration of the oscillatory nature of the NA–EM teleconnection. For smoothness, here the SLPA dataset is temporally smoother than the SLPA dataset on which all other calculations are based. This is done by convolution with [1 2 3 2 1]/9, unlike the [1 1 1]/3 in the rest of the paper. Panel (b) shows the percent of the domain covered by significant correlations at leads 0–28 months. The dashed–dotted line is the field-significance threshold percentage (found by the Monte Carlo calculation). Field-significant leads are shown by open circles. The shaded bars show the four leads used for forecasting. Note the different horizontal axes of (a) and (b).

winter (at 13–14-month lead), the percentage rises yet again in the fall of the previous year (lead $\cong$ 15–17 months) before correlations are finally lost.

Figure 7b suggests that spring and fall are particularly important for the NA–EM teleconnection. While this impression may arise from a seasonal cycle of signal-to-noise ratio, it is tempting to speculate about alternative underlying physics, as follows. Due to winter-long interactions, the subsurface ocean (SSO; between the permanent and seasonal thermoclines) is closest to thermal equilibrium with the atmosphere in late winter. Shortly thereafter, the SSO decouples from the atmosphere as the summer thermocline becomes established. In the following fall, the SSO reconnects to the surface (and thus the atmosphere) as the mixed layer deepens. Given their widely different timescales, the summer-long disconnect affects the SSO and atmosphere differently. Whereas the atmosphere has essentially no spring-to-fall memory, the slowly evolving, near-adiabatic, SSO retains considerable memory of its spring-time state into the following fall. Consequently, the thermodynamic states of the SSO and atmosphere may differ widely in the fall, enhancing surface heat fluxes. To the extent that the atmosphere is sensitive to surface flux forcing, the enhanced surface fluxes force the atmosphere toward a state more consistent with that of the SSO. But the SSO's state will have changed relatively little since the previous spring, when the SSO was nearest to equilibrium with the atmosphere. In other words, the enhanced surface fluxes nudge the atmosphere toward the atmosphere's state in the previous winter. This provides the atmosphere with winter-to-winter persistence, enhancing predictability. We leave the substantiation of the above conjectures to future work.

Within the above simple framework, the SLPA correlation maps $\{\mathbf{p}_l\}$ of all $p < 0.05$ field-significant leads $l$ represent NA SLPA configurations that yield the most $\mathbf{P}_{EM}$ predictability at the chosen lead. Consequently, we project SLPAs at lead $l$ during each year on this lead's correlation map to obtain the predictor time series $\mathbf{r}_l$(yr). For example, for lead = 8.5 months, in which $\mathbf{P}_{EM}$ is hindcasted using MAM SLPAs, we project every year's MAM-mean SLPA field (after detrending over all available MAMs) on the (time independent) 8.5-month lead time correlation map $\mathbf{p}_{8.5}$.

### c. Hindcasting

The SLPA-derived time series $\mathbf{r}_l$(yr) are next combined in multiple regression analyses for hindcasting $\mathbf{P}_{EM}$. The choice of predictor combinations is based on the following considerations. First, of course, we wish to maximize lead time and skill, which dictates using simultaneously as many predictors as are available. Such a strategy also reduces some of $\{\mathbf{r}_l\}$'s noise. On the other hand, the risk of overfitting to noise increases with the number of predictors used. {Considering only the effective number of degrees of freedom in the $\mathbf{P}_{EM}$ timeseries, ~33 [using Eq. (A1)], an upper bound of four or five predictors is defensible.} In addition, because $\{\mathbf{r}_l\}$ represent various phases of the SLPA-damped oscillation discussed above, a given year's evolution of SLPA projection between two leads $l_1$ and $l_2$, $\mathbf{r}_{l_1} \rightarrow \mathbf{r}_{l_2}$, represents additional predictive information. Guided by the above considerations, we use two predictors in a given regression analysis. We construct two hindcasts based on two predictors each, chosen among the four leads with the highest individual hindcasting skills. (This raises a concern of artificial skill. However, since over half of all considered leads prove highly significant, and because each of these significant leads individually yields a $p < 0.01$ significant hindcast, we believe this concern, while reasonable, is unwarranted in this case. This is tested by cross validation and reshuffling below.)

The first hindcast combines SLPA information from the previous JJA and SON (17.5- and 14.5-month leads): the second combines ASO and MAM (15.5- and 8.5-month leads):

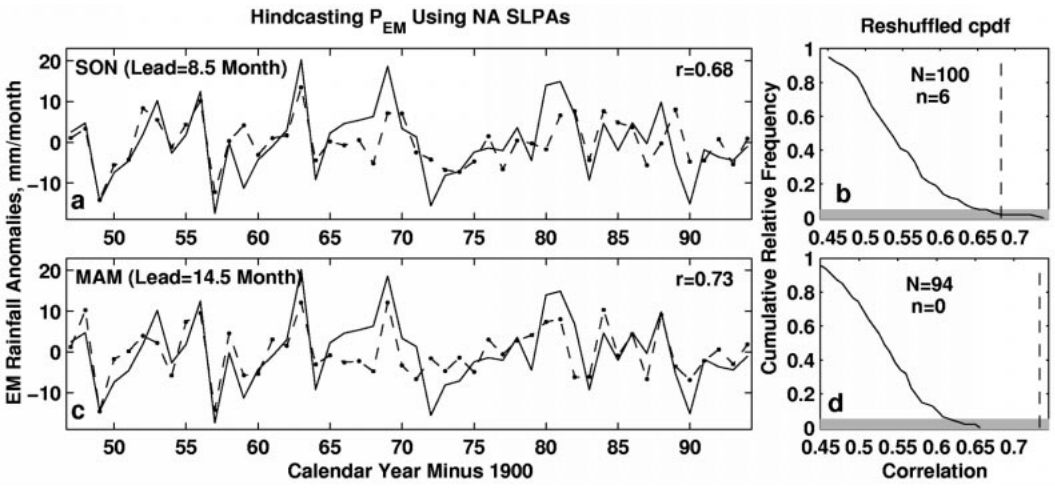


FIG. 8. Hindcasting $\mathbf{P}_{EM}$ using NA SLPAs. Leads and hindcast-observed correlations are indicated. (a), (c) Observations and hindcasts are plotted with solid and dashed with solid circles, respectively. (b), (d) The calculated hindcast–observation correlations (vertical dashed lines) compared with the cumulative pdf of the field-significant subset of the 1000-member reshuffling correlation population (whose size is given by the uppercase $N$). Lowercase $n$ denotes the number of members whose hindcast correlations equaled or exceeded the ones obtained with the original (nonreshuffled) rain record. The shading indicates the $p < 0.05$ region (with respect to the field-significant subset only).

$$\mathbf{B}_{14.5} = (\mathbf{r}_{17.5} \quad \mathbf{r}_{14.5}) \quad \text{and} \quad \mathbf{B}_{8.5} = (\mathbf{r}_{15.5} \quad \mathbf{r}_{8.5}), \qquad (1)$$

where the $N \times 2$ matrices $\mathbf{B}_l$ yield the two regression coefficients in $\hat{\mathbf{x}}$ through a least squares solution of

$$\mathbf{B}_l\mathbf{x} = \mathbf{P}_{EM}. \qquad (2)$$

The two $\mathbf{P}_{EM}$ hindcasts thus obtained [for the two lead combinations given in Eq. (1)] are shown in Figs. 8a and 8b. Estimating *df* using Eq. (A1) and then accounting for the loss of *df* due to the multiple predictors, Fig. 8's correlations are *t*-test significant at $p \sim O(10^{-5})$. In addition, both results prove $p < 0.05$ significant with respect to the reshuffled population (Figs. 8b and 8b, and section e of the appendix; note that computing the $p = 0.05$ using the field-significant members only results in a very conservative significance estimate, because the vast majority of the population members are zero). The difference between the results of the two leads (0.73–0.68) is insignificant ($p = 0.24$) and is probably coincidental.

## 4. Cross validation

The highly significant correlations of Fig. 8 are consistent with the connection between NA SLPAs and EM rainfall variability previously argued for. However, because hindcasting skill often fails to yield a corresponding *forecasting* skill, the latter must be *shown.*

To test forecasting skill, we use cross validation; we withhold from both the SLPA and $\mathbf{P}_{EM}$ datasets some years, repeat the entire analysis described above (beginning with the correlation pattern calculations), and use the $\hat{\mathbf{x}}'$ obtained from solving the deficient (partial) Eq. (2) (where the prime denotes ''deficient'' values based on the retained information only) to forecast $\mathbf{P}_{EM}$ values for the withheld years. Because the deficient correlation patterns are tested for field significance, the analysis can fail even before a forecast is attempted, because it allows for the possibility that the correlation pattern of a given lead will prove field insignificant, and will not be used for forecasting.

We first withhold randomly chosen years (i.e., in general nonsequential) two, three, and four at a time. Since the total number of possible permutations in each of these groups is very large, we sample the populations randomly 1000 time each. The results are shown by Fig. 9 and in Table 1. For a given experiment (number of withheld years) and lead time, we first average all forecasts of a given year. [That is, each of the full dots connected by a dashed line in Fig. 9 represents the average of all forecasts of that year, while the forecasts' spread ($\pm 1.96\sigma$) about that average is indicated by the shading width.] The time series of these means for all years is the ''mean forecast.'' In all experiments, the number of forecasts per year used to construct the mean forecast (Table 1, columns 7–9) is quite high, affording reasonably stable and representative mean forecasts.

The correlations between the mean forecasts and the observed $\mathbf{P}_{EM}$, and their significance, are shown in Table 1, columns 3–4. All are $p < 0.05$ *t*-test significant allowing for persistence and loss of *df* due to two predictors. Unlike the non-cross-validated case, the skill decreases with increasing lead, as expected. As an alternative to the parametric *t* test, Fig. 10 compares the obtained results to an experimental population obtained from repeated cross-validation experiments with randomly reshuffled $\mathbf{P}_{EM}$. Because the results of the 2-, 3-, and 4-yr withholding experiments are very similar (Table 1), Fig. 10 shows only the 3-yr withholding results.

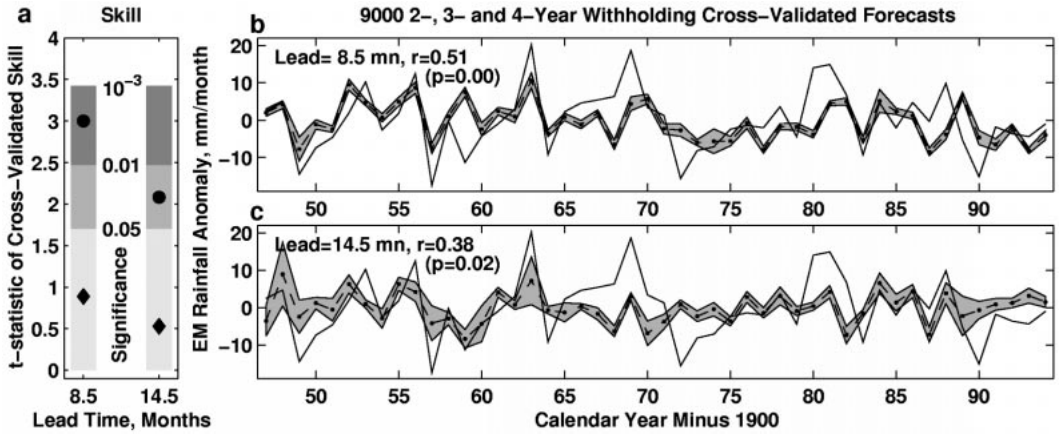


FIG. 9. Cross-validation skill for two lead times; see text for a full description. (a) The results for both leads. Forecasts for withheld years from all withholding experiments are averaged to form a full-length forecasted time series, shown by the dashed lines in (b) and (c). The variability of a given year's forecasts ($\pm 1.96\sigma$) is indicated by the shaded region enclosing the mean forecast (dashed). Also shown are the observed rain record (solid), the leads, and forecast–observation correlations and their significance. The correlations (skills) are transformed into $t$ values (allowing for copersistence and loss of $df$ due to the two predictors), shown in panel (a) by the full circles. Their significance can be visualized in comparison to 0.05, 0.01, and 0.001, indicated by the shadings. A ''null'' persistence forecast based on $\mathbf{P}_{EM}$'s autocorrelation function is shown for both leads (full diamonds).

Figures 10a and 10c show that the obtained correlations clearly exceed the expected $p = 0.05$ limit derived from the reshuffling experiments.

Because correlation analysis takes note primarily of phasing (and not amplitudes), we also analyze the variance, dividing the total variance in $\mathbf{P}_{EM}$'s forecast into the parts due to the forecast and error (see section d of the appendix). The analysis of variance (ANOVA) results are given in Table 1, columns 5 and 6. For the 8.5-month lead, all experiments yield highly significant results (essentially independent of the number of withheld years). The results for the 14.5-month lead are (marginally) significant when two years are withheld and become insignificant when more than two years are withheld; with a lead time of 14.5 months, $\mathbf{P}_{EM}$'s amplitudes are not forecasted very well by the scheme. These conclusions are supported by the reshuffling results shown in Figs. 10b and 10d. The somewhat enhanced skill the scheme displays at forecasting extreme events is encouraging, as those events have disproportionally high societal impacts. Also reassuring is the relatively tight distribution of forecasts about both observed values and forecast means, suggesting that the forecasting skill is not a chance occurrence.

In the final cross-validation experiment, including 37 cases, 25% of the datasets (12 seasons) are withheld *sequentially*. We begin by withholding years 1–12, then 2–13, and so on, until the last experiment, in which years 37–48 are withheld and later forecasted. This sliding withholding window results in year 1 having a single forecast (as part of the year 1–12 withholding experiment), increasing to 12 forecasts for year 12. Years 12–

TABLE 1. Statistics of the multiyear nonsequential cross-validation experiments. First two columns give the number of withheld years and the lead. Third and fourth columns present the correlations between the mean forecast and the observed $\mathbf{P}_{EM}$, and its $t$-test significance. Columns 5 and 6 present the results ($F$ ratios and their significance, respectively) of the analysis of variance as described in the text. Columns 7–9 give the minimum, average, and maximum number of forecasts per year, which are averaged while forming the ''mean forecast'' for the given combination of number of withheld years and lead. Because the results are relatively insensitive to the number of withheld years, the last two rows (marked C) combine all results (of withholding 2, 3, and 4 yr) into a single population.

| | | Mean forecast | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | ANOVA | | No. of forecasts per year | | |
| $N_w$ | Ld | $r$ | $p$ | $F$ | $p$ | Min | Avg | Max |
| 2 | 14.5 | 0.38 | 0.0250 | 3.48 | 0.0466 | 19 | 41 | 59 |
| 2 | 8.5 | 0.51 | 0.0032 | 6.05 | 0.0072 | 29 | 42 | 59 |
| 3 | 14.5 | 0.38 | 0.0244 | 3.23 | 0.0563 | 44 | 62 | 88 |
| 3 | 8.5 | 0.52 | 0.0028 | 6.13 | 0.0068 | 45 | 63 | 88 |
| 4 | 14.5 | 0.39 | 0.0245 | 3.28 | 0.0545 | 66 | 82 | 98 |
| 4 | 8.5 | 0.52 | 0.0028 | 6.15 | 0.0067 | 68 | 83 | 98 |
| C | 14.5 | 0.38 | 0.0234 | 3.29 | 0.0537 | 137 | 185 | 219 |
| C | 8.5 | 0.51 | 0.0029 | 6.12 | 0.0069 | 162 | 187 | 220 |

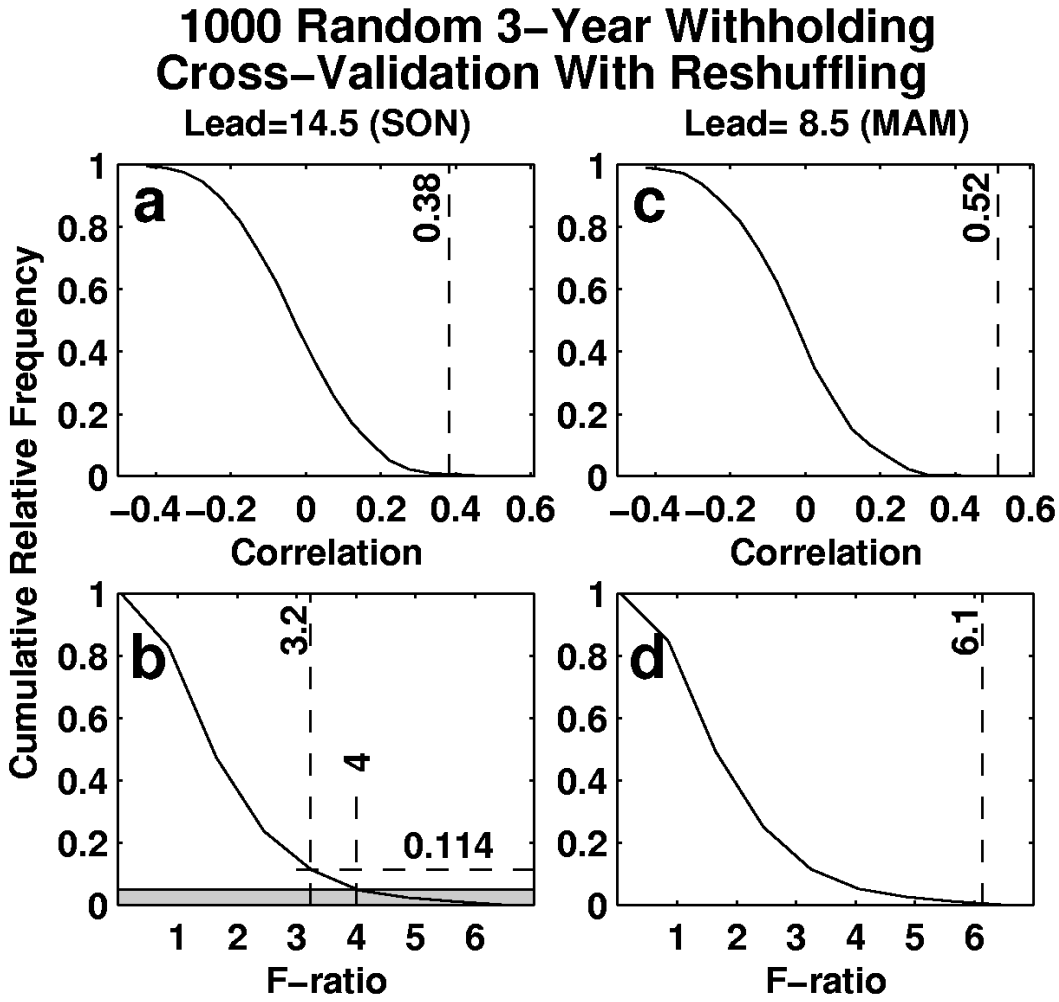


FIG. 10. Reshuffling test of the 3-yr withholding cross-validation results. The left and right columns present the results for 14.5- and 8.5-month leads, respectively. In each of the 1000 reshuffling experiments, a cross-validation test is performed 250 times, a mean forecast based on these 250 results is calculated, and the correlation between this mean forecast timeseries and the reshuffled rain time series is evaluated. The cumulative probability distribution functions of these correlations are shown by the solid curves of (a) and (c). The obtained results without reshuffling are given by the vertical dashed lines. (b), (d) The corresponding *F*-test results.

37 have 12 forecasts each and from year 38 onward the number of forecasts decreases until the last year (48), which again has a single forecast. The results of the 12-yr withholdings are shown in Fig. 11, which presents the mean forecast and the forecast variability about the mean for each of the years with seven or more forecasts (years 7–42). Figure 11 also shows the correlations of the mean forecast with $\mathbf{P}_{\mathrm{EM}}$ (0.41 and 0.51) and their significance allowing for persistence. While the correlations are still quite high and significant, the forecast amplitude is lacking, as is visibly clear in Fig. 11. The analysis of variance yields $F$ ratios for the 14.5- and 8.5-month leads of 2.24 and 4.20, corresponding to $p \simeq 0.135$ and $p \simeq 0.032$. That is, the 14.5-month lead forecasts fail to achieve 0.05 significance, while the 8.5-month lead forecasts are still significant at 0.05, but not at 0.01, and their skill appears limited. We interpret this result not as invalidating the forecasting scheme, but rather as delineating its minimal requirements for adequate operation. Given the nature (spectral characteristics) of EM rainfall and NA SLPAs, this minimum appears to be approximately 40 years' worth of data, a requirement the currently existing observational database exceeds.

## 5. Summary and main conclusions

The main objectives of this paper are to quantify NA-based EM predictability, and to construct a forecasting scheme that exploits this predictability. We use a relatively simple scheme; it is likely that skill can be improved by employing more sophisticated methods.

Sections 3 and 4 address the paper's stated objectives. Significant hindcasting skills are first demon-

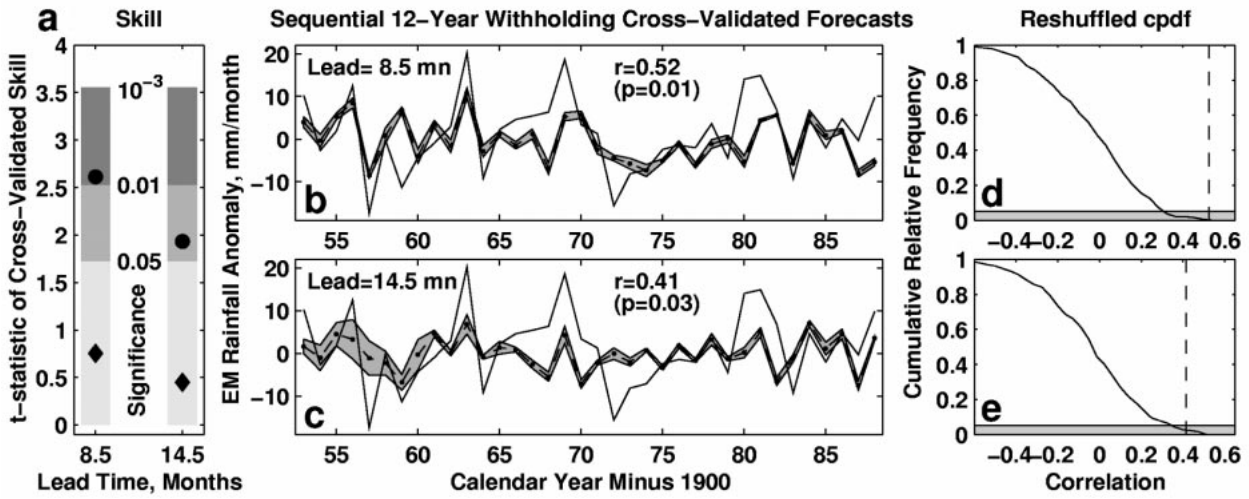


FIG. 11. Sequential 12-yr cross-validation skill for two lead times; see text for a full description. For the 36 years with seven or more different forecasts (corresponding to forecasts of that year as part of different groups of sequential 12 withheld years), the mean forecast is computed (dashed), as well as the variance of the forecasts about that mean, shown by the shading. The leads, correlations of the mean forecast and observed rainfall (solid) over the 36-yr period, and their significances are also shown.

strated, and then tested, using cross validation and random reshuffling. The scheme hindcasts almost 50% of the variance in both 14.5- and 8.5-month leads. Cross-validated levels are 25% and 15%, respectively. With sufficient data ($\geq$40 yr), the forecasts are skillful and reproducible, and their correlations with observations, and internal distribution of variance, are both highly significant.

The forecasts can certainly be further refined. However, the present method performs surprisingly well, providing more than a full year of drought warning, with further refinement in the preceding early spring. Most importantly, the scheme can be implemented operationally by the region's authorities with nontrivial but straightforward modifications.

*Acknowledgments.* B. Farrell was supported by NSF Grant ATM9623539, G. Eshel by the UCAR/NOAA Climate and Global Change postdoctoral fellowship and by the Woods Hole Oceanographic Institution, and M. Cane by NOAA Grant NA56GP0221.

## APPENDIX

## Methods

### *a. Local significance test of plotted fields*

In various figures, we present significance levels of local (at a grid point) anomalies $a$ and correlations $r$, employing the $t$ test [using the Fisher $\zeta$ transform against $N(0, 1)$ yielded essentially indistinguishable results, not shown]. To form $t$ statistics, we use $t = r\sqrt{(df - 2)/(1 - r^2)}$ and $t = (a - \overline{a})/\sigma_a^t$ (Philips 1988), where $\overline{a}$ and $\sigma_a^t$ are $a$'s temporal mean and standard error. In estimating the number of degrees of freedom $df$ (for plotting purposes only), we make the simplifying rather stringent assumption that the decorrelation timescale is 2 yr (i.e., that only every 2-yr pair represents an independent realization; $df = N/2$, where $N$ is the time series length in years). The autocorrelation function of the rain record, as well as those of SLPA at randomly chosen grid points, indicate that this is a lower bound $df$ estimate, thus representing a conservative estimate of significance. To exclude spurious correlations primarily due to temporal trends, correlations are computed after linear trends are removed by a least squares fit. The calculated $t$ values are compared with critical ones ($t_c$) obtained from the $t$ distribution with $df$. When $|t| > t_c$, the value (either anomaly or correlation) is assumed locally significant.

### *b. Local significance test for calculations*

Pattern (and other) calculations require a more rigorous estimate of correlation significance, based on the time series' *actual df.* For these calculations, we allow for copersistence in the time series involved by integrating the autocorrelation functions:

$$df = \frac{N}{1 + 2\mathbf{c}_r^{\mathrm{T}}\mathbf{c}_p}, \tag{A1}$$

where $df$ is the correlation's degrees of freedom, $N$ is the length of the correlated time series, the $\mathbf{c}_r$ and $\mathbf{c}_p$ column vectors are the autocorrelation functions of the two participating time series (typically EM rainfall and SLPA), and superscript T denotes transpose [Livezy and Chen (1983), Eqs. (1) and (2); see also Trenberth (1984), Eq. (2.8)]. To avoid artificial high lag autocorrelations related to the small number of elements remaining at high lags, we restrict $\mathbf{c}$'s maximum lag to $N/2$; the trun-

cated terms are almost invariably very small, as expected. As before, all time series are linearly detrended before correlations are calculated.

### c. Field significance test

Among the central elements of the presented forecasting scheme are correlation patterns of the EM rainfall time series with NA SLPA. Such patterns are established by evaluating the correlation between the EM rainfall and the SLPA time series at each of the domain's grid points individually, and then testing their local significance according to section b of the appendix. Because the domain comprises 4370 grid points, we expect some of them (or a certain areal fraction of the domain) to be locally significant *by chance* (Barnston and Van den Dool 1993; Barnston 1994; Barnston and Kumar 1999; Livezy and Chen 1983). If each of the domain's grid points were independent of all other points, this areal fraction would have been binomially distributed, and the problem would have reduced to a look-up at the cumulative binomial distribution table. However, this spatial independence is clearly not the case for most geophysical fields, nor for all the ones considered here; there is substantial redundancy in the gridpoint representation, related to the field's number of spatial degrees of freedom.

To estimate a given field's spatial degrees of freedom (or the field-significance threshold, the areal fraction that can be expected to prove locally significant by chance), various approaches can be employed. One general approach is based on various forms of eigenanalysis of the data arrays; from the rate of falloff of the (eigen or singular) spectrum, one can attempt to estimate the field's number of spatial degrees of freedom. However, this method can be too sensitive to the temporal coverage and the domain geometry (which together determine the dimensions of the decomposed matrices and, thus, the maximum possible number of modes) and to noise (which can contaminate higher modes). A more general alternative to this method we employ here is Livezy and Chen's (1983) Monte Carlo approach. We fit an autoregressive, moving average ARMA(1,1) model (the most robust and parsimonious choice based on terms larger than lag 1 in both partial autocorrelation and autocorrelation functions being insignificantly different from zero) to the observed $\mathbf{P}_{\mathrm{EM}}$ time series, and generate 2000 synthetic rainfall records with $\mathbf{P}_{\mathrm{EM}}$'s spectral properties and variance. Within the probabilistic framework adapted here, these synthetic time series play the role of other possible realizations drawn from the population that the actual observed $\mathbf{P}_{\mathrm{EM}}$ is a member of. We then perform 2000 correlation analyses, in each of which one of the synthetic rainfall records is correlated with the time series of the tested field at each of the domain's grid points. For each one of the 2000 analyses, we save the areal fraction of the domain covered by locally $p < 0.05$ significant correlations (allowing for poleward convergence of the meridians). The 2000 realizations form an experimental population with a known probability density function (pdf), the likelihood for locally significant correlations covering a given areal fraction of the domain. We then form the cumulative pdf, the likelihood for locally significant correlations covering an area equal to or smaller than a given areal fraction of the domain. From this cumulative pdf, we obtain the areal fraction of the domain covered by locally significant correlations that is exceeded by only 5% of the population, or 100 out of the 2000 tested realizations. This threshold fraction $f^t$ is our estimate of the minimal fraction of the domain that must be covered by locally significant correlations for requiring the rejection of a null hypothesis of no correlation between the time series and the field (i.e., for the field as a whole to be $p < 0.05$ significant). For the SLPA data used in this study, this fraction is 22%–29%, depending on the calendar month.

### d. Analysis of variance

The ANOVA divides the total variance in $\mathbf{P}_{\mathrm{EM}}$'s forecast into the parts due to the forecast and error. With *df* given by Eq. (A1), the regression-related part has two degrees of freedom (for the two predictors) and the error-related part has $df - 3$. Given $\overline{\mathbf{P}_{\mathrm{EM}}} = 0$, the computed variance ratio is $F^c = [(df - 3)(\mathbf{P}^f_{\mathrm{EM}})^{\mathrm{T}}(\mathbf{P}^f_{\mathrm{EM}})][2(\mathbf{P}_{\mathrm{EM}} - \mathbf{P}^f_{\mathrm{EM}})^{\mathrm{T}}(\mathbf{P}_{\mathrm{EM}} - \mathbf{P}^f_{\mathrm{EM}})]^{-1}$ (Middleton 2000), which can be compared to the cumulative $F(2, df - 3)$ distribution. When $F^c > F(2, df - 3)$, a null hypothesis stating that the fits do not contribute to reduction of rms error is rejected.

### e. Reshuffling

Another test the results (cross-validated forecasts or hindcasts) are subjected to is performed by reshuffling the elements (years) of the $\mathbf{P}_{\mathrm{EM}}$ time series randomly. The entire analysis, beginning with the calculation of the correlation patterns, is performed with the reshuffled $\mathbf{P}_{\mathrm{EM}}$, and the results are stored. This calculation is repeated 1000 times (if the tested result is a cross-validation experiment, the full number of cross-validation withholding experiments is conducted for each member of the reshuffling population). The results are combined into a pdf, and the cumulative pdf's are presented, representing a nonparametric significance test.

## REFERENCES

Baker, C. B., J. K. Eischeid, T. R. Karl, and H. F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. Preprints, *Ninth Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 150–155.

Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7,** 1513–1563.

——, and A. Kumar, 1999: Comment on ''A three-year lagged correlation between the North Atlantic oscillation and winter con-

ditions over the North Pacific and North America.'' *Geophys. Res. Lett.,* **26,** 475–476.
——, and H. M. Van den Dool, 1993: A degeneracy in cross-validated skill in regression-based forecasts. *J. Climate,* **6,** 963–977.
Battisti, D. S., U. S. Bhatt, and M. A. Alexander, 1995: A modeling study of the interannual variability in the North Atlantic Ocean. *J. Climate,* **8,** 3067–3083.
Cullen, H. M., and P. B. deMenocal, 2000: North Atlantic influence on Tigris–Euphrates streamflow. *Int. J. Climatol.,* in press.
Curry, R. G., M. S. McCartney, and T. M. Joyce, 1998: Oceanic transport of subpolar climate signals to mid-depth subtropical waters. *Nature,* **391,** 575–577.
daSilva, A., A. C. Young, and S. Levitus, 1994: *Atlas of Surface Marine Data.* Vol. 1: *Algorithms and Procedures.* NOAA Atlas NESDIS 6, U.S. Department of Commerce, Washington, DC.
Eshel, G., and B. F. Farrell, 2000: Mechanisms of eastern Mediterranean rainfall variability. *J. Atmos. Sci.,* **57,** 3219–3232.
Griffies, S. M., and K. A. Bryan, 1997: A predictability study of simulated North Atlantic multidecadal variability. *Climate Dyn.,* **13,** 459–487.
Hansen, D. V., and H. F. Bezdek, 1996: On the nature of decadal anomalies in North Atlantic sea surface temperature. *J. Geophys. Res.,* **101** (C4), 8749–8758.
Held, I. M., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: Theory. *Large-Scale Dynamical Processes in the Atmosphere,* B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 127–168.
Hulme, M., 1992: A 1951–80 global land precipitation climatology for the evaluation of general circulation models. *Climate Dyn.,* **7,** 57–72.
Hurrell, J. W., 1995: Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation. *Science,* **269,** 676–679.
——, and H. van Loon, 1997: Decadal variations in climate associated with the North Atlantic oscillation. *Climatic Change,* **36,** 301–326.
Johansson, A., A. G. Barnston, S. Saha, and H. M. Van den Dool, 1998: On the level and origin of seasonal forecast skill in northern Europe. *J. Atmos. Sci.,* **55,** 103–127.
Joyce, T. M., C. Deser, and M. A. Spall, 2000: The relation between decadal variability of subtropical mode water and the North Atlantic oscillation. *J. Climate,* **13,** 2550–2569.
Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.* **77,** 437–471.
Kushnir, Y., 1994: Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Climate,* **7,** 141–157.
——, 1999: Europe's winter prospects. *Nature,* **398,** 289–291.
——, and J. M. Wallace, 1989: Low-frequency variability in the Northern Hemisphere winter: Geographical distribution, structure, and time-scale dependence. *J. Atmos. Sci.,* **46,** 3122–3142.
——, and I. M. Held, 1996: Equilibrium atmospheric response to North Atlantic SST anomalies. *J. Climate,* **9,** 1208–1220.
Lamb, P. J., and R. A. Peppler, 1987: North Atlantic oscillation: Concept and an application. *Bull. Amer. Meteor. Soc.,* **68,** 1218–1225.
Latif, M., and T. P. Barnett, 1994: Causes of decadal variability over the North Pacific and North America. *Science,* **266,** 634–637.
——, and ——, 1996: Decadal variability over the North Pacific and North America: Dynamics and predictability. *J. Climate,* **9,** 2407–2423.
Livezy, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte-Carlo techniques. *Mon. Wea. Rev.,* **111,** 46–59.
Middleton, G. V., 2000: *Data Analysis in the Earth Sciences Using MATLAB.* Prentice-Hall, 260 pp.
Molinari, R. L., D. A. Mayer, J. F. Festa, and H. S. Bezdek, 1997: Multiyear variability in the near-surface temperature structure of the midlatitude western North Atlantic Ocean. *J. Geophys. Res.,* **102** (C2), 3267–3278.
Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.,* **119,** 269–298.
Philips, J. L., 1988: *How to Think About Statistics.* Freeman, 198 pp.
Rodwell, M. J., and B. J. Hoskins, 1996: Monsoons and the dynamics of deserts. *Quart. J. Roy. Meteor. Soc.,* **122B,** 1385–1404.
——, D. P. Rowell, and C. K. Follan, 1999: Oceanic forcing of the wintertime North Atlantic oscillation and European climate. *Nature,* **398,** 320–323.
Slutz, R. J., S. J. Lubker, J. D. Hiscox, S. D. Woodruff, R. L. Jenne, P. M. Steurer, and J. D. Elms, 1985: Comprehensive Ocean–Atmosphere Data Set; release 1. Climate Research Program, Boulder, CO. [Available from at http://ingrid.ldeo.columbia.edu/SOURCES/.COADS/]
Sutton, R. T., and M. R. Allen, 1997: Decadal predictability of North Atlantic sea surface temperature and climate. *Nature,* **388,** 563–567.
Taylor, A. H., 1996: North–south shifts of the Gulf Stream: Ocean–atmosphere interactions in the North Atlantic. *Int. J. Climatol.,* **16,** 559–583.
——, and J. A. Stephens, 1998: The North Atlantic oscillation and the latitude of the Gulf Stream. *Tellus,* **50A,** 134–142.
Trenberth, K. E., 1984: Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations. *Mon. Wea. Rev.,* **112,** 2359–2368.
——, and C. J. Guillemot, 1998: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalysis. *Climate Dyn.,* **14,** 213–231.
Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Climate,* **5,** 561–576.
Weng, W., and J. D. Neelin, 1998: On the role of ocean–atmosphere interaction in the midlatitude interdecadal variability. *Geophys. Res. Lett.,* **25,** 167–170.