Finding Storm Track Activity Metrics That Are Highly Correlated with Weather Impacts. Part I: Frameworks for Evaluation and Accumulated Track Activity

ALBERT MAN-WAI YAU AND EDMUND KAR-MAN CHANG

School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York

(Manuscript received 29 May 2020, in final form 10 August 2020)

ABSTRACT: In the midlatitudes, storm tracks give rise to much of the high-impact weather, including precipitation and strong winds. Numerous metrics have been used to quantify storm track activity, but there has not been any systematic evaluation of how well different metrics relate to weather impacts. In this study, two frameworks have been developed to provide such evaluations. The first framework quantifies the maximum one-point correlation between weather impacts at each grid point and the assessed storm track metric. The second makes use of canonical correlation analysis to find the best correlated patterns and uses these patterns to hindcast weather impacts based on storm track metric anomalies using a leave-*N*-out cross-validation approach. These two approaches have been applied to assess multiple Eulerian variances and Lagrangian tracking statistics for Europe, using monthly precipitation and a near-surface high-wind index as the assessment criteria. The results indicate that near-surface storm track metrics generally relate more closely to weather impacts than upper-tropospheric metrics. For Eulerian metrics, synoptic time scale eddy kinetic energy at 850 hPa relates strongly to both precipitation and wind impacts. For Lagrangian metrics, a novel metric, the accumulated track activity (ATA), which combines information from both cyclone track frequency and amplitude, is found to be best correlated with weather impacts when spatially filtered 850-hPa vorticity maxima are used to define cyclones. The leading patterns of variability for ATA are presented, demonstrating that this metric exhibits coherent large-scale month-to-month variations that are highly correlated with variations in the mean flow and weather impacts.

KEYWORDS: Extratropical cyclones; Precipitation; Storm tracks; Climate variability; Wind effects

1. Introduction

Extratropical cyclones and their associated fronts are responsible for much of the high impact and extreme weather over the midlatitudes (e.g., Ma and Chang 2017), including precipitation extremes that can lead to inland flooding (Pfahl and Wernli 2012; Kunkel et al. 2012), high wind events that can cause property damages and casualties (e.g., Ashley and Black 2008; Donat et al. 2010), storm surges that give rise to coastal inundating (e.g., Colle et al. 2008), and extreme cold events (e.g., Kocin et al. 1988). All these adverse weather events have caused significant economic losses and fatalities in the past (e.g., Berko et al. 2014), including multiple events that incurred over U.S. $1 billion of damages (NOAA NCEI 2020).

Apart from their weather impacts, synoptic transient eddies that make up the storm tracks are responsible for much of the poleward momentum, heat, and moisture transport across the midlatitudes (e.g., Peixoto and Oort 1992), helping to maintain and shape Earth's climate. Moreover, variations in storm track activity drive and feed back upon variations in the large-scale circulation, including jet shifts (e.g., Lorenz and Hartmann 2001) and the North Atlantic Oscillation (NAO; e.g., Benedict et al. 2004).

Given its huge impacts on weather and climate, numerous metrics have been developed to quantify storm track activity (e.g., Chang et al. 2002; Hoskins and Hodges 2002). These metrics generally fall into two categories: Eulerian eddy variance/covariance statistics (e.g., Blackmon 1976), and Lagrangian cyclone track statistics (e.g., Klein 1958). These two types of metrics provide different perspectives on storm track variability. Eulerian eddy statistics such as geopotential height variance, eddy kinetic energy, and eddy heat and momentum fluxes relate directly to eddy forcing and feed back to the large-scale flow (e.g., Lorenz and Hartmann 2001), while Lagrangian statistics such as cyclone frequency and amplitude relate more closely to daily synoptic weather situations and are more familiar and intuitive to weather forecasters.

The use of so many different metrics by different scientists can give rise to some confusion, in that different metrics may exhibit variability and change that are not always consistent. As an example, Chang et al. (2012) examined future changes in storm track activity as projected by models participating in phase 5 of the Coupled Model Intercomparison Project (CMIP5) and found that in the Northern Hemisphere, storm track activity as quantified by 300-hPa meridional velocity variance indicates a poleward shift and increase in activity, while based on sea level pressure (SLP) variance statistics a decrease in activity is projected. As another example, Chang (2014) showed that even using the same cyclone tracking algorithm to compile cyclone track statistics, how one defines a cyclone—based either on minima in total SLP [as in Mizuta (2012)] or on spatially filtered SLP that filters out the large-scale features and retains only synoptic-scale features [as in Chang et al. (2012) and Hoskins and Hodges (2002)]—can result in future projections that exhibit opposite signs in the Pacific.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-20-0393.s1.

Corresponding author: Edmund Chang, kar.chang@stonybrook.edu

DOI: 10.1175/JCLI-D-20-0393.1

To alleviate this problem, one might argue that one should always examine multiple storm track metrics as done by Hoskins and Hodges (2002), who examined Eulerian and Lagrangian storm track statistics computed based on 12 different variables. In several recent studies we have indeed examined variability and changes found in multiple storm track metrics (e.g., Chang et al. 2012; Chang and Yau 2016). Nevertheless, this is not always practical—even Hoskins and Hodges (2002) could only display selected statistics from some of these variables—and might provide conflicting information, which may be confusing as in the situation described by Chang (2014).

An alternative strategy is to focus on a small number of metrics that have been verified to be highly relevant to one's application. For example, for application to wave–mean flow interactions, one might focus on Eulerian variance and covariance statistics such as eddy heat and momentum fluxes. On the other hand, if the focus is on the weather impacts of storm tracks, one might want to focus on metrics that are highly correlated to variations in precipitation or frequency of high wind events. For example, Chang et al. (2015) showed that observed variability and projected changes in winter precipitation in California are highly correlated with variability in an Eulerian metric defined by bandpass filtered SLP variance over eastern Pacific. Ma and Chang (2017) found that variability in the same metric over North America highly modulates the frequency of occurrence of extreme cold, extreme wind, and extreme precipitation events over the Ohio Valley and the Pacific Northwest. Paciorek et al. (2002) showed that the frequency of strong cyclones is significantly correlated (pointwise) to precipitation and the frequency of strong winds in some regions. Nevertheless, the correlations that Paciorek et al. (2002) found are relatively modest in most locations.

While these studies have found metrics that are significantly correlated with weather impacts, they have only examined a small number of metrics, and thus there are likely other metrics that correlate better with weather impacts. In this study, we propose two frameworks to systematically evaluate different storm track metrics based on their correlation with weather impacts including precipitation and occurrence of strong winds. The first framework involves finding the maximum point-by-point correlation between weather impacts and any storm track metric. Since the relationship between storm track activity and its weather impacts is not necessarily pointwise, we developed an alternative framework using canonical correlation analysis (CCA; see Barnett and Preisendorfer 1987; Chang and Fu 2003) to find the most highly correlated pairs of patterns. In addition, cross-validation is utilized to avoid overfitting due to the limited number of degrees of freedom in time. We have applied these two frameworks to assess multiple Eulerian and Lagrangian metrics and found several metrics that are better correlated with weather impacts than those described above.

Among Eulerian metrics, we found that eddy kinetic energy (EKE) at the 850-hPa level is in most regions better correlated with weather impacts than SLP variance, while variance of pressure vertical velocity ω at 700 hPa is best correlated with precipitation. For Lagrangian metrics, we developed a novel metric—accumulated track activity (ATA)—that combines cyclone track density and amplitude (based on tracking 850-hPa vorticity maxima) together that performs better than either cyclone track density or cyclone amplitude and is less noisy. These, and other, metrics will be described in more detail in section 2. The frameworks we developed will be described in section 3. In this study, we will apply these frameworks to assess the relationship between storm track activity and weather impacts over Europe as quantified by multiple metrics, and the results will be presented in section 4. In section 5, the leading patterns of storm track variability over Europe as quantified by ATA, and their relationship to weather impacts and large-scale circulation, will be described. A summary and discussion will be presented in section 6.

2. Data and metrics

a. Data

The main data source for computing storm track metrics is ERA-Interim reanalysis data (Dee et al. 2011), for the winter months (December–February) from 1979/80 to 2014/15, a total of 108 months of data. Since we are mainly assessing continental-scale relationships, we use 6-hourly data at a horizontal resolution of 2.5° × 2.5°. For more regional-scale analyses higher-resolution data should be used. Variables used include SLP, zonal wind u, meridional wind v, relative vorticity ζ, ω, and geopotential height z at several pressure levels. For weather impacts, the monthly precipitation data from the Global Precipitation Climatology Project (GPCP; Adler et al. 2003) are used. Over continental regions, GPCP data are based on gauge observations, while over oceans satellite data are used. GPCP data are available at a resolution of 2.5° × 2.5°, and this is also the main reason why this horizontal resolution is chosen for our analyses. For wind impacts, in this study we make use of a high-wind index derived from the near-surface 10-m wind product from ERA-Interim (see below).

b. Metrics

1) EULERIAN STORM TRACK METRICS

Many different Eulerian storm track metrics have been used to quantify storm track activity. Blackmon (1976) examined bandpass filtered 500-hPa geopotential height (z500), but many other variance and covariance statistics have been used since that time (e.g., Chang et al. 2002; Hoskins and Hodges 2002). Our previous studies have shown that SLP variance is highly correlated with precipitation and extreme wind events over parts of North America (Chang et al. 2015; Ma and Chang 2017). In this study, we will discuss evaluations of nine different variance statistics, namely SLP, z500, v at 850 hPa (v850) and 300 hPa (v300), eddy kinetic energy at 850 hPa (EKE850) and 300 hPa (EKE300), ζ at 850 hPa (ζ850) and 300 hPa (ζ300), and ω at 700 hPa (ω700).

To extract synoptic time-scale variance statistics, we use the simple 24-h difference filter introduced by Wallace et al. (1988), as follows:

$$\mathrm{Var(SLP)} = \overline{[\mathrm{SLP}(t + 24\,\mathrm{h}) - \mathrm{SLP}(t)]^2}. \quad (1)$$

In (1), the overbar denotes a time average, in this case over each month. As discussed in Wallace et al. (1988), (1) can be

considered a bandpass filter with half power at 1.2 and 6 days, thus highlighting synoptic time scale variability. Note that 24-h SLP change statistics have been widely used to quantify "storminess" (e.g., Feser et al. 2015; Alexander et al. 2005). We have also tested variance computed from several alternative bandpass filters, including filters constructed similar to those used by Blackmon (1976) for highlighting 1–6-, 2–8-, and 1–10-day bands. However, in our limited tests we have not found a filter that improves the correlations over those provided by the simple 24-h difference filter, and thus we have used this filter for all variables.

Note that

$$\mathrm{EKE850} = 1/2[\mathrm{Var(u850)} + \mathrm{Var(v850)}] \quad (2)$$

with the variances as defined by (1) filtered using the 24-h difference filter.

2) Lagrangian storm track metrics

To compile Lagrangian track statistics, we use the feature tracker developed by Hodges (1994, 1999) and used by Hoskins and Hodges (2002). We have used this tracker in several previous studies, including Chang et al. (2012), Chang (2014), and Chang and Yau (2016). In this study, this tracker has been applied to track minima in SLP, maxima in ζ850 and ζ300, and minima in ω700 (following centers of rising motion). The tracked variables are summarized in Table 1. Hoskins and Hodges (2002) applied this tracker by first spatially filtering the data to retain only scales equivalent to T5–T42, and we use the same truncation for ζ850, ζ300, and ω700. For SLP, given the results of Chang (2014), who showed that tracking statistics for SLP are sensitive to removal of large-scale low-frequency background, we define cyclones by two different methods, as follows. The first way (SLP1) is spatially filtered similar to that of Hoskins and Hodges (2002), but with a truncation of T5–T70. A slightly higher wavenumber cutoff of T70 is used for SLP because SLP data are not as noisy as vorticity. For the second way, as in Chang and Yau (2016), based on the suggestion of Donohoe and Battisti (2009), the SLP data are filtered both spatially and temporally, with the low-frequency monthly mean SLP first removed to give submonthly transients, and the remaining perturbations are filtered to retain only T5–T70 (SLP2). Tracking is done for all five variables for all winter months and statistics compiled. As in Hoskins and Hodges (2002), tracks lasting less than 2 days and with displacement of less than 1000 km are removed from the statistics.

Multiple statistics can be derived from the tracking output. Following Guo et al. (2017; see also Grise et al. 2013; Sinclair 1997), cyclone frequency at a grid point is defined as the number of times a cyclone is found within 500 km of the grid point each month. This is similar to the feature density of Hoskins and Hodges (2002). In this definition, slow-moving cyclones can be counted more than once, and fast-moving ones that pass close to a grid point may be missed if they are never within 500 km of the grid point during the four synoptic hours. Cyclone amplitude for each month is the average amplitude of all cyclone centers that are found within the same distance of the grid point. On the other hand, track frequency (similar to track density of Hoskins and Hodges) is defined as the number of cyclone tracks that pass within 500 km of each grid point. In this case each cyclone is counted once and only once. We have also examined track amplitude (average of the maximum amplitude when a cyclone is within 500 km of each grid point) and results are similar to those for cyclone amplitude and will not be discussed here.

Table 1. Summary of tracked features and Lagrangian metrics.

	Tracked features
SLP1	SLP minima spatially filtered to retain T5–T70
SLP2	Similar to SLP1, but with monthly mean removed
ζ850	ζ850 maxima spatially filtered to retain T5–T42
ζ300	ζ300 maxima spatially filtered to retain T5–T42
ω700	ω700 minima (rising motion) spatially filtered to retain T5–T42
	Metrics
Cyclone amplitude	Average value of the amplitude of all cyclone centers that are located within 500 km of each grid point at the 6-hourly synoptic times for the month
Cyclone frequency	Number of times per month when a cyclone center is located within 500 km of each grid point at the 6-hourly synoptic times (each cyclone can be counted multiple times)
Track frequency	Number of times per month when a cyclone track passes within 500 km of each grid point (each cyclone is counted once)
ACA	Accumulated cyclone activity: The sum of the amplitude of all cyclone centers that are located within 500 km of each grid point at the 6-hourly synoptic times (each cyclone can be counted multiple times)
ATA	Accumulated track activity: The sum of the maximum amplitude of every cyclone that passes within 500 km of each grid point (each cyclone is counted once)

Guo et al. (2017) examined how the MJO phases modulated cyclone activity using both Eulerian and Lagrangian storm track metrics. They found that anomalies in Lagrangian statistics including cyclone frequency and amplitude are generally noisier than those for Eulerian statistics. However, they found that by combining cyclone frequency and amplitude into a new metric, which they called the "accumulated cyclone activity" (ACA), the anomalies became less noisy and more like those found in SLP variance statistics. Inspired by their results, we also computed ACA by accumulating the amplitude of all cyclone centers that are found within 500 km of a grid point for each month. As with cyclone frequency, this statistic can be strongly dominated by slow-moving cyclones staying close to a grid box for extended periods of time. Hence we developed an

alternative novel metric, which we call the ''accumulated track activity'' (ATA). For this case, each cyclone (or feature) passing within 500 km of a grid point is only counted once, using its maximum amplitude during the time when the cyclone is within 500 km of the grid point. This amplitude is determined by linearly interpolating the 6-hourly cyclone track positions and amplitudes into hourly values. For each of the five definitions of cyclones, we evaluated five statistics: cyclone frequency, cyclone amplitude, track frequency, ACA, and ATA. Hence we have evaluated a total of 25 different Lagrangian statistics. Each of these statistics is computed on a hemispheric 2.5° × 2.5° grid. More details about these Lagrangian statistics will be presented in section 5 below.

Previous studies (e.g., Neu et al. 2013) have found that tracking statistics can be sensitive to the tracking algorithm used. Here, we tested the sensitivity by using a very simple tracker that we have developed ourselves, simply by linking together features that are closest to each other and within 500 km of each other over consecutive 6-hourly periods. Similar statistics are compiled for tracking ζ850 and SLP1, and the results are compared to those generated using the Hodges tracking algorithm. Statistics generated by our simplistic tracker produced qualitatively similar results, providing similar rankings of the tracking metrics as those generated by the Hodges algorithm, except that the correlations between the storm track metrics and weather impacts are systematically lower (see discussions and figures in the online supplemental material). While the results show some sensitivity to the tracker, it is comforting to see that statistics generated using a more sophisticated tracker generally outperform those generated by a simple tracker, and that the main conclusions are supported by results provided by the simple tracker. We have also tested the sensitivity of the results to the distance used to accumulate cyclone statistics, and found that if we either increase or decrease the distance (from 500 km) by 10%, the resulting correlations between the Lagrangian storm track metrics and weather impacts are practically unchanged.

3) METRICS FOR WEATHER IMPACTS

As mentioned above, for precipitation, monthly GPCP precipitation data are used. We have also used ERA-Interim model generated precipitation product for comparison (not shown) and generally obtained qualitatively similar results, except that correlations between the storm track metrics and ERA-Interim precipitation are generally higher than those with GPCP, especially over the oceans. We assessed monthly precipitation data here because we would like to use an observational dataset that has global coverage. The methods discussed in this paper will be applied to examine precipitation extremes in future studies.

For wind impacts, instead of a fixed wind speed threshold for everywhere, previous studies (e.g., Paciorek et al. 2002; Alexandersson et al. 1998) have suggested that a fixed percentile at each location may be more useful, since building structures will generally be built to adapt to the local climatological wind conditions. Here, we use the value of the 95th percentile wind speed for each month based on ERA-Interim 10-m wind at each grid point to be our strong wind index. As suggested by Paciorek et al. (2002), this is not intended to represent the true wind, but is used as an indicator that strong wind conditions are expected when this index is large. In practice, the sixth strongest wind speed from each month is taken to be the high wind index for that month. Given that anomalies are computed as deviations from monthly climatology, the difference in the number of observations for each month is not expected to give rise to significant biases.

To test the sensitivity to the weather impacts dataset, we have also used monthly precipitation and 6-hourly 10-m winds generated by the National Center of Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data for some of our analyses. Results are very similar to those obtained using ERA-Interim data and will not be presented here.

3. Frameworks for assessing metrics

a. *Maximum correlation*

As discussed above, Chang et al. (2015) found that month-to-month precipitation variations over Southern California are highly correlated to those of Var(SLP) over the eastern Pacific. Here the correlation between monthly mean precipitation anomalies at a grid box over Ireland (53.75°N, 8.75°W) with Var(SLP) at all grid boxes over eastern Atlantic and Europe is shown in Fig. 1a. The maximum correlation is over 0.7 located just to the northwest of the grid box. Similar correlations are shown for a grid box over southern France (43.75°N, 6.25°E) in Fig. 1b. For this grid box, the maximum correlation is slightly lower (<0.7) and with storm track activity located about 1000 km west-southwest of the grid box. This location is physically reasonable, since for a grid box on the south coast of France, cyclones located to its west would lead to southerly flow toward this location, bringing in moisture over the Mediterranean (e.g., Nuissier et al. 2011).

Different relationships are found at different geographical locations. To summarize results for the entire Northern Hemisphere, similar correlation maps are computed for each grid box, and the highest correlation between precipitation anomalies and Var(SLP) within a region of 60° longitude and 20° latitude centered about the grid box is found and plotted at the location of the grid box (Fig. 2a). This map shows the locations where precipitation anomalies are highly correlated with Var(SLP) nearby. We call this maximum correlation score CORMAX below. Conceptually, this is similar to the teleconnectivity map of Wallace and Gutzler (1981), but here we apply this concept to correlate two different variables and find the maximum positive correlation instead of the minimum negative correlation. In general, highest CORMAX (>0.6 or 0.7) is found over the eastern parts of ocean basins extending into the western parts of North America and Europe. CORMAX is also generally high (>0.5) over much of Europe, northeastern North America, the east coast of Asia, and parts of the Middle East and central Asia. CORMAX over oceanic regions is generally lower. However, higher CORMAX is found if we use ERA-Interim reanalysis generated precipitation data instead of GPCP precipitation. Overall, the average value of CORMAX is 0.45 when averaged over the Northern

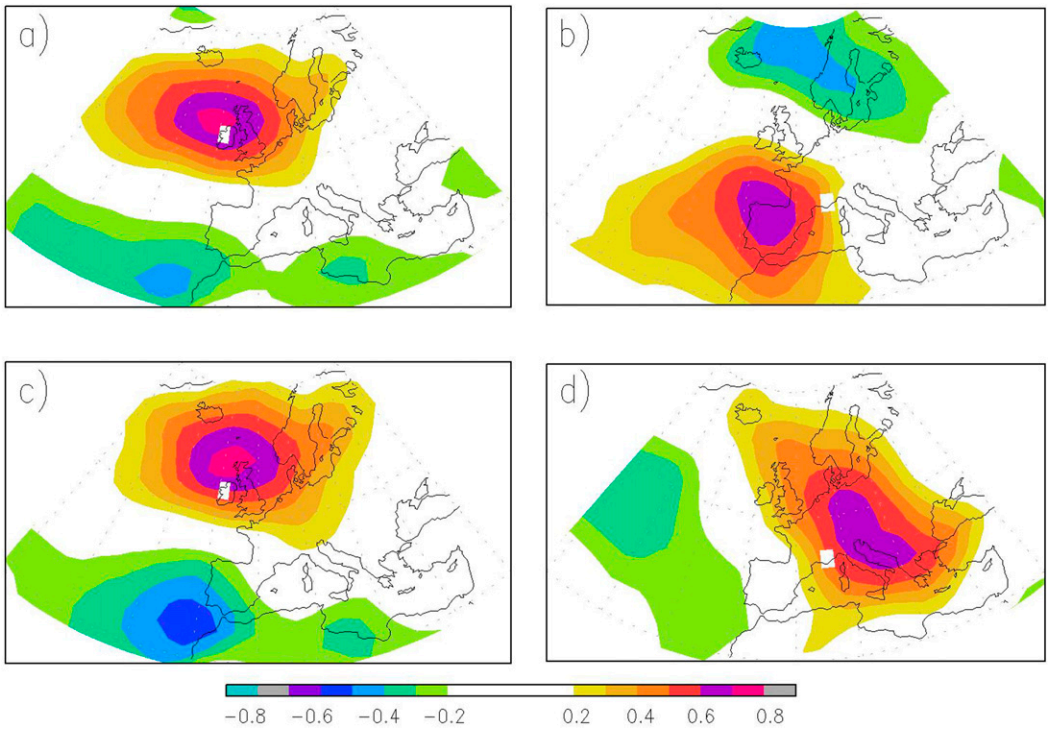

FIG. 1. One point correlation map, between the reference time series and Var(SLP). The reference time series for each panel is (a) GPCP precipitation at 53.75°N, 8.75°W (marked by white grid in the panel; same in the other panels); (b) GPCP precipitation at 43.75°N, 6.25°E; (c) high wind index at 53.75°N, 8.75°W; and (d) high wind index at 43.75°N, 6.25°E.

Hemisphere midlatitudes (30°–75°N), and 0.55 over Europe (10°W–45°E, 35°–70°N). These values increase to 0.53 and 0.59, respectively, if ERA-Interim precipitation is used instead. If NCEP–NCAR reanalysis precipitation is used instead [with Var(SLP) still computed using ERA-Interim data], the averaged CORMAX are 0.51 and 0.57, suggesting that CORMAX is not very sensitive to the reanalysis dataset used.

Similar correlations have been computed using the monthly high wind index (95th-percentile wind speed based on ERA-Interim 10-m wind for each month). Figure 1c shows the correlation of this index for the grid box over Ireland with Var(SLP). Again, the maximum correlation is over 0.7, but for a location just to the north-northeast of the grid box. For the grid box over southern France (Fig. 1d), the maximum correlation of its high wind index is with storm track activity located east of the grid box, suggesting that over this location high winds are more frequent when storms are to the east (e.g., Jiang et al. 2003; see also Pfahl 2014).

The CORMAX between Var(SLP) and the high wind index at all grid boxes are shown in Fig. 2b. Overall, CORMAX is high over much of the ocean basins, with large areas reaching values over 0.7. The average value over the Northern Hemisphere midlatitudes is 0.58, which increases to 0.63 over Europe. Hence Var(SLP) is more strongly correlated to high winds than precipitation. Note that we have also computed this high wind index using NCEP–NCAR reanalysis data and obtained nearly identical averaged correlations.

Paciorek et al. (2002) computed similar correlations between their cyclone count index and their wind index. Their cyclone count index is a nontracking index and counts the number of cyclones per winter at each grid point. Their extreme wind index uses the 95th percentile of four times daily wind speeds at the $\sigma = 0.995$ level (~40 m above the surface) based on NCEP–NCAR reanalysis data. They computed one-point correlation between their cyclone index and extreme wind index, and found relatively modest correlations, with only small regions showing correlations above 0.5. One difference between their method and ours is that they used a fixed displacement between the two fields at all locations to compute the correlations, while we searched for the maximum correlation between the wind index at a grid box and the storm track metric. Nevertheless, if we use a similar cyclone count index based on cyclone frequency for tracking SLP1 instead of Var(SLP), the average value of CORMAX becomes 0.39 for the Northern Hemisphere and 0.42 for Europe, much lower than those based on Var(SLP). Hence Var(SLP) correlates much better with wind impacts than cyclone frequency. The question is whether there are other Lagrangian metrics that correlate better with high wind impacts than cyclone frequency.

Maps similar to Fig. 2 can be generated for any storm track metric and provide one way to assess the relationship between each metric and weather impacts. Nevertheless, it is not clear whether the estimates based on this method represent

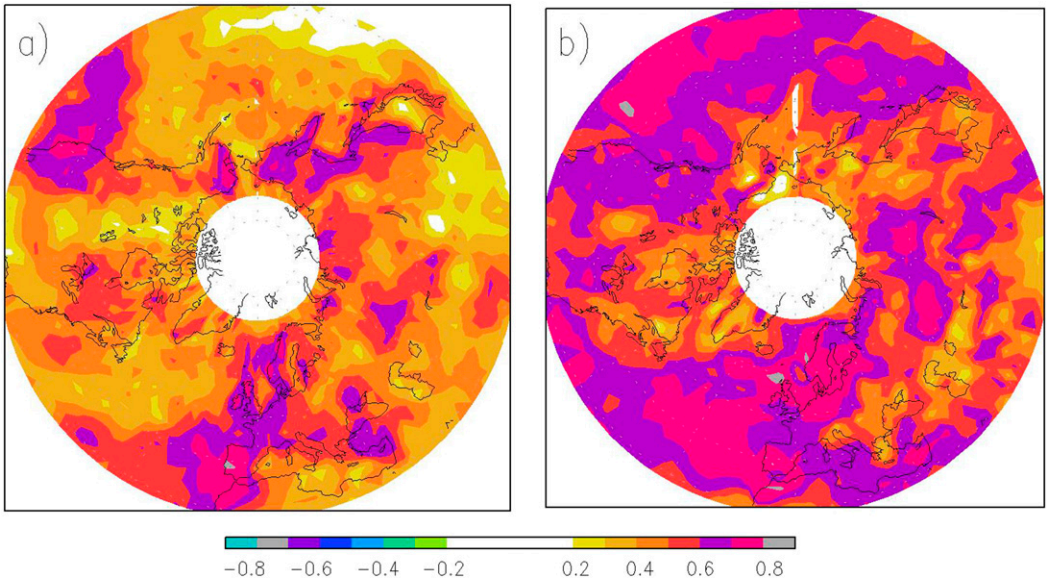

FIG. 2. CORMAX score (see description in text) at each grid box for correlating (a) GPCP precipitation at the grid box to Var(SLP) and (b) high wind index at the grid box to Var(SLP).

conservative guesses or overestimates. On the one hand, since the maximum correlation at each grid point is selected, one might expect that CORMAX represents the upper bound for the correlation between storm track activity and the weather impact and thus potentially an overestimate. Nevertheless, it is possible that precipitation (or high wind) at a grid box may be related to cyclone activity at two (or more) separate locations, hence different areas of cyclone activity may contribute to weather impacts over a single location, with the total storm track impact potentially explaining more of the weather impact than that by the highest correlated grid point alone. Thus an independent estimate based on CCA has been developed to provide an alternative assessment.

b. *CCA assessment framework*

Figure 1 shows that the relationship between Var(SLP) (and other storm track metrics) with weather impacts is not pointwise. Hence we would like to assess correlations between pairs of patterns. One efficient way to do so is CCA. However, since we only have 108 months of data, the number of temporal degrees of freedom is likely less than the number of spatial degrees of freedom, hence there could be spurious remote correlations and this could lead to overfitting (Barnett and Preisendorfer 1987). One way to deal with this is by EOF prefiltering (i.e., performing an EOF truncation prior to conducting the CCA analysis). Nevertheless, it is not always clear what the correct truncation should be, given that the CCA-generated correlations increase monotonically when more EOF pairs are included in the analysis. To overcome this overfitting issue, we conduct a leave-N-out cross-validation analysis ($N = 3$ in our application) to obtain a conservative estimate of the correlation between the two fields.

The procedure is outlined in Fig. 3 for assessing precipitation. A similar procedure is separately conducted for assessing wind impacts. First, a number M, where $0 < M < 36$ (the upper bound is arbitrary as long as it is sufficiently large, but taken to be 35 here), is selected. For each storm track metric [e.g., Var(SLP)] and precipitation, we have 108 pairs of data, one pair for each month. A CCA model is first constructed using 105 pairs of data, leaving out 3 pairs (one season) of data each time. Two EOF analyses are conducted separately for the storm track metric and precipitation. Then both datasets are truncated to retain M leading EOFs. A CCA model is then constructed using these M EOFs following the procedure given in Barnett and Preisendorfer (1987; see also Chang and Fu 2003). Briefly, an $M \times M$ cross-correlation matrix is formed between the leading M PCs, and a singular value decomposition analysis is conducted on this matrix—the resulting M singular values are the canonical correlation coefficients, while the M left and right vector pairs correspond to pairs of patterns whose time series have correlations that are equal to the canonical correlation coefficients. This CCA model is then used to predict precipitation anomalies for the left out 3 months based on the storm track metric anomalous patterns projected onto the M EOFs. This procedure is repeated 35 times, each time leaving out three other pairs (another season) of data to construct a new CCA model. The resultant 108 predicted precipitation fields are then assessed using the observed precipitation anomaly fields. The anomaly correlation coefficient (ACC) is computed at each grid point. ACC is computed at each grid box and averaged over the domain to provide a domain averaged "score." In addition, the fraction of variance explained (FVE) is estimated, as follows:

$$\mathrm{FVE} = 1 - \frac{\mathrm{MSE}}{\mathrm{VAR}}, \tag{3}$$

where MSE is the mean square error between the predicted and observed precipitation, and VAR is the variance of observed

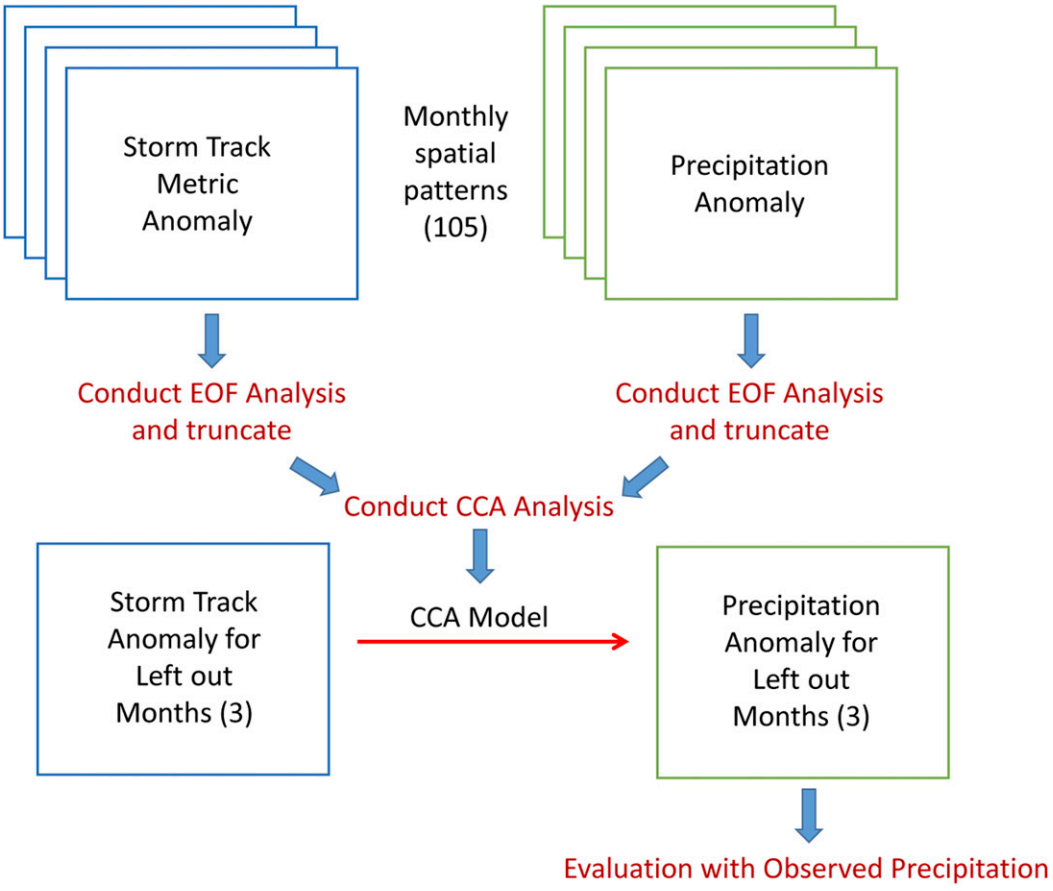

FIG. 3. Schematic of CCA assessment framework (see description in text).

precipitation at each grid point. The value of FVE is capped by 1 and can be negative, with 1 being a perfect prediction. For FVE, we also show its value at individual grid boxes. However, for the domain averaged "score," the MSE and VAR are first averaged over the domain, and FVE for the entire domain is calculated based on (3). Thus these two scores highlight different aspects: the ACC score summarizes the results for all grid boxes without weighing by the precipitation amounts, while the FVE score highlights regions where precipitation variance is large. This entire procedure is then repeated for other values of M, with $0 < M < 36$. The model with the highest averaged score [domain averaged (ACC + FVE)/2] is considered the best fit model, and provides the estimate of the correlation between this particular storm track metric and the weather impact. Practically, this procedure informs us how well we can predict the weather impact if we know the anomalies in this storm track metric.

To illustrate this procedure, we apply this framework to assess the relationship between Var(SLP) and GPCP precipitation. We have tried to apply this framework over the entire hemisphere, but found that regional applications (e.g., over Europe, the Atlantic, the Pacific, etc.) produced better results, likely because that with few (108) temporal degrees of freedom, there are spurious long distance correlations found among the leading hemispheric EOFs that tend to degrade the predictions, and some regional precipitation or storm track variations could be relegated into high-order EOFs and thus be filtered out. Since we are assessing impacts, we would like to first apply this over continental regions. Figure 2 suggests that among the different continental regions, Var(SLP) and precipitation appear to be most tightly related over Europe, and hence in this study we will first apply this framework to Europe. The procedure is applied to the area 20°W–50°E, 25°–80°N, with the results assessed in the subregion 10°W–45°E, 35°–70°N since storm track activity adjacent to the region of interest may still impact the weather over the region (see Fig. 1).

For the relationship between Var(SLP) and GPCP precipitation, both the ACC and FVE scores are maximized at $M = 13$, with ACC = 0.55, FVE = 0.39, and a mean score of 0.47. The dependence of all three scores on M is shown in Fig. 4a. The results are not sensitive to the truncation M; the average score is higher than 0.46 for all M from 9 to 14. This "flat peak" is a shared feature of all the cases we have analyzed. These scores increase rapidly from $M = 1$ to $M = 5$, and then increase slowly to the peak values, after which they slowly decrease likely due to overfitting. Note that the maximum ACC score of 0.55 is very similar to the value of CORMAX found in section 3a for the same region. This suggests that the correlation between the storm track and weather metrics consist of spatially coherent patterns that can be effectively captured by CCA.

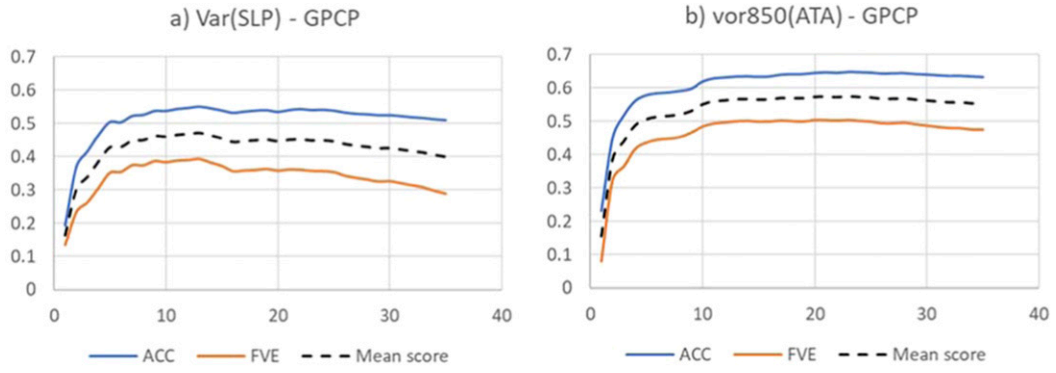

FIG. 4. Dependence of the ACC and FVE scores on the EOF truncation value M (abscissa), (a) using Var(SLP) to hindcast GPCP precipitation and (b) using ATA from tracking ζ850 to hindcast GPCP precipitation.

The spatial patterns for ACC and FVE for $M = 13$ are shown in Figs. 5b and 5c, respectively. Both scores are highest over the Iberian Peninsula, the United Kingdom, and the north shore of France spreading northeastward toward Scandinavia, with ACC scores of generally >0.6 reaching 0.8 in some isolated locations, and explaining up to 60% of the month-to-month precipitation variance. These regions are consistent with the regions with high CORMAX (Fig. 5a). However, there are some regions (e.g., the northern part of France extending into Belgium and the Netherlands) where the ACC is higher than the CORMAX, suggesting that the CORMAX is not necessarily the upper limit of the correlation between Var(SLP) and precipitation.

In section 4, the two frameworks described in this section are applied to other metrics as well as to the high wind index to assess how closely each metric relates to these weather impacts.

4. Results

a. *Eulerian metrics*

The summary statistics (scores averaged over Europe) for the nine Eulerian metrics are presented in Table 2. Results are generally consistent between the two frameworks. For the CCA framework, results based on ACC and FVE are also highly consistent. First, we examine the results for precipitation. Among the metrics evaluated, ω700 is best correlated with precipitation (Table 2). The average CORMAX is 0.67 over Europe using the first framework, while for the CCA framework the average ACC score is 0.65, and it explains 50% (FVE = 0.50) of the variance in the month-to-month precipitation variations over Europe. EKE850 and Var(ζ850) are close seconds, with Var(ζ850) performing slightly better in terms of CORMAX, but EKE850 slightly better in terms of FVE.

Comparing the four pairs of upper-level and near-surface metrics, in all cases near-surface or lower tropospheric metrics perform better than upper-level metrics. Also, at both levels, EKE performs slightly better than the variance of v'. These results also hold for wind impacts.

The geographical distribution of the three scores for EKE850 and Var(ω700) are shown on the middle and right columns of Fig. 5, respectively. All three metrics [including Var(SLP)] generally exhibit high scores over the same locations, with the scores increasing from Var(SLP) to EKE850 to Var(ω700). In particular, over the southwestern part of the Iberian Peninsula, the western part of the British Isles, parts of Denmark, and the west coast of Scandinavia, both EKE850 and Var(ω700) explain over 60% of the month-to-month precipitation variance (Figs. 5f,i).

Results for wind impacts are shown on the right four columns of Table 2 and Fig. 6. For wind impacts, EKE850 performs the best, closely followed by Var(SLP), with Var(ω700) also performing well (Table 2). It is not surprising that EKE850 and Var(SLP) perform well on wind impacts, since surface winds are closely related to lower tropospheric EKE as well as SLP variations through the geostrophic relationship.

The geographical distributions of the three scores for the three metrics are shown in Fig. 6. Again, they generally perform well over similar regions. Both EKE850 and Var(SLP) exhibit >0.7 ACC scores and explain over 50% of the wind index variance over the northern part of continental Europe, the British Isles, and coastal and southern parts of Scandinavia. All three do not perform well over southeastern Europe or Turkey, explaining less than 30% of the variance in these places.

Overall, among variance metrics, Var(ω700) performs best for precipitation, while doing third best for wind impacts. EKE850 does best for wind impacts and is second for precipitation. Var(SLP) performs nearly as well as EKE850 for wind impacts but is not as good for precipitation. Given that ω is usually not readily available from climate models, EKE850 may be a good storm track metric to analyze if one's goal is to link storm track variations to their weather impacts.

b. *Lagrangian metrics*

Summary statistics for selected metrics are shown in Table 3. Summary statistics for all 25 Lagrangian metrics examined are shown in Table S1 in the online supplemental material.

In all cases, track statistics perform better than cyclone (feature) statistics, with track frequency doing better than cyclone frequency, and ATA doing better than ACA in all cases. ATA also always does better than track frequency, and ACA better than cyclone frequency, showing that including information about amplitude is helpful. ATA always performs best among the five Lagrangian metrics. Similar to variance statistics, near-surface metrics perform better than upper-level

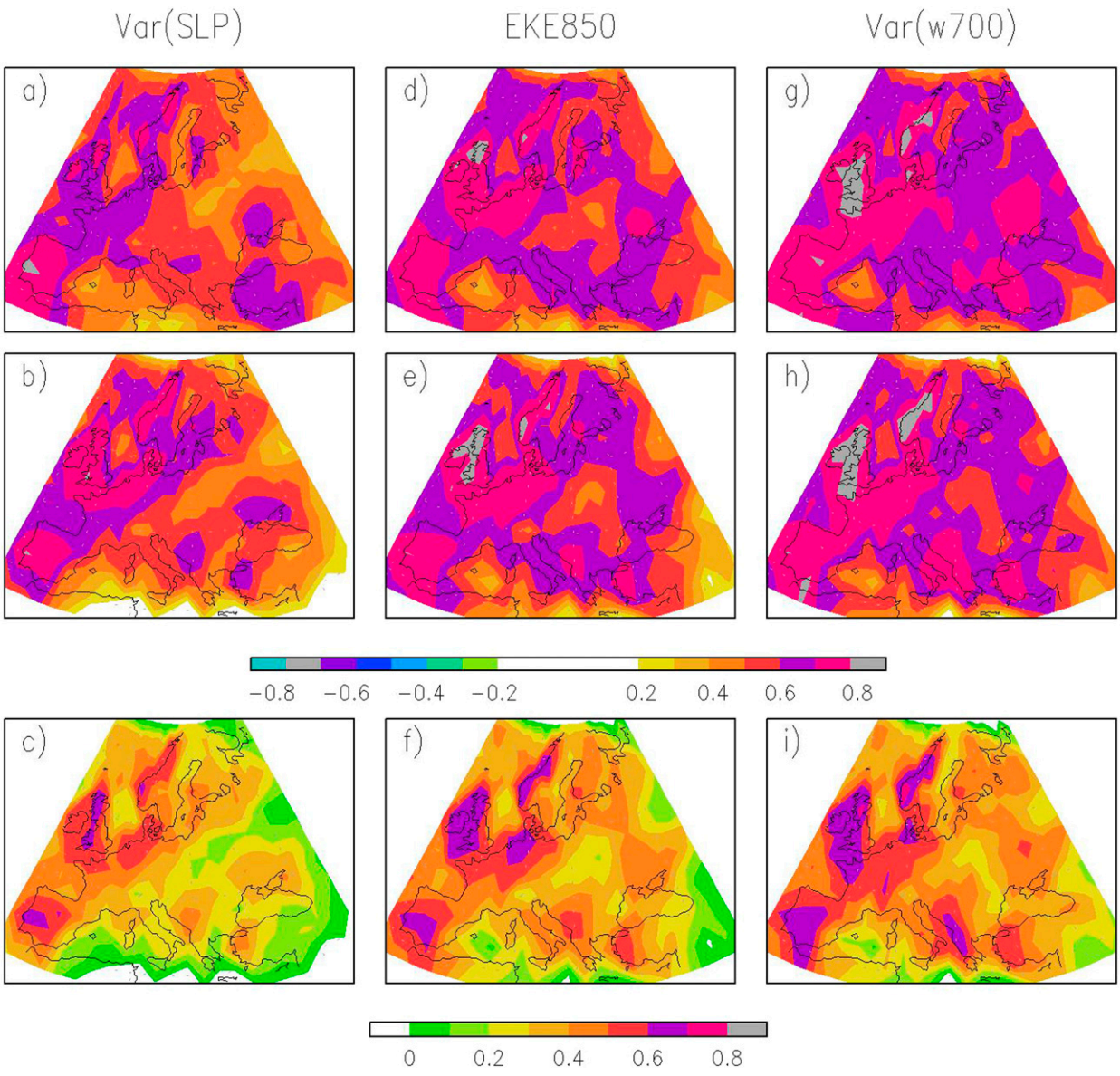

FIG. 5. Results of applying the two frameworks to assess GPCP precipitation for Eulerian metrics: (top) CORMAX score, (middle) ACC score from the CCA framework, and (bottom) FVE score from the CCA framework. Metric assessed is (a)–(c) Var(SLP), (d)–(f) EKE850, and (g)–(i) Var(ω700).

metrics, with ζ850 outperforming ζ300. For precipitation, cyclone amplitude performs poorly, while track frequency performs only slightly worse than ATA, suggesting that it is the frequency of cyclones passing close to a region that controls the monthly precipitation, while cyclone amplitude is secondary, consistent with the results of Pfahl and Wernli (2012) that extreme precipitation is only weakly related to cyclone strength. However, for the wind index, cyclone amplitude performs better than cyclone or track frequency, suggesting that stronger cyclones give rise to stronger wind impacts which is physically reasonable.

Overall, ATA for ζ850 tracks performs best among all 25 metrics for both precipitation and high wind index, with the ACC (0.65) and FVE (50%) scores for precipitation identical to those for Var(ω700), the best Eulerian metric. However, the results for high winds are not as good as those for the leading Eulerian metrics including EKE850 and Var(SLP), with the ACC score being 0.54 and FVE of 36% when averaged over Europe.

Among the two different ways of defining cyclones based on SLP, SLP2 (spatial filtering retaining T5–70, and temporal filtering by subtracting monthly mean) performs much worse than SLP1 (spatial filtering alone with no temporal filtering) in all cases (Table 3; see also Table S1). This suggests that removing the low-frequency component by removing the monthly mean is not beneficial. Nevertheless, more tests should be conducted to examine whether other ways of removing the

TABLE 2. Summary statistics for Eulerian metrics. Maximum value in each column is highlighted by bold fonts.

	GPCP precipitation				High wind index			
		CCA				CCA		
Metric	CORMAX	*M*	ACC	FVE	CORMAX	*M*	ACC	FVE
Var(SLP)	0.55	13	0.55	0.39	0.63	14	0.58	**0.41**
Var(z500)	0.50	13	0.54	0.37	0.55	10	0.53	0.33
Var(v850)	0.61	15	0.61	0.45	0.62	13	0.57	0.38
Var(v300)	0.52	13	0.58	0.41	0.49	10	0.52	0.32
EKE850	0.62	15	0.62	0.46	**0.65**	12	**0.59**	**0.41**
EKE300	0.49	13	0.57	0.40	0.51	10	0.52	0.32
Var(ζ850)	0.63	24	0.62	0.45	0.58	24	0.55	0.36
Var(ζ300)	0.44	12	0.52	0.34	0.44	10	0.47	0.26
Var(ω700)	**0.67**	17	**0.65**	**0.50**	0.60	13	0.58	0.38

low-frequency variability may fare better. Note that tracking for the other three variables (ζ850, ζ300, and ω700) examined in this study is all performed without temporal filtering (similar to SLP1, except that the spatial truncation retains T5–42 instead).

Spatial distribution of the scores for ATA of SLP1, ζ850, and ω700 based on the best truncation models are shown in Fig. 7. Given that the three scores (CORMAX, ACC, and FVE) are highly consistent, only the results for FVE are shown.

In Fig. 7, the top row shows the results for precipitation. Clearly, ATA based on ζ850 (Fig. 7b) performs better than that based on either SLP1 (Fig. 7a) or ω700 (Fig. 7c). Compared to Var(ω700) (Fig. 5i), results are quite similar, except that ATA based on ζ850 does slightly better over the Iberian Peninsula and north of the Black Sea, while Var(ω700) performs better near the Adriatic Sea and close to France. For the high wind index (bottom row of Fig. 7), again ATA based on ζ850 (Fig. 7e) performs better than ATA based on SLP1 (Fig. 7d) or ω700 (Fig. 7f). Comparing to variance statistics (bottom row of Fig. 6), the values of FVE are generally slightly lower, but all metrics display qualitatively similar spatial patterns, with highest scores toward the northwest part of the map and lowest scores toward the southeast. Note that as discussed above, the results are not sensitive to the value of the truncation *M*. For example, the dependence of ACC and FVE on the value of *M* for ATA of ζ850 is shown in Fig. 4b, again showing a very flat "peak," even flatter than that for Var(SLP) (Fig. 4a).

Finally, in Fig. 8, we display results based on other Lagrangian metrics for ζ850, including those based on cyclone amplitude (AMP), track frequency (TRF), and ACA. For precipitation (top row), cyclone amplitude (Fig. 8a) performs poorly, with only limited areas with FVE higher than 40%—over the Iberian Peninsula and near Denmark. ACA (Fig. 8c) does the best among the three. Track frequency (Fig. 8b) performs quite well also, but neither does as well as ATA (Fig. 7b). For high wind index (bottom row), track frequency does not do well (Fig. 8e), with track amplitude now doing better (Fig. 8d) and performing nearly as well as ACA (Fig. 8f). Again, both fall a bit short of the performance of ATA (Fig. 7e). Figure 8 confirms that cyclone amplitude variability is important for wind impacts, while track frequency is more important for precipitation. Nevertheless, a metric that combines both amplitude and frequency, such as ATA, is more highly correlated with weather impacts.

5. Leading patterns of variability of ATA of ζ850 over Europe

In section 4b, we showed that ATA based on tracking positive centers of ζ850 is one of the storm track metrics that is highly correlated to weather impacts (especially precipitation) over Europe. Since ATA is a novel metric that we have developed, in this section we will examine some of its properties including its leading patterns of variability. More information about ATA in other geographical locations will be presented in Part 2 of this series.

Figure 9a shows the winter climatology of ATA. Maximum activity extends from the downstream end of the Atlantic storm track into northern Europe and western Russia, with a secondary maximum over the Mediterranean. This distribution highly resembles that of track frequency (Fig. 9c), but modulated by amplitude (Fig. 9d), which leads to more emphasis on the Atlantic storm track and lower amplitude over Russia and the Mediterranean. Month-to-month variability (Fig. 9b) highlights the same regions of action. Month-to-month variations in ATA is highly correlated with that of track frequency, with correlation coefficient of >0.8 over most regions (not shown). Correlation with track amplitude is more modest but is statistically significant over most regions (Fig. 9f), with the highest correlation of >0.5 over the eastern Atlantic and parts of the Mediterranean. Note that correlation between track frequency and amplitude (Fig. 9e) is generally low, only displaying isolated patches of significant correlation especially over the eastern Atlantic.

The leading EOF patterns of monthly variations in ATA, as well as patterns of related fields, are shown by correlation maps computed using the leading PCs of ATA variability. While the EOF analysis is conducted over Europe, the correlations are shown over the half-hemisphere centered over Europe to explore the upstream and downstream linkages. Note that the sign of the EOFs is arbitrary. Also note that we are not suggesting that these EOFs represent different modes of variability, but that they just represent the most frequently observed patterns of anomalies over this region.

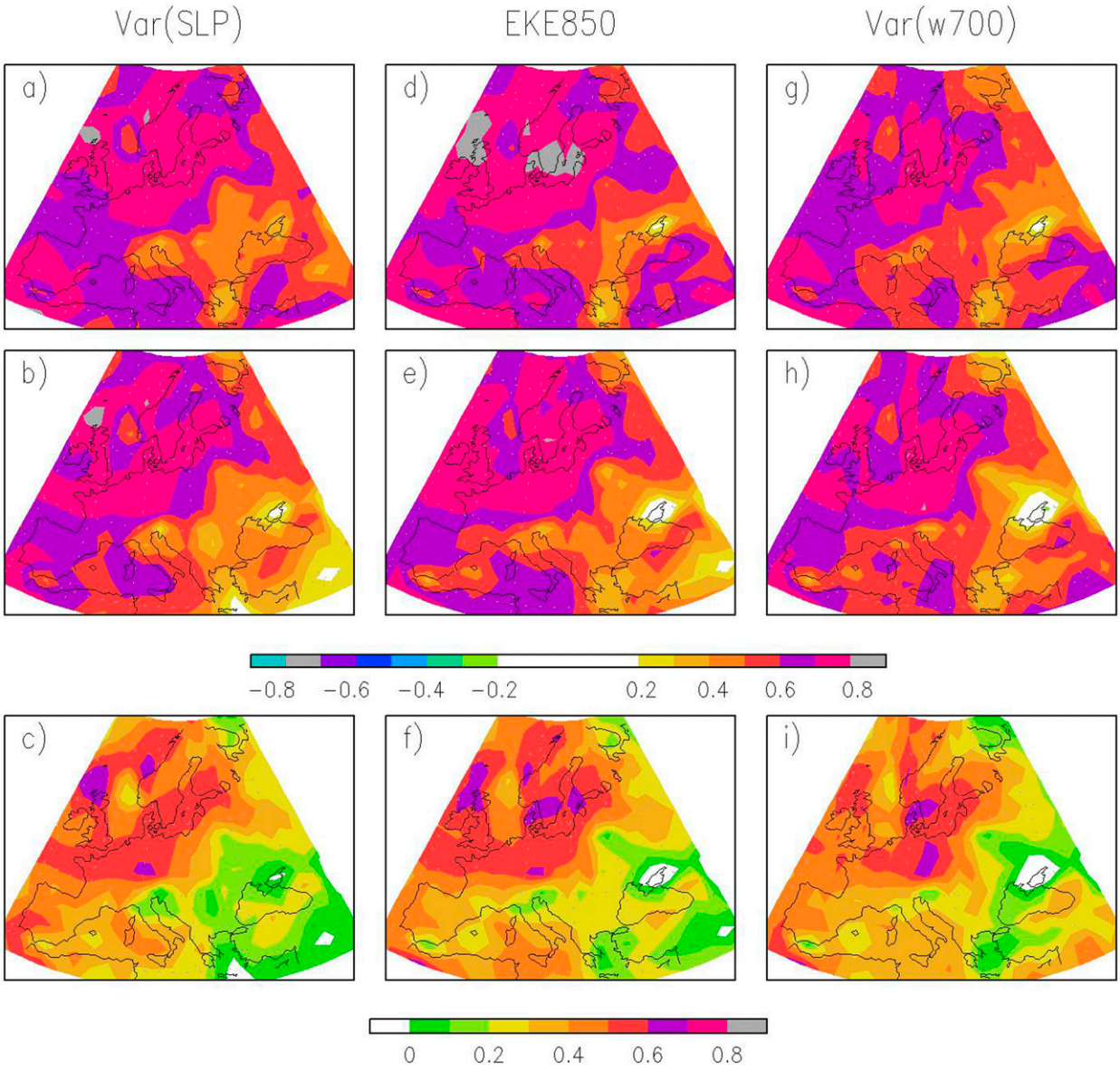

FIG. 6. As in Fig. 5, but for the high wind index.

EOF1 accounts for 21.0% of the variance and exhibits mainly a dipolar structure (Fig. 10a, shading). The northern branch of positive anomalies extends across the northern part of the United Kingdom into Scandinavia, while the southern branch with opposite sign extends into the Iberian Peninsula. Farther south, another branch can be seen with positive anomalies over the subtropical eastern Atlantic as well as to the southeast of the Caspian Sea. Given that Europe is located at the downstream side of the Atlantic storm track, it is not surprising that the leading EOF (as well as EOF2) is dominated by variations of the Atlantic storm track over eastern Atlantic (see also Fig. 9b). Anomalies in track frequency (Fig. 10b, shading) show a highly consistent pattern. Anomalies in track amplitude (Fig. 10b, contours) are not as widespread, but coincide with the peaks of the dipolar structure, contributing to the high correlations between ATA and PC1 over these regions (Fig. 10a). The ATA anomalies are highly consistent with those of the large-scale flow. Zonal wind at 300 hPa generally shows the same sign of anomalies but shifted slightly to the south (Fig. 10a, contours), while 500-hPa geopotential height (Fig. 10c, contours) shows patterns consistent with the zonal wind anomalies. The zonal wind anomalies are quite similar to those associated with the NAO, which many previous studies (e.g., Rogers 1997; Chang 2009) have shown to strongly influence storm track activity over the Atlantic and Europe. In fact, the correlation between PC1 and the NAO index[1] is 0.65.

[1] The NAO index is a rotated-EOF based index obtained from the Climate Prediction Center (see acknowledgments).

TABLE 3. Selected summary statistics for Lagrangian metrics. Maximum value in each column is highlighted by bold fonts. For full summary statistics for all 25 Lagrangian metrics, see Table S1 in the online supplemental material.

		GPCP precipitation				High wind index			
			CCA				CCA		
Tracked feature	Metric	CORMAX	*M*	ACC	FVE	CORMAX	*M*	ACC	FVE
SLP1	Amplitude	0.46	12	0.49	0.31	0.48	10	0.50	0.30
	Cyclone frequency	0.55	17	0.56	0.37	0.43	10	0.43	0.24
	Track frequency	0.58	16	0.59	0.41	0.47	14	0.46	0.26
	ACA	0.57	15	0.56	0.38	0.49	11	0.48	0.29
	ATA	0.61	22	0.61	0.44	0.54	18	0.53	0.34
SLP2	ATA	0.51	11	0.52	0.36	0.51	10	0.51	0.33
ζ850	Amplitude	0.43	25	0.50	0.30	0.49	13	0.50	0.31
	Cyclone frequency	0.56	16	0.60	0.42	0.41	16	0.45	0.25
	Track frequency	0.59	21	0.61	0.46	0.47	11	0.48	0.28
	ACA	0.61	16	0.63	0.47	0.49	13	0.51	0.32
	ATA	**0.64**	23	**0.65**	**0.50**	**0.56**	13	**0.54**	**0.36**
ζ300	ATA	0.55	13	0.59	0.43	0.50	13	0.51	0.32
ω700	ATA	0.60	15	0.61	0.44	0.48	10	0.46	0.27

The ATA anomalies are also largely consistent with the anomalies in 850-hPa EKE (Fig. 10c, shading), although the EKE anomalies appear to be slightly broader in spatial scale and do not show positive anomalies over subtropical eastern Atlantic. Correlations with other storm track metrics, including variance in SLP, are also similar (not shown). Correlations with weather impacts are shown in Fig. 10d. Precipitation anomalies (shading) display a dipolar structure very similar to that displayed by ATA. The high wind index also shows a dipole over a similar region, but also shows an extended positive region over the subtropics over a region larger than that displayed by the ATA anomalies.

Similar patterns for EOF2 (accounting for 20.2% of the variance) based on correlations with PC2 are shown in the right

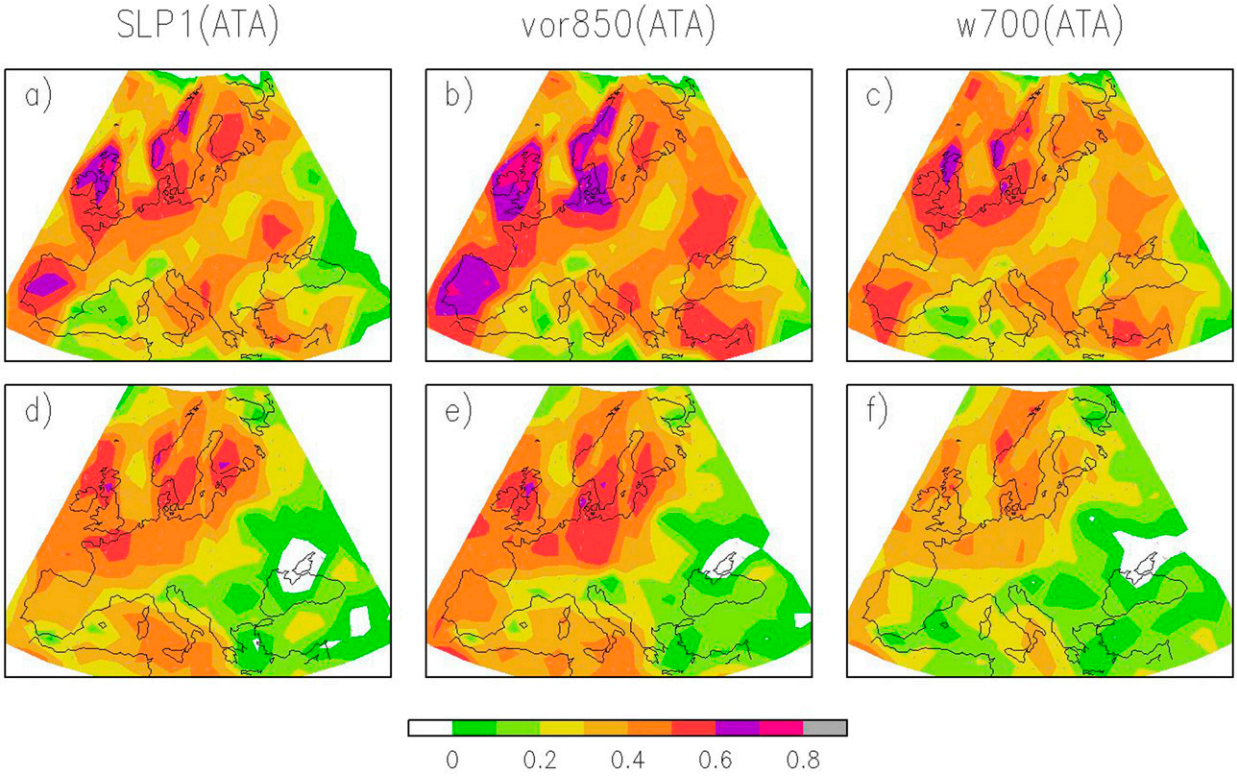

FIG. 7. FVE score from applying the CCA framework to assess Lagrangian tracking metrics, for (top) GPCP precipitation and (bottom) high wind index. The metrics assessed are ATA derived from tracking: (a),(d) SLP1, (b),(e) ζ850, and (c),(f) ω700.

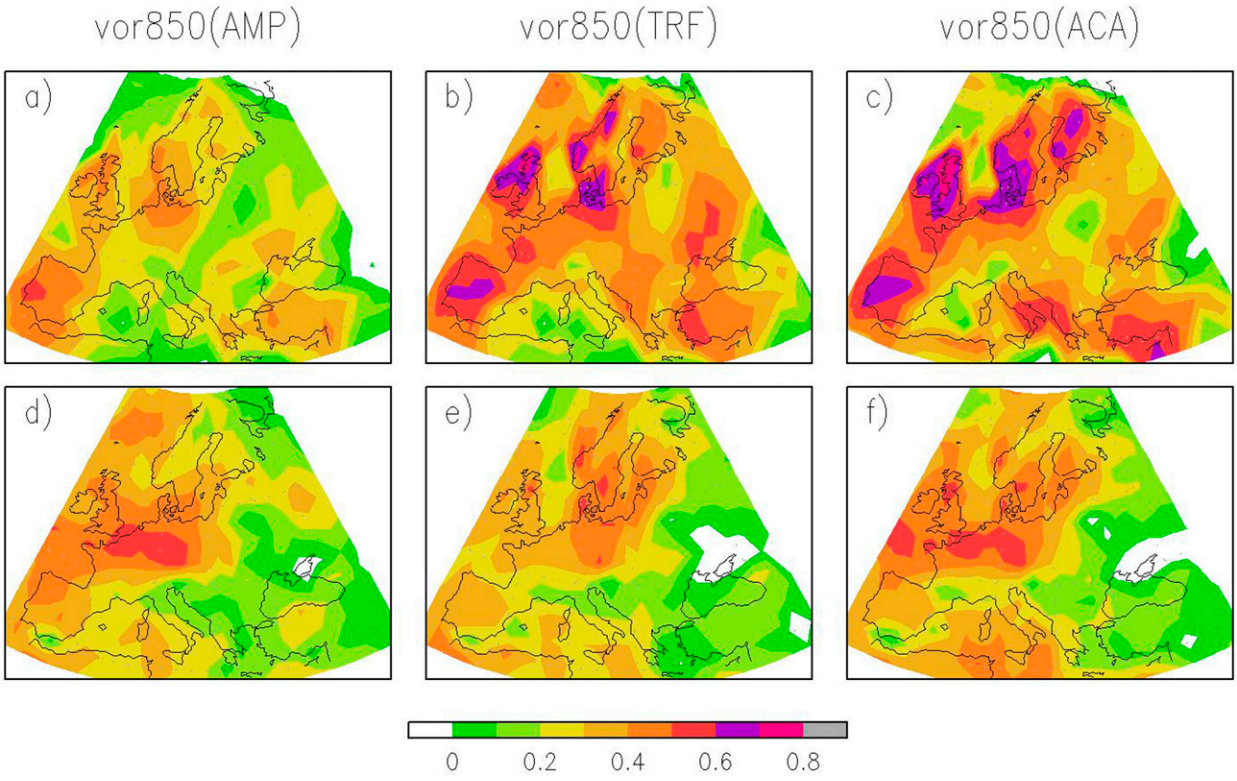

FIG. 8. As in Fig. 7, but applied to different metrics, all derived from tracking ζ850: (a),(d) cyclone amplitude, (b),(e) track frequency, and (c),(f) ACA.

panels of Fig. 10. The ATA anomalies (Fig. 10e) exhibit a dipole that is shifted poleward with respect to that of EOF1. Relationships to other parameters are very similar, except that zonal wind anomalies (Fig. 10e, contours) display a significant negative anomaly over North Africa that is not accompanied by any significant anomalies in storm track activity or weather impacts, likely because there is little storm track activity over that region. Otherwise, over the midlatitudes, for both EOF1 and EOF2, Fig. 10 shows that ATA is highly correlated with cyclone track frequency and less so with cyclone amplitude (Fig. 10f). It is also highly correlated with variations in the large-scale flow (Figs. 10e,g), other storm track metrics including 850-hPa EKE (Fig. 10g), and significant weather impacts (Fig. 10h).

Selected correlations for EOFs 3–5 are shown in Fig. 11. For each of these EOFs, we show the correlations between the PC with ATA and 300-hPa zonal wind in the left panel, and the correlations with weather impacts including precipitation and high wind index on the right. In most cases, enhanced (reduced) ATA is generally associated with enhanced (reduced) zonal flow in its vicinity, showing consistency between variations in ATA with baroclinicity. Nevertheless, EOF5 shows significant ATA variations over southern Europe and the Mediterranean (Fig. 11c), with only rather weak signals in the large-scale flow, suggesting that this variability could reflect internal chaotic variability in the storm track itself. All three EOFs show significant weather impacts of the same sign to nearby storm track anomalies (right panels), confirming the close relationship between ATA and weather impacts.

6. Summary and discussion

In the midlatitudes, storm tracks are responsible for high-impact weather over most regions, especially during the cool season. Thus it is useful to quantify storm track variability using metrics that are strongly related to weather impacts. While previous studies have shown that several storm track metrics, such as variance in SLP or ω, are highly correlated with weather impacts such as precipitation, to date no systematic study has been conducted to evaluate multiple storm track metrics. In this study, we have developed two independent frameworks to quantify how well each storm track metric relates to weather impacts including precipitation and high winds. The first framework searches for the maximum correlation (which we call CORMAX) between weather impacts and the storm track metric, making use of the monthly time series of weather impact at each location to compute a one-point correlation map between the weather impact and the storm track metric. The second framework evaluates pattern correlations using CCA. A leave-N-out cross-validation approach ($N = 3$ in this study) is used to prevent overfitting due to the limited number of temporal degrees of freedom. A CCA model is constructed based on the remaining storm track/weather impact pairs and used to hindcast the weather impact for the three left-out months using the storm track anomalies

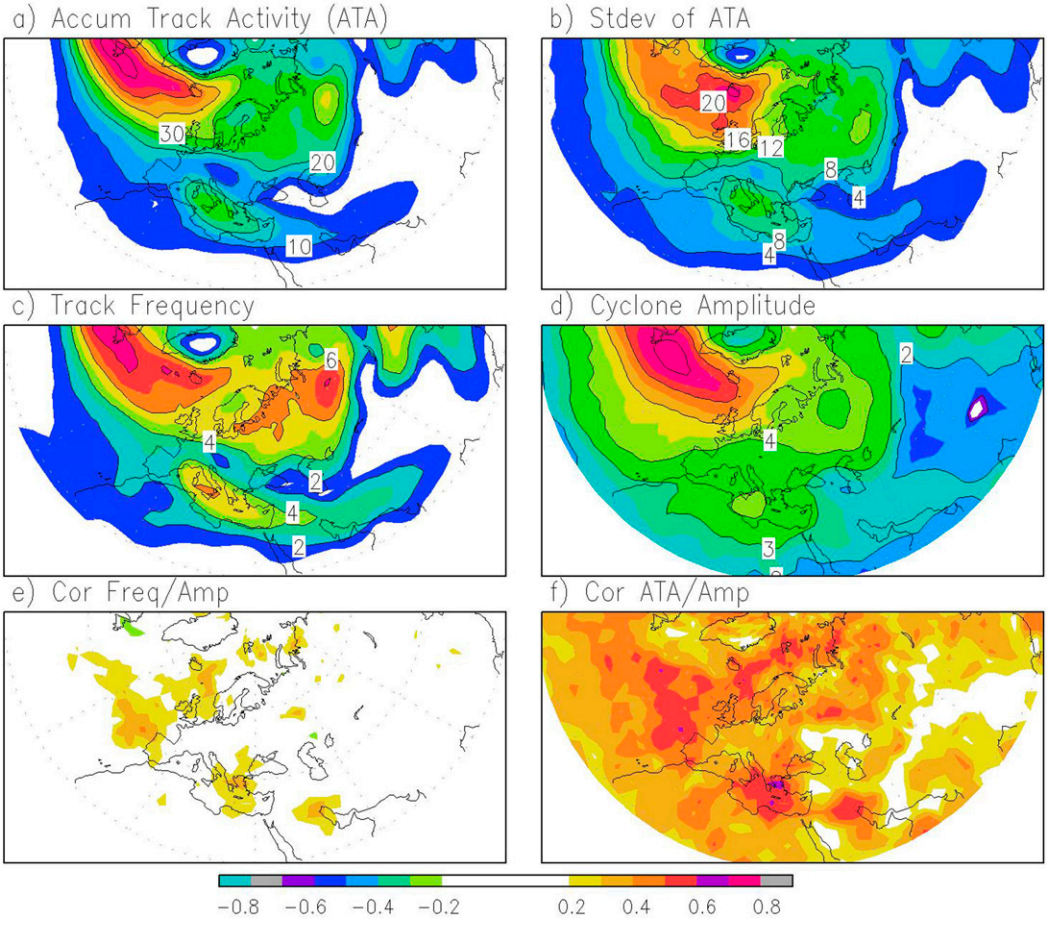

FIG. 9. Climatological distribution of Lagrangian metrics derived from tracking ζ850. (a) ATA ($10^{-5}\,s^{-1}$ $month^{-1}$). (b) Standard deviation of month-to-month variations in ATA ($10^{-5}\,s^{-1}$ $month^{-1}$). (c) Track frequency ($month^{-1}$). (d) Cyclone amplitude ($10^{-5}\,s^{-1}$). (e) Correlation between month-to-month variations in track frequency and cyclone amplitude at each grid point. (f) As in (e), but between ATA and cyclone amplitude.

for these months. Results are quantified by the ACC and FVE scores by comparing the hindcast weather impacts to those observed. Results from the two frameworks are largely consistent, with the ACC scores based on the CCA framework very similar to the CORMAX scores based on the first framework. The consistency between these two frameworks gives us confidence that we are likely not overestimating the correlations between the storm track metrics and weather impacts.

Among the nine Eulerian variance metrics that we examined, most of them correlate strongly with weather impacts. In general, near-surface or lower tropospheric metrics such as variances in SLP, 850-hPa v' and ζ', and 850-hPa EKE, are more closely related to weather impacts than their upper-level counterparts such as 500-hPa z and 300-hPa v', ζ', and EKE. EKE also performs slightly better than the variance of v'. For precipitation, the variance of 700-hPa ω' performs best, which is not surprising given the close association between precipitation and vertical motion. EKE and the variance of ζ' at 850 hPa also correlate highly with precipitation. For winds, 850-hPa EKE and SLP variance perform best, which is reasonable given the close relationship between wind speed and EKE, as well as the quasigeostrophic relationship between wind and pressure. the variance of ω at 700 hPa also relates quite well with wind impacts. Given that EKE is generally less noisy compared to ω, and is more readily available from climate model data archives such as CMIP5 or CMIP6, EKE at 850 hPa is a good candidate to be used as an Eulerian metric to quantify storm track variability and change.

For Lagrangian tracking statistics, ATA, a novel metric that we developed that combines information from track frequency and amplitude variations, does best in all cases. Cyclone amplitude variations do not capture a lot of precipitation variability but do quite well in capturing wind impacts. On the other hand, track frequency does quite well for precipitation but not as well for wind impacts. Among the five different definitions of cyclones, spatially filtered (T5–42) relative vorticity ζ maxima at the 850-hPa level provides statistics that are best correlated with both GPCP precipitation and the high wind index. Our limited testing suggests that

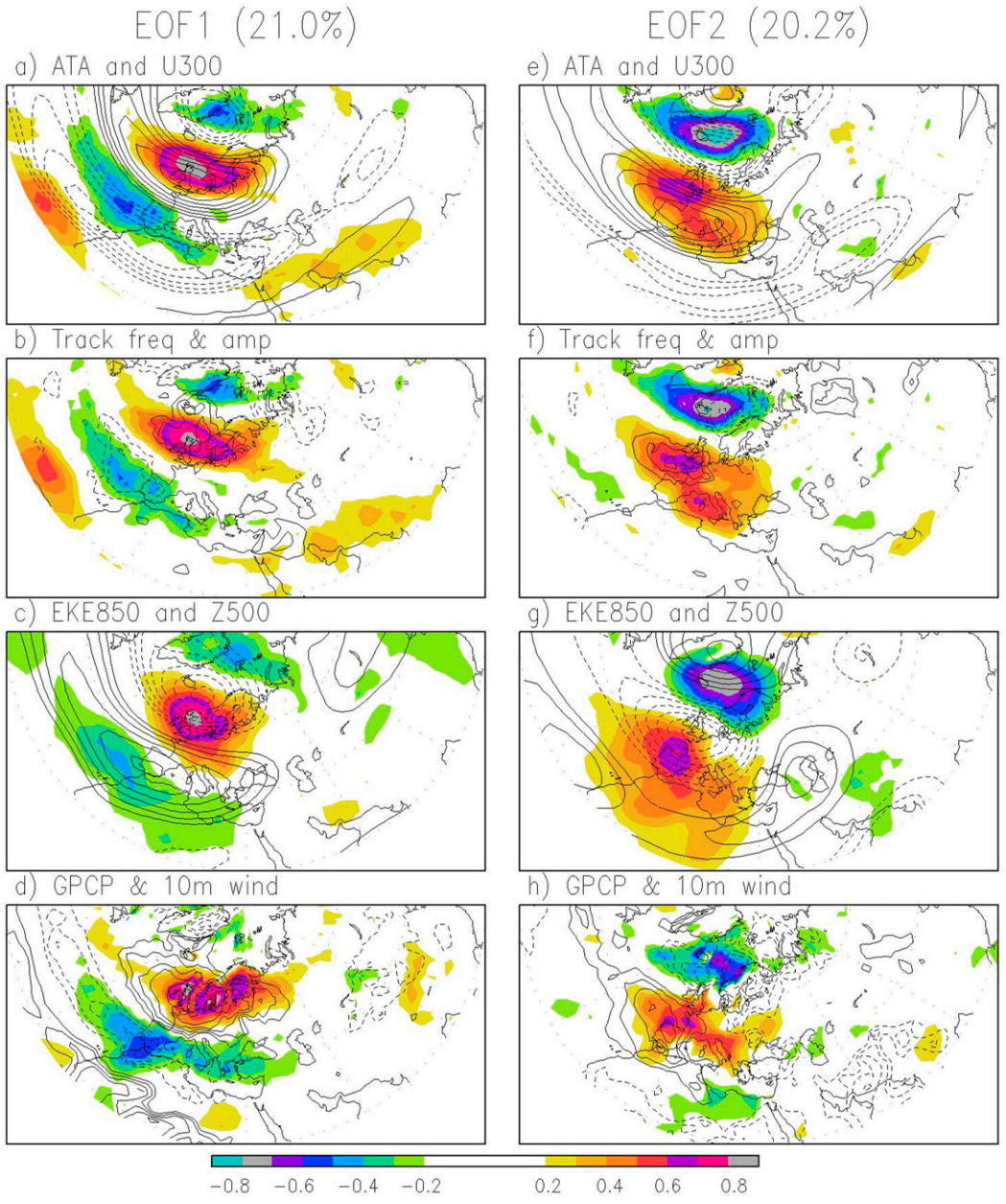

FIG. 10. Correlation patterns associated with the two leading EOFs of ATA derived from tracking ζ850, showing correlations with (left) PC1 and (right) PC2, for (a) ATA (shading) and 300-hPa monthly mean zonal wind (contours; contour interval 0.1 with absolute values below 0.2 not shown; same for other panels), (b) track frequency (shading) and cyclone amplitude (contours), (c) EKE850 (shading) and 500-hPa monthly mean geopotential height (contours), and (d) GPCP precipitation (shading) and high-wind index (contours).

spatial filtering alone to remove large spatial scale background (and small-scale noise for vorticity) provides statistics that perform better than applying both spatial and time filtering by removing large-scale background and the monthly mean. Similar tests removing monthly means from ζ850 also produce statistics that are inferior to just spatial filtering alone (not shown). Nevertheless, more tests should be conducted to see how the results are sensitive to different spatial and temporal filtering.

One potential caveat is that previous studies have found that Lagrangian tracking statistics are somewhat sensitive to the tracking algorithm used. Here, we tested the sensitivity by using a very simple tracker that we have developed ourselves. The tracker and the results are discussed in the supplemental material (see supplemental text S1). We show there that while the statistics exhibit some differences, ATA computed from statistics using the simple tracker is highly correlated with ATA discussed in section 5, with the leading five EOFs being

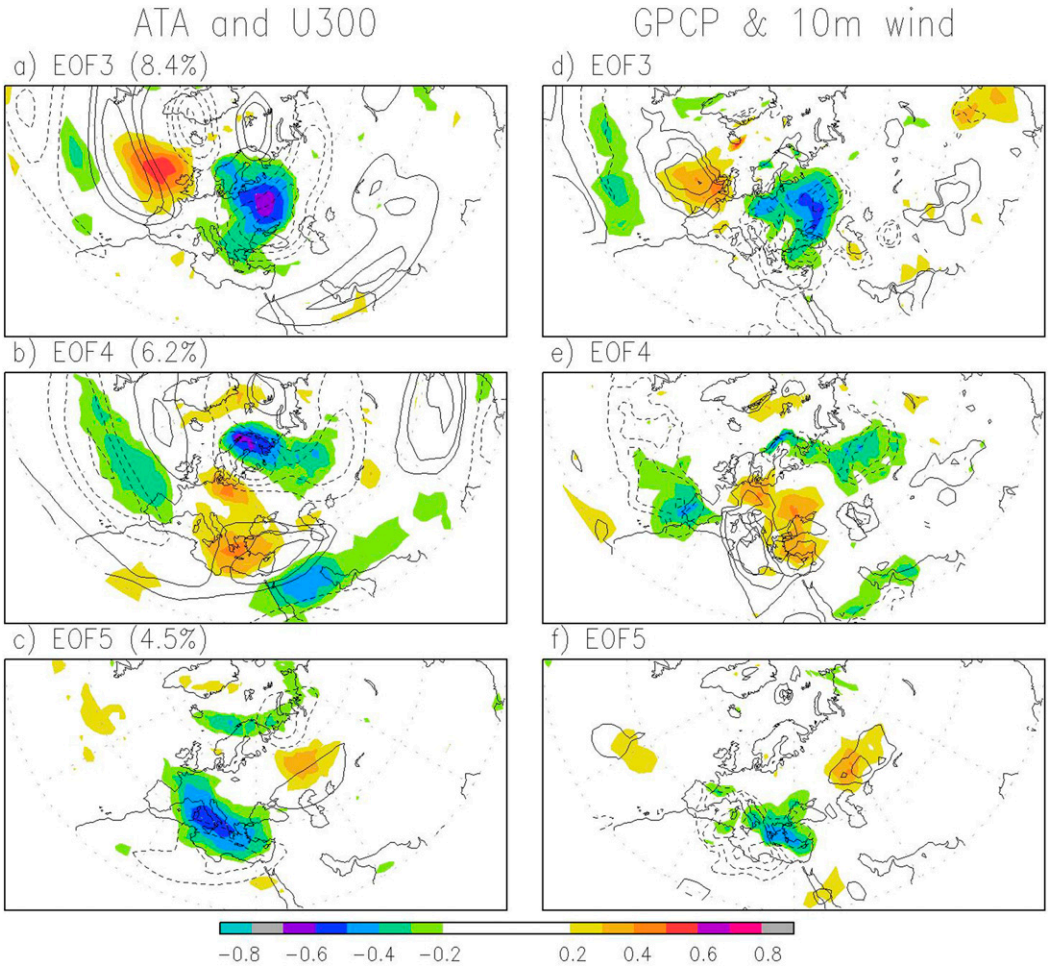

FIG. 11. Correlation patterns for EOFs 3–5 of ATA derived from tracking ζ850. Correlations with (top) PC3, (middle) PC4, and (bottom) PC5 for (a)–(c) ATA (shading) and 300-hPa monthly mean zonal wind and (d)–(f) GPCP precipitation (shading) and high-wind index (contours).

very similar—except that the order of EOFs 1 and 2 is reversed. ATA derived from the simple tracker also does very well in predicting weather impacts, performing only slightly worse than that derived from the more sophisticated tracker used in the main body of the paper. Hence we are confident that our results are not very sensitive to the tracker used.

While our focus is on the link between the storm track metrics and weather impacts, another consideration might be how predictable these metrics are. To date, few studies have assessed how well climate models can predict storm track variability. For Eulerian variance metrics, Yang et al. (2015) and Zheng et al. (2019) showed that the climate prediction model of the Geophysical Fluid Dynamics Laboratory and the models participating in the North American Multi-Model Ensemble can predict storm track variability [in terms of a metric similar to Var(SLP)] related to El Niño–Southern Oscillation (ENSO) at least out to several months in advance, while Zheng et al. (2019, 2020, manuscript submitted to *Wea. Forecasting*) showed that models participating in the Seasonal to Subseasonal Experiment and the Subseasonal Experiment can predict Var(SLP) with some skill out to weeks 3–4. For the subseasonal case, the source of predictability appears to be ENSO and polar vortex variations.

For Lagrangian track statistics, Lukens and Berbery (2019) assessed how well the National Oceanic and Atmospheric Administration (NOAA) Coupled Forecast System version 2 (CFSv2) can predict cyclone frequency and amplitude by tracking lower tropospheric potential vorticity (PV) anomalies. Their results suggested that the prediction skill for a single member is rather limited, with the root-mean-square error nearly the same as the standard deviation at most locations, even after bias corrections. Nevertheless, Zheng et al. (2019, 2020, manuscript submitted to *Wea. Forecasting*) suggested that combining different ensemble members into either a single model or a multimodel ensemble can significantly improve the prediction skill for variance statistics, and we hypothesize that even track statistics can be predicted to certain extent using ensemble prediction techniques.

As far as we know, no studies have assessed how well current climate models can predict EKE850 or ATA, and we plan to do that next. Nevertheless, our results suggest that ATA exhibits coherent large-scale variations (see Figs. 10 and 11) and is

highly correlated with variability in the large-scale flow and other variance statistics; hence, we expect that ATA is likely to be better predicted than other more noisy Lagrangian track statistics such as cyclone amplitude or cyclone frequency.

Another potential application is to examine storm track changes under climate change projections. Previous studies have tried to link model projected precipitation changes to changes in Var(SLP) (Chang 2014; Chang et al. 2015), the number and strength of cyclones (Zappa et al. 2015), or track frequency and intensity (Osburn et al. 2018). Here we have found that EKE850 and ATA are more highly correlated with precipitation and wind impacts than Var(SLP) and track frequency/amplitude, and hence it will be of interest to reassess the weather impacts of projected changes in storm tracks by analyzing projected changes in EKE850 and ATA.

Finally, in this study, we developed two frameworks to assess how well storm track metrics are related to weather impacts, and developed a novel Lagrangian storm track metric, ATA, that displays coherent large-scale variability and is highly correlated to weather impacts. Nevertheless, since we have not exhaustively tested all available storm track metrics, we do not claim that EKE850 and ATA are the ''best'' storm track metrics. In fact, the ''best'' metric may depend on the geographical location or season, although thus far our experience is that among Lagrangian cyclone statistics, ATA derived from ζ850 always performs better than the other Lagrangian metrics. This will be discussed in more details in Part 2 of this series. The frameworks we have developed can be used to assess other metrics. In addition, these frameworks can be adopted to assess other metrics using different criteria. Given that a large number of metrics have been used to quantify storm track variability, the frameworks developed in this study can be used to help pick out a small number of metrics that are closely related to the objective of each individual user.

Acknowledgments. This work is supported by NSF Grant AGS-1261311 and NASA Grant NNX16AG32G. EC is also supported by NOAA Grant NA19OAR4310283. The authors thank Dr. Kevin Hodges for providing the code for the feature tracker. The ERA-Interim data are available from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim, the GPCP data are available from https://psl.noaa.gov/data/gridded/data.gpcp.html, and the NAO index is available from http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml.

REFERENCES

Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Alexander, L. V., S. F. B. Tett, and T. Jonsson, 2005: Recent observed changes in severe storms over the United Kingdom and Iceland. *Geophys. Res. Lett.*, **32**, L13704, https://doi.org/10.1029/2005GL022371.

Alexandersson, H., T. Schmith, K. Iden, and H. Tuomenvirta, 1998: Long-term variations of the storm climate over NW Europe. *Global Atmos. Ocean Syst.*, **6**, 97–120.

Ashley, W. S., and A. W. Black, 2008: Fatalities associated with nonconvective high-wind events in the United States. *J. Appl. Meteor. Climatol.*, **47**, 717–725, https://doi.org/10.1175/2007JAMC1689.1.

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, **115**, 1825–1850, https://doi.org/10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2.

Benedict, J. J., S. Lee, and S. B. Feldstein, 2004: Synoptic view of the North Atlantic Oscillation. *J. Atmos. Sci.*, **61**, 121–144, https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2.

Berko, J., D. D. Ingram, S. Saha, and J. D. Parker, 2014: Deaths attributed to heat, cold, and other weather events in the United States, 2006–2010. National Health Statistics Rep. 76, National Center for Health Statistics, 15 pp.

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33**, 1607–1623, https://doi.org/10.1175/1520-0469(1976)033<1607:ACSSOT>2.0.CO;2.

Chang, E. K. M., 2009: Are band-pass variance statistics useful measures of storm track activity? Re-examining storm track variability associated with the NAO using multiple storm track measures. *Climate Dyn.*, **33**, 277–296, https://doi.org/10.1007/s00382-009-0532-9.

——, 2014: Impacts of background field removal on CMIP5 projected changes in Pacific winter cyclone activity. *J. Geophys. Res. Atmos.*, **119**, 4626–4639, https://doi.org/10.1002/2013JD020746.

——, and Y. F. Fu, 2003: Using mean flow change as a proxy to infer interdecadal storm track variability. *J. Climate*, **16**, 2178–2196, https://doi.org/10.1175/2773.1.

——, and A. M. W. Yau, 2016: Northern Hemisphere winter storm track trends since 1959 derived from multiple reanalysis datasets. *Climate Dyn.*, **47**, 1435–1454, https://doi.org/10.1007/s00382-015-2911-8.

——, S. Lee, and K. L. Swanson, 2002: Storm track dynamics. *J. Climate*, **15**, 2163–2183, https://doi.org/10.1175/1520-0442(2002)015<02163:STD>2.0.CO;2.

——, Y. Guo, and X. Xia, 2012: CMIP5 multimodel ensemble projection of storm track change under global warming. *J. Geophys. Res.*, **117**, D23118, https://doi.org/10.1029/2012JD018578.

——, C. Zheng, P. Lanigan, A. M. W. Yau, and J. D. Neelin, 2015: Significant modulation of variability and projected change in California winter precipitation by extratropical cyclone activity. *Geophys. Res. Lett.*, **42**, 5983–5991, https://doi.org/10.1002/2015GL064424.

Colle, B. A., F. Buonaiuto, M. J. Bowman, R. E. Wilson, R. Flood, R. Hunter, A. Mintz, and D. Hill, 2008: New York City's vulnerability to coastal flooding. *Bull. Amer. Meteor. Soc.*, **89**, 829–842, https://doi.org/10.1175/2007BAMS2401.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim Reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Donohoe, A., and D. S. Battisti, 2009: The amplitude asymmetry between synoptic cyclones and anticyclones: Implications for filtering methods in feature tracking. *Mon. Wea. Rev.*, **137**, 3874–3887, https://doi.org/10.1175/2009MWR2837.1.

Donat, M. G., G. C. Leckebusch, J. G. Pinto, and U. Ulbrich, 2010: Examination of wind storms over central Europe with respect to circulation weather types and NAO phases. *Int. J. Climatol.*, **30**, 1289–1300, https://doi.org/10.1002/joc.1982.

Feser, F., M. Barcikowska, O. Krueger, F. Schenk, R. Weisse, and L. Xia, 2015: Storminess over the North Atlantic and northwestern Europe—A review. *Quart. J. Roy. Meteor. Soc.*, **141**, 350–382, https://doi.org/10.1002/qj.2364.

Grise, K. M., S.-W. Son, and J. R. Gyakum, 2013: Intraseasonal and interannual variability in North American storm tracks and its relationship to equatorial Pacific variability. *Mon. Wea. Rev.*, **141**, 3610–3625, https://doi.org/10.1175/MWR-D-12-00322.1.

Guo, Y., T. Shinoda, J. Lin, and E. K. M. Chang, 2017: Variations of Northern Hemisphere storm track and extratropical cyclone activity associated with the Madden–Julian oscillation. *J. Climate*, **30**, 4799–4818, https://doi.org/10.1175/JCLI-D-16-0513.1.

Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.*, **122**, 2573–2586, https://doi.org/10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2.

——, 1999: Adaptive constraints for feature tracking. *Mon. Wea. Rev.*, **127**, 1362–1373, https://doi.org/10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2.

Hoskins, B. J., and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061, https://doi.org/10.1175/1520-0469(2002)059<1041:NPOTNH>2.0.CO;2.

Jiang, Q., R. B. Smith, and J. D. Doyle, 2003: The nature of the mistral: Observations and modeling of two MAP events. *Quart. J. Roy. Meteor. Soc.*, **129**, 857–875, https://doi.org/10.1256/qj.02.21.

Klein, W. H., 1958: The frequency of cyclones and anticyclones in relation to the mean circulation. *J. Meteor.*, **15**, 98–102, https://doi.org/10.1175/1520-0469(1958)015<0098:TFOCAA>2.0.CO;2.

Kocin, P. J., A. D. Weiss, and J. J. Wagner, 1988: The great Arctic outbreak and east coast blizzard of February 1899. *Wea. Forecasting*, **3**, 305–318, https://doi.org/10.1175/1520-0434(1988)003<0305:TGAOAE>2.0.CO;2.

Kunkel, K. E., D. R. Easterling, D. A. R. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *J. Hydrometeor.*, **13**, 1131–1141, https://doi.org/10.1175/JHM-D-11-0108.1.

Lorenz, D. J., and D. L. Hartmann, 2001: Eddy–zonal flow feedback in the Southern Hemisphere. *J. Atmos. Sci.*, **58**, 3312–3327, https://doi.org/10.1175/1520-0469(2001)058<3312:EZFFIT>2.0.CO;2.

Lukens, K. E., and E. H. Berbery, 2019: Winter storm tracks and related weather in the NCEP Climate Forecast System weeks 3–4 reforecasts for North America. *Wea. Forecasting*, **34**, 751–772, https://doi.org/10.1175/WAF-D-18-0113.1.

Ma, C. G., and E. K. M. Chang, 2017: Impacts of storm track variations on wintertime extreme weather events over the continental United States. *J. Climate*, **30**, 4601–4624, https://doi.org/10.1175/JCLI-D-16-0560.1.

Mizuta, R., 2012: Intensification of extratropical cyclones associated with the polar jet change in the CMIP5 global warming projections. *Geophys. Res. Lett.*, **39**, L19797, https://doi.org/10.1029/2012GL053032.

Neu, U., and Coauthors, 2013: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bull. Amer. Meteor. Soc.*, **94**, 529–547, https://doi.org/10.1175/BAMS-D-11-00154.1.

NOAA NCEI, 2020: NOAA National Centers for Environmental Information (NCEI) U.S. billion-dollar weather and climate disasters. Accessed April 2020, https://doi.org/10.25921/stkw-7w73.

Nuissier, O., B. Joly, A. Joly, V. Ducrocq, and P. Arbogast, 2011: A statistical downscaling to identify the large-scale circulation patterns associated with heavy precipitation events over southern France. *Quart. J. Roy. Meteor. Soc.*, **137**, 1812–1827, https://doi.org/10.1002/qj.866.

Osburn, L., K. Keay, and J. L. Catto, 2018: Projected change in wintertime precipitation in California using projected changes in extratropical cyclone activity. *J. Climate*, **31**, 3451–3466, https://doi.org/10.1175/JCLI-D-17-0556.1.

Paciorek, C. J., J. S. Risbey, V. Ventura, and R. D. Rosen, 2002: Multiple indices of Northern Hemisphere cyclone activity, winters 1949–99. *J. Climate*, **15**, 1573–1590, https://doi.org/10.1175/1520-0442(2002)015<1573:MIONHC>2.0.CO;2.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*. American Institute of Physics, 520 pp.

Pfahl, S., 2014: Characterising the relationship between weather extremes in Europe and synoptic circulation features. *Nat. Hazards Earth Syst. Sci.*, **14**, 1461–1475, https://doi.org/10.5194/nhess-14-1461-2014.

——, and H. Wernli, 2012: Quantifying the relevance of cyclones for precipitation extremes. *J. Climate*, **25**, 6770–6780, https://doi.org/10.1175/JCLI-D-11-00705.1.

Rogers, J. C., 1997: North Atlantic storm track variability and its association to the North Atlantic Oscillation and climate variability of northern Europe. *J. Climate*, **10**, 1635–1647, https://doi.org/10.1175/1520-0442(1997)010<1635:NASTVA>2.0.CO;2.

Sinclair, M. R., 1997: Objective identification of cyclones and their circulation intensity and climatology. *Wea. Forecasting*, **12**, 595–612, https://doi.org/10.1175/1520-0434(1997)012<0595:OIOCAT>2.0.CO;2.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

——, G.-H. Lim, and M. L. Blackmon, 1988: Relationship between cyclone tracks, anticyclone tracks, and baroclinic waveguides. *J. Atmos. Sci.*, **45**, 439–462, https://doi.org/10.1175/1520-0469(1988)045<0439:RBCTAT>2.0.CO;2.

Yang, X., and Coauthors, 2015: Seasonal predictability of extratropical storm tracks in GFDL's high-resolution climate prediction model. *J. Climate*, **28**, 3592–3611, https://doi.org/10.1175/JCLI-D-14-00517.1.

Zappa, G., M. K. Hawcroft, L. Shaffrey, E. Black, and D. J. Brayshaw, 2015: Extratropical cyclones and the projected decline of winter Mediterranean precipitation in the CMIP5 models. *Climate Dyn.*, **45**, 1727–1738, https://doi.org/10.1007/s00382-014-2426-8.

Zheng, C., E. K. M. Chang, H. Kim, M. Zhang, and W. Wang, 2019: Subseasonal to seasonal prediction of wintertime Northern Hemisphere extratropical cyclone activity by S2S and NMME models. *J. Geophys. Res. Atmos.*, **124**, 12 057–12 077, https://doi.org/10.1029/2019JD031252.