SPECIAL
Waves to Weather (W2W)
COLLECTION

Sampling Error Correction Evaluated Using a Convective-Scale 1000-Member Ensemble

TOBIAS NECKER

Hans-Ertel Centre for Weather Research, Ludwig-Maximilians-University, Munich, Germany, and Institut für Meteorologie und Geophysik, Universität Wien, Vienna, Austria

MARTIN WEISSMANN

Hans-Ertel Centre for Weather Research, Deutscher Wetterdienst, Munich, Germany, and Institut für Meteorologie und Geophysik, Universität Wien, Vienna, Austria

YVONNE RUCKSTUHL

Meteorological Institute, Ludwig-Maximilians-University, Munich, Germany

JEFFREY ANDERSON

Data Assimilation Research Section, NCAR, Boulder, Colorado

TAKEMASA MIYOSHI

RIKEN Center for Computational Science, Kobe, Japan

(Manuscript received 16 May 2019, in final form 26 November 2019)

ABSTRACT

State-of-the-art ensemble prediction systems usually provide ensembles with only 20–250 members for estimating the uncertainty of the forecast and its spatial and spatiotemporal covariance. Given that the degrees of freedom of atmospheric models are several magnitudes higher, the estimates are therefore substantially affected by sampling errors. For error covariances, spurious correlations lead to random sampling errors, but also a systematic overestimation of the correlation. A common approach to mitigate the impact of sampling errors for data assimilation is to localize correlations. However, this is a challenging task given that physical correlations in the atmosphere can extend over long distances. Besides data assimilation, sampling errors pose an issue for the investigation of spatiotemporal correlations using ensemble sensitivity analysis. Our study evaluates a statistical approach for correcting sampling errors. The applied sampling error correction is a lookup table–based approach and therefore computationally very efficient. We show that this approach substantially improves both the estimates of spatial correlations for data assimilation as well as spatiotemporal correlations for ensemble sensitivity analysis. The evaluation is performed using the first convective-scale 1000-member ensemble simulation for central Europe. Correlations of the 1000-member ensemble forecast serve as truth to assess the performance of the sampling error correction for smaller subsets of the full ensemble. The sampling error correction strongly reduced both random and systematic errors for all evaluated variables, ensemble sizes, and lead times.

1. Introduction

The evolution of numerical weather prediction (NWP) and the exploration of the chaotic behavior of weather in the 1960s (Lorenz 1963) are the starting points of present-day ensemble prediction. The European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP) produced their first operational ensemble forecasts in the early 1990s. Nowadays, most operational weather services maintain ensemble systems to gain probabilistic information using various ensemble configurations. The applied ensemble size to some extent depends on the primary purpose of the ensemble [e.g., estimating forecast uncertainty (variances) or estimating error covariances for data assimilation (DA)], but also on the

Corresponding author: Tobias Necker, tobias.necker@univie.ac.at

DOI: 10.1175/MWR-D-19-0154.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

available computing power. Therefore, the number of ensemble members is a trade-off between the required ensemble size and computational resources. Operational ensemble sizes range from about 20 up to 250 members (Houtekamer et al. 2014; Caron and Buehner 2018; Gustafsson et al. 2018), which is very small compared to the number of degrees of freedom in the model. All state-of-the-art ensemble applications, therefore, have to deal with sampling errors.

In DA, ensemble Kalman filter algorithms (Evensen 1994) or hybrid variational/ensemble approaches rely on accurate estimates of error correlations and covariances. To reduce the effect of spurious correlations, localization techniques are usually applied (Houtekamer and Mitchell 1998; van Leeuwen 1999; Houtekamer and Mitchell 2001). Localization is a physically motivated approach, which cuts off or damps spatial correlations after a certain distance. However, the choice of localization length scale is an intrinsically difficult task given that physical correlations in the atmosphere can extend horizontally over thousands of kilometers and vertically throughout the troposphere and even into the stratosphere. Particularly, vertical localization is a challenging issue for ensemble DA (Lei et al. 2018) as several observation types (e.g., passive satellite observations) provide vertically integrated information of the atmosphere.

Our study evaluates the sampling error correction (SEC) introduced by Anderson (2012) using the first convective-scale 1000-member ensemble simulation for an area in central Europe. The SEC statistically corrects for the overestimation of correlations due to undersampling and was designed to reduce the need for localization in ensemble Kalman filter DA. The approach comes down to a lookup table calculated using the Monte Carlo technique. The SEC is explicitly applied to spatial correlations to evaluate its potential for ensemble data assimilation. In this context, the SEC is compared to a standard distance-based localization using a Gaspari–Cohn function (Gaspari and Cohn 1999). Furthermore, the SEC is applied to spatiotemporal correlations to evaluate its potential for ensemble sensitivity analysis (ESA).

ESA was first introduced by Ancell and Hakim (2007) and is an efficient method to explore probabilistic datasets by investigating linear relations between a forecast metric and initial quantities. ESA has been applied for various synoptic-scale case studies (e.g., Hakim and Torn 2008; Torn and Hakim 2009; Torn 2010; Hanley et al. 2013; Barrett et al. 2015). Recently, Bednarczyk and Ancell (2015), Wile et al. (2015), and Limpert and Houston (2018) showed that ESA also seems to provide reasonable results for the analysis of convective-scale simulations. However, these studies could not quantify potential errors due to spurious correlations as no larger ensemble was available for comparison. Several previous studies attempted to account for undersampling by applying a confidence test that excludes insignificant correlations (Torn and Hakim 2008). However, this approach may also exclude small physical correlations, which can lead to systematic effects and is therefore not ideal for quantitative analysis of sensitivities. Our study compares the SEC to results using a confidence test. Another approach to reduce sampling errors performing ESA is to apply a standard distance-based localization (Gaspari and Cohn 1999). Hacker and Lei (2015) mitigate sampling errors by using a hierarchical ensemble filter to estimate an appropriate weighting factor for the regression coefficient as proposed by Anderson (2007).

Accurate probabilistic forecasts are especially required in convective-scale forecasting, which aims at predicting local weather phenomena and the occurrence of extreme weather events that are often related to atmospheric convection (Miyoshi et al. 2016a; Gustafsson et al. 2018). For this purpose, many weather centers now deploy convection-permitting NWP and ensemble prediction systems with a grid spacing of a few kilometers (Bouttier et al. 2016; Hagelin et al. 2017; Gustafsson et al. 2018). The chaotic nature of convection, however, leads to significantly lower predictability and distinctly different error characteristics compared to global large-scale weather forecasts (Hohenegger and Schaer 2007). Consequently, it is difficult to sample forecast errors and their covariances with low-dimensional ensemble systems.

Our study uses a unique convection-permitting 1000-member ensemble simulation with a horizontal grid spacing of 3 km centered over Germany, which provides reliable estimates of correlations. Necker et al. (2020, hereafter N20) evaluated the general performance of the 1000-member ensemble, compared the simulation to the convective-scale regional ensemble system of Deutscher Wetterdienst, and demonstrated that correlations obtained from ESA can be used to estimate the potential impact of different observable quantities on primary forecast quantities such as precipitation. For this reason, precipitation is considered as the forecast response function calculating spatiotemporal correlations. The 1000-member ensemble is considered as truth to quantify sampling errors for different ensemble sizes and to evaluate approaches that can be used to mitigate sampling errors. Several previous studies similarly used a large ensemble for studying sampling errors of smaller subsets (Hamill et al. 2001; Bannister et al. 2017).

This paper is outlined as follows: section 2 introduces ensemble sensitivity analysis and methods that are

considered to account for sampling errors. Section 3 summarizes the setup and properties of the 1000-member simulation. Section 4 presents a qualitative and quantitative analysis of correlations obtained for different ensemble sizes and evaluates the SEC using spatial and spatiotemporal correlations. The evaluation includes a comparison of the SEC to a confidence test and to a standard distance-based localization approach. Conclusions are provided in section 5.

2. Methods

a. Sampling error correction (SEC)

Let $\hat{r}$ denote the sample correlation between quantities x and J. Then

$$\hat{r} = \frac{\mathrm{cov}_m(\mathbf{J}, \mathbf{x})}{\sqrt{\mathrm{var}_m(\mathbf{J})\mathrm{var}_m(\mathbf{x})}}, \tag{1}$$

where $\mathbf{x}$ and $\mathbf{J}$ are vectors containing m ensemble estimates of x and J, respectively; cov_m denotes the sample covariance; and var_m denotes the sample variance. Following Anderson (2012) the sampling error corrected correlation, $\hat{r}_{\mathrm{sec}}$ given $\hat{r}$ is subject to the ensemble size m and an appropriate prior distribution of the true correlation coefficient r. The SEC statistically corrects for the overestimation of correlations caused by sampling errors. Its offline computation is based on Monte Carlo simulations that approximate the likelihood of $\hat{r}$. The final result is a separate lookup table for each pair of ensemble size and prior distribution, comprising 200 bins ranging from -1 to 1 that maps a sample correlation $\hat{r}$ to its corresponding sampling error corrected correlation $\hat{r}_{\mathrm{sec}}$. In this paper, we use the uniform distribution $U(-1, 1)$ as the default prior distribution. Additionally, the impact of more informative prior distributions on the performance of the SEC is evaluated in section 5e. Figure 1 shows the sampling error corrected correlation $\hat{r}_{\mathrm{sec}}$ as a function of the sample correlation $\hat{r}$ for different ensemble sizes and a uniform prior. For example, applying the SEC using a 40-member ensemble, a sample correlation of 0.5 is corrected to approximately 0.42. This study mainly uses the SEC table provided by the Data Assimilation Research Testbed (DART; Anderson et al. 2009) that is based on a uniform prior. In the following, we assume that sampling errors in the 1000-member ensemble are negligible and the large ensemble therefore can be seen as ''truth'' to assess the performance of the SEC.

b. Application to ensemble sensitivity analysis (ESA)

Following Ancell and Hakim (2007), the sensitivity S of a forecast metric J with respect to a state variable x

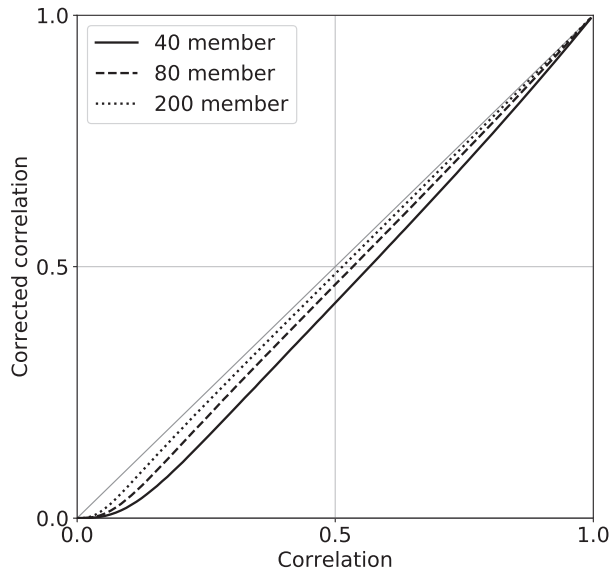

FIG. 1. Absolute sampling error corrected correlation $|\hat{r}_{\mathrm{sec}}|$ as a function of absolute sample correlation $|\hat{r}|$ using different ensemble sizes and a uniform prior.

can be approximated using spatiotemporal correlations from the ensemble:

$$S = \frac{\partial J}{\partial x} \approx \frac{\mathrm{cov}_m(\mathbf{J}, \mathbf{x})}{\mathrm{var}_m(\mathbf{x})} = \hat{r}\frac{\sqrt{\mathrm{var}_m(\mathbf{J})}}{\sqrt{\mathrm{var}_m(\mathbf{x})}}. \tag{2}$$

The sampling error corrected sensitivity S_{sec} can then be obtained using a lookup table by substituting $\hat{r}$ with the sampling error corrected correlation $\hat{r}_{\mathrm{sec}}$:

$$S_{\mathrm{sec}} = \hat{r}_{\mathrm{sec}}\frac{\sqrt{\mathrm{var}_m(\mathbf{J})}}{\sqrt{\mathrm{var}_m(\mathbf{x})}}. \tag{3}$$

In this paper, we use hourly precipitation averaged over a box of 40×40 grid points as forecast metric J. The state variable x can be any quantity of interest.

A common approach to suppress spurious correlations is based on a confidence test. Torn and Hakim (2008) first introduced the confidence test in combination with ESA. Their study examined if a state variable x is able to produce a statistically significant change in the forecast metric J:

$$\left|\frac{\mathrm{cov}_m(\mathbf{J}, \mathbf{x})}{\mathrm{var}_m(\mathbf{x})}\right| > \delta_s, \tag{4}$$

where δ_s is the confidence interval on the linear regression coefficient. The approach aims to exclude statistically insignificant sensitivities by rejecting the null hypothesis that there is no correlation between the forecast metric and the state variable with predefined confidence. In this

manuscript, we apply a Student's t test with a 95% confidence level (T95). The 1000-member ensemble is used to compare the performance of the SEC with that of the T95. Insignificant correlations are not considered and excluded from the analysis.

c. *Application to ensemble and hybrid data assimilation*

NWP data assimilation schemes combine observations with a short-term model forecast to achieve an optimal estimate of the atmospheric state. How the spatially sparse observational information is distributed in space is determined by sample correlations that are obtained from the ensemble. However, the ensemble size is usually too small to sample all possible states. Consequently, spurious correlations caused by undersampling strongly degrade the initial conditions. In this context, the sampling error correction of Anderson (2012, 2016) provides an alternative to constant covariance localization length scales that are usually applied. For this purpose, we use the 1000-member ensemble to evaluate the effect of the sampling error correction applied to spatial correlations for different variables and ensemble sizes in section 5. Furthermore, the SEC is compared to a standard distance-based localization using a Gaspari–Cohn function (LOC) (Gaspari and Cohn 1999). In this study, localization scales are fixed horizontally to 100 km and vertically to $\ln(p) = 0.3$ based on the 1000-member ensemble DA setup (N20) and previous convective-scale DA studies (Lange and Janjić 2016; Necker et al. 2018). While the SEC is a simple statistical correction method, it should be noted that its application for data assimilation is only straight forward in ensemble and hybrid data assimilation schemes that calculate covariances explicitly.

3. Experiments

a. *1000-member ensemble simulation*

The initial conditions (IC) for the simulation are obtained from a 1000-member ensemble DA experiment with a horizontal grid spacing of 15 km that has been spun up for one week. Boundary conditions (BC) were generated from the NCEP 20-member Global Ensemble Forecast System (GEFS, NCEP). The GEFS ensemble is used 50 times and combined with 1000 additional random perturbations. Atmospheric states for the computation of random perturbations were obtained from the Climate Forecast System Reanalysis (CFSR) dataset (Saha et al. 2010) in the period between 2006 and 2009. All simulations use the Scalable Computing for Advanced Library and Environment Regional Model (SCALE-RM) (Nishizawa et al. 2015; Sato et al. 2015) and have been computed on the K-Computer in Kobe, Japan (Miyoshi et al. 2016b). SCALE-RM is set up in two different domains both centered over Germany (Fig. 2a). The data assimilation cycling (CY) is done in the outer domain, which has 100 × 100 grid points, 31 vertical levels and a grid spacing of 15 km. The applied data assimilation method is a localized ensemble transform Kalman filter (LETKF) (Hunt et al. 2007) that assimilates conventional observations in a 3-hourly cycling using the SCALE-LETKF system (Miyoshi et al. 2016b; Lien et al. 2017). The convective-scale forecasts (FC) are performed in the inner domain. This domain has 350 × 250 grid points with 30 vertical levels and a grid spacing of 3 km. The convective-scale analysis ensembles are generated by downscaling. The 3-km forecasts are driven by 15 km mesh size forecasts carried out in the outer cycling domain. In total, we computed 10 different 14-h 1000-member ensemble forecasts initialized every 12 h from 0000 UTC 29 May to 1200 UTC 2 June 2016. Further details on the 1000-member ensemble simulation are provided in N20.

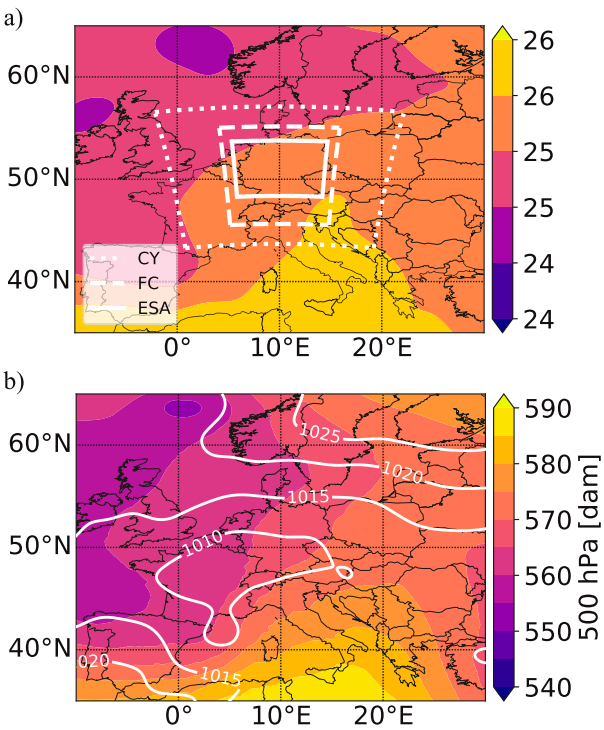

FIG. 2. Synoptic overview using ECMWF IFS analysis. (a) Temperature at 500 hPa (shaded, K) as well as borders of the cycling domain (CY; white dotted), the forecast domain (FC; white dashed), and the ESA domain (ESA; white solid). (b) Geopotential height at 500 hPa (shaded, dam) and sea level pressure (white contour, hPa) at 0000 UTC 29 May 2016.

b. *Synoptic overview*

During the 5-day period from 29 May to 3 June 2016, Europe was influenced by an atmospheric blocking

situation over the Atlantic (Piper et al. 2016). During the blocking, an upper-level trough developed leading to a cutoff low over central Europe. On 29 May this low pressure system was located over France moving eastward toward Germany (Fig. 2b) including advection of warm and moist air masses from southern Europe toward central Europe (Fig. 2a). As a consequence, all evaluated days are characterized by synoptic instabilities that featured strong convective lifting causing extreme weather events accompanied by flash floods, landslides, hail, and tornadoes. The entire period revealed weak pressure gradients and low wind speeds in the midtroposphere and consequently exceptionally strong precipitation rates in some regions.

For the visual analysis in section 4, we present sensitivities calculated for a nocturnal precipitation event that occurred on 29 May 2016. During that time, the upper-level trough approached Germany leading to advection of positive vorticity as well as warm and moist air masses at midtropospheric levels. This exceptional period with several high-impact weather events has also been the subject of several other studies (Rasp et al. 2018; Necker et al. 2018; Keil et al. 2019; Bachmann et al. 2019).

c. *Ensemble sensitivity analysis setup*

Our study evaluates spatiotemporal correlations in the context of ESA. Correlations are calculated using hourly precipitation as forecast response function. The precipitation metric is averaged over boxes of 40 × 40 grid points (see the box in Fig. 3a) to account for the model resolvable scale of precipitation. The 1-h forecast is used as the initial state for the ensemble sensitivity calculation to avoid potential spinup effects within the first hour of the model integration (N20). Furthermore, a slightly reduced domain is used for the ESA calculations that extend over an area of 200 × 200 grid points located in the center of the forecast domain ESA (Fig. 2a) to exclude potential nesting effects. All results are compared using four different ensemble sizes (40, 80, 200, and 1000 members) and various atmospheric variables. Smaller ensembles are generated by subsampling from the 1000-member ensemble such that the GEFS members are equally represented within the subsets. This means that the 40-member subset contains each member of the GEFS BC two times. The 80-member subset consists of the 40-member subset plus 40 additional members, and the 200-member subset combines the 80-member subset plus 120 additional members.

4. Spatiotemporal correlations

This section evaluates the SEC for spatiotemporal correlations and compares its effect to that of a confidence test, which has often been applied to reduce sampling errors in previous studies.

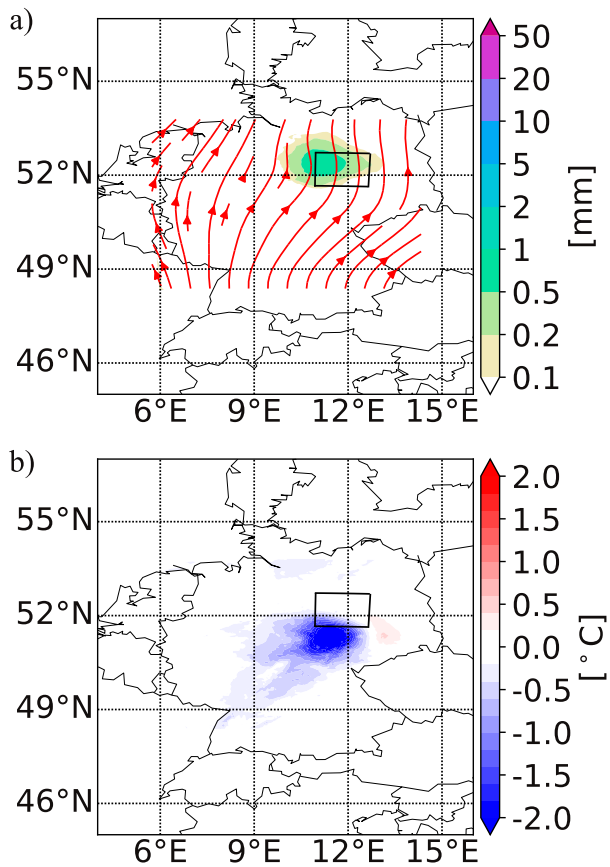

FIG. 3. (a) 1000-member ensemble mean precipitation and streamlines of 500-hPa wind (0400 UTC 29 May 2016). (b) Initial 2-m temperature anomaly calculated comparing the ensemble mean 2-m temperature of the 100 members with strongest and 100 members with weakest precipitation inside the forecast response function (0100 UTC 29 May 2016).

a. *Example of correlation fields*

We start with a qualitative analysis of spatiotemporal correlation for the first forecast initialized at 0000 UTC 29 May 2016. Figure 4 displays sensitivities of the 3-h precipitation forecast (Fig. 3a) to the initial 2-m temperature field calculated for different ensemble sizes and with different sampling error approaches. The differences compared to the 1000-member ensemble correlation (Fig. 4a) illustrate the effect of sampling errors.

The 1000-member ensemble shows strong negative correlations of precipitation to the initial 2-m temperatures in a region southwest of the response function. These negative correlations are related to evaporative cooling caused by precipitation resulting in colder surface temperatures in this area. Clustering the 100 members with the strongest and weakest precipitation inside the

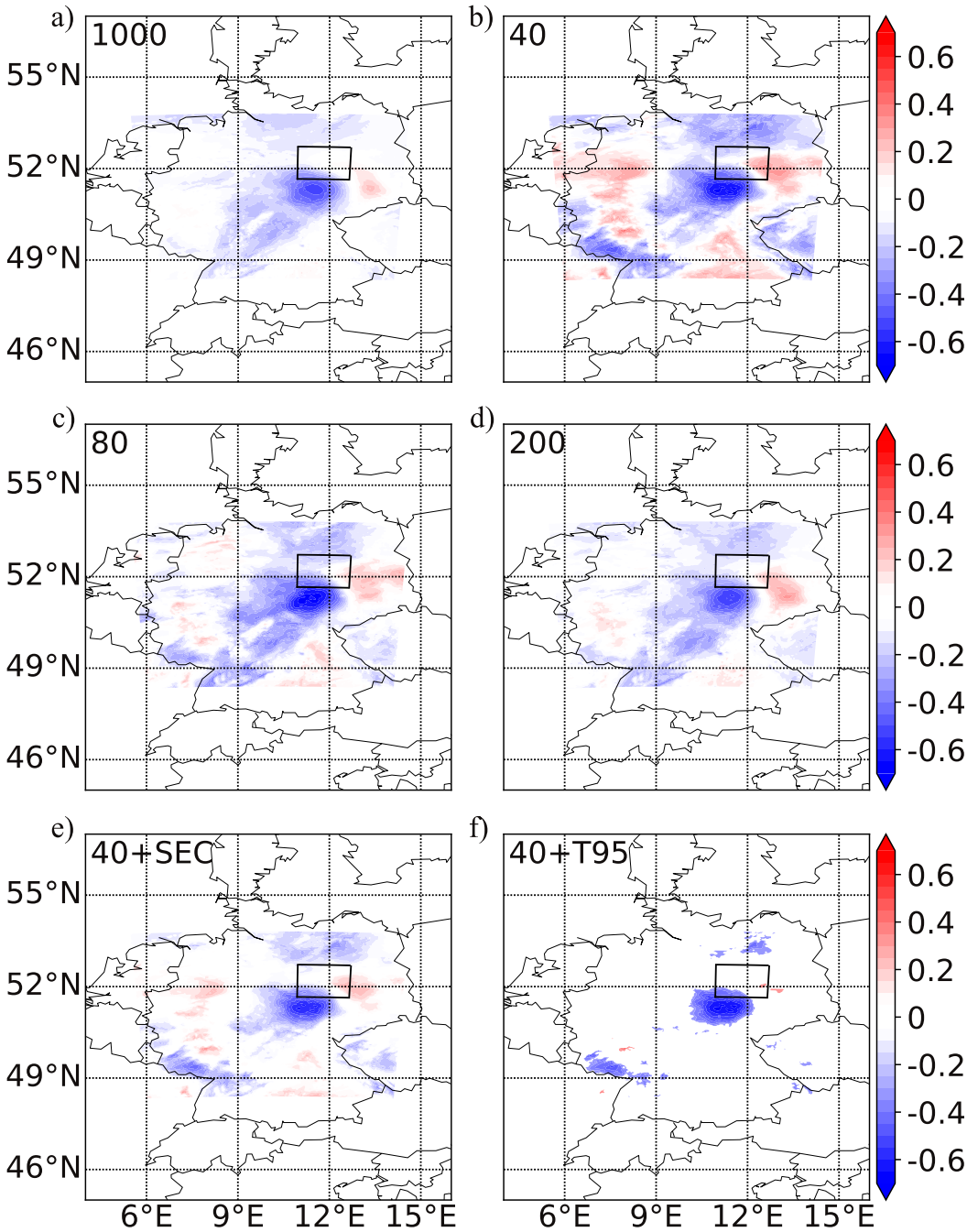

FIG. 4. Correlation of the 3-h precipitation forecast to the initial 2-m temperature field at 0100 UTC 29 May 2016 for different ensemble configurations: (a) 1000 members, (b) 40 members, (c) 80 members, (d) 200 members, (e) 40 members with SEC, and (f) 40 members with T95.

response function reveals a temperature anomaly in the initial surface temperature field (Fig. 3b) that matches the area of negative sensitivities. The southwesterly tail of negative correlations roughly marks the track of the precipitating systems during the night (Fig. 3a). This corresponds to the southwesterly wind indicated by streamlines in Fig. 3a. The region with positive correlations southeast of the response function is related to a westward shift of precipitation in some of the ensemble members. This effect is stronger for shorter lead times (not shown).

In contrast, the 40-member ensemble correlation field (Fig. 4b) exhibits various spurious correlations

in the south and west of the domain. Furthermore, the small ensemble systematically overestimates the amplitude of sensitivities in several locations. Increasing the ensemble size to 80 or 200 ensemble member (Figs. 4c,d) systematically reduces the number of spurious correlations at larger distances from the precipitation event. However, some small positive spurious correlations are still visible for the 200-member ensemble.

Figure 4e shows the 40-member ensemble correlation field corrected with the SEC. The SEC is able to reduce several spurious correlations and also corrects the amplitude of the strongest negative correlations. However, it also affects the tail southwest of the area of maximum correlation. Applying the confidence test (T95) to the 40-member ensemble correlation field (Fig. 4f) removes all correlations approximately smaller than ±0.25 and returns an incomplete correlation field. Compared to the SEC, the confidence test eliminates nearly all positive correlations and also removes the entire tail. Nevertheless, some spurious correlations at the French–German border remain as those exhibit comparably large correlation values. Furthermore, the T95 does not correct the amplitude of the strongest correlation. Results for other variables are overall similar (not shown).

b. Correlation distribution

Figure 5 shows four different correlation frequency distributions. The histograms are calculated using correlations from all 10 available 3-h lead-time forecasts and 2-m temperature as the target state variable. The distribution of the 40-member ensemble nearly resamples the shape of a normal distribution peaking slightly shifted toward negative values. The 1000-member ensemble distribution peaks at a similar position but showing an approximately three times higher amplitude combined with a smaller width. Applying the SEC to the 40-member ensemble correlations improves the distribution substantially. The width and the amplitude of the peak are now similar to the 1000-member ensemble but slightly shifted toward zero. The shift of the peak originates from the assumed uniform prior $U(-1, 1)$ and could be reduced by using a more informed prior assumption when calculating the systematic error correction offline. This does especially make sense for highly positively or negatively correlated fields. Both, a climatological prior (Anderson 2016) or a prior obtained from a larger ensemble can be used to generate a more specified table (see section 5e).

Filtering all unreliable 40-member ensemble correlations using the confidence test (T95) changes the distribution fundamentally. The confidence test removes all sensitivities smaller than approximately ±0.25 and therefore discards the majority of correlations. Comparing

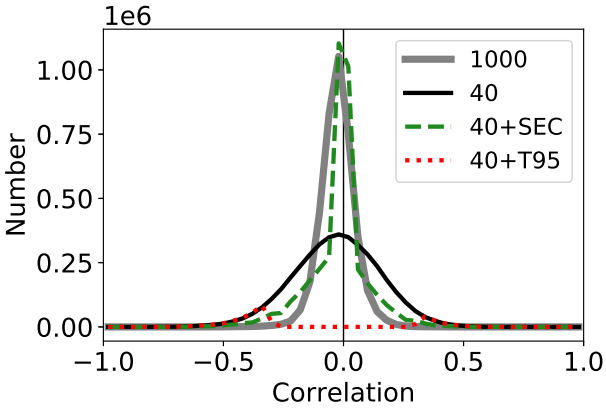

FIG. 5. Frequency distributions for correlations of the 3-h precipitation forecast to the initial 2-m temperature field using all 10 forecasts. 1000-member ensemble correlations (bold solid gray) and 40-member ensemble (solid black) including SEC (green dashed) or T95 (red dashed).

both approaches, the SEC substantially improves the distribution whereas the application of the T95 leads to an unrealistic distribution of correlations. The effect is similar for correlation distributions of other variables (not shown).

c. Sampling error as function of correlation value

Figure 6 presents the mean absolute correlation error as a function of the 40-member (left column) and 1000-member ensemble correlation (right column). The sampling error of 2-m temperature (Fig. 6a) evaluated as a function of the 40-member ensemble correlation is smallest for small correlation values and largest for large correlations. The SEC systematically reduces the sampling error independent of the strength of the correlation. The improvements achieved by the SEC correspond to the impact that can be expected according to the correction curve (see Fig. 1). The performance of the SEC is similar for other variables (Figs. 6c,e).

The 40-member ensemble sampling error for 2-m temperature (Fig. 6b) plotted as a function of the 1000-member ensemble correlation is smallest for large negative correlation values and largest for strong positive correlations. Applying the SEC largely reduces the error for small correlation values but slightly degrades the performance for large positive correlations. However, results for large correlation values should be treated with caution as there are only a few data points (see frequency distribution in Fig. 5). The absolute error obtained for correlations of precipitation with 500-hPa temperature (Fig. 6d) looks similar as for surface temperature. Again, the SEC mainly improves small correlation values, whereas for 500-hPa zonal wind (Fig. 6f) improvements are visible for the entire range of correlation values. For very small correlation values, the SEC almost

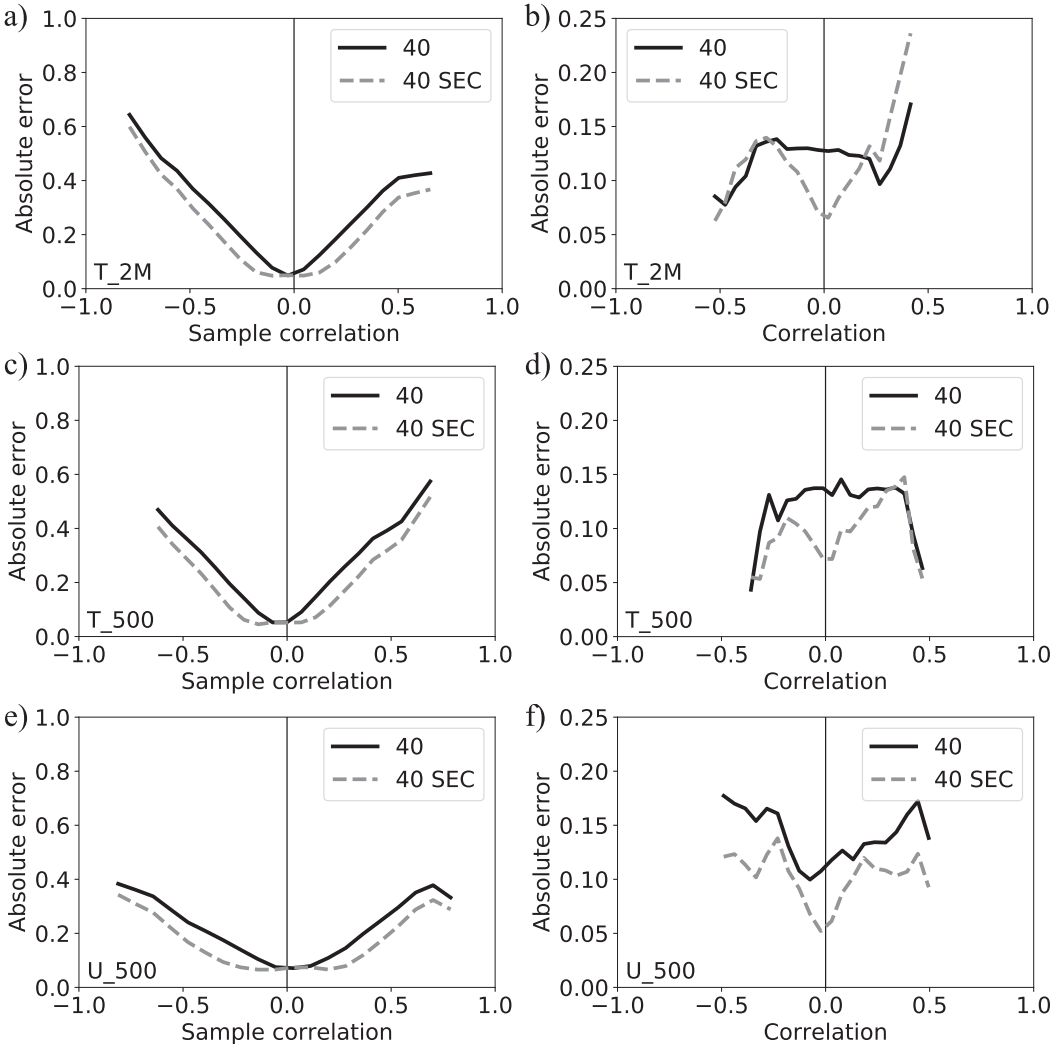

FIG. 6. Mean absolute error of the 40-member sample correlation (solid black) and sampling error corrected correlation (gray dashed) as a function of the (a),(c),(e) 40-member and (b),(d),(f) 1000-member ensemble correlation. Correlations of the 3h precipitation forecast to initial (a),(b) 2-m temperature; (c),(d) 500-hPa temperature; and (e),(f) 500-hPa zonal wind using all 10 forecasts.

halves the sampling error compared to the 1000-member ensemble correlation. However, one should keep in mind the relatively small sample of evaluated large correlation values of the 1000-member ensemble.

In summary, the SEC based on a uniform prior has its strongest effect on small correlation values, which seems reasonable considering the correction function displayed in Fig. 1. For larger correlation values, the effect of the SEC gets smaller and differs depending on the considered variable.

d. Sensitivity to ensemble size

Figure 7a presents the time-averaged root-mean-square error (RMSE) of correlations as a function of ensemble size and investigates the same correlations as shown in the previous two sections (precipitation correlated with 2-m temperature). Here, the RMSE of a 40-member ensemble is given by

$$\mathrm{RMSE}_{40} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{r}_{40,n} - r_n)^2},$$

where we assume $r = \hat{r}_{1000}$ and N is the number of grid points in the domain. The RMSE is calculated using correlations obtained for the full 1000-member ensemble for verification. The RMSE of the 40-member ensemble is approximately 0.16. Doubling the sample size up to 80 members reduces the RMSE by about 30% whereas increasing the sample size by the factor of 5 up to 200 members lowers the RMSE by more than 50%.

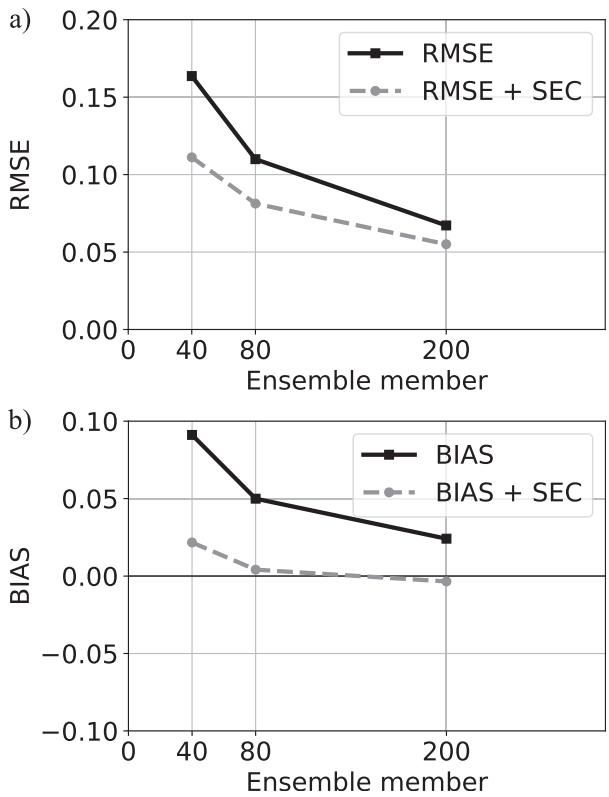

FIG. 7. Time-averaged (a) root-mean-square error and (b) magnitude bias of correlations with and without SEC compared to 1000 members evaluated for different ensemble subsets. Spatiotemporal correlations of precipitation to 2-m temperature.

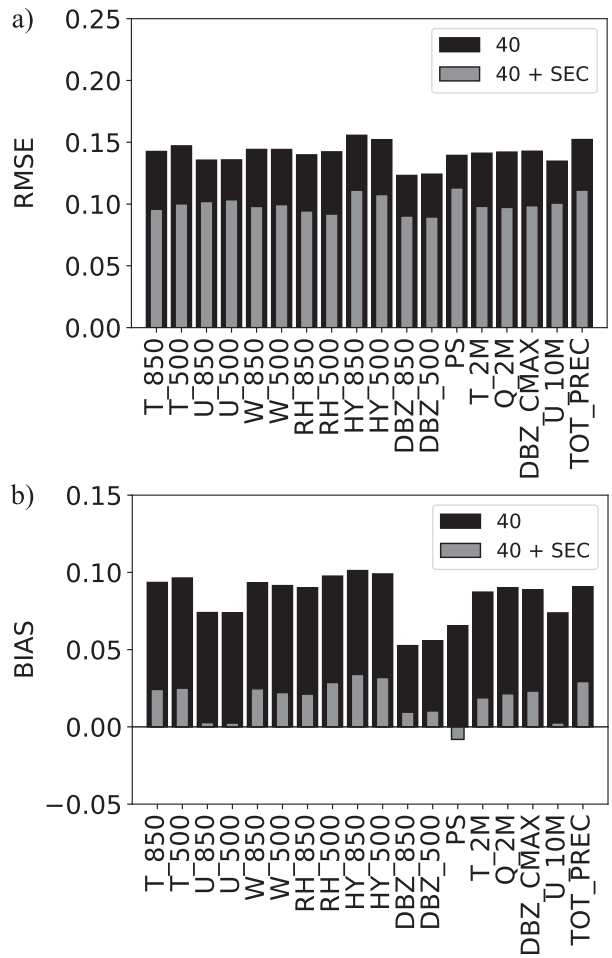

FIG. 8. Time-averaged (a) root-mean-square error and (b) magnitude bias of 40-member precipitation correlation to various variables with and without SEC. A list of variable abbreviations is provided in the appendix of this manuscript.

For small ensemble samples, the SEC strongly improves the performance. Applying the SEC to the 40-member ensemble subset even achieves slightly better results than doubling the ensemble size. The reduction of RMSE due to the SEC decreases with increasing ensemble size. Nevertheless, the 200-member RMSE is still reduced by about 15% by the SEC.

Figure 7b shows the corresponding time-averaged difference of the mean absolute correlation (BIAS) compared to the 1000-member ensemble for all six configurations. Here,

$$\mathrm{BIAS}_{40} = \frac{1}{N}\left(\sum_{n=1}^{N} |\hat{r}_{40,n}| - \sum_{n=1}^{N} |r_n|\right).$$

Similar to the RMSE, the BIAS decreases with increasing ensemble size and applying the SEC largely reduces the BIAS. For nearly all subsets, the BIAS almost vanishes. For larger subsets, the SEC also reduces the bias causing a change in sign. Nevertheless, the improvements due to the SEC are substantial and visible for all variables. Different prior assumptions used for computing the SEC table could presumably improve the results further.

e. *Sensitivity to variable*

Figure 8a presents the RMSE for 40-member correlations of precipitation to various initial quantities. The black and gray bars displayed for 2-m temperature coincide with the markers of the 40-member ensemble shown in Fig. 7a. The RMSE for all variables ranges from approximately 0.13 to 0.18. As discussed for 2-m temperature, correcting the correlations using the SEC substantially reduces the RMSE independent of the chosen variable. The improvements range from about 20% to 30% and are smallest for sea level pressure (PS).

Examining the BIAS (Fig. 8b), sea level pressure is the only variable that exhibits a change in sign of the bias. This is likely related to the structure of the correlation field, which is homogeneously distributed over the entire domain as sea level pressure exhibits a fairly smooth large-scale field consisting of mainly small negative

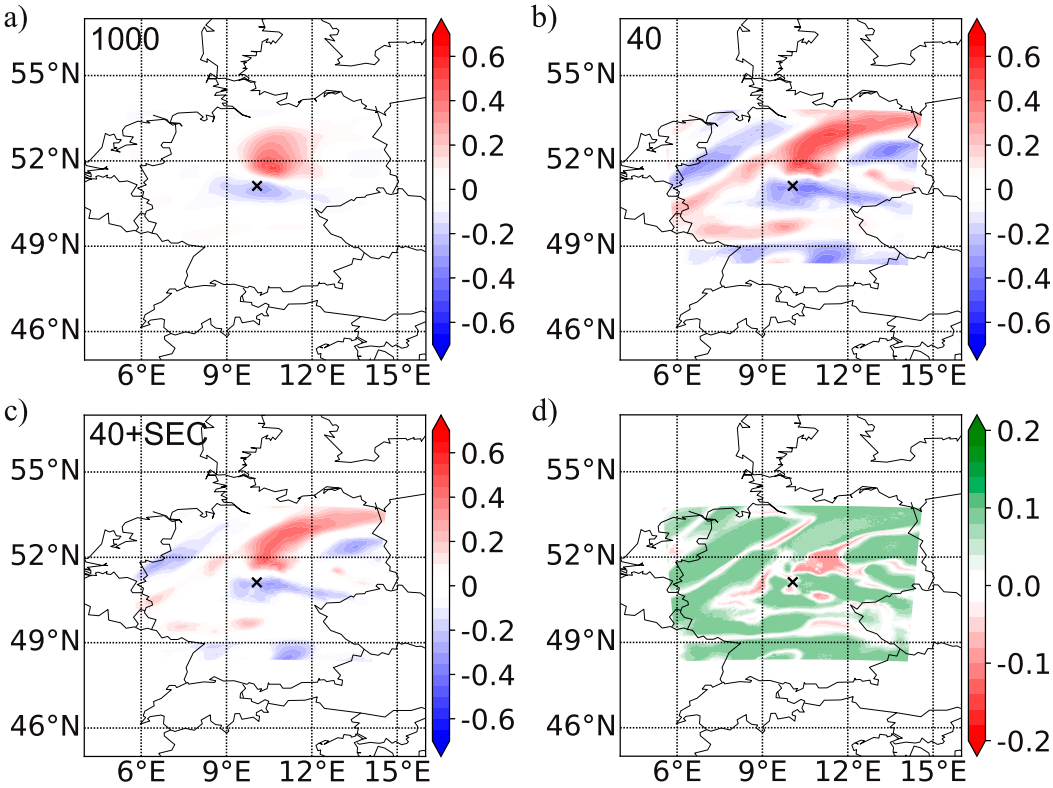

FIG. 9. Cross correlation of 500-hPa temperature at a single grid point (black marker) to 500-hPa specific humidity in the ESA domain at 0100 UTC 29 May 2016 for (a) 1000 members, (b) 40 members, and (c) 40 members including SEC as well as (d) changes in correlation field due to the SEC (green—error reduction).

correlations. The SEC systematically reduces the BIAS for all variables and works most efficiently for zonal wind. Examining the impact of the SEC on the 80 and 200-member ensemble correlations (not shown), the systematic reduction of the BIAS relatively increases with increasing ensemble size leading to changes in sign as discussed for 2-m temperature (Fig. 7b). Nevertheless, the reduction of BIAS is substantial for all investigated ensembles sizes and variables.

Further sensitivity studies have been conducted that are not shown in this manuscript. These experiments targeted the sensitivity of the SEC to the precipitation metric kernel size, the choice of the ensemble subset as well as the dependence on forecast lead time. However, these sensitivity studies are not discussed here as these experiments did not reveal any fundamentally different results.

5. Spatial correlations

This section investigates the impact of the SEC on spatial correlations that are crucial for ensemble or hybrid DA. Results are shown for the correlation of temperature to various model variables. Spatial correlations are calculated using 1-h forecasts, which is similar to taking the first guess during hourly cycling. The performance of the SEC is compared to a standard distance-based localization approach (LOC). Furthermore, the sensitivity of the SEC to different prior assumptions is examined.

a. *Example of spatial correlations*

Figure 9a displays horizontal cross correlations of 500-hPa temperature at a single grid point to 500-hPa specific humidity at every grid point in the domain. The correlation pattern is a dipole showing a negative correlation in the vicinity and a positive correlation to the north of the response function. Except for the dipole, no other considerable correlations are visible. Examining the 40-member correlations (Fig. 9b), various spurious correlations show up all over the domain, similar as discussed for spatiotemporal correlation in section 4a. To some degree, the dipole is still indicated by the strongest correlations. Applying the SEC (Fig. 9c) reduces the number of spurious correlations strongly and reveals the dipole more distinctly. Overall, the SEC can reduce the sampling error for the majority of grid points (Fig. 9d) showing slightly increased errors only in

some small areas. The improvements are consistent for spatial correlations to other variables (not shown) and agree with the results obtained for spatiotemporal correlations considering a precipitation-based response function.

b. Horizontal correlation

Below, horizontal correlations are averaged using the 10 available 1-h 1000-member forecasts. Each ensemble forecast is evaluated with nine gridpoint size metrics that are evenly distributed in the domain with a distance of 50 grid points (150 km) to neighboring metrics and boundaries. In total, 90 correlation fields are examined for each variable pair. The mean absolute correlation (MAC) and error (MAE) for a given spatial distance are defined as

$$\mathrm{MAC}_m = \frac{1}{N}\sum_{n=1}^{N} |\hat{r}_m^n|$$

and

$$\mathrm{MAE}_m = \frac{1}{N}\sum_{n=1}^{N} |\hat{r}_m^n - r^n|,$$

where again we assume $r = \hat{r}_{1000}^n$. The correlations are binned into annuli, each with a width of 13 km; N specifies the number of grid points in each annulus.

Figure 10a shows the mean absolute correlation of 2-m temperature to 2-m temperatures and Fig. 10b the corresponding error with and without SEC or LOC. The 1000-member ensemble exhibits a correlation of nearly 1 in the close vicinity of the response function dropping to a value of about 0.4 at a distance of 100 km. Up to 100 km, the 40-member ensemble correlation coincides with the 1000-member ensemble correlation. Farther away, the 40-member ensemble systematically overestimates the mean absolute correlation due to spurious correlations. The mean absolute correlation error (Fig. 10b) strongly increases up to a distance of 100 km, which roughly matches the radius of horizontal localization applied in regional DA systems. For distances larger than 100 km, the sampling error keeps increasing, but slower compared to the vicinity of the response function. Applying the SEC increases the error close to the response function slightly, but strongly reduces the error at larger distances. Similar changes are visible for the mean correlation. Especially for distances larger than 150 km, the sampling error corrected 40-member mean absolute correlation almost coincides with the 1000-member correlation.

The mean absolute cross correlation of 2-m temperature to 10-m zonal wind (Fig. 10c) and 2-m temperature to near-surface humidity (Fig. 10e) show similar results. Both variables exhibit the strongest correlation in the near vicinity dropping to a constant value of approximately 0.2 at a distance of 150 km. The mean absolute errors (Figs. 10d,f) slightly change with distance showing a similar absolute value for large distances as found in Fig. 10b. However, the relative error is larger considering the weak mean absolute correlation for these pairs. Applying the SEC substantially improves both the mean and error of the spatial cross correlations. The SEC performs best for distances larger than 100 km reducing the error of the humidity cross correlation by up to 40%.

The correlation of 2-m temperature to sea level pressure (Fig. 10g) is weaker compared to spatial correlations discussed previously. Mean absolute correlation and error (Fig. 10h) hardly change with distance. Due to sampling errors, the 40-member mean correlation is twice as large as the 1000-member mean correlation. The SEC substantially improves the 40-member mean correlation, which is now close to the 1000-member mean correlation. The absolute error decreases by approximately 20%.

Applying a distance-based localization (LOC), the mean absolute correlation drops to zero at a distance of 200 km for all variable combinations (Figs. 10a,c,e,g). For short distances, the LOC overestimates the mean absolute correlation for the majority of variables while it systematically underestimates the mean absolute correlation for large distances. Using different localization scales for different variables could improve the performance of the LOC. Overall, the SEC is able to match the 1000-member mean absolute correlation best.

For correlations of 2-m temperature to 2-m temperatures, the LOC increases the error for distances shorter than 200 km (Fig. 10b). For cross correlation to other variables, no degradation by the LOC is found (Figs. 10d,f,h). For most variables, the SEC performs best on short distances while the LOC seems to outperform the SEC for distances larger than 250 km.

Figure 11 shows the mean absolute correlation and error as a function of horizontal distance using correlations of 500-hPa temperature to different upper-tropospheric variables. Both the 1000 and 40-member ensemble correlation decline consistently examining spatial correlations of 500-hPa temperature (Fig. 11a). The magnitude of the correlation is larger than for all other discussed quantities. Furthermore, the 40-member mean absolute correlation error is smaller, grows less rapidly and does not appear saturated at a horizontal distance of 500 km (Fig. 11b). In contrast to other variables, applying the SEC degrades the performance for the entire spatial range. The mean absolute correlation is now underestimated, and the error increases correspondingly.

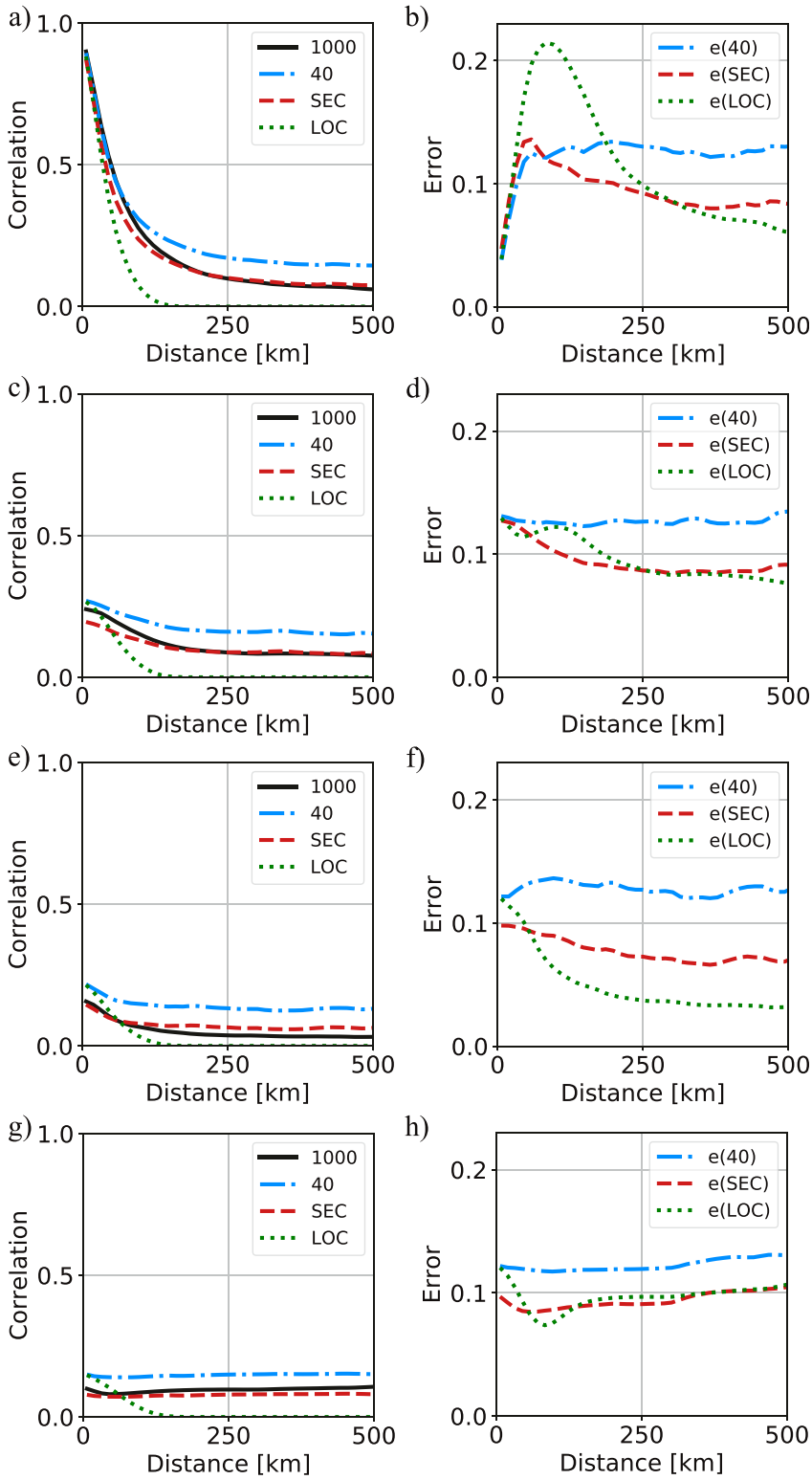

FIG. 10. Mean absolute (left) correlation and (right) error as function of spatial distance (km) for different ensembles with and without SEC or LOC. Correlation of 2-m temperature to (a),(b) 2-m temperature, (c),(d) 10-m zonal wind, (e),(f) 925-hPa specific humidity and (g),(h) sea level pressure.

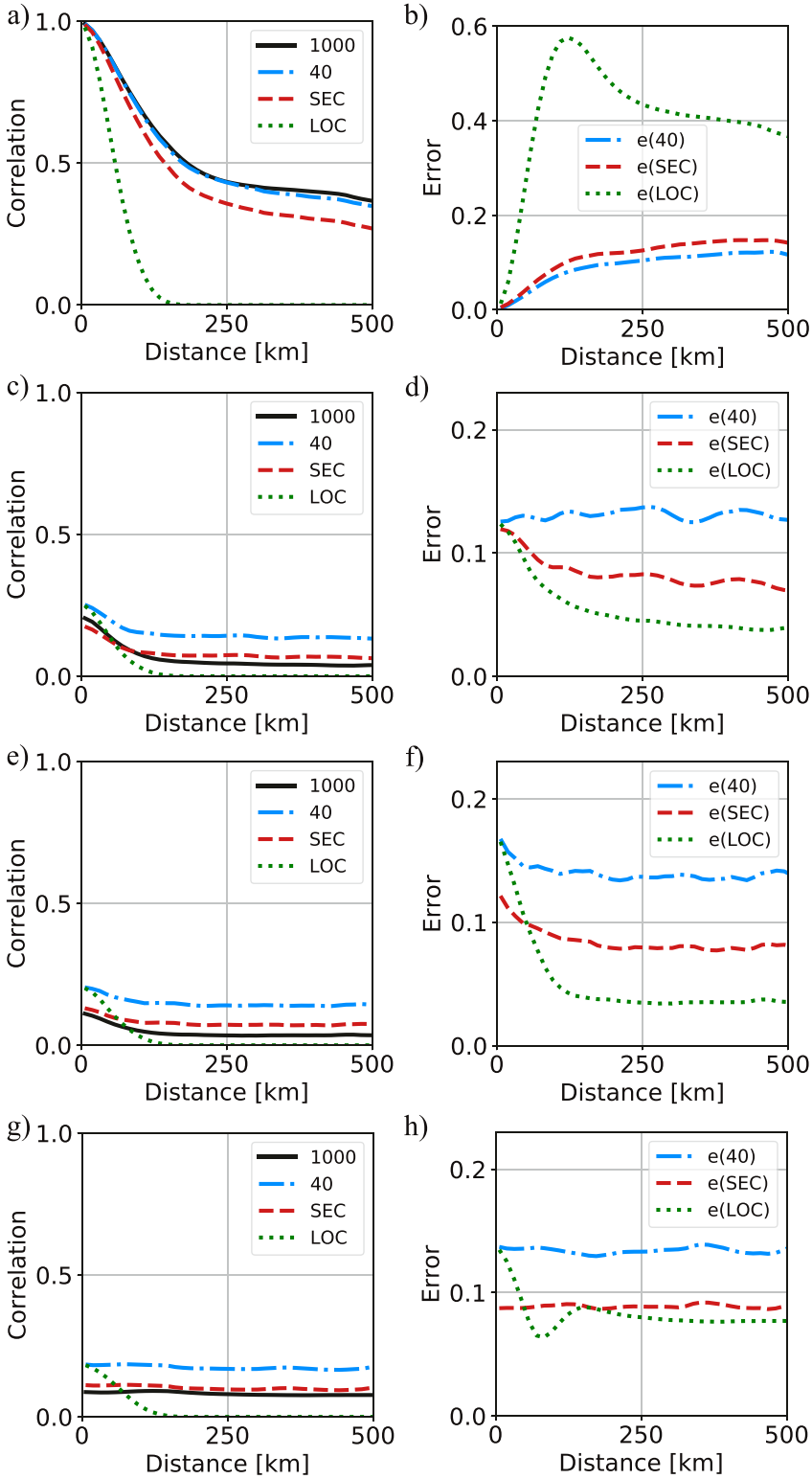

FIG. 11. As in Fig. 10, but spatial correlation of 500-hPa temperature to (a),(b) 500-hPa temperature, (c),(d) 500-hPa specific humidity, (e),(f) 500-hPa hydrometeors, and (g),(h) 500-hPa zonal wind.

The negative impact of the SEC is caused by an insufficient prior assumption, which affects the behavior of the SEC. In this case, a uniformly distributed prior appears to be unsuitable. The impact of different priors is discussed in more detail in section 5e.

Figures 11c and 11d analyze horizontal cross correlations of 500-hPa temperature to 500-hPa specific humidity. Again, the mean absolute correlation decreases with increasing distance. The SEC reduces both mean and error showing an improved performance far from the response function. Cross correlations of 500-hPa temperature to 500-hPa hydrometeors (Fig. 11e) are weaker compared to cross correlations of temperature and humidity. As before, the SEC reduces the error (Fig. 11f) while it slightly overadjusts the mean absolute correlation. The results for cross correlations of 500-hPa temperature to 500-hPa zonal wind (Figs. 11g,h) are similar as discussed for cross correlations of 2-m temperature to sea level pressure (Fig. 10) although the mean absolute cross correlations and errors are slightly larger in this case.

The LOC substantially degrades correlations of 500-hPa temperature to 500-hPa temperature (Figs. 11a,b). For tropospheric temperature, a considerably larger localization scale is required compared to other variable pairs. Again, the SEC performs best evaluating the mean absolute correlation independent of the variable (Figs. 11a,c,e,g). For cross correlations of 500-hPa temperature to other variables the LOC performs best for distances larger than 100 km (Figs. 11d,f,h).

Overall, the SEC reduces the sampling error for the majority of horizontal (cross) correlations using a uniformly distributed prior as applied in this case. Furthermore, the SEC shows a large impact on cross correlations and distances larger than 100 km. Only strongly positively correlated variables revealed ambiguous results. Yet, this can be addressed by a different prior assumption or the exclusion of these variables from the correction. The LOC performs best on distances larger than about 250 km. On short distances, the SEC outperforms the LOC. Given this, a combination of both approaches seems to be most promising.

c. *Vertical correlation*

Vertical correlations are evaluated using a single 1000-member ensemble forecast on 1300 UTC 30 May 2016 and in total 40 000 vertical profiles. For vertical correlations, we focus on spatial correlations of 500-hPa temperature to 20 different pressure levels and four different variables. Figure 12 shows 1000-member ensemble mean correlation and the RMSE of vertical temperature correlations comparing the 40- and 1000-member ensemble for different configurations. The RMSE of the temperature correlated with itself is zero at the 500-hPa response level (Fig. 12a) as both 40 and 1000 members exhibit a correlation of 1. The RMSE of the 40-member ensemble correlation increases to a value of 0.15 at a vertical distance of 100 hPa and seems to be saturated for distances larger than 150 hPa. The error applying the SEC increases slower and saturates earlier reducing the relative error far from the response level up to 30%. Only at 350 hPa, the SEC increases the RMSE as the 40-member ensemble subset on average slightly underestimates the true correlation (not shown). For vertical correlations, the SEC performs better than the LOC in the vicinity of the response level while the LOC seems to be more beneficial with increasing distance.

Figure 12b shows the RMSE for vertical cross correlations of temperature at 500 hPa to specific humidity in the entire tropospheric column. Compared to the previous example, the RMSE for the 40-member ensemble does not exhibit a local minimum at 500 hPa and hardly changes with height. Applying the SEC reduces the RMSE at all levels, but the reduction is smallest at the 500-hPa response level. The RMSE reduction increases up to a vertical distance of 150 hPa and again hardly changes far from the response level. Evaluating vertical cross correlations of temperature to hydrometeors (Fig. 12c) or zonal wind (Fig. 12d) the effect of the SEC is independent of the vertical distance and the SEC substantially reduces the RMSE at all levels by about 30%.

In general, the impact of the SEC is largest for vertical cross correlations and far from the response level. As for horizontal correlations, the SEC outperforms the LOC on short distances while the LOC appears to be more beneficial with increasing distance (in case of no strong long-range correlations). The error is roughly symmetric comparing results above and below the response level. On average, the SEC efficiently reduces the overestimation of the true correlation due to spurious correlations. We hypothesize that the SEC becomes advantageous if correlations extend over the full vertical profile of the atmosphere (e.g., for passive satellite observations). In such situations, localization techniques are potentially dangerous as they damp or eliminate correlations after a certain distance. The same applies to cloud information, which can affect the surface as well as the entire tropospheric column by modified radiative processes.

d. *Sampling error correction as function of correlation value*

Figure 13 displays the 2D correlation frequency distribution comparing the 1000-member ensemble spatial correlations with corresponding spatial correlations obtained for different ensemble subsets. Each analysis includes approximately 38 million spatial correlations

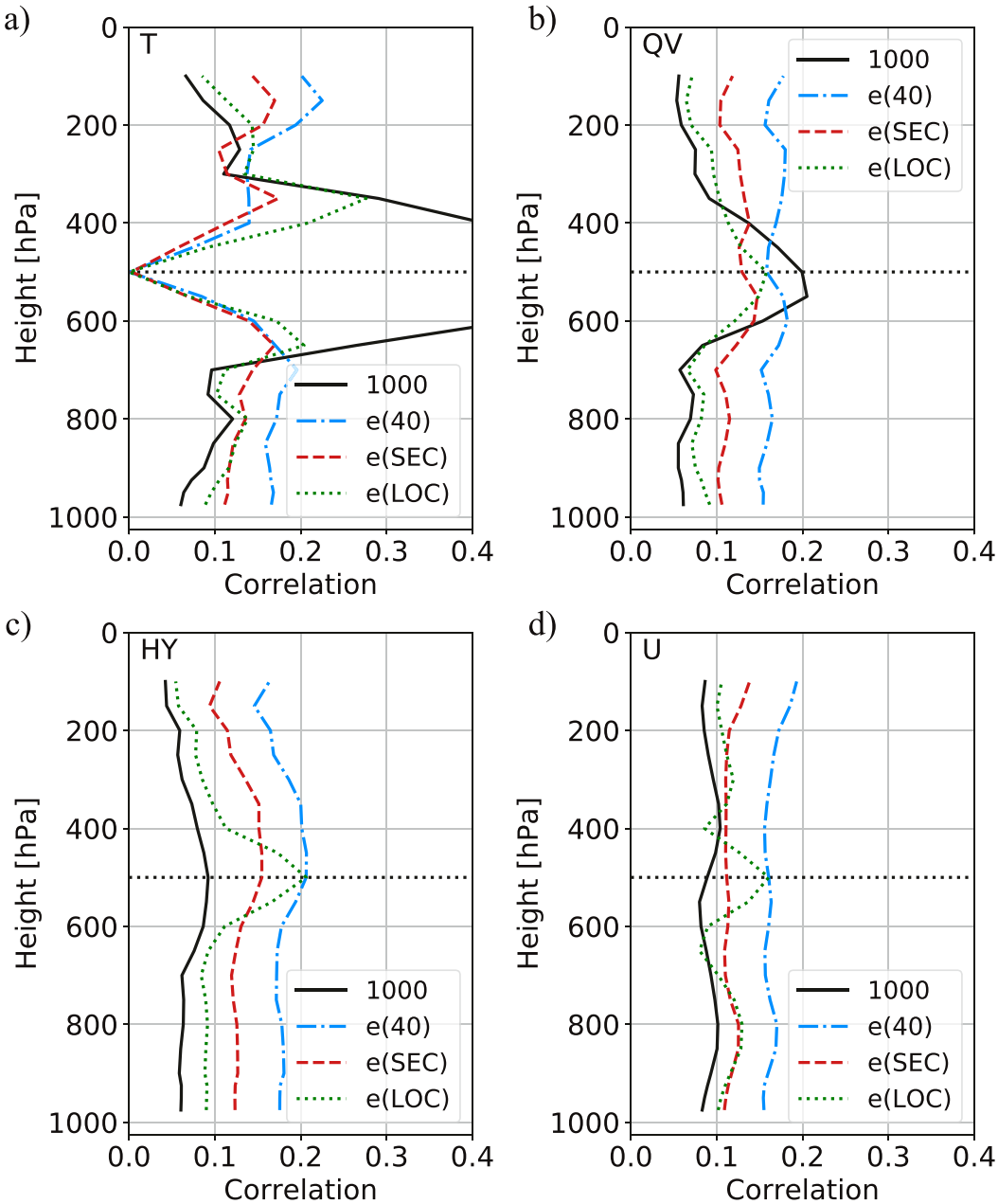

FIG. 12. Root-mean-square error of the 40-member correlation compared to the 1000-member correlation with (red, dotted) and without (blue, dashed) SEC and for a distance-based localization (LOC). Correlation of 500-hPa temperature to (a) temperature, (b) specific humidity, (c) hydrometeors, and (d) zonal wind at different height levels. RMSE averaged over 40 000 vertical profiles. Note: The black solid line displays the mean absolute correlation for 1000 members.

of temperatures at 500 hPa to various other variables. Each frequency distribution exhibits a maximum at small correlation values. Positive correlations range from 0 up to 1, while the largest negative 1000-member correlation is approximately −0.5. For the 40-member ensemble (Fig. 13a), the maximum around zero is elongated in the horizontal direction indicating the overestimation of small correlations due to spurious correlations. Applying the SEC reduces this overestimation systematically and changes the pattern of the frequency distribution (Fig. 13b). The distribution peaking around zero is now narrow and extends vertically. The Pearson correlation coefficient between both correlation samples is displayed in the corner of

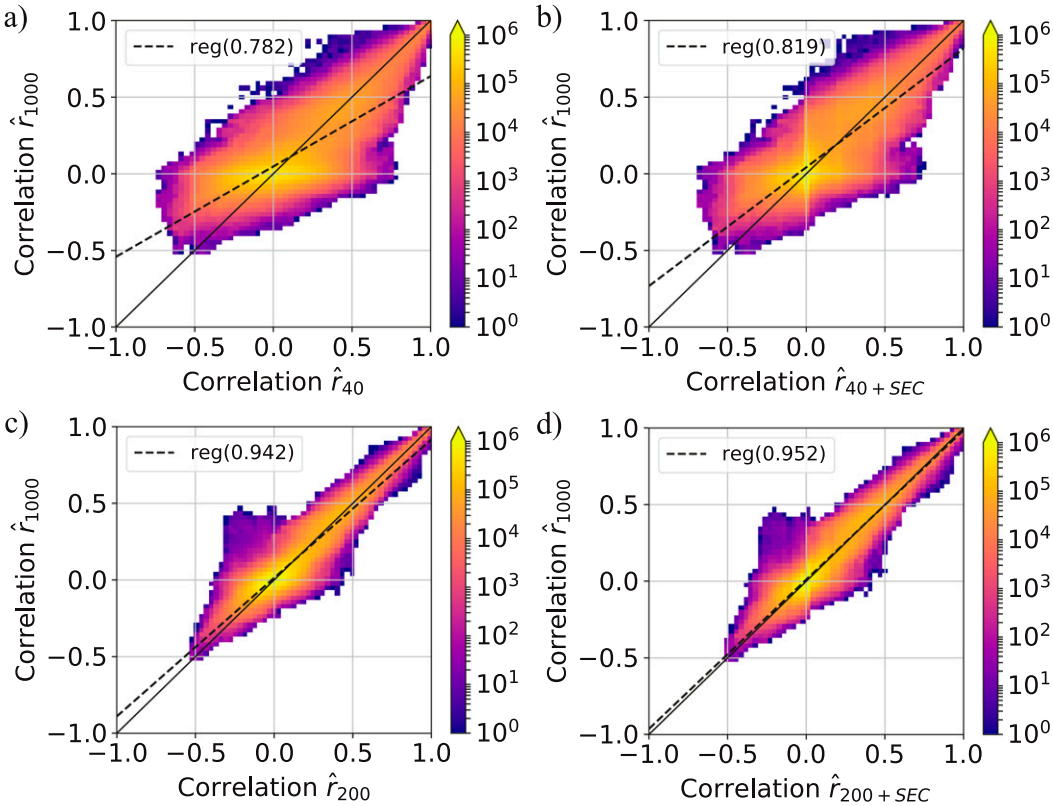

FIG. 13. Two-dimensional frequency distribution comparing correlations of the 1000-member ensemble and different subsets: (a) 40 members, (b) 40 members with SEC, (c) 200 members, and (d) 200 members with SEC. The analysis includes about 38 million spatial correlations of temperature at 500 hPa to temperature, specific humidity, hydrometeors, zonal wind, sea level pressure, and precipitation. Slope of the linear regression fit (dashed line).

each subfigure to facilitate the comparison. Plotting the linear regression line (dashed line) reveals the impact of the SEC as it improves both the slope and the intersection as the SEC reduces the magnitude bias. Overall, the SEC improves the performance of the 40-member ensemble by about 5% using the Pearson correlation as a measure.

Figure 13c shows the frequency distribution comparing 200 with 1000 members. Using 200 members substantially reduces sampling errors for the entire range of correlation values. Increasing the ensemble size by a factor of 5 especially improves the estimation of small correlation values. The 200-member ensemble exhibits a maximum offset of approximately 0.4, which is substantially less than found for 40 members. Applying the SEC (Fig. 13d) again improves the frequency distribution systematically. The absolute impact is smaller compared to 40 members, but the improvements are particularly visible for small correlations as well as in the slope of the linear regression line.

Considering that the SEC showed ambiguous results for some highly correlated variables (section 5b), it is important to assess if the SEC systematically fails for large correlation values. Figure 14a shows the change in the absolute correlation error caused by the SEC as a function of the 40-member absolute correlation value. The frequency distribution again reveals the greatest improvements for small correlation values. Both, negative and positive impacts mainly exhibit the strength of the maximum possible adjustment that is indicated by the dashed line and derived from the correcting function. Examining the average improvement, the SEC systematically improves the results independent of the amplitude of the 40-member correlation value. Overall, the SEC improves about three-quarter of the correlations.

Figure 14b shows the same data as before but now distributed as a function of the 1000-member absolute correlation value. Again, the main improvements are observed for small correlation values, and the overall impact is beneficial. However, the impact of the SEC seems to be detrimental for 1000-member correlation values larger than 0.25. Similar behavior is seen for vertical correlations (not shown). However, as the true

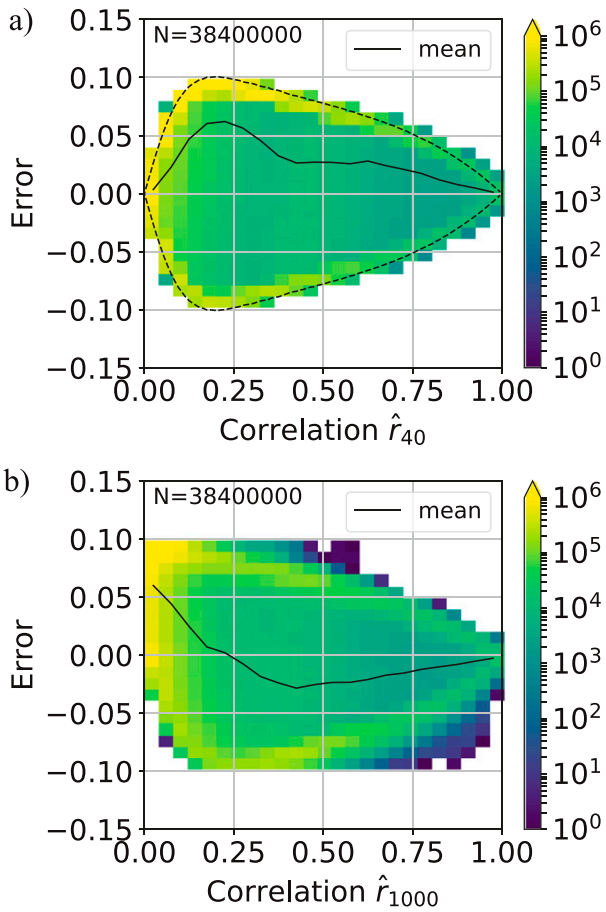

FIG. 14. Frequency distribution of error reduction δe applying the SEC to a 40-member ensemble as a function of the absolute value of the (a) 40-member or (b) 1000-member correlation. The solid black line shows the average change and the dashed line sketches the maximum expected adjustment, which is restricted by the correction function. The analysis considers the same correlations as in Fig. 13. The SEC improves δe for 72.3% of the correlations [$\delta e = |(\hat{r}_{40} - r)| - |(\hat{r}_{40+\mathrm{SEC}} - r)|$, where $r = \hat{r}_{1000}$].

correlation is usually unknown, it is difficult to use this behavior to improve such cases. Overall, results suggest that based on the available information from the small ensemble (Fig. 14a) the SEC should be applied to all correlations.

e. *Comparison of different priors*

As discussed in section 5b, the uniform prior appears to be unsuitable for variables exhibiting strong positive correlations. Already Anderson (2012) suggested using a more informative prior to improve the performance of the SEC in such cases. Recently, Anderson (2016) examined the impact of a climatological prior. This subsection evaluates the impact of different priors on the performance of the SEC. Here, we compare the impact of three different more informative priors: First, a perfect prior based on the 1000-member ensemble ($p1000$). Second, a climatological prior that is obtained using the 40-member ensemble correlations ($p0040$). Third, a distance-dependent prior based on 40-member sample correlations of short-range correlations (<100).

Figure 15b displays the sampling error corrected correlation $\hat{r}_{\mathrm{sec}}$ as a function of the sample correlation $\hat{r}$ for the four different priors. Note that the assumed priors do not cover the whole range $[-1, 1]$ (see Fig. 15a), which affects the (nonnegligible) sample space of $\hat{r}$. One can see that especially for smaller positive sample correlations, the correction is sensitive to the prior. Figure 15c shows the error of $\hat{r}_{\mathrm{sec}}$ for the different priors as a function of distance (same as Fig. 11b). Using the "true" prior ($p1000$) clearly outperforms both the corrected sample correlations obtained with a uniform prior (uniform) and the uncorrected sample correlations (40) for distances larger than 150 km. However, this true prior is in practice rarely available. We therefore also constructed a prior from the sample correlations ($p0040$), which performs slightly worse than $p1000$, but still significantly outperforms both uniform and 40 for larger distances.

For distances smaller than 150 km, neither $p1000$ nor $p0040$ are able to reduce the sampling errors. This motivates the use of a separate prior for correlations corresponding to a distance smaller than a certain threshold. Figure 15c suggests this threshold should be 100 km. However, the information displayed in Fig. 15b is generally not available and therefore we suggest setting the threshold equal to the localization radius, which in this case is also 100 km. The resulting corrected sample correlations (<100) have a smaller error than those corresponding to $p1000$ and $p0040$ for distances smaller than approximately 150 km. A clear advantage with respect to the uncorrected sample correlations (40) is present for distances between 80 and 180 km.

The results discussed in this subsection motivate the use of distance-based priors. Given knowledge on the prior distributions, a different SEC table could be computed with an additional dimension for horizontal or vertical distance.

6. Conclusions

The sampling error correction (SEC) described by Anderson (2012) is evaluated applying the first convective-scale 1000-member ensemble simulation over central Europe. This unique dataset consists of 10 available 1000-member ensemble forecasts with 3-km mesh size and has been computed using the Japanese SCALE-RM model and a LETKF-based DA system (N20). The SEC

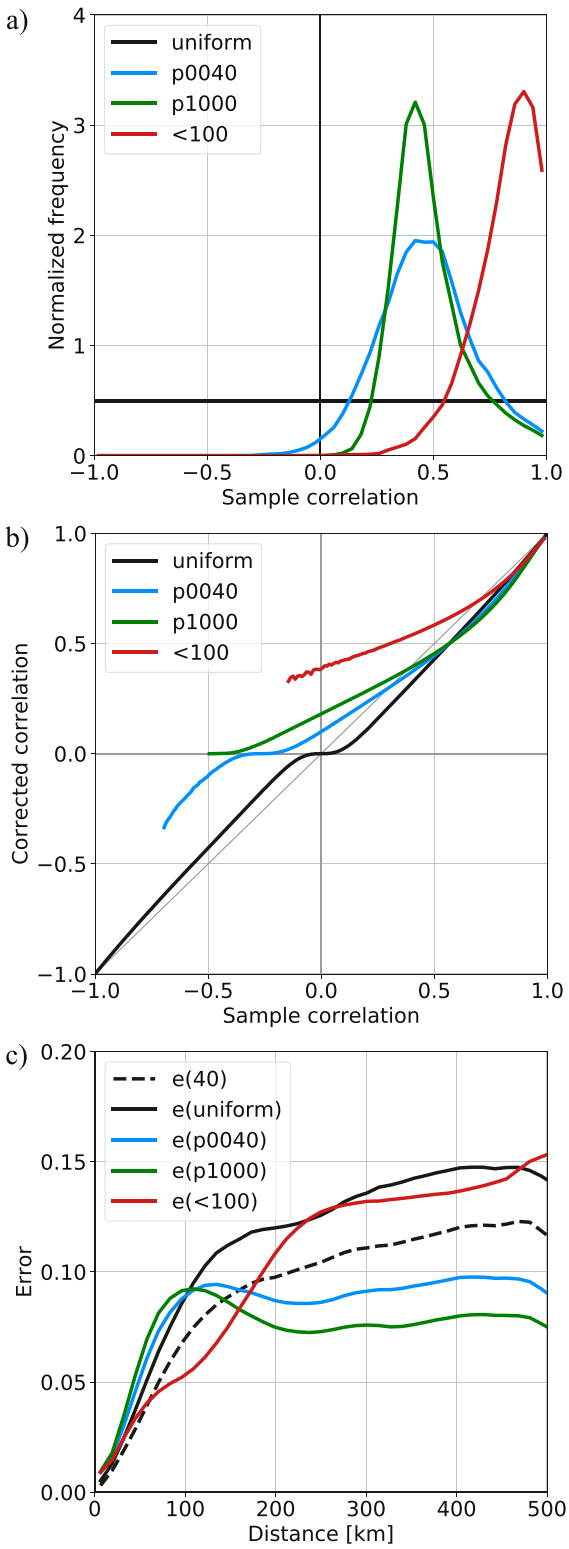

FIG. 15. (a) Different normalized priors and (b) the resulting sampling error correction. (c) Mean absolute error as function of spatial distance (km) for different priors and correlations of 500-hPa temperature to 500-hPa temperature (same as Fig. 11b). Black = uniform prior, blue = 40-member ensemble climatological prior, green = 1000-member ensemble prior, red = 40-member ensemble distance dependent prior, and black dashed = reference/no SEC.

is a simple lookup table–based approach, which is calculated using a Monte Carlo technique. If the lookup table is already computed for a target ensemble size and prior distribution, no additional information is needed to correct for sampling errors. Our study evaluates the SEC for spatiotemporal correlations that are important for ensemble sensitivity analysis (ESA; Ancell and Hakim 2007) and for spatial correlations that are crucial for ensemble and hybrid DA systems. For the application to ESA, the SEC is compared to a confidence test (T95; Torn and Hakim 2008). A confidence test is a commonly used approach to exclude spurious correlations in ESA. In the context of DA, the SEC is compared to a standard distance-based localization with a Gaspari–Cohn function (LOC; Gaspari and Cohn 1999). In addition, the impact of different prior assumptions on the SEC is examined. The 1000-member ensemble correlations are taken as a reference to assess the performance in all experiments. Furthermore, different subsets of the 1000-member ensemble are used to quantify sampling errors in a convective-scale NWP modeling system.

Examples of correlation fields demonstrate that the 1000-member ensemble provides physically meaningful correlations that are hardly affected by sampling errors while smaller subsets reveal spurious correlations. The 40-member ensemble subset is able to qualitatively indicate regions of maximum correlation in short-range convective-scale forecasts. However, small ensembles overestimate the magnitude of the majority of correlations due to spurious correlations. Increasing the ensemble size up to 80 or 200 members substantially reduces spurious correlations. This agrees with the results of Wile et al. (2015) who performed ESA on 4-km resolution using a 96-member ensemble and different subsets.

A confidence test can eliminate some spurious correlations by rejecting small insignificant correlations. However, it also eliminates small true correlations. This behavior is especially visible examining the frequency distribution of correlation values. While this is useful for a qualitative analysis of spatiotemporal correlations, the associated removal of weak correlations can lead to systematic errors and is therefore not optimal for a quantitative analysis. In contrast to the t test, the SEC is able to reduce spurious correlations while still allowing for small correlations. The SEC corrects spurious correlations independently of the strength of the correlation

←

and by this substantially improves the frequency distribution. Similar to the confidence test, the SEC has its largest impact on small correlations. Overall, the SEC is appropriate for both the qualitative and quantitative interpretation of correlations. The SEC is beneficial for all evaluated ensemble sizes and variable combinations. The mean absolute correlation bias, as well as the RMSE of correlations, are substantially reduced independent of the ensemble size. For spatiotemporal correlations, the 40-member ensemble applying SEC even outperforms the 80 member ensemble as the RMSE is reduced by up to 30% and the magnitude bias almost vanishes.

Spatial correlations are calculated to investigate sampling errors in ensemble DA. In the vertical, the SEC systematically reduces the RMSE in the entire tropospheric column independent of height. The reduction is largest far from the response level, the impact slightly decreases for distances smaller than 150 hPa and is smallest close to the response level. Compared to the Gaspari–Cohn localization the SEC works best in the vicinity of the response level. For the presented examples the localization performs better with increasing distance. However, it should be noted that the SEC allows for correlations far from the response level in contrast to operational localization techniques, which damp or exclude long-range correlations. This is potentially crucial for the assimilation of nonlocal observations (e.g., cloud, satellite radiance, or pressure) in a data assimilation scheme with observation-space localization such as an LETKF.

Horizontally, the SEC efficiently improves the estimation of the mean absolute correlation and mitigates the overestimation of the absolute correlation using small ensembles. Furthermore, it reduces the mean absolute error for most variable pairs and performs best on large distances. On short distances, the SEC performs better than a standard distance-based localization (LOC). However, the uniform prior $U(-1, 1)$, which is assumed in the calculation of the default SEC table, appears unsuitable for highly correlated variables. For instance, horizontal correlations of temperatures in the troposphere are already sufficiently well estimated by a very small ensemble sample and therefore hardly affected by sampling errors. This issue can be addressed by using an informative prior assumption of the correlation distribution. An improved prior can be obtained from a forecast climatology or from a large ensemble sample (e.g., the 1000-member ensemble used in this study). In particular, a distance-dependent prior can further improve the performance of the SEC. A combination of the SEC and standard localization techniques should also be considered. As shown, the SEC performs best on relatively short distances while the LOC performs best for long-range correlations.

Sensitivity studies on the ensemble size with a uniform prior show that sampling error corrected spatial correlations using 200 members are already very close to correlations obtained for 1000 members. For horizontal correlations, the SEC increases the correlation between the 40-member and 1000-member ensemble approximately by 5% and by 1% using 200 members, respectively. Using 200 members to estimate error covariances in convective-scale DA seems to be a reasonable choice thinking of the achieved accuracy and the computational cost compared to 1000 members.

Overall, the results strongly encourage to use the evaluated sampling correction for ensemble data assimilation systems and ensemble sensitivity analysis. Similarly, it could be applied in the framework of calculating ensemble forecast sensitivity to observation impact (Kalnay et al. 2012; Sommer and Weissmann 2014, 2016; Buehner et al. 2018). As the method is already implemented in DART, its application is technically simple. Further improvements could be achieved by using more informed prior assumptions, which should and will be the subject of future studies.

Acknowledgments. The authors want to thank the RIKEN DA group for their support with the K-computer system as well as Leonhard Scheck, Stefan Geiss, and Juan Ruiz for their contributions. We are also grateful to the reviewers for their suggestions, which helped to improve the manuscript. The open source project and Python package ''xarray'' (Hoyer and Hamman 2017) has been used to process data computing the correlations. Furthermore, we appreciate that Greg Hakim and Julia Keller provided their code for ensemble sensitivity analysis. This study was carried out in the Hans-Ertel-Centre for Weather Research (Weissmann et al. 2014; Simmer et al. 2016). This German research network of universities, research institutes, and DWD is funded by the BMVI (Federal Ministry of Transport, Building, and Urban Development). This research used computational resources of the K computer provided by the RIKEN Center for Computational Science through the HPCI System Research project (Project ID:ra000015, ra001011). Finally, this study was partly funded by the Transregional Collaborative Research Center SFB/TRR 165 ''Waves to Weather'' funded by the German Science Foundation (DFG).

APPENDIX

List of Variable Abbreviations

TOT_PREC	Hourly accumulated precipitation
*T*_2M	2-m temperature

*U*_10M	10-m zonal wind
PS	Sea level pressure
DBZ_CMAX	Column maximum radar reflectivity
*T*_500	500-hPa temperature
*U*_500	500-hPa zonal wind
*W*_500	500-hPa vertical wind
QV_500	500-hPa specific humidity
HY_500	500-hPa hydrometeors
DBZ_500	500-hPa radar reflectivity

REFERENCES

Ancell, B., and G. J. Hakim, 2007: Comparing adjoint-and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, https://doi.org/10.1175/2007MWR1904.1.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111, https://doi.org/10.1016/j.physd.2006.02.011.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1.

——, 2016: Reducing correlation sampling error in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **144**, 913–925, https://doi.org/10.1175/MWR-D-15-0052.1.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Bachmann, K., C. Keil, G. C. Craig, M. Weissmann, and C. A. Welzbacher, 2019: Predictability of deep convection in idealized and operational forecasts: Effects of radar data assimilation, orography and synoptic weather regime. *Mon. Wea. Rev.*, **148**, 63–81, https://doi.org/10.1175/MWR-D-19-0045.1.

Bannister, R. N., S. Migliorini, A. C. Rudd, and L. H. Baker, 2017: Methods of investigating forecast error sensitivity to ensemble size in a limited-area convection-permitting ensemble. *Geosci. Model Dev. Discuss.*, https://doi.org/10.5194/gmd-2017-260.

Barrett, A. I., S. L. Gray, D. J. Kirshbaum, N. M. Roberts, D. M. Schultz, and J. G. Fairman, 2015: Synoptic versus orographic control on stationary convective banding. *Quart. J. Roy. Meteor. Soc.*, **141**, 1101–1113, https://doi.org/10.1002/qj.2409.

Bednarczyk, C. N., and B. C. Ancell, 2015: Ensemble sensitivity analysis applied to a southern plains convective event. *Mon. Wea. Rev.*, **143**, 230–249, https://doi.org/10.1175/MWR-D-13-00321.1.

Bouttier, F., L. Raynaud, O. Nuissier, and B. Menetrier, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quart. J. Roy. Meteor. Soc.*, **142**, 390–403, https://doi.org/10.1002/qj.2622.

Buehner, M., P. Du, and J. Bedard, 2018: A new approach for estimating the observation impact in ensemble–variational data assimilation. *Mon. Wea. Rev.*, **146**, 447–465, https://doi.org/10.1175/MWR-D-17-0252.1.

Caron, J.-F., and M. Buehner, 2018: Scale-dependent background error covariance localization: Evaluation in a global deterministic weather forecasting system. *Mon. Wea. Rev.*, **146**, 1367–1381, https://doi.org/10.1175/MWR-D-17-0369.1.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1218–1256, https://doi.org/10.1002/qj.3179.

Hacker, J. P., and L. Lei, 2015: Multivariate ensemble sensitivity with localization. *Mon. Wea. Rev.*, **143**, 2013–2027, https://doi.org/10.1175/MWR-D-14-00309.1.

Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met office convective-scale ensemble, MOGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **143**, 2846–2861, https://doi.org/10.1002/qj.3135.

Hakim, G. J., and R. D. Torn, 2008: Ensemble synoptic analysis. *Synoptic–Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders*, *Meteor. Monogr.*, No. 33, Amer. Meteor. Soc., 147–162, https://doi.org/10.1175/0065-9401-33.55.147.

Hamill, T., J. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Hanley, K. E., D. J. Kirshbaum, N. M. Roberts, and G. Leoncini, 2013: Sensitivities of a squall line over Central Europe in a convective-scale ensemble. *Mon. Wea. Rev.*, **141**, 112–133, https://doi.org/10.1175/MWR-D-12-00013.1.

Hohenegger, C., and C. Schaer, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478, https://doi.org/10.1175/2007JAS2143.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

——, X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1143–1162, https://doi.org/10.1175/MWR-D-13-00138.1.

Hoyer, S., and J. J. Hamman, 2017: xarray: N-D labeled arrays and datasets in python. *J. Open Res. Software*, **5**, 10, https://doi.org/10.5334/jors.148.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D*, **230**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus*, **64A**, 18462, https://doi.org/10.3402/TELLUSA.v64i0.18462.

Keil, C., F. Baur, K. Bachmann, S. Rasp, L. Schneider, and C. Barthlott, 2019: Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Quart. J. Roy. Meteor. Soc.*, **145**, 3102–3115, https://doi.org/10.1002/qj.3607.

Lange, H., and T. Janjić, 2016: Assimilation of Mode-S EHS aircraft observations in COSMO-KENDA. *Mon. Wea. Rev.*, **144**, 1697–1711, https://doi.org/10.1175/MWR-D-15-0112.1.

Lei, L., J. S. Whitaker, and C. Bishop, 2018: Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. *J. Adv. Model. Earth Syst.*, **10**, 3221–3232, https://doi.org/10.1029/2018MS001468.

Lien, G.-Y., T. Miyoshi, S. Nishizawa, R. Yoshida, H. Yashiro, S. A. Adachi, T. Yamaura, and H. Tomita, 2017: The near-real-time SCALE-LETKF system: A case of the September 2015 Kanto-Tohoku heavy rainfall. *SOLA*, **13**, 1–6, https://doi.org/10.2151/SOLA.2017-001.

Limpert, G. L., and A. L. Houston, 2018: Ensemble sensitivity analysis for targeted observations of supercell thunderstorms. *Mon. Wea. Rev.*, **146**, 1705–1721, https://doi.org/10.1175/MWR-D-17-0029.1.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Miyoshi, T., and Coauthors, 2016a: Big data assimilation revolutionizing severe weather prediction. *Bull. Amer. Meteor. Soc.*, **97**, 1347–1354, https://doi.org/10.1175/BAMS-D-15-00144.1.

——, and Coauthors, 2016b: "Big data assimilation" toward post-petascale severe weather prediction: An overview and progress. *Proc. IEEE*, **104**, 2155–2179, https://doi.org/10.1109/JPROC.2016.2602560.

Necker, T., M. Weissmann, and M. Sommer, 2018: The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quart. J. Roy. Meteor. Soc.*, **144**, 1667–1680, https://doi.org/10.1002/qj.3390.

——, S. Geiss, M. Weissmann, J. Ruiz, T. Miyoshi, and G.-Y. Lien, 2020: A convective-scale 1000-member ensemble simulation and potential applications. *Quart. J. Roy. Meteor. Soc.*, https://doi.org/10.1002/qj.3744, in press.

Nishizawa, S., H. Yashiro, Y. Sato, Y. Miyamoto, and H. Tomita, 2015: Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations. *Geosci. Model Dev.*, **8**, 3393–3419, https://doi.org/10.5194/gmd-8-3393-2015.

Piper, D., M. Kunz, F. Ehmele, S. Mohr, B. Mühr, A. Kron, and J. Daniell, 2016: Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany. Part 1: Meteorological background. *Nat. Hazards Earth Syst. Sci.*, **16**, 2835–2850, https://doi.org/10.5194/nhess-16-2835-2016.

Rasp, S., T. Selz, and G. C. Craig, 2018: Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *J. Atmos. Sci.*, **75**, 691–706, https://doi.org/10.1175/JAS-D-17-0258.1.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Sato, Y., S. Nishizawa, H. Yashiro, Y. Miyamoto, Y. Kajikawa, and H. Tomita, 2015: Impacts of cloud microphysics on trade wind cumulus: Which cloud microphysics processes contribute to the diversity in a large eddy simulation? *Prog. Earth Planet. Sci.*, **2**, 23, https://doi.org/10.1186/s40645-015-0053-6.

Simmer, C., and Coauthors, 2016: Herz: The German Hans-Ertel Centre for Weather Research. *Bull. Amer. Meteor. Soc.*, **97**, 1057–1068, https://doi.org/10.1175/BAMS-D-13-00227.1.

Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **140**, 2672–2679, https://doi.org/10.1002/qj.2343.

——, and ——, 2016: Ensemble-based approximation of observation impact using an observation-based verification metric. *Tellus*, **68A**, 27885, https://doi.org/10.3402/TELLUSA.v68.27885.

Torn, R. D., 2010: Ensemble-based sensitivity analysis applied to African easterly waves. *Wea. Forecasting*, **25**, 61–78, https://doi.org/10.1175/2009WAF2222255.1.

——, and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

——, and ——, 2009: Initial condition sensitivity of western Pacific extratropical transitions determined using ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **137**, 3388–3406, https://doi.org/10.1175/2009MWR2879.1.

van Leeuwen, P. J., 1999: Comment on data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **127**, 1374–1377, https://doi.org/10.1175/1520-0493(1999)127<1374:CODAUA>2.0.CO;2.

Weissmann, M., and Coauthors, 2014: Initial phase of the Hans-Ertel Centre for Weather Research–A virtual centre at the interface of basic and applied weather and climate research. *Meteor. Z.*, **23**, 193–208, https://doi.org/10.1127/0941-2948/2014/0558.

Wile, S. M., J. P. Hacker, and K. H. Chilcoat, 2015: The potential utility of high-resolution ensemble sensitivity analysis for observation placement during weak flow in complex terrain. *Wea. Forecasting*, **30**, 1521–1536, https://doi.org/10.1175/WAF-D-14-00066.1.