Identifying Causes of Short-Range Forecast Errors in Maximum Temperature during Recent Central European Heatwaves Using the ECMWF-IFS Ensemble

ALEXANDER LEMBURG and ANDREAS H. FINK

Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract: In the last few years, central Europe faced a number of severe, record-breaking heatwaves. Previous studies focused on predictability of heatwaves on medium-range to subseasonal time scales (5–30 days). However, also short-range (3-day) forecasts of maximum temperature (Tmax) can exhibit substantial errors even on larger spatial scales. This study investigates the causes of short-range forecast errors in Tmax over central Europe for the summers of 2015–20 using the 50-member ensemble of the operational ECMWF-IFS (ECMWF-ENS). The 3-day forecast errors, individually calculated for each ensemble member with respect to a 0–18-h control forecast, are fed into a multivariate linear regression model to study the relative importance of different error sources. Outside of heatwaves, errors in Tmax forecasts are predominantly caused by incorrectly predicted downwelling shortwave radiation, mainly due to errors in low cloud cover. During heatwaves, ECMWF-ENS exhibits a systematic underestimation of Tmax (−0.4 K), which is exacerbated under clear-sky and low wind conditions, and other error sources gain importance: the second most important error source is over- or underestimation of nocturnal temperatures in the residual layer. Additional Lagrangian trajectory analysis for the years 2018–20 (due to limited data availability) suggests a link to accumulating errors in near-surface diabatic heating of air masses associated with forecast errors in residence time over land and cloud cover. Regionally, other physical processes can be of dominant importance during heatwaves. Coastal regions are influenced by errors in near-surface wind whereas errors in soil moisture are more important in southeastern parts of central Europe.

Keywords: Temperature; Heat wave; Short-range prediction; Forecast verification/skill; Diabatic heating; Boundary layer

1. Introduction

Reliable and timely prediction of weather extremes is of immense value to avoid or at least mitigate damage to humans, economy, and infrastructure. In times of global warming, forecasting of extremely high temperatures and severe droughts becomes increasingly important. In Europe, notably, the August of 2003 with more than 700,000 excess deaths (Robine et al. 2008) has been stuck in the memory and accelerated research on this topic (e.g., Black et al. 2004; Fink et al. 2004; García-Herrera et al. 2010). During the 2010s, multiple summers followed where many parts of Europe were stricken by periods of severe heat and dryness. Extreme events such as the Russian heatwave in 2010 (Trenberth and Fasullo 2012; Dole et al. 2011) or the record-breaking heatwave in June–July 2019 (Sousa et al. 2020; Ma et al. 2020) are projected to become significantly more frequent over all parts of Europe in the next decades (Collins et al. 2013). To prepare against such extreme hot and dry weather periods, a timely as possible prediction of this kind of extreme weather is desirable. Hence, a majority of recent studies put an emphasis on predictability of European heatwaves on time scales between medium-range (5 days) and seasonal (>30 days) forecasting (e.g., Matsueda 2011; Lavaysse et al. 2019; Weisheimer et al. 2011; Fragogoulidis et al. 2018; Zschenderlein et al. 2018; Miralles et al. 2014).

Comparably less attention was attributed to forecast quality on much shorter time scales of up to 3 days. However, accurate short-range forecasts may also be of high value and are often the basis for decision-making in more regionalized and specific heat warnings (Casanueva et al. 2019). In contrast to medium-range forecasts, errors in large-scale circulation are thought to be small for 3-day lead time. Indeed, averaged over the Northern Hemispheric extratropics, the RMSE of the 500-hPa geopotential field in ECMWF-IFS (for both the deterministic run
and the 50-member ensemble mean) amounts to less than 20 m—about a third of that of the 6-day forecast (Haiden et al. 2021). In addition, a study by Rodwell et al. (2013) shows that even for 6-day lead time, ECMWF-IFS nowadays rarely exhibits forecast busts regarding large-scale synoptic patterns over Europe.

Thus, when investigating errors in short-range forecasts, the focus should be shifted from large-scale dynamics toward more regional or local-scale processes. Of particular importance is undoubtedly the accurate prediction of diabatic processes near the surface or within the boundary layer. The first example coming to mind is certainly over- or underestimation of downwelling shortwave radiation at the surface due to errors in simulating cloudiness. In summer, the effects of dense cloud cover on the surface energy budget and corresponding turbulent heat fluxes can cut normal daytime warming by half (Dai et al. 1999).

Another important factor controlling the daily maximum 2-m temperature (Tmax) is the partitioning of surface turbulent heat fluxes into sensible and latent heat fluxes. To first order, this partitioning, generally expressed via the Bowen ratio, is governed by soil moisture. Within idealized simulations conducted with a boundary layer model, Lockart et al. (2013) found Tmax differences of up to 5 K between a very dry soil and a wet soil on a cloud-free summer day with little wind. Therefore, errors in prediction of soil moisture may considerably affect Tmax forecasts, particularly in undisturbed, high insolation conditions. With regard to short-range forecasts, however, large-scale systematic errors in soil moisture are unlikely. Instead, soil moisture deviations will likely be linked to falsely predicted previous rainfall events, which may lead to rather spatially heterogeneous and small-scale forecast error patterns.

In many cases, multiple potential error sources may of course occur in combination, which shall be underlined by the following example scenario: An unpredicted mesoscale convective system on forecast day 2 not only interrupts radiative heating of the surface in some regions and thereby lessens sensible heating of overlying air masses. It further increases soil moisture which may eventually influence sensible heating of overlying air masses as well as local cloudiness on forecast day 3. This example stresses the importance of tracking the explicit history of air masses with regard to diabatic heating, an idea introduced before in the context of heatwaves (e.g., Quinting and Reeder 2017; Zschenderlein et al. 2019). In other words, any substantial over- or underestimation of diabatic heating on forecast day 1 and 2 in remote regions may also impact the quality of local Tmax forecasts for day 3—even when the prediction of most local diabatic processes is perfectly accurate.

Despite the assumption of large-scale dynamics being an unlikely error source on the investigated time scale, not only diabatic processes and their history, but also errors in regional near-surface circulation may play a role in 3-day forecasts. Particularly near the coast, even a 10-m wind direction forecast that is only off by some degrees may decide on whether the region remains in hot continental air or whether advection of cooler maritime air interrupts a heatwave (Ramamurthy et al. 2017).

To our best knowledge, the relative importance of above-mentioned potential error sources for summertime short-range Tmax forecasts has never been quantified before for central Europe. In this study we therefore extensively investigate potential physical causes of 3-day Tmax forecast errors. Our domain of interest is the central European region (black outline in Fig. 1), which is subdivided into five regions for more detailed analyses. All ocean points and coastal grid points with a land–sea fraction of less than 90% are excluded from analysis.

![Fig. 1. Research domain of this study. Colored shading indicates terrain elevation in meters above sea level. The entire central European domain (in thick black outline) is subdivided into five regions for more detailed analyses. All ocean points and coastal grid points with a land–sea fraction of less than 90% are excluded from analysis.](image)

### 2. Data and methods

#### a. Data and variables of interest

For this study, we use the full 50 perturbed member ensemble forecast of ECMWF-IFS, the so-called ECMWF-ENS. For most analyses, we concentrate on 3-day forecasts which
may comprise forecast hours 60–78 h depending on the respective variable of interest. To determine forecast errors, we stay within the model space by evaluating 3-day forecasts against the respective forecast days’ 0–18-h forecast of the ECMWF-ENS control run (undisturbed ensemble member with same resolution), here called “quasi-analysis.” Unless explicitly stated otherwise, forecast errors are always calculated by subtracting the quasi-analysis from each ensemble member individually, and each thus obtained ensemble forecast error will be included in further analyses. We do not compare forecasts against reanalyses due to reasons of physical consistency and to exclude a possible general Tmax bias of the model climate from analysis. Data are downloaded from ECMWF’s MARS archive in form of hourly instantaneous or 6-hourly integrated values at 0.25° spatial resolution. As mentioned in the section 1, we investigate the recent six summers 2015–20, focusing exclusively on June–July–August (JJA). Due to the retrieval of ECMWF forecasts in monthly data packages, the 3-day forecasts are always missing the first three days of June which are therefore excluded from analysis. It is further important to point out that ECMWF-IFS had one major update in spring 2016 (IFS cycle 41r2) featuring a switch to an octahedral grid. This update left the spectral truncation unchanged but increased gridpoint space resolution to more accurately represent physical processes and advection. For the Lagrangian-type analysis later on, we have the opportunity to use previously saved forecasts of all 50 ECMWF-ENS members, saved on model levels at a resolution of 1° in space and 6 h in time. However, these extensive data necessary to calculate trajectories have only been stored for the years 2018–20; such data do not exist for the years before, unfortunately.

Variables of interest in this study—those hypothesized to affect Tmx forecasts—and further details about them are listed in Table 1. As mentioned in the introduction, this selection encompasses mainly variables that are closely tied to local diabatic heating of near-surface air masses as well as the near-surface flow field represented by u10m and v10m. To capture the effects of cloudiness, the focus within the scope of this study is set on 0600–1800 UTC integrated surface downward shortwave radiation (SWDS). The rationale for this choice is that—in contrast to clouds and rainfall—errors in SWDS are somewhat coherent over larger spatial scales and nearly normally distributed. SWDS errors are of course mainly associated with shading by either non-precipitating but also precipitating clouds. We therefore want to stress that the SWDS impact on Tmax may also contain effects of cooling of near-surface air due to evaporation of precipitation. We further consider the midday temperature at 700 hPa (T700-12UTC) to check whether temperature forecast errors are only affecting the planetary boundary layer (PBL) or whether they also extend into the free troposphere which could suggest involvement of synoptic-scale disturbances. Most importantly, the selection of variables is limited to those that clearly cause Tmax errors but are unlikely to be also affected by Tmax themselves. For instance, errors in Tmax cannot influence soil moisture in the morning whereas the height of the PBL may stand in a complex two-way relationship with Tmax and is therefore disregarded despite high correlations.

For the trajectory-based investigation, we focus on the temporal evolution of air parcels’ potential temperatures as well as cloud cover and land sea fraction traced along trajectory paths.

b. Heatwave detection

Heatwaves are detected via a percentile-based objective algorithm that was introduced for the calculation of the so-called HWMI in Russo et al. (2014). Our exact technique is very similar to the one applied by Zschenderlein et al. (2019). We use hourly ERA-5 data (Hersbach et al. 2020) from 1979 to 2019 at 0.25° spatial resolution and determine maximum 2-m temperature values for each day. After detrending the data series at each grid point individually, we calculate the local 90th percentile. In the next step, we identify heatwaves by demanding that in at least 10% of the central European domain, Tmax has to surpass the local 90th percentile for at least 3 days. In comparison to earlier studies our definition of a heatwave is less stringent (90th percentile instead of 95th) resulting in a higher number of heatwave days. This choice is motivated by the need for a reasonable number of heatwave days to obtain significant results when comparing heatwave days to regular summer days. An overview of all identified heatwaves during summers 2015–20 is presented in Fig. 2. Averaged over all six investigated summers, about one-sixth of all JJA days are classified as central European heatwave days.

c. Statistical analyses within the Eulerian framework

The multiple error sources explaining false Tmax forecasts are explored with the help of two different statistical approaches. First we quantify forecast errors in a spatially integrated perspective by calculating pattern correlation coefficients between Tmax forecast errors fields and error fields of any variable of interest. We then divide the available data into non-HW and HW days according to our previously described algorithm. The pattern correlation analysis we then apply is schematically described in Fig. 3. For each available day and ensemble forecast member, we calculate pattern correlation coefficients between forecast errors in Tmax and another variable of interest over the depicted central European domain containing only land points. To increase the signal-to-noise ratio, not all grid points of this domain will be included in the computation; at grid points where local anomalies in either Tmax or the other variable of interest lie within the respective middle quintile (within the 40th–60th percentile range), the anomalies are set to missing value and thereby excluded from the pattern correlation. If more than two thirds of grid points within our domain have been set to missing by this technique due to generally low forecast errors, the respective ensemble member of the respective day is removed from analysis. Finally, statistics of all available pattern correlation coefficients are depicted via standard box and whisker plots.

In a second approach we identify the respective importance of possible error sources in a local gridpoint-based perspective. For this, we make use of a multivariate linear regression model (MLRM) that is applied to each grid point in our research domain individually. The advantage of this method is
Table 1. Overview over the data used in this study and the variables of interest. If not explicitly noted otherwise, we always refer to the 3-day forecast (e.g., integration time 66–78 h for the Tmax forecast, +72 h for the u10m forecast as the considered ECMWF-ENS simulations are initialized at 1200 UTC) and the forecast error is always evaluated against the quasi-analysis (e.g., integration time 6–18 h for the Tmax forecast, +12 h for the u10m forecast as the considered ECMWF-IFS forecast is initialized at 0000 UTC). Cloud cover forecast data were only available in appropriate temporal resolution for the summers 2017–20.

<table>
<thead>
<tr>
<th></th>
<th>Forecast dataset</th>
<th>Evaluation dataset</th>
<th>Dataset for trajectory calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>ECMWF-IFS</td>
<td>ECMWF-IFS</td>
<td>ECMWF-IFS</td>
</tr>
<tr>
<td>Time period</td>
<td>JJA 2015–20 (= 534 days)</td>
<td>JJA 2015–20 (= 534 days)</td>
<td>JJA 2018–20 (= 267 days)</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>0.25°</td>
<td>0.25°</td>
<td>1°</td>
</tr>
<tr>
<td>Temporal resolution</td>
<td>1 h</td>
<td>1 h</td>
<td>6 h</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable (acronym)</th>
<th>Vertical level</th>
<th>Instantaneous time/integration time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables of interest—Eulerian analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max temperature (Tmax)</td>
<td>2 m</td>
<td>Max (0600–1800 UTC)</td>
</tr>
<tr>
<td>Min temperature (Tmin)</td>
<td>2 m</td>
<td>Min (1800 UTC day before–0600 UTC)</td>
</tr>
<tr>
<td>Max temperature on day before (Tmax-day2)</td>
<td>2 m</td>
<td>Max (0600–1800 UTC day before)</td>
</tr>
<tr>
<td>Surface downward shortwave radiation (SWDS)</td>
<td>Surface</td>
<td>Mean (0600–1800 UTC)</td>
</tr>
<tr>
<td>Sensible heat flux (SSHF)</td>
<td>Near surface</td>
<td>Mean (0600–1200 UTC)</td>
</tr>
<tr>
<td>Latent heat flux (SLHF)</td>
<td>Near surface</td>
<td>Mean (0600–1200 UTC)</td>
</tr>
<tr>
<td>Volumetric soil moisture (SM)</td>
<td>First soil level (0–7 cm)</td>
<td>0600 UTC</td>
</tr>
<tr>
<td>(Residual layer) temperature (T925-00 UTC)</td>
<td>925 hPa</td>
<td>0000 UTC</td>
</tr>
<tr>
<td>(Free tropospheric) temperature (T700-12 UTC)</td>
<td>700 hPa</td>
<td>1200 UTC</td>
</tr>
<tr>
<td>Near-surface zonal wind (u10m)</td>
<td>10 m</td>
<td>1200 UTC</td>
</tr>
<tr>
<td>Near-surface meridional wind (v10m)</td>
<td>10 m</td>
<td>1200 UTC</td>
</tr>
<tr>
<td>Total cloud cover (TCC)—Only 2017–20</td>
<td>Vertically integrated</td>
<td>Mean (1100–1400 UTC)</td>
</tr>
<tr>
<td>Low cloud cover (LCC)—Only 2017–20</td>
<td>Vertically integrated</td>
<td>Mean (1100–1400 UTC)</td>
</tr>
<tr>
<td>High cloud cover (HCC)—Only 2017–20</td>
<td>Vertically integrated</td>
<td>Mean (1100–1400 UTC)</td>
</tr>
</tbody>
</table>

| Variables of interest—Lagrangian analysis |
| Potential temperature (Tpot) | Air parcel height | 6-hourly |
| Land–sea fraction (LSF) | Surface | 6-hourly |
| Total cloud cover (TCC) | Vertically integrated | 6-hourly |

that we can easily compare relative importance of different physical processes against each other. The predictand in our model is always the 3-day forecast error in Tmax. We use SWDS, soil moisture, T925-00 UTC, T700-12 UTC, and 10-m wind components as predictors. This selection of predictors was decided upon after an initial testing phase. We aimed to optimize the selection such that $R^2$ received a substantial increase after inclusion of each additional variable while at the same time the variance inflation factor (VIF), a measure for collinearity among predictors, stays at a considerably low level below 2.5 (Johnston et al. 2018) for each variable. As was the case for the pattern correlation analysis, both the temporal and the ensemble space dimension will be treated equally by being merged into one large sample of individual forecast error pairs between Tmax forecast error and errors in another quantity.

The MLRM is programmed in R where we use the package “relaimpo” to quantify the relative importance of each predictor variable with the help of the so-called “lmig” metric. This metric is based on the sequential $R^2$ approach but takes care of the dependence on orderings by averaging over all possible orderings (Lindeman et al. 1980).

In an additional analysis step, we attempt to eliminate the often dominant role of errors in SWDS in order to investigate the leading error sources in times of a reasonably good SWDS forecast. To do so, we randomly subsample the available data in an iterative way, where the subsample size gets decreased gradually until the correlation between SWDS and Tmax errors vanishes ($r$ lower than 0.03). Due to the high number of iteration steps required, we only apply this method to ensemble means and no longer to all 50 members of the ECMWF-ENS.

d. Lagrangian analysis

As outlined in the introduction, the history of an air mass may be an important piece of the puzzle to understand false Tmax forecasts. In this context, Lagrangian analysis via backward trajectories is generally a useful tool. However, turbulence and entrainment/detrainment within the PBL leads to a subgrid scale mixing of air masses, which can render a Lagrangian tracking of PBL air masses highly unreliable, particularly if the input data are not available in high spatial and temporal resolution (Stohl 1998). Therefore we restrict ourselves to analyzing statistics of a large number of cases regarding average airmass origin.
as well as average diabatic heating rates and do not put high value in single trajectories.

The previously described ECMWF-ENS forecast data on model levels, available for the years 2018–20, indeed allows us to create a large ensemble of 72-h backward trajectories using the Lagrangian analysis tool LAGRANTO (Sprenger and Wernli 2015). We suppose that it is not appropriate to evaluate a Lagrangian analysis over a large geographical area, i.e., to pool all obtained trajectories for the entire central European domain. Therefore we split up the trajectory starting region into the subregions depicted in Fig. 1. For each sub region, we track about 20 individual “air parcels” each starting 0.5 apart at a height of 25 hPa over ground, initiated with ECMWF-ENS data from time step +72 h, which provides the 3-day forecast for 1200 UTC for the respective day of interest. We only consider trajectories that start over land. To obtain the travel history of each air parcel, i.e., the backward trajectory, we then go back in time such that the next input for the backward trajectory is the forecast at forecast hour +66 h and so on. We then calculate average trajectory statistics with regard to target region-relative airmass origin, travel distance, and travel time over land as well as average diabatic heating rates of air parcels along a large sample of trajectories. Due to the availability of all 50 ECMWF-ENS members, we are then able to systematically investigate why some ensemble members end up with higher or lower maximum temperatures than others.

3. Results

With an emphasis on heatwaves, we will first briefly review whether ECMWF-ENS displays any systematic 3-day Tmax forecasts errors. Thereafter, we quantify potential causes of erroneous Tmax forecasts by means of two different statistical methods. The end of this section features a trajectory analysis aimed at characterizing the role of airmass origin and heating history for Tmax forecast errors.

a. Tmax forecast biases

Figure 4 depicts the multiyear JJA (2015–20) mean and ensemble-averaged error of the 3-day Tmax forecasts of ECMWF-ENS, as determined with respect to the quasi-analysis. For simplicity we will refer to this systematic forecast error as Tmax “bias” although this term is most often used in the context of a model evaluation against observations. For non-HW days, ECMWF-ENS performs well in all central European regions with a spatially averaged Tmax bias close to zero in our domain of interest (Fig. 4a). On heatwave days, however, ECMWF-ENS displays a slight cold bias of $-0.43$ K (Fig. 4b). This cold bias is more strongly pronounced over eastern parts of our research domain.

As heatwaves are generally characterized by a high number of clear-sky days with high SWDS, we apply the bias analysis to a subset of days for both non-HW and HW days (Figs. 4c,d). This subset consists of days on which area-averaged SWDS exceeds either the heatwave or non-heatwave related 75th percentile values and on which the area-averaged absolute SWDS forecast error does not exceed the 25th percentile. Interestingly, non-HW days now also show a cold bias of $-0.39$ K and while the HW bias also grows to $-0.65$ K, the difference in Tmax bias between non-HW and HW days actually becomes smaller. Thus, the slight underestimation of Tmax in the
ECMWF-ENS ensemble mean is mainly present on undisturbed high-insolation days.

Finally, we consider a low wind subset consisting of days with area-averaged 10-m wind speed below the 25th percentile and an ensemble mean error of less than the median error in both zonal and meridional components simultaneously (Figs. 4e,f). Interestingly, the contrast between non-HW and HW days grows in low wind situations. Whereas the area-averaged error in the 3-day Tmax forecast amounts to $0.27$ K in non-HW days, the Tmax underestimation increases to about $0.85$ K within heatwaves.

In summary, ECMWF-ENS exhibits a slight cold bias of $-0.43$ K in its 3-day Tmax forecasts for heatwave periods. This heatwave-specific bias increases substantially in high insolation and low wind conditions, under which it also shows through to a lesser extent on non-HW days. Haiden et al. (2018) noted that ECMWF-IFS sometimes underestimates superadiabatic lapse rates in the surface layer, leading to a slight cold bias in summertime Tmax forecasts. Schmieder et al. (2019) further found an underestimation of the daily amplitude in 2-m temperatures due to overestimation of land–atmosphere fluxes. However, both these studies evaluated forecasts against observations and reanalyses. Moreover, a distinction between heatwaves and regular summer days was also missing. It is therefore questionable whether the found heat-specific cold bias stems from the similar mechanisms as those suggested by these authors.

Keeping in mind the data availability-related limitation in the Lagrangian analysis following later this paper, we have further calculated the biases for the 2018–20 period only, shown in Fig. S1 in the online supplemental material. Overall, our main findings do not change substantially as we then obtain—averaged over the central European domain—a non-heatwave bias of $-0.09$ K and a bias of $-0.55$ K for heatwaves.

b. Identifying Tmax forecast error sources and their relative importance

1) PATTERN CORRELATION ANALYSIS

Not surprisingly, forecast errors in SWDS are the dominant error source for Tmax forecasts at 3-day lead time. The dominance of SWDS as error source is particularly pronounced outside of heatwaves where the median pattern correlation coefficient amounts to nearly 0.7, whereas it stays slightly below 0.5 during heatwaves (Fig. 5a). An explanation of why the importance of SWDS forecast errors is higher for HW days than for non-HW days is likely related to the overall more stable character of heatwaves with lots of cloud-free periods and a generally lower chance of a significant misforecast of cloudiness.

Error patterns in surface sensible heat fluxes exhibit the second strongest linear relationship to Tmax forecast error patterns, with a median pattern correlation of up to 0.5 for non-HW days and 0.3 on HW days. It is, however, likely that a large portion of this correlation arises due to the fact that surface turbulent heat fluxes are strongly controlled by the energy input from SWDS. The fact that errors in latent heat fluxes are also positively correlated to Tmax errors, albeit to a lesser extent, underline this hypothesis.

The spatial error pattern relationships between Tmax and Bowen ratio are generally weaker than those between Tmax and SWDS.
surface heat fluxes. This confirms that the high correlations for sensible heat fluxes are to some extent associated with the error in SWDS. Interestingly, the forecast error in Bowen ratio appears to be somewhat more important on non-heatwave days than on heatwave days. To first order, Bowen ratio should be controlled by soil moisture. Indeed, the pattern correlations to Tmax forecast errors are similar for both Bowen ratio and soil moisture, with the sign switched, of course. However, in contrast to Bowen ratio and therefore somewhat contrary to our expectation, soil moisture represents a slightly more important error source on HW days than on non-HW days. Within the scope of this paper we have not explored the reasons for this in greater detail.

For all particular cases, we find that errors in simulated 2-m minimum temperature—albeit statistically significant due to the large sample size—play no substantial role for Tmax forecast errors. Comparably more important are errors in nocturnal residual layer temperature, which are represented by errors in T925-00UTC. During heatwaves, the error patterns of this quantity are considerably correlated with Tmax error patterns, reaching a median of 0.25 which is in this case only ranked slightly behind SWDS and the turbulent heat fluxes. As the residual layer can be considered a remnant of the previous day’s PBL, we assume that errors in T925-00UTC are a rather good proxy for errors in sensible heating of PBL air masses on the day before. At a first sight, this argument is not strongly supported by the finding that the error pattern correlation between the 72-h Tmax forecast error and the Tmax error on the day before is substantially lower. However, with each hour going further back, it will be less likely to find substantial correlations within an Eulerian analysis. After all, in the presence of advection, differences in sensible heating of PBL air could have occurred in remote areas on the
day before. Therefore, a causal link between Tmax forecast errors on day 3 and erroneous sensible heating on day 2 may still be partly visible in the local residual layer temperature via this Eulerian-type analysis, but certainly less when checking error fields of Tmax or sensible heating for day 2. The need for a nonlocal Lagrangian point of view—particularly in the context of the found Tmax–T925-00UTC error relationship—is therefore one of the main motivations of the backward trajectory analysis presented later on.

Errors in the temperature of the free troposphere (T700-12UTC) are comparably weakly correlated with Tmax forecast errors and there is no discernible difference between HW and non-HW days. We argue that this result is consistent with our assumption of a subordinate role of forecast errors with respect to the large-scale circulation. In an additional analysis of the predicted 500-hPa fields, we have indeed confirmed that considerable forecast busts in terms of large-scale features are indeed rare in the 2015–20 period, particularly during heatwaves (Fig. S2).

Figure 5b shows results for the same kind of error pattern correlation analysis as in Fig. 5a, but now applied to the error in SWDS. Within the scope of this paper, we are not aiming to assess in greater detail by what kind of cloud type forecast errors in SWDS are mainly caused. Because cloud cover forecast data were not available in adequate temporal resolution for the whole research period, this analysis had to be limited to the summers 2017–20. What appears to be quite a robust result, however, despite this limitation, is that errors in midday low cloud cover explain a large portion of the SWDS error. Error patterns in high cloud cover are in comparison much less correlated to SWDS errors. Overall, there is no clearly distinguishable difference between non-HW and HW days. It is up to future work to investigate more thoroughly what kind of clouds on which spatial scale represent dominant error sources and whether precipitating and/or deep convective clouds play a major role. Moreover, slight errors in the SWDS forecast (when one evaluates against observations) may be related to ECMWF-ENS’s use of climatological rather than prognostic aerosol.

It is worth pointing out that the pattern correlation analysis applied here may favor physical processes associated with larger spatial scales. Due to a certain level of noise, locally important effects due to soil moisture, for instance, may be underestimated against the effects of a wide band of clouds which is spatially coherent and will also cause a substantial effect on regional Tmax. Moreover, we want to stress again a potential additional effect of rainfall-associated evaporative cooling that may affect the SWDS–Tmax error relationship.

As was the case for the forecast bias, an additional analysis was performed only for the period 2018–20. Overall, all previously discussed qualitative differences between non-HW and HW days remain unchanged (Fig. S3). Some minor quantitative differences to the longer 6-yr period mainly show up during heatwaves: In the summers 2018–20, SWDS is slightly less correlated to Tmax forecast errors while the Tmax–T925-00UTC error pattern correlations increase to a median value of 0.3.
2) RELATIVE IMPORTANCE OF THE SIX MOST IMPORTANT ERROR SOURCES

We now apply a multivariate linear regression model (MLRM) to the six summers 2015–20, making again use of a large sample by using 534 (days) times 50 (ensemble members) forecast error fields as input. For both non-HW and HW days, the MLRM and its six predictor variables are able to explain roughly 50%–65% of the variance in the Tmax error in most regions (Figs. 6a and 7a). This grid point-based analysis of Tmax error sources and their relative importance generally confirms the results of the spatial pattern correlation method shown earlier. Outside of heatwaves, errors in SWDS forecast are clearly the dominant cause of Tmax forecast errors (Fig. 6b). This statistical result, based on a sample consisting of 445 non-HW days from 2015 to 2020 with 50 ensemble members each, is true for all grid points in our central European domain.

The second most important error source with explained variance of up to 30% locally, is either T925-00UTC or T700-12UTC, depending on the respective region (Figs. 6d,e). Of course, when interpreting the role of T700-12UTC, caution has to be applied near the Alps because of elevated terrain. Interestingly, the relative importance of the error in soil moisture does not exceed 20% anywhere in central Europe on non-HW days. Even less important for Tmax forecast errors outside of heatwaves are errors in local near-surface circulation. Only near the coast of Poland the relative importance of v10m reaches 20%.

During heatwaves, SWDS is still the dominant error source for Tmax predictions but only if the spatial average over the research domain is considered (Fig. 7b). Regionally, two other variables are now at least as important. In parts of central Germany and in particular in the eastern CE region, the error in T925-00UTC now explains up to 45% of the Tmax forecast error (Fig. 7d). Along the coasts of Germany and Poland, errors in v10m are the dominant source of error during heatwaves with relative importance of up to around 55% (Fig. 7g). Similar to non-HW days, the under- or overestimation of soil moisture does not pose an important error source anywhere in central Europe. This does not necessarily imply that there is no strong coupling between errors in soil moisture and errors in Tmax. We assume, however, that in many cases with an existing soil moisture–Tmax coupling, concomitant errors in SWDS may partially mask the comparably smaller evaporative cooling-related effect of the Bowen ratio on near-surface temperature. Likewise, other significant sources of Tmax forecast errors may also be obscured by concurrent errors in SWDS. To circumvent this issue and shed light on other error sources, which may become important in cases with a nearly perfect SWDS forecast, we apply a subsampling technique which systematically reduces the sample size until the linear relationship between SWDS and Tmax forecast errors vanishes.
Results of the MLRM applied on this subsample are depicted in Fig. 8. The amount of explained variance remains satisfactorily high with the only exception being Southern Germany (where SWDS errors are the dominant error source) and some smaller regions in different parts of the domain. In absence of SWDS errors, three main causes of Tmax over- or underestimations can be identified within central European heatwaves: most important in the largest fraction of central Europe is the error in T925-00UTC (Fig. 8d). Whereas northern Germany’s Tmax forecast is less affected by this variable’s error, parts of western Germany and large parts of eastern central Europe exhibit a relative importance of 50%–70%. The aforementioned hypothesis of the role of soil moisture being obscured by concurrent errors in SWDS is supported by the analysis of the subsample. Now errors in soil moisture are the dominant error source in some regions across the center and southeast of the central European region (Fig. 8c). In coastal areas of Germany and Polad, forecast errors in v10m dominate every other error cause (Fig. 8g). This is consistent with the results of the previously investigated full sample suggesting that during heatwaves errors in v10m are very important near the Baltic Sea coast regardless of whether a concomitant SWDS error exists or not.

To assess the robustness of our findings, we performed some additional sensitivity tests, the results of which are presented in the supplemental information. First of all, it was checked whether qualitative statements about relative importance are irrespective of the treatment of ensemble members as equally weighted members within a merged sample of all available days and ensemble members. We find that using just the ensemble mean instead and thereby reducing the sample size by a factor of 50 does not change results substantially (Figs. S4 and 5). As a second test, we remove outliers from the large sample by only considering data where the predictor Tmax is within the respective 5th and 95th percentile ranges (Figs. S6 and 7). In this case, the total explained variance by the MLRM declines considerably to values below 50% at all grid points, for both non-HW and HW days. While the general spatial patterns and the order of relative importance of predictors do not change substantially, there is an overall increase in the relative importance of T925-00UTC and soil moisture at the expense of SWDS, particularly for HW days. This is probably related to the fact that Tmax forecast errors in the outer percentile margins are linked to substantial SWDS misforecasts in most cases. Finally, we test the sensitivity against the sample period by recreating Figs. 6 and 7 with the same MLRM configuration, but now only for the 2018–20 period (Figs. S8 and S9). For non-HW days, the results barely change compared to the total 2015–20 period. For heatwaves, we find an overall increase in the relative importance of both T925-00UTC and soil moisture which comes at the expense of a reduced importance of SWDS.

In summary, both Eulerian-type analyses clearly demonstrate that 3-day Tmax forecast errors are predominantly associated with forecast errors in downward shortwave radiation, mainly due to errors in low cloud cover. However, for heatwaves the importance of SWDS drops substantially when compared to regular summer days. Further remarkable is the
finding of T925-00UTC clearly becoming the second most important error source in heatwaves. This may be a surprising result, but we found it to be quite robust with respect to the conducted sensitivity tests.

Investigating physical processes causing recent heatwaves, Miralles et al. (2014) pointed out that heat generated during daytime is often preserved in an anomalous kilometers-deep atmospheric layer, ready to be mixed into near-surface layers again on the following day. Forecast errors in residual layer temperature may therefore indeed serve as a good proxy for errors in diabatic heating of near-surface air occurring one day earlier in some distant region. The next section will therefore feature a Lagrangian trajectory-based analysis allowing a closer look at how airmass origin and travel history may impact Tmax forecasts in the region of interest.

c. Lagrangian perspective on Tmax forecast errors

1) DIFFERENCES IN DIABATIC HEATING BETWEEN 10 WARMEST AND 10 COLDEST ENSEMBLE MEMBERS

Using a large ensemble of backward trajectories for the time period 2018–20, we now aim to understand the extent to which too low or too high Tmax predictions depend on the origin of air masses and respective diabatic heating rates along the trajectory path. To this end, we systematically compare trajectories of the 10 ensemble members with the lowest 72-h Tmax forecast for the respective considered day (averaged over respective CE subregions) against those 10 with the highest Tmax forecast. In this analysis, in which we calculate statistics over all available 210 non-HW and 48 HW days, respectively, an emphasis is placed on air parcel potential temperature as well as trajectory-traced total cloud cover and land sea fraction. For the sake of brevity, we focus on two subregions where the Lagrangian analysis shows pronounced differences compared to each other—western and eastern CE. Qualitatively, the results for the northern CE region closely resemble those for the western CE region described below, and the southern region is largely similar in its characteristics to the eastern CE region. For completeness, the results for those regions are provided in the same style and order as below in Figs. S10–S13.

Outside of heatwaves, temperature differences between warm and cold ensemble members in the target region are only weakly related to differences in diabatic heating along the trajectory path—with little differences between western and eastern central Europe. This result is obtained when comparing the red and blue lines depicting the temporal evolution of average air-parcel tracked potential temperature in Figs. 9a and 10a. To a rather large extent, differences in potential temperature already exist at 272 h. In the context of backward trajectories, this simply means that two different air masses end up at the region of interest due to small errors in large-scale dynamics adding up over the span of 72 h. An additional substantial spread in potential temperatures only occurs on the final forecast day 3, meaning that local errors in the
Diabatic heating play a major role for false Tmax forecast in the region of interest.

During heatwaves, the picture changes substantially, particularly for western CE. In this region, accumulated differences in diabatic heating are now one important explanation for why some ensemble members end up with higher near-surface temperatures than others on HW days. Beginning with a relatively small initial potential temperature difference, the difference between the warm and the cold cluster gradually increases over the 72 h of the forecast (Fig. 9b). Interestingly, anomalous diabatic heating turns out to be most pronounced on forecast day 2 whereas the additional spread in potential temperature between the 10 coldest/warmest cluster is comparably smaller on forecast day 3. A different picture exists for heatwaves in eastern parts of central Europe (Fig. 10b). Here, the difference in Tmax forecast between the 10 warmest and 10 coldest members appears to be much more associated with a considerable ensemble spread in near-surface diabatic heating appearing only at forecast day 3. In contrast to the western parts, only little differences in diabatic heating develop on forecast day 2 already.

Differences in adiabatic compression/expansion do not play any important role in explaining Tmax forecast errors. For both subregions, the temporal evolution of the averaged airmass pressure level looks nearly identical with no substantial differences between the cold and warm cluster irrespective of whether non-HW or HW days are considered (not shown).

What could be the possible reason for heatwave-related diabatic heating differences between cold and warm ensemble members? Due to limited data availability, this question cannot be fully answered in this paper. One variable at our disposal is total cloud cover which temporal evolution along trajectory paths is plotted in Figs. 9c, 9d, 10c, and 10d, respectively. In agreement with expectation, significant differences exist between cold and warm ensemble members. Cold ensemble members display—in absolute terms—up to about 15% more cloud cover. This observation is, however, very similar in non-HW days. This is why cloud cover differences and corresponding changes in sensible heating of the overlying air mass are unlikely to fully explain the previously described behavior. In eastern CE, differences in traced total cloud cover...
are overall similar but slightly more pronounced, especially near the target region (i.e., on forecast day 3).

2) THE ROLE OF AIRMASS ORIGIN AND TRAVEL HISTORY

In the introduction of this paper, we argue that errors in the prediction of the large-scale flow are an unlikely error source in 3-day Tmax forecasts. However, even subtle differences in large-scale pressure pattern among ensemble realizations may lead to strongly diverging trajectory paths, even within a relatively short time span. Therefore, ensemble spread in Tmax forecasts may also depend on airmass origin. A different airmass origin, which is directly visible as potential temperature gap at \(2^\text{nd} \) day in Figs. 9a, 9b, 10a, and 10b, may of course further influence diabatic heating of air parcels along the trajectory path toward the target region. Figure 11 shows a clear split in trajectory origin between the clusters of the 10 warmest and 10 coldest ensembles for both regions, the more maritime-influenced western CE as well as the more continental eastern CE region. For both HW and non-HW days, trajectories associated with colder forecasts tend to originate more from northwesterly regions (relative to the ensemble mean, not in absolute terms) whereas trajectories belonging to warmer members show an anomalous southeasterly origin in the majority of cases. In the scatterplot, the “center of mass” of the cold and warm clusters’ relative longitude–latitude displacement does not differ much between regions outside of heatwaves. Within heatwaves, however, the distance between the cold and warm cluster’s “center of mass” is larger for western CE (Fig. 11b) than for eastern CE (Fig. 11d). Hence, western CE features larger intra-ensemble spread with regard to airmass origin. Given its proximity to North Sea and English Channel, this means therefore also a stronger likelihood of a coexistence of realizations where some trajectories remain over land whereas others travel over ocean for some time before arriving in western CE. In summer, diabatic heating of overlying air masses may of course strongly depend on whether they travel over ocean or continental land.

This intuitive hypothesis has been stressed before in a study by Santos et al. (2015) who found a strong role of high residence time over land and associated enhanced near-surface diabatic heating for heatwaves over the Iberian Peninsula. In our study for central Europe, this hypothesis is also supported by Figs. 12a and 12b, which depict differences in forecast-accumulated trajectory travel time over land for all subregions. On non-HW days, 3-day backward trajectories of cold ensemble members are generally associated with some 0–4 h less travel time over land whereas air masses of warm ensemble members reside over land some 0–4 h longer than the ensemble mean. This picture is very similar for all central European subregions. Within heatwaves, however, clear contrasts emerge: whereas no substantial changes are found for the northern and western region, the two ensemble clusters grow closer in terms of airmass residence time over land for the southern and eastern subregions. Synoptic situations associated with central European heatwaves are often characterized by weak pressure gradients (Spensberger et al. 2020). We therefore expect overall longer airmass residence over continental

**Fig. 10.** As in Fig. 9, but for the eastern CE region. In contrast to the western region, the eastern region includes 20 trajectory starting points instead of 17.
areas especially toward the southeast. This is indeed the case in our analysis as well (see green crosshair symbols depicting absolute ensemble mean values). Hence, heatwaves may offer less opportunity for diabatic heating differences to emerge from different residence time over land. Only toward the northwest, the proximity to North Sea and English Channel may substantially affect the time air masses reside over land even if the large-scale flow is nearly stagnant.

To put this result into context with previously discussed differences in diabatic heating history, Figs. 12c–f show again the temporal evolution of air parcel potential temperature similar to Fig. 9. This time, however, we do not split the analysis into a warm and a cold cluster. Instead, we pick the 10 ensemble members with the highest (brown) and lowest (dark blue) trajectory travel time over land, respectively. By comparing Fig. 12d against Fig. 9b, it is evident that for the western CE region travel time over land may to large extent explain the widening gap in potential temperature between warm and cold ensemble members. The only marked difference exists on forecast day 3, where the widening gap in potential temperature is not present. The emergence of a temperature difference at forecast day 2 already, which closely resembles the one between the warmest and coldest ENS cluster in 9b, is extremely unlikely to exist by chance. We tested this via a randomized recreation of Fig. 12d, performed 5000 times in a row as follows: for each calendar day of either the non-heatwave or heatwave period, we randomly select 10 ensemble members for two subsample sets while making sure that both sets are always disjoint. Whereas some cases turned up where both random clusters develop significant temperature differences to the 50-member ensemble mean, there is, however, not a single case where the characteristic diverging course develops with a comparable signal amplitude as in Fig. 12d.

Thus, for western CE heatwaves, ensemble spread in Tmax forecast is considerably determined by travel history of arriving air masses and associated anomalous heating in remote regions, but it may be further amplified by local diabatic heating errors, likely due to cloud forecast errors. In contrast,

**FIG. 11.** Origin of air masses during non-heatwave and heatwave days for western and eastern CE region. The scatterplots depict the origin of air masses in form of a relative displacement of backward trajectories at $-48 \text{ h}$ (i.e., at forecast hour $+24 \text{ h}$) compared to the start region (at forecast hour $+72 \text{ h}$) in pseudo longitude–latitude coordinates. (top) Results for the western CE subregion, for (a) non-HW and (b) HW days. (bottom) Results for the eastern CE region, again for (c) non-HW and (d) HW days. Each point in the respective scatterplots represents the “spatially averaged” mean of all $\approx20$ trajectories initiated in the respective subregion of one particular ensemble member at a particular day (time period 2018–20). Blue dots denote the 10 coldest ensemble members whereas red dots show the 10 warmest ensemble members, respectively. Dots in transparent colors do not differ significantly from the ensemble mean at 5% level. A filled star denotes the mean value of all data points deviating significantly from the ensemble mean for the respective warm/cold cluster whereas an unfilled star represents the mean value for all data points irrespective of significance. The percentages given in each quadrant of the plots denote the fraction of non-heatwave-related (blue) and heatwave-related (red) data points within the respective quadrant.
Divergent Tmax forecasts during heatwaves in the eastern CE region are associated with a comparably stronger influence of local errors in diabatic heating on forecast day 3. As a complimentary analysis, we calculate the average intersection fraction between the 10 warmest (coldest) ENS cluster and the respective clusters with highest (lowest) travel time over land; in other words, how many ensemble members are found in both, the subsample of the 10 warmest (coldest) and in the subsample of the 10 ensembles with the highest (lowest) travel time over land? In both, the western and the eastern CE region, this average intersection fraction amounts to about one-third outside of heatwaves (Table 2). Within

---

**FIG. 12.** The role of overland travel time differences for Tmax over- or underestimations based on 72-h backward trajectories calculated from ECMWF-ENS 72-h forecasts for the summers 2018–20. Differences in average trajectory travel distance to the 50-member ensemble mean (depicted by green crosshair symbols) for the respective 10 warmest (red boxes)/coldest (blue boxes) ECMWF-ENS members for (a) non-HW and (b) HW days in form of a classic box-and-whisker plot (5th, 25th, 50th, 75th, 95th percentiles highlighted, cross depicts the mean). Western CE: Mean differences in trajectory-traced potential temperature between the respective 10 members with highest (brown) and lowest (dark blue) travel time over land and the 50-member ensemble mean, for (c) 210 non-HW and (d) 48 HW days, respectively. Shadings denote the interquartile range. Before calculating the interquartile and significance statistics, the respective differences were averaged “spatially” over all trajectories initiated over the region of interest. Solid lines depict significant difference against the ensemble mean at the 5% level. (e),(f) As in (c) and (d), but for the eastern CE subregion.
heatwaves, this fraction remains virtually unchanged in western CE. In the eastern CE domain, however, we find a statistically significant drop in the intersection fraction, meaning that airmass residence time over land becomes a less important factor for Tmax forecasts errors in this more continental part of central Europe.

4. Summary and conclusions

This study examines causes of short-range (3-day) forecast errors in maximum 2-m temperature (Tmax) over central Europe (CE) for the summers 2015–20. To this end, ECMWF-ENS forecasts with 3-day lead time are evaluated against short 0–18-h integrations of the ECMWF-ENS control run for the day of interest, to which we refer to as quasi-analysis. First we investigate whether 3-day ECMWF-ENS forecasts display any systematic Tmax forecast error with respect to the quasi-analysis.

- During heatwaves, 3-day ECMWF-ENS forecasts show—evaluated against the model’s quasi-analysis—a slight cold bias over CE (−0.43 K) that is more pronounced on clear-sky days (−0.65 K) and low wind days (−0.85 K)
- For non-heatwave days, there exists no bias in Tmax forecasts for CE in ECMWF-ENS (with respect to the model’s quasi-analysis), although some cold bias also shows through in high insolation and low wind situations

We then systematically assess the role of different local physical processes that may impact Tmax forecasts on the 3-day time scale, utilizing both a pattern correlation-based method and a multivariate linear regression model. The main inferences can be summarized as follows:

- In ECMWF-ENS’s 3-day forecasts, Tmax forecast errors are predominantly linked to an over- or underprediction of shortwave radiation reaching the surface (SWDS), more so outside heatwaves than within heatwaves
- Errors in the partitioning of surface heat fluxes (Bowen ratio), which are strongly governed by soil moisture, generally play a much lesser role for Tmax forecast errors
- In heatwaves, errors in nocturnal residual layer temperature are a substantial and regionally often the most or second most important source of error (and much more important than Tmin or Tmax on the day before)

- On the regional scale, Tmax forecasts in heatwaves can be predominantly affected by errors in 10-m wind near the coast whereas further inland errors are in most regions mainly tied to errors in nocturnal residual temperature and to a lesser extent to errors in soil moisture

In summary, short-range forecasts errors of summertime maximum temperature over central Europe are primarily caused by over- or underestimation of shortwave irradiance, mainly due to erroneously predicted low cloud cover. However, the dominance of this error source diminishes substantially during heatwaves. To some extent, this finding may be related to a rather trivial explanation: in general, heatwaves feature extended periods of stable and cloud-free weather conditions. In the extreme case of both forecast and quasi-analysis displaying zero cloud cover, errors in shortwave radiation would then of course be almost eliminated as potential error source. On the other hand, reduced importance of SWDS errors in heatwaves also points to other error sources gaining importance such as near-surface wind near the coasts. Moreover, errors in Tmax may not exclusively be caused by errors in local diabatic heating on forecast day 3, which is clearly suggested by the high relative importance of errors in nocturnal residual layer temperatures during heatwaves.

To compliment the Eulerian-type statistical analyses we also adopted a Lagrangian view. Using a large ensemble of trajectories based on available ECMWF-ENS model data for summers 2018–20, we investigate how the 10 warmest and 10 coldest ENS members differ in terms of airmass origin and diabatic heating history:

- Particularly during heatwaves, air masses within the planetary boundary layer (PBL) may be subject to over- or underestimations of diabatic heating that may accumulate until forecast day 3
- In western CE, largest intra-ensemble spread in diabatic heating occurs on forecast day 2 whereas eastern CE air parcels display the largest spread in potential temperature mainly due to local diabatic heating differences on forecast day 3
- Intra-ensemble differences in diabatic heating of PBL air masses are linked to geographic origin (colder/warmer ENS = more northwesterly/southeasterly origin), cloud
cover traced along trajectories and in particular airmass residence/travel time over land

- Outside of heatwaves, warmer/cooler ECMWF-ENS 72-h Tmax forecasts coincide with longer/shorter residence time of air masses above land areas in all CE regions; during heatwaves, this observation still holds for western CE but becomes much less valid for eastern parts of CE

These findings generally agree with Zschenderlein et al. (2019) who found—although their Europe-focused study has not explicitly targeted forecast errors—that diabatic heating of near-surface air masses proves to be an important contribution to near-surface temperature changes on lead times up to 3 days. The role of residence time over land in more maritime regions has been highlighted before by Santos et al. (2015) and Hochman et al. (2021), albeit for the Mediterranean region. We want to stress that the limited time period of 2018–20 in the Lagrangian analysis may compromise the representativeness of these results. Although the sample size is high due to the inclusion of a 50-member ensemble, some of the reported findings may be tied to specific synoptic characteristics of heatwaves in the considered time period. This may also be partly true for the Eulerian-type analyses over the longer period 2015–20 as we saw some minor quantitatively changes of our results when the sample period was limited to 2018–20.

Future work may therefore aim to extend this study, either temporally or with an emphasis on the question of whether the presented results are also transferable to other regions outside of central Europe. In addition, further research may address the question of which type of clouds are mainly affecting SWDS and Tmax forecasts and how strong the sole effect of SWDS shading is compared to the effects of rainfall-associated evaporative cooling. With the heatwave-specific cold bias in mind that is aggravated in sunny and calm conditions, one may further investigate more deeply on the following question: Is there a general tendency in ECMWF-ENS to underestimate the storage of excess heat in the nocturnal residual layer under heatwave conditions? Aircraft observations analyzed by Zhang et al. (2020) have shown that the residual layer is much more pronounced in heatwave conditions. Likewise, Miralles et al. (2014) hinted at its potentially underestimated role in heat accumulation. Hence, looking more deeply into these possible issues may certainly be valuable. Tmax forecast errors associated with such errors in accumulated diabatic heating may be quite small, though, and improvements in this area fall more into the category of a minor bias improvement. Actual forecast busts where the 3-day forecast may be off by some 5–10 K are of course mainly associated with over- or underestimations of cloud cover. This crucial quantity is indeed still the one for which the forecast skill of the ECMWF-IFS forecasts drops below zero first, namely, at around forecast day 3 (Haiden et al. 2015). Thus, the quality of summertime 3-day Tmax forecasts will in our opinion benefit the most from improvements in prediction of convection and cloudiness.

Acknowledgments. This research has been supported by the Deutsche Forschungsgemeinschaft (Grant SFB/TRR 165, “Waves to Weather”) and conducted within the subproject C4: “Predictability of European heat waves.” We wish to thank Christian Grams and his working group “Large-scale dynamics and predictability” for storing ECMWF-ENS forecast data on model levels, which made possible the Lagrangian analysis in this study. We also thank the three anonymous reviewers for their very helpful comments and suggestions for improvement.

Data availability statement. The ECMWF-ENS data used in this study are available from ECMWF’s MARS archive (https://apps.ecmwf.int/datasets/). The ERA5 data used for the detection of heatwave days can be freely downloaded under the following URL: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview. ECMWF-ENS data on model levels are stored by the working group of Christian Grams at the Karlsruhe Institute of Technology and may be provided upon request (christian.grams@kit.edu). Scripts used to generate the plots of this paper can be provided by the corresponding author upon request (alexander.lemburg@kit.edu).

REFERENCES


Haiden, T., R. Forbes, M. Alhgrimm, and A. Bozzo, 2015: The skill of ECMWF cloudiness forecasts. ECMWF Newsletter, No. 143, ECMWF, Reading, United Kingdom, 14–19.

—–, I. Sandu, G. Balsamo, G. Arduini, and A. Beljaars, 2018: Addressing biases in near-surface forecasts. ECMWF Newsletter, No. 157, ECMWF, Reading, United Kingdom, 20–25.