Development of a combined empirical index for a 5-day forecast of heavy precipitation over the Bernese Alps

NGUYEN, Liliane, et al.

Abstract

The Bernese Alps are a region that is very prone for the initiation of thunderstorms. In fact, the flow and convergence of air and water vapor from the Swiss Plateau to the Swiss Alps is frequently favouring the formation of isolated rainfall events, which then may cause loss and damage in settlements. Due to the complex topography of the Bernese Alps, the forecasting and nowcasting of heavy convective precipitation remain challenging. A critical need therefore exists for the development of new forecasting tools so as to improve the predictability of convective precipitation events, also with the aim to alert first responders and to subsequently reduce damage. This study aims at developing an empirical index for the forecasting of heavy precipitation events in the Bernese Alps by using two reanalysis datasets, ECMWF’s ERA-Interim and NASA’s MERRA-2; in addition, the ICON-EU model is employed here to test and verify the index for the 2018 summer period. Our approach is based on the calculation of several convective indices as well as on the assessment of their relative forecast skills using a dichotomous scheme. The [...]
Development of a combined empirical index for a 5-day forecast of heavy precipitation over the Bernese Alps

Liliane Nguyen, Mario Rohrer, Manfred Schwarb, Markus Stoffel

Abstract

The Bernese Alps are a region that is very prone for the initiation of thunderstorms. In fact, the flow and convergence of air and water vapor from the Swiss Plateau to the Swiss Alps is frequently favouring the formation of isolated rainfall events, which then may cause loss and damage in settlements. Due to the complex topography of the Bernese Alps, the forecasting and nowcasting of heavy convective precipitation remains challenging. A critical need therefore exists for the development of new forecasting tools so as to improve the predictability of convective precipitation events, also with the aim to alert first responders and to subsequently reduce damage. This study aims at developing an empirical index for the forecasting of heavy precipitation events in the Bernese Alps by using two reanalysis datasets, ECMWF's ERA-Interim and NASA's MERRA-2; in addition, the ICON-EU model is employed here to test and verify the index for the 2018 summer period. Our approach is based on the calculation of several convective indices as well as on the assessment of their relative skill using a dichotomous scheme. The Heavy Precipitation Index (HPI) is then defined by combining the best performing combination of convective indices. HPI is aimed at forecasting heavy precipitation events over the Bernese Alps. We show that the combination of several indices, including DCI or KI, have a better capability to forecast heavy precipitation in the Bernese Alps than has the commonly used CAPE. Therefore, HPI should be seen as a pre-alert index when it comes to assist first responders in situations of crisis and in the process of decision making.

1. Introduction

As a result of its hilly and mountainous landscapes, major parts of Switzerland are very vulnerable to the occurrence of extreme weather phenomena, and the subsequent triggering of natural hazard events such as (torrential) floods, landslides, debris flows, or snow avalanches (Badoux et al., 2016; Stoffel and Corona, 2018a; Voumard et al., 2018). Extreme weather and subsequent hazard events can develop from short-lived, yet intense thunderstorms (Graham et al., 2012; Stoffel et al., 2014; Giorgi et al., 2016; Scherrer et al., 2016), mostly in summer, or from longer-lasting, yet less intense stratiform rainfall episodes, predominantly in autumn (Massacand et al., 1998; Giannakaki and Martius, 2015). The negative impacts of intense rainfall events are sometimes further enhanced by rain-on-snow events, leading to massive runoff and floods (Beniston and Stoffel, 2016; Morán-Tejeda et al., 2016; Fehlmann et al., 2019; Stoffel and Corona, 2018b).

Heavy thunderstorms are one of the major causes of extreme summer weather events in the alpine region (Badoux et al., 2016; Andres and Badoux, 2018), particularly at foothills, like in the Bernese Alps. This area is prone to moist upvalley and upslope flow from the Swiss plateau, which favours the initiation of thunderstorms with localized heavy rainfall events (Graham et al., 2012). These phenomena are typically of rather short duration as well as stationary over an area, and can thus lead to the occurrence of severe natural disasters in the form of (flash) floods, debris flows or shallow landslides (Ruiz-Villanueva et al., 2018; Schauwecker et al., 2019; Fehlmann et al., 2019).

Although the spatial resolution of global numerical weather prediction (NWP) models has improved considerably in the recent past (Bauer et al., 2015), the forecasting and nowcasting of heavy convective precipitation events remain challenging (Weckwerth et al., 2011; Wilson et al., 1998). This limitation is related to the fact that
NWPs may not yet resolve local topographic effects in the initiation of heavy convective precipitation events (Giorgi et al., 2016).

The general preconditions for the formation of convective storms are well known and require a coincidence of (1) a conditional instability, (2) a moist layer of sufficient depth in the lower or mid troposphere, and (3) a source of lift capable to initiate convection (Doswell, 1987; Johns and Doswell, 1992). To quantify the two first ingredients to form a convective storm (omitting the lifting mechanism), various thermodynamic parameters have been developed with the aim to improve the forecast of convective storms; they are often referred to as convective parameters or indices in the literature.

Convective parameters or indices indicate the potential of an air mass to develop into a thunderstorm. Parameters that have been employed to predict thunderstorms include the Lifted Index (Galway, 1956), K Index (George, 2014) or the original and modified Showalter Index for alpine regions (Showalter, 1947; Steinacker, 1977). To de- velop heavy precipitation events, a sufficient amount of Convective Available Potential Energy (CAPE) is needed. CAPE is among the most popular indices and has been used in several studies (e.g. Brooks et al. (2003), Brooks et al. (2007), Craven et al. (2004), Doswell and Evans (2003), Markowski et al. (2002), Rasmussen and Blanchard (1998)).

When combined with its opposing parameter, the Convective Inhibition (CIN), CAPE can characterize the convective potential of the atmosphere. However, several studies have shown that both CAPE and CIN are performing poorly in mountainous areas (Graham et al., 2009; Graham et al., 2012). Moreover, it has been commonly acknowledged that any convective index considered in isolation has only a rather limited performance. Doswell and Schultz (2006) state that no single index has the potential to properly forecast thunderstorms, because most indices only measure a very specific part of the atmospheric stability. Therefore, several indices, such as the Severe Weather Threat Index (Miller, 1972), have combined a suite of indices using statistical regression. In Switzerland, Huntrieser et al. (1997) suggested their own SWISS Index by investigating the relationship of different thermodynamic and kinematic parameters to the initiation of thunderstorms.

The best thermodynamic and kinematic parameters calculated from the radiosounding in Payerne (started at 0000 and 1200 UTC) were combined with the Payerne radiosounding data. The Bernese Alps have been a subject of Adelboden, Interlaken, and Grimsel are in the Bernese Alps.

In recent years, research increasingly used NWP to avoid the limitations given by the radiosondes (Korologou et al., 2014; Steinheimer and Haiden, 2007). The use of NWP allows the forecast of vertical atmospheric profiles for each grid point with a higher spatio-temporal resolution and offers a longer lead-time. However, in current NWP, only few convective indices are available to date, such as CAPE and CIN; combinations of indices are, by contrast, still scarce.

Our work has been undertaken under the European Horizon 2020 research project ANYWHERE (http://anywhere-h2020.eu/), whose aim was to create a Pan-European multi-hazard emergency management and response platform to empower exposed responder institutions and citizens to enhance their anticipation and pro-active capacity to face extreme and high-impact weather and climate events. The purpose of this study is to provide the Natural Hazards Division of the Canton of Bern (Switzerland) with a simple and fully automatic index to forecast heavy precipitation in the Bernese Alps by warning them up to five days in advance. Our working hypothesis is that NWP still has limited skills in forecasting heavy precipitation events, and that other prognostic variables remain more accurate for the time being, including those used to calculate convective indices. Therefore, the aim of this study is to (i) identify the best performing convective indices for the forecast of heavy precipitation over the Bernese Alps, (ii) compare different combinations of convective indices with the aim to develop a new empirical index aimed at forecasting heavy precipitation events, (iii) assess the ability of this new index to forecast heavy precipitation using an NWP model, so as to (iv) obtain a “Heavy Precipitation Index” (hereafter referred to as HPI) yielding high Hit Rates up to five day ahead of an event. As forecast and re-forecast data are typically short time-series and often discontinuous, we decided to employ reanalysis datasets to define a new combined index which is optimal in forecasting future heavy precipitation events. To gain confidence in the ability of this new index to forecast heavy precipitation events with NWP, we used two reanalysis datasets, the ERA-Interim reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) as well as the MERRA-2 dataset from the National Aeronautics and Space Administration (NASA). The relative forecast skills of the indices was then assessed in a dichotomous (yes/no) scheme. Based on the best combination of convective indices, we defined a new Heavy Precipitation Index in a final step. The ability of this new empirical index to forecast heavy precipitation was tested with the ICON-EU model from the German Weather Service (DWD) and the merged precipitation-radar CombiPrecip dataset from MeteoSwiss.

2. Data

2.1. Precipitation-gauge data

The Bernese Alps have been affected by intense rainfall events in the past (Collier and Lilley, 1994; Huntrieser et al., 1997; Schauwecker et al., 2019; Fehlmann et al., 2019). To develop the Heavy Precipitation Index, we used five rain-gauge stations from the SwissMetNet network of MeteoSwiss for which data is available for at least the past 30 years, namely Adelboden (ABO), Bern (BER), Grimsel (GRH), Interlaken (INT), and Napf (NAP). This automatic monitoring network provides precipitation data at ten-minute intervals. Table 1 describes the five rain-gauge stations with the number of heavy and non-heavy precipitation events between 1980 and 2015. Heavy precipitation events are defined here as events for which aggregated 6-hourly precipitation totals exceeded or were equal to a return period of two years at the rain-gauge station, based on values provided by MeteoSwiss calculated on the longest period available. Fig. 1 presents the location of the five rain-gauge stations: the stations of Bern and Napf are located on the Swiss Plateau, whereas the stations of Adelboden, Interlaken, and Grimsel are in the Bernese Alps.

2.2. Thermodynamic and kinematic parameters from objective reanalysis

Several existing thermodynamic and kinematic parameters (or convective indices) were then calculated for the five rain-gauge stations using the ERA-Interim reanalysis from ECMWF (Dee et al., 2011) and

Table 1

<table>
<thead>
<tr>
<th>N</th>
<th>NAME</th>
<th>ID</th>
<th>Lon. [°]</th>
<th>Lat. [°]</th>
<th>Alt. [m]</th>
<th>HP events</th>
<th>No-HP event</th>
<th>Total events</th>
<th>HP threshold [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adelboden</td>
<td>ABO</td>
<td>7.57</td>
<td>46.50</td>
<td>1326</td>
<td>26</td>
<td>46880</td>
<td>46906</td>
<td>≥33</td>
</tr>
<tr>
<td>2</td>
<td>Bern</td>
<td>BER</td>
<td>7.47</td>
<td>46.98</td>
<td>552</td>
<td>32</td>
<td>51100</td>
<td>51132</td>
<td>≥34</td>
</tr>
<tr>
<td>3</td>
<td>Grimsel</td>
<td>GRH</td>
<td>8.33</td>
<td>46.57</td>
<td>1980</td>
<td>24</td>
<td>38958</td>
<td>38982</td>
<td>≥36</td>
</tr>
<tr>
<td>4</td>
<td>Interlaken</td>
<td>INT</td>
<td>7.87</td>
<td>46.67</td>
<td>577</td>
<td>22</td>
<td>51110</td>
<td>51132</td>
<td>≥32</td>
</tr>
<tr>
<td>5</td>
<td>Napf</td>
<td>NAP</td>
<td>7.93</td>
<td>47.00</td>
<td>1403</td>
<td>30</td>
<td>51102</td>
<td>51132</td>
<td>≥39</td>
</tr>
</tbody>
</table>
the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) reanalysis from NASA (Rienecker et al., 2011), resulting in 19 convective indices for each of the five rain-gauge stations (Table 2).

ERA-Interim is an intermediate version for the next generation of reanalysis products and developed by ECMWF. It contains atmospheric and surface parameters, and has temporal and spatial resolutions of 6 h and 0.75 degrees, respectively; the dataset covers the period back to 1979 and contains 60 vertical levels from the surface up to 0.1 hPa.

MERRA-2 reanalysis is developed by NASA and was introduced to replace the original MERRA dataset because of the advances made in the assimilation system that enable assimilation of modern hyperspectral radiance and microwave observations, along with GPS-Radio Occultation datasets. It is a 3-hourly reanalysis product and is available since 1980. It has a native resolution of 0.5° lat × 0.625° long on 72 hybrid sigma/pressure levels.

Both reanalysis products have a rather coarse spatial resolution. The spatial resolution of ERA-Interim is approximately 80 km, and that of MERRA-2 is about 50 km, thus yielding only a small number of grid points over our area of interest. Moreover, and as a result of the coarse resolution, topography is flattened in both datasets over this mountainous region. Nonetheless, despite their coarse resolution, the convective indices derived from both reanalysis datasets may have skills, Taszarek et al. (2018) have, for instance, demonstrated that ERA-Interim reanalysis estimates parameters can describe boundary layer moisture and mid-tropospheric lapse rates well reasonably well. At the same time, Taszarek et al. (2018) in their comparison of convective parameters from radiosounding measurements and ERA-Interim datasets, illustrated that reanalysis data would still have significant limitations regarding instability and low-level shear parameters. However, the authors were also concluding that many errors remaining in the ERA-Interim datasets may not affect the qualitative interpretation of results, even if quantitative differences may occur whenever reanalysis-based thresholds are applied directly to observed soundings.

2.3. Validation datasets

To test and verify the ability of the HPI developed in this study to predict heavy precipitation up to five days ahead of an event, we recomputed it from the NWP model ICON-EU (ICOshedral Nonhydrostatic Europe) and verified it against the high-resolution

Table 2

<table>
<thead>
<tr>
<th>N</th>
<th>Convective index</th>
<th>ID</th>
<th>Unit</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Convective Available Potential Energy</td>
<td>CAPE</td>
<td>J/kg</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>Convective Inhibition</td>
<td>CIN</td>
<td>J/kg</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>downdraft CAPE</td>
<td>DCAPE</td>
<td>J/kg</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>Lifted Index</td>
<td>LI</td>
<td>°C</td>
<td>Galway (1956)</td>
</tr>
<tr>
<td>5</td>
<td>Showalter Index</td>
<td>SHOWI</td>
<td>°C</td>
<td>Showalter (1947)</td>
</tr>
<tr>
<td>6</td>
<td>K Index</td>
<td>KI</td>
<td>°C</td>
<td>George (2014)</td>
</tr>
<tr>
<td>7</td>
<td>mod. K Index</td>
<td>modKI</td>
<td>°C</td>
<td>Charba (1977)</td>
</tr>
<tr>
<td>8</td>
<td>Thompson Index</td>
<td>THOMI</td>
<td>°C</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>mod. Thompson Index</td>
<td>modTHOMI</td>
<td>°C</td>
<td>–</td>
</tr>
<tr>
<td>10</td>
<td>Total Totals</td>
<td>TT</td>
<td>°C</td>
<td>Miller (1972)</td>
</tr>
<tr>
<td>11</td>
<td>mod. Total Totals</td>
<td>modTT</td>
<td>°C</td>
<td>–</td>
</tr>
<tr>
<td>12</td>
<td>Deep Layer Shear</td>
<td>DLS</td>
<td>m/s</td>
<td>–</td>
</tr>
<tr>
<td>13</td>
<td>Craven Parameter/1000</td>
<td>Craven</td>
<td>m3/s3</td>
<td>–</td>
</tr>
<tr>
<td>14</td>
<td>Bulk Richardson Number</td>
<td>BRN</td>
<td>1</td>
<td>Weisman and Klemp (1982)</td>
</tr>
<tr>
<td>15</td>
<td>SWISS Index</td>
<td>SWISS</td>
<td>1</td>
<td>Huntrieser et al. (1997)</td>
</tr>
<tr>
<td>16</td>
<td>Severe Weather Threat Index</td>
<td>SWEAT</td>
<td>1</td>
<td>Miller (1972)</td>
</tr>
<tr>
<td>17</td>
<td>Deep Convective Index</td>
<td>DCI</td>
<td>°C</td>
<td>Barlow (1993)</td>
</tr>
<tr>
<td>18</td>
<td>Cross Total Index</td>
<td>CrossT</td>
<td>°C</td>
<td>Miller (1972)</td>
</tr>
<tr>
<td>19</td>
<td>Vertical Totals</td>
<td>VT</td>
<td>°C</td>
<td>Miller (1972)</td>
</tr>
</tbody>
</table>
precipitation product CombiPrecip provided by MeteoSwiss.

ICON-EU is the newest model based on the global forecast model ICON; it is a joint project of the German Weather Service (DWD) and the Max Planck Institute for Meteorology (MPI-M) (Zängl et al., 2015; Reinert et al., 2018). ICON-EU is a higher-resolution regional forecast model for Europe and is nested within the ICON global model. The native model grid has a horizontal grid spacing of 6.5 km, the output grid has a spacing of 0.0625° (∼7 km). In the vertical dimension, ICON-EU relies on 60 levels up to a height of 22.5 km. Forecast runs are performed four times per day, the forecast extends to a maximum of 180 h starting at 00 and 12 UTC and to 120 h for the runs starting at 06 and 18 UTC.

CombiPrecip is the result of precipitation estimated through the merging of radar and rain-gauge station data. CombiPrecip therefore combines information from two sources, i.e. (a) hourly aggregation of rain-gauge precipitation measurements at the SwissMetNet stations and (b) radar precipitation estimates from the radar composite precipitation product (Sideris et al., 2014). CombiPrecip yields hourly accumulated rainfall data since 2005 at a spatial resolution of 1 km².

3. Methods

3.1. Development of a Heavy Precipitation Index

The empirical index developed in this study, named Heavy Precipitation Index (HPI), was created from a combination of the best performing convective indices presented in Section 2.2. After an evaluation of existing indices and their performance, a combination of indices specific to each rain-gauge station was realized by using the ERA-Interim and MERRA-2 reanalysis datasets. Then, different combinations of indices were investigated and the best overall indices were chosen to create an algorithm forming the HPI.

In a first step, 19 convective indices were calculated with the atmospheric parameters from the ERA-Interim and MERRA-2 datasets for the five rain-gauge stations: temperature, relative humidity, specific humidity, horizontal components of wind, and geopotential height. Performance of the convective indices was evaluated with the help of a statistical method commonly used in forecast verification, i.e. the Relative Operating Characteristic or Receiver Operating Characteristic diagram (ROC diagram) (Mason, 1982). The ROC diagram plot shows the Hit Rate (H) (Y-axis) against the False Alarm Rate (F) (X-axis) for different decision thresholds. The ROC diagram gives the advantage to see the trade-off between sensitivity and specificity for different possible threshold. In the ROC diagram, data were entered into a 2 × 2 contingency table of forecasts and observations with four elements a to d based on whether an event was observed (YES/NO) and predicted (YES/NO) (Table 3).

The a table entry gives the number of event forecasts that correspond to event observations, or the number of hits; entry b shows the number of event forecasts that do not correspond to observed events, or the number of false alarms; entry c is the number of no-event forecasts corresponding to observed events, or the number of misses; and entry d illustrates the number of no-event forecasts corresponding to no-events in observations, or the number of correct negatives. An event was observed as soon as precipitation events measured at the rain-gauge stations exceeded or were equal to a return period of two years within the next 12 h using the 6-hourly precipitation data. On this basis, we calculated Hit Rates and False Alarm Rates as follows:

\[
H = \frac{a}{a + c}
\]

(1)

\[
F = \frac{b}{b + d}
\]

(2)

The Hit Rate \( H \) is the proportion of occurrences that were correctly forecasted and the False Alarm Rate \( F \) is the proportion of non-occurrences that were incorrectly forecasted. The ROC diagram is a graph of the Hit Rate against the False Alarm Rate as the decision threshold varies, with the False Alarm Rate plotted on the X-axis and the Hit Rate on the Y-axis. Therefore, each threshold generated a new 2 × 2 contingency table with new values for \( H \) and \( F \). The threshold for each convective index was set at 5% intervals between 0 and 100%, which is fine enough for the study of heavy precipitation as we are looking here for events with a return period of 2 years. The ability of an index to forecast heavy precipitation events was then evaluated based on the area located under the ROC curve (AUC). Thus, the larger this area, the better the forecast. An area of 1 indicates a perfect Hit Rate with no false alarm. Based on the AUC, the best performing indices were identified and thereafter used to create the combination of convective indices for each station.

The threshold for each of the best performing indices was set when a perfect Hit Rate was reached. A cross-validation method was then used to test the thresholds. To this end, a 10-fold cross-validation was applied. The advantage of this method was that all observations were used at least once in the training and validation sub-sampling. Thereafter, a combination of indices was built for each of the five stations with the best performing indices. By combining too many indices, the number of hits may reduce. Because the index aims at not missing any events, it seemed appropriate to test different numbers of indices to be included in the final HPI. In our case, only one combination was tested. The best performing indices were ordered based on their True Skill Score (TSS), from higher to lower score. TSS is also sometimes referred to as the Hansen Kuipers’ score and is in fact a verification measure yielding the ability of the forecast to distinguish event occurrences from non-occurrences (Hansen and Kuipers, 1965). The score has a range of −1 to +1, with 0 representing no skill and 1 representing a perfect score. In this study, TSS was calculated based on the outcomes of the contingency table as follows:

\[
TSS = H - F = \frac{(ad - bc)}{(a + c)(b + d)}
\]

(3)

The same cross-validation method was used to test this combination with different numbers of indices. The indices and the number of indices to be included in the index were chosen on the basis of the True Skill Score (TSS) resulting from the cross-validation for each station, so that heavy precipitation events were not missed on the one hand, but without issuing too many false alarms on the other hand.

Based on the results of each station, the indices with a good overall performance and a stable threshold over the stations and reanalysis were then chosen to develop an algorithm to form the HPI that we applied over the Bernese Alps.

3.2. Forecast verification methods

The performance of the Heavy Precipitation Index was assessed based on the ICON-EU model and real-time merged radar-rain-gauge precipitation records. We chose the summer period of 2018 (1 May to 31 August, 2018) to test the new index. Several heatwaves occurred during this period, and resulted in the fifth warmest May, the fourth warmest June, and fifth warmest July (fourth for some regions) on record since the beginning of systematic measurements in Switzerland in 1864 (MeteoSwiss, 2018). Moreover, and although the summer 2018

Table 3

<table>
<thead>
<tr>
<th>Event forecast</th>
<th>Event observed</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>a (Hit)</td>
<td>b (False alarm)</td>
<td>c (Miss)</td>
</tr>
</tbody>
</table>
The ability of the new index to predict heavy precipitation was assessed with a dichotomous forecast verification method using the $2 \times 2$ contingency table of forecasts and observations. In a first step, the convective indices chosen for the Heavy Precipitation Index were calculated with the ICON-EU model. To this end, new thresholds had to be estimated for each of the chosen indices based on data from the summers of 2016 and 2017. Similar to the differences between reanalysis and radiosoundings (Taszarek et al., 2018), we could not simply use the thresholds as provided by the reanalysis in this operational forecast model. Therefore, we adapted new thresholds for the ICON-EU model from the thresholds found in the reanalysis in a first guess approach so as to not consider the main goal of this study, i.e. to maximize the Hit Rate under the condition of not increasing the False Alarm Rate. Table 4 presents the new thresholds obtained for the Heavy Precipitation Index. Then, a contingency table was derived for each grid cell of the ICON-EU model. An event was forecasted if all convective indices included in the Heavy Precipitation Index predicted an event. To simplify the implementation, a single precipitation threshold was used. Therefore, an event was observed when CombiPrecip presented precipitation totals greater than 20 mm within the ICON-EU model grid cell. This value was chosen based on the average precipitation value found for a return period of two years based on several rain-gauge records from the area of interest.

Finally, the performance of the Heavy Precipitation Index was evaluated with the skill scores calculated from the contingency tables, plus two additional skill scores presented hereafter. The False Alarm Ratio (FAR) is the fraction of the predicted “yes” events that actually did not occur, and is calculated as follows:

$$FAR = \frac{b}{(b + a)}$$

(4)

FAR is sensitive to false alarms, but ignores misses. Values range from 0 to 1, where 0 indicates a perfect score. The last skill score used was the Heidke Skill Score (HSS). HSS measures the fractional improvement of the chance forecast over the standard forecast, and is calculated as follows:

$$HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}$$

(5)

Negative Values indicate that the chance forecast is better, 0 mean no skill, and a perfect forecast obtains a HSS of 1.

4. Results

4.1. Heavy Precipitation Index as evaluated by ERA-Interim

Performance of the 19 convective indices calculated with the ERA-Interim reanalysis was assessed for the five rain-gauge stations using a dichotomous forecast verification. A ROC diagram was drawn for each index and for each station. Their ability to forecast heavy precipitation events was based on the AUC from the respective ROC diagrams. Table 5 presents the AUCs calculated for each convective index and for each station. The two last columns show the mean of AUC. Values greater or equal to 0.75 (≥ 0.75) are emphasized in bold. Convective indices for Grimsel showed rather poor ability to forecast heavy precipitation. For this station, only the Deep Layer Shear (DLS) yielded high AUC values. By contrast, Interlaken showed the highest AUC values for most convective indices. Due to the poor ability of the convective indices to forecast heavy precipitation at Grimsel, we decided to omit this station from the development of the new index.

The mean AUC values were higher when several convective indices were considered. CAPE, CIN, TT, modTT, Craven, BRN and VT even showed AUC values exceeding 0.75. The highest AUC values were found for the KI and THOMI, and the lowest values for the modKI Index. As a result, seven convective indices showed a good ability to forecast heavy precipitation events for the stations of Adelboden, Bern, Interlaken, and Napf, namely the Lifted Index (LI), Showalter Index (SHOWI), K Index (KI), Thompson Index (THOMI), SWISS Index (SWISS), Severe Weather Threat Index (SWEAT) and Deep Convective Index (DCI), of which four are combinations of indices.

Fig. 2 presents the ROC diagram of the seven best performing convective indices for each station against CAPE. Results show that CAPE has less skills than other indices when using the ERA-Interim reanalysis data. For CAPE, the perfect Hit Rate could only be reached at the cost of a maximum False Alarm Rate, whereas the perfect Hit Rate for the seven best performing indices was reached already for limited False Alarm Rates, resulting in higher AUC values.

For CAPE, the perfect Hit Rate was reached when the perfect False Alarm Rate also reached as well. The perfect Hit Rate of the seven best performing indices was, by contrast, reached while the False Alarm Rate remained low. Resulting in a higher AUC value.

A 10-fold cross-validation approach was then realized to (i) set and validate the thresholds of each convective index, as well as to (ii) determine the best number of indices to include in the HPI. Fig. 3a presents the mean TSS resulting from the cross-validation for each of the seven convective indices and rain-gauge stations. Best TSS values were found for the station of Interlaken. On the other hand, the station of Adelboden showed low TSS values, especially for LI, SHOWI, and SWISS Index. The DCI gives the highest TSS values, going up to 0.77.

Fig. 3b presents the mean TSS of the Heavy Precipitation Index when including different numbers of indices resulting from the cross-validation for each station. To reduce the number of false alarms, an event was only forecasted by the HPI when all the indices included individually forecasted an event. The order of the included indices were based on their respective TSS calculated in the training set, going from high to low. Overall, the new index has higher values than each of the single indices alone, especially for Adelboden where the TSS has been greatly improved. Highest TSS scores for Adelboden and Bern were found when including two indices in the Heavy Precipitation Index. Best TSS values for Interlaken were found when the Heavy Precipitation Index only included one index. For Napf, the best TSS values were observed when including three indices. When including more than three indices, TSS values started to drop. Fig. 3c presents the mean Hit Rate resulting from the cross-validation for each convective index and rain-gauge station. Results show a high Hit Rate for all indices and stations, indicating that the thresholds set for each index correctly identified a great number of heavy precipitation events. Fig. 3d presents the mean Hit Rate of the new index when including different numbers of indices resulting from the cross-validation performed for each station. Results show that the highest Hit Rate values were found by including up to three indices in the Heavy Precipitation Index. When including more than three indices, the capacity of the HPI to identify heavy precipitation events started to decrease and became even lower than for one single index.

Fig. 3 shows the combination of convective indices yielding a good compromise between TSS and Hit Rate with indices. We used the three indices with the highest TSS score to create the HPI for each station. Table 6 presents the index combination for each station and thresholds for each index. Here, the DCI and THOMI were part of the HPI at each of the stations. The KI was included for Interlaken and Napf. The LI,
SWEAT and SHOWI are only found once for the stations of Bern, Adelboden, and Interlaken, respectively.

4.2. Heavy Precipitation Index as evaluated by MERRA-2

The methodology described for ERA-Interim was also applied to the MERRA-2 dataset. Table 5 presents the AUC calculated for each
convective index and for each station. Similarly to the results obtained with the ERA-Interim datasets, several convective indices showed higher mean AUC values as soon as the station of Grimsel was excluded from the analysis. Moreover, the same seven convective indices (LI, SHOWI, KI, THOMI, SWISS, SWEAT, and DCI) showed the best ability to forecast heavy precipitation events for the remaining stations. The AUC of the seven convective indices calculated with MERRA-2 yielded slightly better results when compared to those obtained with the ERA-Interim dataset (Table 5). The highest mean AUC values found with MERRA-2 were obtained with the KI, THOMI, and DCI, whereas the modKI provided the lowest AUC values.

Fig. 4 presents the ROC diagram obtained with the seven best performing convective indices against CAPE for each rain-gauge station. Similarly to ERA-Interim, results show that CAPE has lower skills than the aforementioned indices when using the MERRA-2 reanalysis. Fig. 5a presents the mean TSS resulting from the cross-validation for each of the seven convective indices and rain-gauge stations. The DCI reached the highest TSS values for all stations. By contrast to the results obtained with the ERA-Interim dataset, TSS values are better overall, especially for the station of Adelboden for which the highest values were recorded. The lowest values were found for the station of Bern. Fig. 5b presents mean TSS for the HPI by including different numbers of indices resulting from the cross-validation for each station. Here, the HPI had a better ability to forecast heavy precipitation events than when using a single index. TSS values of the HPI were higher than the best performing indices, i.e. the DCI. For the station of Adelboden and Napf, the highest TSS values were found by including either six or all seven indices in the HPI. For the station of Bern, the highest TSS values were found when including between five and seven indices. On the other hand, the highest TSS values for Interlaken were reached if only one index was included. The lowest TSS values were found by including up to two indices in the new index. In opposition to TSS values found with the ERA-Interim datasets, TSS values tend to increase when the number of indices in the HPI is increased. Fig. 5c presents the mean Hit Rate resulting from the cross-validation for each convective index and station. Results show high Hit Rate values, indicating that the thresholds set for each index were able to correctly identify a greater number of heavy precipitation events than in the ERA-Interim dataset. The station of Adelboden and Interlaken even had a perfect Hit Rate score with the SWISS Index and the DCI, respectively. Fig. 5d presents the mean Hit Rate of the HPI when including different numbers of indices resulting from the cross-validation for each station. Similarly to results from the ERA-Interim dataset in Fig. 3d, results show that the highest Hit Rate values for the station of Adelboden, Bern, and Interlaken were found when including only one index in the HPI. On the other hand, the highest TSS value found for Napf was obtained by including up to three indices. In opposition to the TSS values found in Fig. 5b, Hit Rate values tended to decrease when including more indices in the Heavy Precipitation Index.

Table 5 shows that the optimal combination of convective indices yielding the best compromise between the TSS and Hit Rate was less obvious to be identified in the case of the MERRA-2 analysis as compared to the results of the ERA-Interim datasets. The best compromise between TSS and Hit Rate results from the inclusion of three indices in the HPI. Therefore, the three indices with the highest TSS score were used to create the HPI for each station. Table 6 presents the optimal
combination for each station with the thresholds of each index included. Similarly to results from the ERA-Interim analysis in Table 6, DCI and THOMI were found in the HPI of each station. KI was present in the HPI of Interlaken and Napf, whereas LI and SWEAT were found only once for the stations of Adelboden and Bern, respectively.

4.3. New Heavy Precipitation Index

For the ease of use at the operational level, a new Heavy Precipitation Index for the Bernese Alps is deducted from the three best performing convective indices over all stations and across (Table 6). By doing so, we first excluded indices with too little predictive skill. The three indices included in the Heavy Precipitation Index are presented in Table 7 with their respective thresholds, and included the DCI, KI, and SHOWI.

DCI showed a particularly good ability to forecast heavy precipitation events in both ERA-Interim and MERRA-2 reanalysis for all rain-gauge stations. Moreover, its threshold remained stable through both reanalyses and all stations. In addition to DCI, we selected KI and SHOWI, as these two indices were included in the optimal combinations for some of the rain-gauges and as their thresholds moreover showed a very stable performance in the cross-validation.

The second index chosen was the KI. Although it has only been selected for some stations, the thresholds as evaluated by reanalysis remained very stable. The last index included was SHOWI, as similarly to KI – the thresholds of this index remained very stable.

Thresholds of the three indices were set to the highest level evaluated by reanalysis in order to keep the number of false alarm low.

The new HPI produces categorical forecasts in terms of distinct warning levels. A first category is set in cases for which none of the indices included in the optimal index is forecasting an event. In this case, the level of warning is set to “unlikely”. The second category is defined if one of the indices included in the optimal index is forecasting an event. The level of warning is set to “possible” in this case. In the same way, the third and fourth category are launched if two and all three indices are forecasting an event, respectively. The level of warning is therefore set to “probable” for the third and to “most likely” for the fourth category.

Fig. 6 presents an example of a map produced by the HPI using MERRA-2 over the Bernese Alps on 9 October 2012 at 0 UTC. The illustration shows the different levels of warning for heavy precipitation events, and compares them with the radar image produced by MeteoSwiss (right) on 9 October 2012 at 1 UTC+1. Between the 7 and 10 October 2012, a moist air mass stagnated over Switzerland. The Bernese, Valais, as well as the Grisons Alps were hit by persistent precipitation. On 9 October 2012, the Interlaken rain gauge station recorded 37.1 mm, Adelboden 28.6 mm, Berne 27.4 mm, and Napf 33.5 mm of rainfall, respectively. The optimal HPI issued warnings for heavy precipitation events over the entire Bernese Alps (and neighbouring regions). The severity of the warning decreased over the Alps with the lowest warning level over eastern Valais. When comparing the map given by HPI and the radar image from the same date, one can recognize good similarities in the shape, with the only difference that the area affected by precipitation in the radar image is located slightly farther west than forecasted by HPI.

4.4. Performance of the Heavy Precipitation Index

The performance of the HPI was tested with skill scores calculated for each grid point of the ICON-EU model over an area over Switzerland for summer 2018 (1 May to 31 August 2018) and for lead times ranging from 6 up to 120 h. By way of example, Figs. 7 and 8 present maps of the resulting Hit Rate, False Alarm Rate, True Skill Score and Heidke Skill Score for lead times of 72 and 120 h, respectively, showing that skill scores calculated do not differ significantly between different lead times. The Hit Rate shows high skills over the Bernese Alps (up to a
perfect score), but the score decreases over the Swiss Plateau and particularly over the western part of Valais. In opposition, the False Alarm Rate shows higher scores for the Swiss Plateau, but poor scores over the Bernese Alps and over mountainous areas in general. As for the True Skill Score, values are low in general, only the northern range of the Alps is showing a slightly better score, with values of up to 0.5. The Heidke Skill Score is indicating no skills, or very little skills, with values ranging from $-0.1$ up to 0.25 in the Jura Mountains.

Fig. 9 presents the mean values as obtained from contingency tables for the HPI with a total of 123 predictions. Values of mean hits and misses are consistent across lead times. The mean of missed events is around 0 and the mean hit events count is around 3.5. By contrast, the mean number of false alarms and correct negatives varied with differing lead times. The mean false alarm count tends to decrease with shorter lead times. For a lead time of 120 h, an average of 60 false alarms was produced by HPI over 123 predictions. This value decreased down to 57 with a lead time of 72 h, and 50 for a lead time of 6 h. The mean correct negative count tended to increase towards shorter lead times. For a lead time of 120 h, an average of 58 correct negatives was predicted. This number increased with shorter lead times, with values up to 69 with a lead time of 6 h.

Fig. 10 compares the mean values of the HPI contingency tables with the ones from the three best performing indices (i.e. DCI, KI, and SHOWI) for an area over the Bernese Alps. To validate our working hypothesis in which the precipitation forecast is less accurate than other prognostic variables from the NWP model, the HPI is also compared with the quantitative precipitation forecasts (QPFs) from the ICON-EU model. As a result, QPFs present a higher mean of correct negative values, but no events were forecast. In addition, all indices presented similar mean hits and misses. On average, the DCI presented the highest false alarm count whereas the KI showed smaller false alarm counts. Concerning SHOWI, and for a lead time of 120 h, the mean false alarm count was slightly higher than the mean false alarm count of HPI. But with shorter lead times, the difference of false alarm between SHOWI and HPI increased. The mean false alarm was reduced with HPI as compared to the three best performing convective indices. Similar results were found for the mean correct negative count. On average, the correct negative count of HPI was higher than the ones obtained for the three best performing convective indices. This ability increased further with shorter lead times.

5. Discussion

Convective indices are a popular tool to identify preconditions for the initiation of convective storms (de Coning et al., 2011; Gascon et al., 2015; Haklander and Van Delden, 2003; Huntrieser et al., 1997; Korologou et al., 2014; Merino et al., 2013). Several indices and combined indices were therefore developed in the past to increase the predictability of thunderstorms. These single and combined indices were usually developed for a specific region and did not necessarily perform well in other environments (Gascón et al., 2015; Graham et al., 2009). For instance, CAPE is among the most popular convective indices used to predict thunderstorms. It is also one in only a few convective parameters that is pre-calculated in NWP models. However, in our study focused over the Bernese Alps, CAPE showed lower skills than other indices in both ERA-Interim and MERRA-2 reanalysis products.
This finding is in line with Graham et al. (2009, 2012) who showed that CAPE is performing poorly in complex topography. Consequently, CAPE also turned out inadequate for the Bernese Alps when compared to other indices which have a better ability to predict heavy precipitation as DCI. Moreover, our results also revealed that convective indices are of limited use in complex mountainous topography. A poor ability to forecast heavy precipitation was found for the rain-gauge station of Grimsel for all temperature-based indices in both reanalysis datasets, and only the wind-based DLS parameter showed a good performance. The location of the Grimsel station can explain these results, as it is in fact located at high altitude within the Alps; it is therefore more susceptible to upslope flow precipitation events rather than convective processes. A common ingredient for heavy orographic rainfall is a very moist low-level jet (Lin et al., 2001). DLS, indicating a change of wind between the surface and an altitude of 6 km, may translate this phenomenon. We find that in regions where upslope flow dominate, temperature-based indices might not work well and wind-based indices could be more suitable. As a result, it appears that temperature-based indices have a good ability to distinguish events where convective processes are prevailing in our study area.

Due to the complexity of the Bernese Alps, convective precipitation can be affected by a multitude of processes such, including upslope flow, thus leading to mixed mechanisms. This makes the distinction between convective precipitation and other precipitation types very difficult. To overcome this difficulty, the thresholds of each index were defined based on a perfect Hit Rate so as to take account of all heavy convective precipitation events. The drawback of this threshold is that it increases the number of false alarms, which in turn may lead to the abandonment of the tool by the user. Results showed that the combination of the best performing indices increased TSS values, therefore indicating that the number of false alarms could be reduced. This was also confirmed by the results of the forecast verification, where the mean false alarm count of the HPI was smaller than that derived from the three best performing convective indices. The threshold set for each index showed almost perfect Hit Rate values in both reanalysis datasets, indicating that a great number of events were correctly forecasted by each index individually. However, mean Hit Rate values calculated with MERRA-2 showed higher values than those obtained with the ERA-Interim data, indicating that indices calculated with MERRA-2 were missing fewer events. This can be explained by the fact that the MERRA-2 reanalysis datasets have a higher spatial resolution which may improve the accuracy of indices. On the other hand, when considering each index individually, TSS values were found to be low, especially for ERA-Interim. This indicates that each index considered in isolation had a poor ability to distinguish observed events from observed non-events. By combining indices, mean TSS values increased, indicating that the number of false alarms could be reduced.

This finding is in line with Graham et al. (2009, 2012) who showed that CAPE is performing poorly in complex topography. Consequently, CAPE also turned out inadequate for the Bernese Alps when compared to other indices which have a better ability to predict heavy precipitation as DCI.

Moreover, our results also revealed that convective indices are of limited use in complex mountainous topography. A poor ability to forecast heavy precipitation was found for the rain-gauge station of Grimsel for all temperature-based indices in both reanalysis datasets, and only the wind-based DLS parameter showed a good performance. The location of the Grimsel station can explain these results, as it is in fact located at high altitude within the Alps; it is therefore more susceptible to upslope flow precipitation events rather than convective processes. A common ingredient for heavy orographic rainfall is a very moist low-level jet (Lin et al., 2001). DLS, indicating a change of wind between the surface and an altitude of 6 km, may translate this phenomenon. We find that in regions where upslope flow dominate, temperature-based indices might not work well and wind-based indices could be more suitable. As a result, it appears that temperature-based indices have a good ability to distinguish events where convective processes are prevailing in our study area.

Due to the complexity of the Bernese Alps, convective precipitation can be affected by a multitude of processes such, including upslope flow, thus leading to mixed mechanisms. This makes the distinction between convective precipitation and other precipitation types very difficult. To overcome this difficulty, the thresholds of each index were defined based on a perfect Hit Rate so as to take account of all heavy convective precipitation events. The drawback of this threshold is that it increases the number of false alarms, which in turn may lead to the abandonment of the tool by the user. Results showed that the combination of the best performing indices increased TSS values, therefore indicating that the number of false alarms could be reduced. This was also confirmed by the results of the forecast verification, where the mean false alarm count of the HPI was smaller than that derived from the three best performing convective indices. The threshold set for each index showed almost perfect Hit Rate values in both reanalysis datasets, indicating that a great number of events were correctly forecasted by each index individually. However, mean Hit Rate values calculated with MERRA-2 showed higher values than those obtained with the ERA-Interim data, indicating that indices calculated with MERRA-2 were missing fewer events. This can be explained by the fact that the MERRA-2 reanalysis datasets have a higher spatial resolution which may improve the accuracy of indices. On the other hand, when considering each index individually, TSS values were found to be low, especially for ERA-Interim. This indicates that each index considered in isolation had a poor ability to distinguish observed events from observed non-events. By combining indices, mean TSS values increased, indicating that the number of false alarms could be reduced.

This finding is in line with Graham et al. (2009, 2012) who showed that CAPE is performing poorly in complex topography. Consequently, CAPE also turned out inadequate for the Bernese Alps when compared to other indices which have a better ability to predict heavy precipitation as DCI.
values increased, showing that by combining indices, the distinction between observed events from observed non-events can be improved and the amount of false alarms can be reduced. Moreover, when combining different numbers of indices in the HPI, results revealed that mean Hit Rate values decreased slightly with the number of indices included, especially for the analysis based on ERA-Interim. Mean TSS values for MERRA-2 tended to increase if the number of indices included increased as well, but tended to decrease as soon as more than three indices were included in ERA-Interim. By combining too many indices, the combined index tended to miss more events than one index alone and the distinction between observed and observed non-events was not necessarily improved, especially for ERA-Interim. Therefore, the optimal HPI identified for the Bernese Alps was formed with three indices. The thresholds set for each index were close to the thresholds found by Huntrieser et al. (1997) for the Swiss Plateau and by Kunz (2007) for the northern parts of Baden-Württemberg in South-West Germany. The area investigated by Kunz (2007) is also characterized by complex topography with rolling terrain and low mountain ranges of

Fig. 6. Example of a map produced by the new optimal index using MERRA-2 (left) over the Bernese Alps with its different levels of warning for heavy precipitation events on 00 UTC 9 October 2012 at 0 UTC, compared with the 6-hourly accumulated CombiPrecip (right) on 9 October 2012 at 00 UTC. In white, when the occurrence of heavy precipitation is unlikely; in yellow, when the occurrence is possible; in orange, when the occurrence is probable; and in red, when the occurrence is most likely. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Map of the Hit Rate, False Alarm Rate, Success Ratio, False Alarm Ratio, True Skill Score and Heidke Skill Score for the Heavy Precipitation Index using the ICON-EU model for a 72 h lead time over Switzerland.
the Black Forest and Swabian Jura. Mean TSS values from the cross-validation indicate that the distinction between observed and observed non-events remains difficult. This can be explained by the fact that not all of the observed events were initiated by convection, but also by upslope flow processes or other mechanisms. Consequently, HPI aims at not missing any events rather than predicting them.

To apply a forecast verification, the thresholds of each index needed to be adapted. In our case, new thresholds were defined for the ICON-EU model based on the values defined during the assessment with re-analysis products. The thresholds were defined primarily for the Bernese Alps with the key aim to produce a new index designed for mountainous regions. Results from the forecast verification showed that the new index was performing better over mountainous terrain, such as the Alps and the Jura Mountains, where the Hit Rate and TSS presented better skills than in the other, less mountainous regions. Also, the aim of the new index at not missing any events was fulfilled for most parts of the Bernese Alps where perfect Hit Rate scores were obtained. However, this perfect score comes with a higher number of false alarm, as shown by the FAR. Noteworthy, perfect Hit Rates are, general speaking, sparse among poor scores over the Alps. As a result, it appears that HPI cannot be generalized over mountainous regions. Furthermore, the HPI is highly sensitive to the study area and to the atmospheric model used, and thresholds of the indices need to be adapted for other regions. Last but not least, results from the forecast verification also point to a rather stable performance of HPI and the three best performing convective indices with lead time up to 120 h when using the ICON-EU model, and thus point to a rather limited sensitivity of convective indices to model lead times.

![Fig. 8. Map of the Hit Rate, False Alarm Rate, Success Ratio, False Alarm Ratio, True Skill Score and Heidke Skill Score for the Heavy Precipitation Index using the ICON-EU model for a 120 h lead time over Switzerland.](image)

![Fig. 9. Mean Hits, False alarms, Misses and Correct negative from the contingency table for the Heavy Precipitation Index over the bernese Alps for 6-hourly lead time up to 120 h over 123 predictions.](image)
6. Conclusions

In this study, a new index is proposed for the forecast of heavy precipitation events in the Bernese Alps using an empirical methodology. The index does not claim to explain the physical processes involved in the development of heavy precipitation events, but rather aims at informing decision makers and first responders of potentially forthcoming heavy precipitation with a simple and fully automatic product. The best performing convective indices were first identified and then combined to reduce the number of false alarms. In the process of developing the HPI, several indices proved more adequate to forecast heavy precipitation in the study area than the commonly used CAPE. HPI also showed to have better skills in predicting heavy precipitation than single indices, resulting in an overall smaller number of false alarms.

The ability of HPI to distinguish observed and observed non-events remains, however, difficult, as HPI was designed at not missing any event. As a result, the number of false alarms remains high, which may possibly limit the use of HPI. This high number of false alarms is mainly due to the complex and mixed origin of events occurring in this area (being both convective and upslope flow) and the fact that the indices thresholds are set at a perfect Hit Rate.

Current NWP - even convection permitting models as COSMO-1 or COSMO-E – still struggle to forecast convective precipitation events (Blizáň et al., 2017; Klasa et al., 2018). In our study, the ICON-EU model has shown a poor ability to forecast heavy precipitation as well. Nonetheless, and by using more accurate prognostic variables from the same model, such as those used to calculate the convective indices, it seems possible to develop a complementary tool. Indeed, the results presented in this study demonstrate that the new HPI remains quite stable up to a lead time of 120 h. Noteworthy, the HPI is not meant to replace any other tools, but rather aims at adding a simple, additional index to assist first responders to stay vigilant, by alerting them up to five days in advance that a heavy precipitation event may possibly occur. As the False Alarm Rate of the HPI was rather high, the tool should not be used for direct warnings, but rather to issue pre-alerts to first responders. Based on such a pre-alert, the responsible agencies can increase preparedness by including further risk monitoring and ensuring operational readiness. A few hours before the expected event, nowcasting tools should be used in any case to prepare for intervention if needed.

CRediT authorship contribution statement


Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgements

This work has been supported by the European Horizon 2020 research project ANYWHERE (EC-HORIZON2020-PR700099-ANYWHERE). The authors would like to thank MeteoSwiss for providing precipitation data, CombiPrecip and radar images. They would also like to thank all the institutions that opened their reanalysis datasets and NWP products. ERA-Interim were retrieved from the ECMWF MARS server. MERRA2 data were retrieved from the ftp server provided by the Goddard Earth Sciences Data and Information Services Center. ICON-EU data were retrieved from the open data server of the DWD website. They would also like to thank the three anonymous reviewers whose comments/suggestions helped improve and clarify this manuscript.

References
