<table>
<thead>
<tr>
<th>Title</th>
<th>Can Lagrangian extrapolation of radar fields be used for precipitation nowcasting over complex Alpine orography?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Mandapaka, Pradeep V.; Germann, Urs.; Panziera, Luca.; Hering, Alessandro.</td>
</tr>
<tr>
<td>Date</td>
<td>2012</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/8181">http://hdl.handle.net/10220/8181</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2012 American Meteorological Society. This paper was published in Weather and Forecasting and is made available as an electronic reprint (preprint) with permission of American Meteorological Society. The paper can be found at DOI: [<a href="http://dx.doi.org/10.1175/WAF-D-11-00050.1">http://dx.doi.org/10.1175/WAF-D-11-00050.1</a>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.</td>
</tr>
</tbody>
</table>
Can Lagrangian Extrapolation of Radar Fields Be Used for Precipitation Nowcasting over Complex Alpine Orography?

PRADEEP V. MANDAPAKA, URS GERMANN, LUCA PANZIERA, AND ALESSANDRO HERING

MeteoSwiss, Locarno Monti, Switzerland

(Manuscript received 25 April 2011, in final form 11 July 2011)

ABSTRACT

In this study, a Lagrangian radar echo extrapolation scheme (MAPLE) was tested for use in very short-term forecasting of precipitation over a complex orographic region. The high-resolution forecasts from MAPLE for lead times of 5 min–5 h are evaluated against the radar observations for 20 summer rainfall events by employing a series of categorical, continuous, and neighborhood evaluation techniques. The verification results are then compared with those from Eulerian persistence and high-resolution numerical weather prediction model [the Consortium for Small-scale Modeling model (COSMO2)] forecasts. The forecasts from the MAPLE model clearly outperformed Eulerian persistence forecasts for all the lead times, and had better skill compared to COSMO2 up to lead time of 3 h on average. The results also showed that the predictability achieved from the MAPLE model depends on the spatial structure of the precipitation patterns. This study is a first implementation of the MAPLE model over a complex Alpine region. In addition to comprehensive evaluation of precipitation forecast products, some open questions related to the nowcasting of rainfall over a complex terrain are discussed.

1. Introduction

The hydrologic response of a river basin to a rainfall event is a result of complex nonlinear interactions between rainfall, hillslope processes, and the river network. For example, aggregation and attenuation of flows within the river network play a key role in determining the shape of the streamflow hydrographs (e.g., Menabde et al. 2001; Mantilla et al. 2006). For smaller mountainous basins such as those in the Alps, the location and timing of the rainfall pattern with respect to the hillslopes and river network are very critical. The response times of such basins are often very short, ranging from subhourly to 12 h (e.g., Ahrens et al. 2003; Zillgens et al. 2005) and require very short-term (up to 6 h), high-resolution quantitative precipitation forecasts (QPF) for flash-flood and debris flow forecasting. The high-resolution very short-term QPFs are also required for nonhydrological purposes such as aviation.

Since weather radars provide good areal coverage with high resolution, many radar-based very short-term QPF (also referred to as nowcasting) techniques were developed over the years (e.g., Austin and Bellon 1974; Seo and Smith 1992; Andrieu et al. 1996; Dolcine et al. 1997; Mecklenburg et al. 2000; Pierce et al. 2000; Germann and Zawadzki 2002; Mueller et al. 2003; Seed 2003; Bowler et al. 2004, 2006). The radar echo extrapolation is one of the earlier and most widely used very short-term QPF techniques. The forecasts from the extrapolation-based models meet aforementioned stringent requirements such as high space–time resolution and short lead times. In addition, due to the relative simplicity of the extrapolation techniques, the corresponding forecasts can be generated with high frequency and in near–real time to provide emergency management authorities sufficient time to issue an alert.

The extrapolation model can be expressed in mathematical form as

\[ \Psi(t_0 + \tau, x) = \Psi(t_0, x - \alpha), \]

where \( \Psi(t_0, x - \alpha) \) is the observed precipitation field for time \( t_0 \) at location \( x - \alpha \) and \( \Psi(t_0 + \tau, x) \) is the forecast precipitation field for time lag \( \tau \) at location \( x \). The estimation of the displacement vector \( \alpha \) is an important part of any extrapolation scheme. Techniques such as cell tracking and variational echo tracking have been employed in various radar echo extrapolation models. The
technical details of the tracking algorithms, and the skill of the corresponding forecasts, have been reported in the literature for a single-radar domain to continental scale (e.g., Bellon and Austin 1978; Tsonis and Austin 1981; Bellon and Austin 1984; Germann and Zawadzki 2002; Hering et al. 2004; Berenguer et al. 2005; Reyniers 2008; Sokol et al. 2009; Novák et al. 2009).

The extrapolation-based forecasts were shown to have good skill up to lead times of 1–2 h. However, their skill decreases rapidly with lead time because they do not account for the initiation, growth, and dissipation of precipitation patterns (e.g., Browning and Collier 1989; Wilson 2004). On the other hand, numerical weather prediction (NWP) models have lower skill at short lead times but are expected to have better skill than the extrapolation forecasts at longer lead times due to better representation of atmospheric physics. The crossover lead time beyond which NWP model forecasts have better skill than the extrapolation forecasts was estimated to be in the range of 3–8 h (e.g., Lin et al. 2005; Vasic et al. 2007). In a recent study, Panziera et al. (2011) compared the skill of an analog-based nowcasting technique, Eulerian persistence, and a numerical weather prediction model [the Consortium for Small-scale Modeling model covering the Alpine region (COSMO2)] over the 7830 km² Lago Maggiore region in the southern Alps. They reported better skill of analog-based nowcasting approach over the COSMO2 model up to lead times of 4 h.

The continuous improvement in the ability of the NWP models to resolve the small-scale atmospheric processes prompted researchers to integrate radar data with NWP models to better forecast storm dynamics at short lead times. One approach was to assimilate information from the radar and satellite observations (e.g., boundary layer convergence, radar echo advection) into the NWP models (e.g., Pierce et al. 2000; Mueller et al. 2003; Bowler et al. 2004; Leuenberger 2005; Bowler et al. 2006; Liang et al. 2010). Another approach was to blend the output from NWP models and radar-based QPF systems to improve the accuracy at short space–time scales (e.g., Golding 1998; Atencia et al. 2010). In a recent study (Kober et al. 2011), a probabilistic nowcasting method was blended with the high-resolution NWP model forecasts and the resulting forecast products were evaluated over a large domain covering Germany.

Despite significant advances in the aforementioned approaches of merging radar-based observations and forecasts into NWP models, the resulting forecasts are not up to the desired level of accuracy for scales of thunderstorms and flash floods (e.g., Wilson 2004; Pierce et al. 2004; Lin et al. 2005; Wilson et al. 2010). For example, Wilson et al. (2010) compared QPFs from various sources for the Beijing Summer Olympics project and reported that the radar-based extrapolation forecasts have much better skill than those from the blended NWP-based models.

\section*{a. Objectives}

At MeteoSwiss, our aim is to develop an operational heuristic tool for the nowcasting of storm advection, growth, and dissipation for the region characterized by complex orography. Advection is an important component of overall storm evolution; therefore, we intend to use Lagrangian extrapolation of radar fields as the primary component, and introduce storm growth and dissipation in a statistical manner. Before we set out toward the development of the above heuristic tool, it is important to quantify the performance of extrapolation forecasts rigorously over complex terrain. In this study we focus on the extrapolation component and our objectives are twofold:

1) To quantify the skill achieved by Lagrangian extrapolation of radar fields over the complex region covered by the MeteoSwiss radar composite (620 × 620 km² domain shown in Fig. 1). Figure 1 shows the radar quality mask, which will be discussed later in section 3.

2) To compare the skill of extrapolation forecasts with corresponding Eulerian persistence forecasts.
baselined scenario) and high-resolution NWP model forecasts (complex scenario), and we estimate the crossover lead time.

The extrapolation scheme selected for the analysis is McGill Algorithm for Precipitation nowcasting using Lagrangian Extrapolation (MAPLE) developed by Germann and Zawadzki (2002). The NWP model selected for the analysis is COSMO2, run by MeteoSwiss as part of the Consortium for Small-scale Modeling (e.g., Steppeler et al. 2003; Leuenberger and Rossa 2003). Germann et al. (2006b) evaluated MAPLE using 2720 × 2720 km\(^2\), 4-km resolution radar-reflectivity composites in the United States. This study continues in the same vein but performs a rigorous evaluation of the extrapolation and NWP forecasts for a large domain broadly centered on Switzerland. To the best of our knowledge, no such comparison of extrapolation and NWP model forecasts was carried out over complex Alpine terrain.

b. Radar-rainfall uncertainties

Before proceeding further, a note of caution regarding the radar-rainfall estimation errors: it is well known that radar-rainfall estimates are affected by uncertainties from various sources (e.g., Austin 1987; Krajewski and Smith 2002; Germann et al. 2006a; Mandapaka et al. 2009, 2010; Villarini and Krajewski 2010). Some of the challenges associated with the radar-rainfall estimation in mountainous regions such as Switzerland are severe beam shielding, strong ground clutter, brightband contamination, beam overshooting, and difficult operating conditions (Germann et al. 2006a). Villarini and Krajewski (2010) provided an exhaustive review of various sources of uncertainties in the radar-rainfall estimates. Over the years, several studies evaluated radar-rainfall products, proposed models for residual errors, and employed them to represent radar-rainfall observational uncertainties in the form of ensembles. [See Mandapaka and Germann (2010) for a review.]

Likewise, the ensemble framework is also widely used to characterize uncertainties in QPF models (e.g., errors in the initial conditions, model structure). Germann and Zawadzki (2004) proposed a “local Lagrangian” approach to obtaining probabilistic forecasts. In a recent study, Berenguer et al. (2011) proposed a probabilistic Lagrangian extrapolation forecast technique to account for the uncertainties in the temporal evolution of rainfall patterns. However, much work needs to be carried out regarding the superposition of the radar-rainfall estimation errors and the QPF errors. In this study we focused on deterministic forecasts from MAPLE and COSMO2 and considered radar-rainfall observations to be the reference rainfall fields. Propagation of rainfall measurement uncertainties through the forecast chain and quantifying the overall errors in the QPF model output are beyond the scope of this study.

c. Outline

The forecast setup consisting of MAPLE, Eulerian persistence, and COSMO2 models is briefly described in section 2. The radar-rainfall data are presented in section 3 and their space–time characteristics are presented in section 4. The forecast verification methodology and the corresponding skill scores are described in section 5. The MAPLE, COSMO2, and Eulerian forecasts are evaluated against the radar observations using a variety of verification approaches in section 6. A discussion on the orographic effects and possible ways of incorporating them is presented in section 7 followed by our conclusions in section 8.

2. Forecast setup

a. Overview of MAPLE

In this section, we provide only a brief description of the MAPLE algorithm, which was originally developed by the radar group at the McGill University, in Montreal, Quebec, Canada. The algorithm and its skill in forecasting rainfall events over the continental United States have been well documented in a series of papers (e.g., Germann and Zawadzki 2002, 2004; Turner et al. 2004; Germann et al. 2006b). Recently, a thorough sensitivity analysis and a real-time verification of the MAPLE model were carried out using a radar network in South Korea (Bellon et al. 2010; Lee et al. 2010). The MAPLE forecast setup primarily consists of two steps:

1) velocity field estimation using the variational echo tracking algorithm, and
2) extrapolation of the current radar image, honoring the motion field estimated in the above step.

1) VARIATIONAL ECHO TRACKING

The variational echo tracking (VET) algorithm was proposed by Laroche and Zawadzki (1994) to estimate the three-dimensional wind field from the single-Doppler clear-air echoes. The technique was later modified by Germann and Zawadzki (2002) to obtain the velocity field from the radar-rainfall composites. The technique minimizes the following cost function:

\[ I_{VET}(u) = J_\Psi + J_2. \] (2)

The first term \( J_\Psi \) is the sum of squares of residuals of the conservation equation:
$$J_\Psi = \int_\Omega \beta(x) \left[ \Psi(t_0, x) - \Psi(t_0 - \Delta t, x - u \Delta t) \right]^2 \, dx \, dy,$$

where $\beta(x)$ is the weighting factor representing the data quality. The second term in the equation $J_2$ is a smoothness penalty function, written as

$$J_2 = \gamma \int_\Omega \left( \frac{\partial^2 u}{\partial x^2} \right)^2 + \left( \frac{\partial^2 v}{\partial y^2} \right)^2 + 2 \left( \frac{\partial^2 u}{\partial x \partial y} \right)^2$$

$$+ \left( \frac{\partial^2 v}{\partial x^2} \right)^2 + \left( \frac{\partial^2 v}{\partial x \partial y} \right)^2 \, dx \, dy,$$

where $\gamma$ is a constant weight and $u$ and $v$ are the x and y components of the velocity field $u$, respectively. Similar to Germann and Zawadzki (2002), a conjugate-gradient algorithm was used in the global minimization of the above cost function. To avoid converging toward secondary minima, Germann and Zawadzki (2002) followed the scaling-guess procedure (Laroche and Zawadzki 1994), where the VET algorithm is iteratively initiated and run at multiple scales, starting with coarser scales and gradually moving down to finer scales. We adopted the same approach and estimated the velocity fields starting with a coarse resolution of $600 \times 600$ km$^2$ (one vector for the entire domain), then moved to a resolution of $100 \times 100$ km$^2$ ($6 \times 6$ vectors), and to a final resolution of $25 \times 25$ km$^2$ ($24 \times 24$ vectors). The procedure was repeated every 5 min using 30 min of past radar-reflectivity composites. All the computations were carried out using radar-reflectivity fields in the dBZ scale, and the pixels with reflectivity value $<10$ dBZ (rain rate of 0.1 mm h$^{-1}$) were considered to be nonrainy. However, it should be noted that, wherever necessary, we employed the reflectivity ($Z$)–rain rate ($R$) relationship $Z = 316R^{1.5}$ to convert reflectivity to rain rate.

2) RADAR ECHO EXTRAPOLATION

There are various ways in which the velocity field from the VET technique can be used in the extrapolation of radar echoes. Germann and Zawadzki (2002) described the following four methods: 1) a constant vector forward scheme, 2) a constant vector backward scheme, 3) a semi-Lagrangian forward scheme, and 4) a semi-Lagrangian backward scheme. In this study we employed a semi-Lagrangian backward scheme in the radar echo extrapolation module. In the semi-Lagrangian backward scheme the origin of a parcel that would end up at a particular grid point in the forecast field is determined by following the streamlines upstream in space and backward in time. A semi-Lagrangian scheme also allows for the rotation in the displacement vector $\alpha$ for the extrapolation up to a certain lead time, $\tau$.

Given the velocity field $u$ at the time instant $t_0$ from the VET algorithm, the lead time $\tau$ is divided into $N$ time steps of length $\Delta t$, and $\alpha$ is iteratively obtained as follows:

$$\alpha = \Delta t u \left( t_0, x - \frac{\alpha}{2} \right).$$

The origin of the parcel determined using the backward scheme may not always coincide with a grid point in the current reflectivity field. In such a scenario, bilinear interpolation was carried out to estimate the reflectivity value at the origin. In addition, for longer lead times, it is possible that the backward scheme locates the origin of a parcel to be outside the domain covered by the radar composite. Such pixels are flagged in the forecasts and are not considered in the forecast evaluation.

b. Eulerian persistence

In the Eulerian persistence, the current radar-rainfall image is taken to be a forecast for all the lead times. According to Eq. (1), the displacement vector $\alpha$ is equal to zero. The approach is simple and is usually taken as a baseline system in the forecast evaluation studies.

c. NWP model: COSMO2

The COSMO model was developed as a part of a major cooperative research effort between several national weather services in Europe. It is a high-resolution, limited-area, nonhydrostatic numerical weather prediction model with the radar-rainfall observations assimilated using a latent heat nudging scheme. There are many configurations of the model, such as the COSMO-developed Limited Area Ensemble Prediction System (COSMO-LEPS), the COSMO model covering the eastern Atlantic and Europe (COSMO-EU), and the COSMO model covering central and western Europe (COSMO7), and COSMO2. Some of the main differences between the above COSMO versions include their spatial resolution, updating frequency, forecast range, and the spatial extent for which they are available. For example, COSMO2 has a resolution of 2.2 km, runs every 3 h, and has a forecast range of 24 h. For more information, please refer to the COSMO model documentation available online (http://cosmo-model.org).

3. Data catalog

Radar-rainfall data of 20 summer (June–August) rain-fall events between 2005 and 2010 are used to evaluate the forecasts. We limited the selection of the events to the
summer season to reduce the effects of seasonality on the verification results. The instantaneous reflectivity measurements from the MeteoSwiss network of three Doppler radars undergo several steps of data processing and quality control before being merged into reflectivity and rain-rate composites on a rectangular grid of 1-km spatial resolution. Some of the major steps in the data processing and quality control chain are the identification and mitigation of ground clutter, correction for the vertical variability of the radar-reflectivity, polar-Cartesian grid transformation, and the creation of the reflectivity composite (e.g., Joss and Lee 1995; Germann and Joss 2002; Germann et al. 2006a).

The 20 summer events form a total of 854 h of rainfall. Table 1 lists some basic features of the events while Fig. 2 shows the spatial distribution of event-scale accumulation fields. Hereafter, the events will be referred to by the names mentioned in the first column of Table 1. Although the selection of events was limited to the summer months of June, July, and August, the events displayed space–time structures that are quite different from each other (Table 1 and Fig. 2). For example, event 05231 is characterized by quasi-stationary behavior with widespread rainfall for several hours while event 10157 contains several small and intense convective showers lasting for a short time. Two of the selected events (05231 and 07220) resulted in catastrophic flooding (e.g., Beniston 2006; Jaun et al. 2008; Schmutz et al. 2008).

Even after various quality checks, the radar dataset is not free of errors and data artifacts. One such artifact can be clearly noticed in the accumulation fields in the southwest direction (Fig. 2). To account for such artifacts (e.g., ground clutter, blocking by mountains), we performed a probability-based analysis using 854 h of radar-reflectivity composites. For each pixel within the composite field, we estimated the probability that the reflectivity is greater than 10 dBZ. A pixel affected by the blocking would result in a very low probability of detection, whereas a pixel affected by the ground clutter results in a high probability of detection. After trying out several cutoff points, the coverage and ground clutter probability thresholds were chosen as 2% and 75%, respectively. It should be noted that the above brute-force approach is only aimed at identifying residual artifacts in the reflectivity composites after the rigorous physically based quality control process.

Around 28 000 pixels had a probability of detection <2% and therefore were classified as being affected by the radar beam blockage. While most of these pixels lie far from the location of the radar, some of them are related to the blocking in the southwest direction of La Dôle radar in western Switzerland. None of the pixels had a probability of detection >75%. We repeated the analysis for different probability thresholds. The number of pixels was sensitive to the lower threshold (~22 000 for 1% to ~32 000 for 5%) but not to the upper threshold. We decided to use the threshold of 2% and excluded 28 000 pixels from the remainder of our analysis. The radar quality mask obtained using the above probability analysis is shown in Fig. 1.

### Table 1. List showing the beginning and end times, duration, and total accumulation for each event. The events for which COSMO2 data were available (not available) are indicated by a Y (N) in the last column.

<table>
<thead>
<tr>
<th>Event ID</th>
<th>Begin</th>
<th>End</th>
<th>Duration (h)</th>
<th>Accumulation (mm)</th>
<th>COSMO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>05231</td>
<td>1005 UTC 18 Aug 2005</td>
<td>0200 UTC 23 Aug 2005</td>
<td>88.0</td>
<td>297.5</td>
<td>N</td>
</tr>
<tr>
<td>06175</td>
<td>0905 UTC 24 Jun 2006</td>
<td>0200 UTC 26 Jun 2006</td>
<td>41.0</td>
<td>156.8</td>
<td>N</td>
</tr>
<tr>
<td>06185</td>
<td>0905 UTC 4 Jul 2006</td>
<td>1800 UTC 6 Jul 2006</td>
<td>33.0</td>
<td>207.1</td>
<td>N</td>
</tr>
<tr>
<td>06202</td>
<td>0905 UTC 21 Jul 2006</td>
<td>1800 UTC 22 Jul 2006</td>
<td>33.0</td>
<td>104.7</td>
<td>N</td>
</tr>
<tr>
<td>07156</td>
<td>1035 UTC 5 Jun 2007</td>
<td>2330 UTC 5 Jun 2007</td>
<td>13.0</td>
<td>54.9</td>
<td>N</td>
</tr>
<tr>
<td>07171</td>
<td>1235 UTC 20 Jun 2007</td>
<td>1300 UTC 22 Jun 2007</td>
<td>24.5</td>
<td>193.9</td>
<td>N</td>
</tr>
<tr>
<td>07189</td>
<td>0105 UTC 8 Jul 2007</td>
<td>1200 UTC 9 Jul 2007</td>
<td>35.0</td>
<td>139.2</td>
<td>N</td>
</tr>
<tr>
<td>07220</td>
<td>1105 UTC 8 Aug 2007</td>
<td>1630 UTC 9 Aug 2007</td>
<td>53.5</td>
<td>258.2</td>
<td>Y</td>
</tr>
<tr>
<td>07240</td>
<td>0535 UTC 28 Aug 2007</td>
<td>0530 UTC 31 Aug 2007</td>
<td>72.0</td>
<td>263.4</td>
<td>Y</td>
</tr>
<tr>
<td>08181</td>
<td>1135 UTC 29 Jun 2008</td>
<td>2330 UTC 29 Jun 2008</td>
<td>12.0</td>
<td>72.3</td>
<td>Y</td>
</tr>
<tr>
<td>08194</td>
<td>0105 UTC 12 Jul 2008</td>
<td>0700 UTC 14 Jul 2008</td>
<td>54.0</td>
<td>233.0</td>
<td>Y</td>
</tr>
<tr>
<td>08214</td>
<td>0105 UTC 1 Aug 2008</td>
<td>0130 UTC 2 Aug 2008</td>
<td>23.5</td>
<td>111.5</td>
<td>Y</td>
</tr>
<tr>
<td>08227</td>
<td>1305 UTC 14 Aug 2008</td>
<td>1900 UTC 15 Aug 2008</td>
<td>30.0</td>
<td>140.6</td>
<td>Y</td>
</tr>
<tr>
<td>09176</td>
<td>1305 UTC 25 Jun 2009</td>
<td>1800 UTC 27 Jun 2009</td>
<td>29.0</td>
<td>133.3</td>
<td>Y</td>
</tr>
<tr>
<td>09215</td>
<td>0105 UTC 3 Aug 2009</td>
<td>1500 UTC 3 Aug 2009</td>
<td>14.0</td>
<td>46.5</td>
<td>Y</td>
</tr>
<tr>
<td>09219</td>
<td>0105 UTC 7 Aug 2009</td>
<td>1830 UTC 8 Aug 2009</td>
<td>41.5</td>
<td>129.2</td>
<td>Y</td>
</tr>
<tr>
<td>09221</td>
<td>0710 UTC 9 Aug 2009</td>
<td>0800 UTC 10 Aug 2009</td>
<td>24.9</td>
<td>100.5</td>
<td>Y</td>
</tr>
<tr>
<td>10157</td>
<td>0905 UTC 6 Jun 2010</td>
<td>0400 UTC 7 Jun 2010</td>
<td>19.0</td>
<td>86.6</td>
<td>N</td>
</tr>
<tr>
<td>10183</td>
<td>1005 UTC 1 Jul 2010</td>
<td>1800 UTC 4 Jul 2010</td>
<td>56.0</td>
<td>212.8</td>
<td>Y</td>
</tr>
<tr>
<td>10191</td>
<td>1005 UTC 10 Jul 2010</td>
<td>1850 UTC 12 Jul 2010</td>
<td>56.8</td>
<td>244.5</td>
<td>Y</td>
</tr>
</tbody>
</table>
Fig. 2. Spatial distribution of event-scale rainfall accumulations for each of the selected events over the 620 × 620 km² domain broadly centered over Switzerland. The event ID and the duration are indicated in each panel.
4. Event characteristics

In this section we characterize the spatial structure of each of the selected rainfall events. Better characterization of the rainfall events will lead to better understanding of the forecast errors. Our focus is mainly on the spatial dependence and the banded structure of the reflectivity patterns. The spatial dependence and the associated anisotropy can be characterized in the form of two-dimensional correlations (e.g., Kessler 1966; Velasco-Forero et al. 2009) or a variogram (e.g., Miniscloux et al. 2001; Berne et al. 2009). We selected the variogram approach as it is robust to the effects of sampling and the nonnormality of the process.

Let \( Z(x, y) \) be the \( 620 \times 620 \) reflectivity field for which the variogram needs to be estimated. For all interpixel combinations within the quality mask (Fig. 1), we estimated the two-dimensional variogram \( \gamma(h_x, h_y) \) as follows (e.g., Cressie 1993):

\[
\gamma(h_x, h_y) = \frac{1}{n_h} \sum_{i=1}^{n_h} \left[ Z(x_i, y_i) - Z(x_j, y_j) \right]^2\, ,
\]

where \( h_x = x_i - x_j, h_y = y_i - y_j \), and \( n_h \) is the number of pixels separated by distance \( h = \sqrt{h_x^2 + h_y^2} \). We limited the estimation of the variogram to the \( 80 \text{ km} \times 80 \text{ km} \) grid. That is, \( h_x \) and \( h_y \) in the above equation range from -40 to 40. An example rain-rate field at a particular time step within event 07189 and the corresponding 2D variogram are shown in Fig. 3. Although it is sufficient to display the variogram field for just two quadrants (because of symmetry), we show it for all four quadrants for better visualization. The elliptical shape of the variogram field implies the presence of anisotropy, whereas the ellipticity of the variogram is a measure of anisotropy of the process.

The 2D variogram was then fitted with an elliptical Gaussian function of the form

\[
F(x, y) = A_0 + A_1 e^{-U/2},
\]

where \( U \) is the elliptical function defined as

\[
x' = x \cos \theta - y \sin \theta \\
y' = x \sin \theta + y \cos \theta.
\]

In the above equation, \( 2a \) and \( 2b \) are the major and minor axis lengths in the unrotated \( X \) and \( Y \), respectively, and \( \theta \) denotes the clockwise rotation of the ellipse in radians. The parameters were estimated using the Levenberg–Marquardt algorithm (e.g., Press et al. 1992). Figure 3 also shows the fitted elliptical function to the 2D variogram. The variogram and the corresponding parameters were estimated only for those time steps within the event for which the rainy area is at least 4000 km\(^2\). This is done to reduce the effects of small sample size on the estimation of the variogram. Later in this study, we relate the event characteristics to the predictability of forecasts from MAPLE.
5. Forecast evaluation framework

Several verification techniques were proposed in the literature to characterize the forecast performance. However, no single verification technique gives a complete picture of the forecast performance. The verification scores can be broadly classified into

1) those that focus on the ability of forecasts to reproduce different classes of rainfall intensities (e.g., Gilbert skill score),
2) those that utilize the continuous spectrum of available forecast–observation pairs (e.g., mean absolute error),
3) those that consider certain neighborhoods while performing the evaluation (e.g., upscaling, fractions skill score),
4) those that compare the spatial structure of forecasts and observations (e.g., variograms), and
5) those that characterize how well a specific object is reproduced in the forecast (object-oriented evaluation).

We selected a combination of categorical, continuous, and neighborhood verification strategies to evaluate the MAPLE, Eulerian, and COSMO2 forecasts. The evaluation was performed at the finest space and time resolutions, that is, 1 km in space and 5 min in time. This section describes the forecast evaluation framework, followed by a presentation of the results in section 6.

a. Categorical verification

In this study we limit the categorical evaluation to binary rain–no-rain patterns. All the pixels with reflectivity values below a certain threshold \( Z_t \) were considered to be nonrainy and those with \( Z \geq Z_t \) were considered to be rainy. For each lead time, we applied \( Z_t \) to all the observed and forecast fields within an event, and prepared a \( 2 \times 2 \) contingency table (Table 2) consisting of the numbers of hits \((a)\), false alarms \((b)\), misses \((c)\), and correct negatives \((d)\). The above procedure was then repeated for each event, and for three reflectivity thresholds: 10 dBZ \( (R = 0.1 \text{ mm h}^{-1}) \), 25 dBZ \( (R = 1.0 \text{ mm h}^{-1}) \), and 35 dBZ \( (R = 5.0 \text{ mm h}^{-1}) \). The contingency tables were then used to compute the skill scores. It should be noted that the contingency table and the binary evaluation results are independent of the scale used \((\text{mm h}^{-1} \text{ versus dBZ})\) as long as the pixel values and thresholds are converted appropriately using a \( Z-R \) relation.

Some of the desirable properties of categorical scores are equitability, transpose symmetry, and independence from the base rate \([p = (a + c)/n]\) of the observations (e.g., Stephenson 2000; Göber et al. 2004; Jolliffe 2008; Hogan et al. 2009, 2010). An equitable skill score accounts for any improved level of skill in the forecasting system just by chance occurrence (random forecasts). It can be written as (e.g., Hogan et al. 2009)

\[
\frac{X - X_r}{X_p - X_r},
\]

where \( X \) is a function of the contingency table \((a, b, c, d)\), \( X_p \) is the value of \( X \) for a perfect forecast, and \( X_r \) is a function of a random contingency table \((a_r, b_r, c_r, d_r)\); Table 3). The elements of the random contingency table were obtained for each event using equations given by Stephenson (2000). The base rate \( p \) is an indicator of the sample climatology of the observations. A skill score that is highly sensitive to the base rate yields very small values for rare events, making it difficult to compare the performance of forecasts for such extreme cases. Therefore, it is important to select a skill score that is independent of the base rate of the event.

We selected frequency bias, probability of detection, false alarm ratio, Gilbert skill score, odds ratio, and the symmetric extreme dependence score in our categorical evaluation. The mathematical expressions of each verification score and the possible range of these scores are shown in Table 4. Frequency bias (FB) is defined as the ratio of the number of pixels in the forecasts to the number of pixels in the observed fields exceeding a given threshold. Probability of detection (POD), false alarm ratio (FAR), and Gilbert skill score (GSS) have been widely used in the literature. The odds ratio is defined as the ratio of the odds of a correct detection compared to an incorrect detection of precipitation over a certain threshold (e.g., Stephenson 2000; Göber et al. 2004). Following Hogan et al. (2009), we included elements of the random contingency table while estimating the odds ratio (Table 4). This definition of the odds ratio is more

| Table 2. Two-dimensional contingency table. |
|------------------|------------------|
|                 | \( R \geq R_t \) | \( R < R_t \) |
| \( R \geq R_t \) | Hit \((a)\)       | Miss \((c)\) |
| \( R < R_t \)   | False alarm \((b)\) | Correct negative \((d)\) |

| Table 3. Two-dimensional random contingency table. |
|------------------|------------------|
|                 | \( R \geq R_t \) | \( R < R_t \) |
| \( R \geq R_t \) | \( a_r = (a + b)(a + c)/n \) | \( c_r = (a + c)(c + d)/n \) |
| \( R < R_t \)   | \( b_r = (a + b)(b + d)/n \) | \( d_r = (b + d)(c + d)/n \) |
Table 4. Categorical scores.

<table>
<thead>
<tr>
<th>Verification score</th>
<th>Formula</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency bias (FB)</td>
<td>$a + b$</td>
<td>$[0, \infty]$</td>
</tr>
<tr>
<td>Probability of detection (POD)</td>
<td>$\frac{a}{a + c}$</td>
<td>$[0, 1]$</td>
</tr>
<tr>
<td>False alarm ratio (FAR)</td>
<td>$\frac{b}{a + b}$</td>
<td>$[0, 1]$</td>
</tr>
<tr>
<td>Gilbert skill score (GSS)</td>
<td>$\frac{a - a_s}{a + b + c - a_r}$</td>
<td>$[-1/3, 1]$</td>
</tr>
<tr>
<td>Symmetric extreme dependency score (SEDS)</td>
<td>$\frac{\log(a/n)}{\log(a/n)} - 1$</td>
<td>$[0, 1]$</td>
</tr>
<tr>
<td>Log odds ratio (LOR)</td>
<td>$\frac{\log(ad - be)}{\log(ac - bd)}$</td>
<td>$[-\infty, \infty]$</td>
</tr>
</tbody>
</table>

general and consistent with the concept of a generalized skill score given by Eq. (9). If the logarithm of the odds ratio (LOR) is less than 0, it means that the forecasts have no added value compared to the climatological mean. A symmetric extreme dependence score (SEDS) was proposed by Hogan et al. (2009). They demonstrated that the SEDS is a robust verification score particularly for the extreme events.

Although FB, POD and FAR do not meet requirements such as equitability, transpose symmetry, and independence to base rate, we included them in our categorical evaluation to allow for a comparison of the results of the present study with those reported in the literature. On the other hand, the GSS, SEDS, and LOR are equitable (for large samples) and transpose symmetric, whereas SEDS and LOR are independent of the base rate.

b. Continuous verification

Continuous verification implies the use of the entire spectrum of the rainfall intensity scale to characterize the forecast performance. Similar to the categorical scores, there are several continuous verification measures that can be used to assess the quality of a forecast. Some of the most widely used measures are correlation, mean absolute error, and root-mean-square error. We selected mean absolute error (MAE) and correlation coefficient for this study and estimated the correlation coefficient using the following equation (e.g., Germann and Zawadzki 2002):

$$r = \frac{\sum_n (Z_O Z_F)}{\left[\sum_n (Z_O)^2 \sum_n (Z_F)^2\right]^{1/2}}, \quad (10)$$

where $n$ denotes the number of pixels in the quality mask, $r$ is the correlation coefficient, and $Z_O$ and $Z_F$ are the observed and forecast reflectivity fields (in dBZ scale), respectively. It should be noted that the COSMO2 forecasts were converted to dBZ scale using the $Z-R$ relation $Z = 316 R^{1.5}$. For an exponential decay of correlation coefficient with lead time, the lifetime of the forecast can be defined as the times at which the correlation drops to $1/e$ (e.g., Germann and Zawadzki 2002). We followed the same definition and estimated the lifetime of the MAPLE, Eulerian, and COSMO2 forecasts.

c. Neighborhood verification

The verification scores discussed to this point were based on pixel-to-pixel comparisons of high-resolution gridded forecast and observed rainfall fields. With large uncertainties in the observations at fine scales, the pixel-to-pixel comparison at high resolution is restrictive and may not be a better indicator of forecast performance. In recent years, studies have employed verification scores that take pixels from a neighborhood into account while evaluating the gridded forecasts [See Ebert (2008, 2009) for a review of neighborhood verification techniques.]

We employed upscaling and fractions skill score in this study. Upscaling is perhaps the most straightforward and widely used neighborhood verification approach. It involves the gradual coarsening of forecast and observed fields and estimating the skill. Fractions skill score was proposed by Roberts and Lean (2008), and employed in many recent studies (e.g., Mittermaier and Roberts 2010; Weusthoff et al. 2010; Zacharov and Rezacova 2010). For a given neighborhood around a pixel, let $P_O$ and $P_F$ be the fractional observed and forecast rain areas. The fractions skill score (FSS) was then estimated using

$$FBS = \frac{1}{n_n} \sum_{n_s} (P_F - P_O)^2,$$

$$FSS = 1 - \frac{1}{n_n} \frac{\text{FBS}}{\sum_{n_s} P_O^2 + \sum_{n_s} P_F^2}, \quad (11)$$

where $n_n$ indicates the number of neighborhoods in the study region. The fractions skill score ranges from 0 to 1, with 0 indicating no skill and 1 indicating perfect skill. The FSS is lowest at the finest-scale neighborhood and increases with the increase in the size of the neighborhood.

6. Forecast evaluation results

As mentioned in section 3, we selected 854 h of radar-rainfall data corresponding to 20 summer events (2005–10) for this study. For each time step within the event, we estimated the storm motion field using the VET algorithm described in section 2. Figure 4 shows a sample radar field.
and the corresponding storm motion field for different time steps and different events. The reflectivity field was then extrapolated by following the velocity field up to a lead time of 5 h with a time resolution of 5 min. The Eulerian forecasts were generated in the same manner except that the velocity field was set to zero. Thus, we have MAPLE and Eulerian forecasts every 5 min with a spatial resolution of 1 km and lead times varying from 5 min to 5 h for the 20 selected events. The COSMO2 rainfall forecasts were only available starting from July 2007. Therefore, we split our discussion of the results into two parts. In the first part, we compare the verification scores of the MAPLE and Eulerian persistence forecasts using 20 events from 2005 to 2010, and in the second part we compare the verification scores of the MAPLE, COSMO2, and Eulerian persistence forecasts using 12 events from 2007 to 2010.

a. MAPLE and Eulerian forecasts: 20 events

Figure 5 shows a comparison of the spatial distribution of event-scale MAPLE (3-h lead time) and radar accumulations for four events. Such a comparison gives an overall idea about the performance of the forecasting tool. It can be seen from Fig. 5 that the MAPLE model does a good job in forecasting the spatial patterns of the event accumulations even for a lead time of 3 h. Although there are differences between the MAPLE forecasts and observations when looking at specific locations,
the general pattern of the rainfall accumulations from MAPLE is similar to the observations. We then quantified the skill in the MAPLE forecasts at the highest resolution of 5 min in time and 1 km in space using the verification approaches discussed in the previous section. Hereafter, all the verifications except the estimation of MAE were carried out on the reflectivity values in the dBZ scale.

1) **Binary Evaluation**

For each combination of lead time and $Z_t$ threshold, all the observation–forecast reflectivity pairs within an event were arranged in the form of contingency tables (Tables 2 and 3), and the verification scores were estimated using equations given in Table 4. For a certain $Z_t$, the performance of the forecasts for each event can be characterized by plotting the verification score as a function of lead time. We obtained such verification score versus lead time curves for each of the 20 events considered in this study. To gain insights into the storm-to-storm variability in the forecast skill, we estimated the average, minimum, and maximum verification scores of the 20 events for each lead time.

The variations of the FB, POD, and FAR with the lead time are shown in Fig. 6. The solid dark line indicates the average verification score from the MAPLE forecasts while the gray lines indicate the minimum and maximum scores (from 20 events). For $Z_t = 10$ dBZ, the average FB value varied from 0.99 at 5 min to 0.88 at 5 h. The behavior is similar for $Z_t = 25$ dBZ, where the average FB decreased from 0.96 at 5 min to 0.88 at 5-h lead time (Fig. 6). However, for $Z_t$ of 35 dBZ, the average FB increased slightly from 0.96 at 5 min to 1.02 at 5 h. From Fig. 6, it can also be seen that the storm-to-storm variability in the FB values is larger for higher thresholds. The average FB of Eulerian persistence forecasts is very close to 1 for all of the lead times. This is to be expected of the Eulerian persistence forecasts as they are theoretically unbiased.

Figure 6 further shows that the skill of the MAPLE forecasts as characterized by POD and FAR decreased with the increase in the lead time and $Z_t$. As pointed out by Golding (1998), the decrease in forecast skill with the increase in lead time is a reflection of the fact that in a chaotic system such as rainfall, there is an inevitable loss of information with lead time, whereas the decrease in forecast skill with the increase in $Z_t$ is due to smaller localized precipitation areas at higher thresholds. From Fig. 6, it can also be seen that the interstorm variability in the verification scores decreased with the increase in $Z_t$, meaning that the forecast skill at higher thresholds is consistently lower for all the events. Later in this section, we analyze in detail the relation between the event characteristics and the skill in MAPLE forecasts. The skill in the Eulerian persistence forecasts as characterized by POD and FAR is lower than that for the MAPLE forecasts for all lead times and for $Z_t = 10$ and 25 dBZ. The behavior is similar for a higher
threshold of 35 dB but only up to a lead time of 3 h. Beyond a lead time of 3 h, the POD and FAR of the MAPLE and Eulerian forecasts are almost identical (Fig. 6).

GSS, SEDS, and LOR decreased with the increase in lead time and $Z_t$ (Fig. 7). For $Z_t = 10$ and 25 dBZ, all three scores suggest better skill in the MAPLE forecasts compared to that of the Eulerian forecasts for all lead times. For $Z_t = 35$ dBZ, MAPLE outperforms the Eulerian forecasts only for lead times up to 3.5 h. The logarithm of the odds ratio for both the Eulerian and MAPLE forecasts is mostly above zero suggesting that the forecasts indeed have added value compared to the climatological forecasts. Similar to POD and FAR, the interstorm variability in GSS decreased with the increase in $Z_t$ (Fig. 7).

2) CONTINUOUS EVALUATION

We now proceed to the continuous evaluation and estimate mean absolute error (MAE) and the correlation coefficient between the forecast and observed fields. A comparison of MAE in the Eulerian and MAPLE forecasts is shown in Fig. 8. The MAE in the Eulerian forecasts is larger than the MAE in the MAPLE forecasts for all lead times. The 20-event average MAE of the MAPLE forecasts ranges from 0.14 mm h$^{-1}$ for a lead time of 5 min to about 0.66 mm h$^{-1}$ for a lead time of 5 h. Comparing with the mean observed rain rate of

![Fig. 6. FB, POD, and FAR of the MAPLE and Eulerian forecasts as a function of lead time for three rain–no-rain thresholds. For each event, a verification score vs lead-time curve was estimated by comparing all of the 5-min observed and forecasted fields for all of the lead times. The solid dark line represents the 20-event average, and the solid gray lines characterize the event-to-event variability in MAPLE skill.](image-url)
~4.4 mm h\(^{-1}\), the MAE in MAPLE is about 3.2\% at a lead time of 5 min up to 15\% at a lead time of 5 h. Figure 8 also shows the interstorm variability of the MAE in MAPLE forecasts. For instance, the MAE in MAPLE ranges from 8.4\% to 29.8\% of the observed mean rain rate at a lead time of 3 h.

The correlation coefficient between the observed and forecast reflectivity fields was estimated for each event and all lead times using Eq. (10). The average, minimum, and maximum correlation coefficients of 20 events were then obtained for each lead time. The solid dark line in Fig. 9 represents the variation of the 20-event average correlation coefficient with lead time whereas the solid gray lines indicate the variation of the minimum and maximum correlations with lead time. The correlations between the MAPLE forecasts and observations displayed an exponential decay with lead time for all the events. Figure 9 also shows the 20-event average correlation coefficients for the Eulerian persistence forecasts.

It may be recalled (from section 5) that for an exponential decay of correlation with lead time, the lifetime of the forecasts is defined as the lead time at which the correlation drops to 1/e (e.g., Germann and Zawadzki 2002). We fitted the correlations obtained for each event with an exponential function of the form \(\exp[-(t/t_\tau)]\), where \(\tau\) is the lead time and \(t_\tau\) is the lifetime. The average lifetime for the MAPLE forecasts was found to be around 3 h. It is lowest (2 h) for event 10183 and highest (6.5 h) for event 07220. The lifetime for the Eulerian forecasts was found to be around 2.1 h. That is, we gain a buffer of approximately 1 h in lifetime by moving from...
an Eulerian to a Lagrangian (MAPLE) forecast space. It should be noted that the result depends on such factors as the geography of the region (e.g., complex orography versus flat terrain), the meteorological situation (e.g., convective versus stratiform events), and the data characteristics (e.g., space–time resolution).

Based on an analysis of 1424 h of warm season rainfall events over the continental United States, Germann et al. (2006b) reported the average lifetime of the MAPLE forecasts to be 5.1 h. Their study showed that the lifetimes are longer over the midwestern United States with long-lived mesoscale precipitation systems, and shorter over Texas and Florida, where there is pronounced convective activity. They also showed that the average lifetime of Eulerian persistence forecasts is around 2.9 h. That is, they obtained a gain of about 2.2 h by moving from an Eulerian to a semi-Lagrangian extrapolation approach. One reason for the relatively shorter gain of 1 h obtained in the current study compared to that of Germann et al. (2006b) could be the presence of complex orography and its impacts on the meteorological situation. As mentioned in the introduction, extrapolation techniques do not completely account for changes in the precipitation structure due to the orographic enhancement and dissipation.

The correlations shown in Fig. 9 were estimated by pooling together all of the 5-min observed and forecasted fields within an event. Therefore, we have one correlation function for each event and the corresponding lifetime represents the predictability of the event. To gain some insight into the variability of the lifetime within the event, we estimated the correlations for 5 min–5 h forecasts issued at each time step within the event. The lifetime at each time step was then estimated by fitting an exponential function. Figure 10 shows the distribution of lifetimes for each of the 20 selected events in the form of box plots. Events 07156 and 10183 display the least variability in lifetime whereas event 07740 has the highest variability and skewness. From Fig. 10, we can see that the lifetimes within some events are as high as 12 h.

Germann et al. (2006b) showed that the lifetime at each time step within the event depends on the precipitation area of the initial field. To further understand this dependence, we characterized the spatial structure of the initial reflectivity field at each time step using the spatial variogram described in section 4. The spatial variogram was then fitted with an elliptical Gaussian function [Eqs. (7) and (8)]. That is, the spatial variogram of the reflectivity field at each time step was characterized in the form of the major and minor axes, as well as the eccentricity of the elliptical Gaussian function (see Fig. 3 for an example). The correlations between the
observed and MAPLE forecast fields at lead times of 1, 2, and 3 h are plotted against the variogram characteristics of the initial reflectivity field in Fig. 11. The correlations were found to be higher for the reflectivity fields characterized by longer major and minor axes. It can be said that the MAPLE model results in better forecasts when the initial reflectivity fields are smoother in space. No such conclusion can be drawn for the relation between the eccentricity and the skill of the MAPLE forecasts.

3) NEIGHBORHOOD EVALUATION

The verification results discussed so far have been obtained by comparing pixel-to-pixel values at highest resolution. In the previous section, we noted that complex orography is one of the reasons for the shorter lifetimes of the MAPLE forecasts over Switzerland compared to those over the United States. However, the size of the domain and the data resolution [2720 × 2720 km² with 4-km resolution in Germann et al. (2006b) versus 620 × 620 km² with 1-km resolution in the current study] may also play a key role in the characterization of forecast skill. For example, in a recent study Ruzanski and Chandrasekar (2011, manuscript submitted to J. Appl. Meteor. Climatol.) used high-resolution (0.2 km) radar fields with a spatial extent of 140 × 140 km² and reported the lifetimes of the Eulerian and Lagrangian forecasts to be around 15 and 20 min, respectively.

To tease out the effects of resolution on the estimated lifetimes, we repeated the evaluation experiment by gradually upscaling the forecasts and observations to different resolutions. Figure 12 shows the average correlations between the MAPLE forecasts and observations as a function of lead time for resolutions of 4 and 20 km. It can be seen from Fig. 12 that evaluation results are sensitive to the resolution. The average lifetime of the MAPLE forecasts is about 3.5 h for 4-km resolution and about 3.75 h for 20-km resolution. However, the lifetimes are still shorter than those reported for the 4-km resolution U.S. radar data. Therefore, it can be said that complex orography is a major factor for shorter lifetimes observed in the current study.

In addition to the correlations, we characterized the resolution dependence of forecasted rainy areas. However, we did not use upscaling for this purpose as the resulting smoothing severely affects the estimation of binary scores such as POD and GSS, particularly at coarse resolutions and higher thresholds. Instead, we used the fractions skill score described in section 5c. For each square neighborhood centered on pixels in observed and forecasted fields, the fractional observed (PO) and forecasted (PF) rainy areas were estimated by counting the number of pixels exceeding a given Zₜ value. FSS (section 5c) was then estimated using Eq. (11). Figure 13 shows the variation of the average FSS of the MAPLE and Eulerian persistence forecasts with the size of the spatial neighborhood for a fixed lead time of 3 h and different Zₜ thresholds. The FSS was found to be lowest for the domain size of 3 km and increased with the size of the neighborhood. A nonzero FSS value for the smallest neighborhood size of 3 km implies reasonable skill even at that scale. For all of the spatial scales, and for all three values of Zₜ, the average FSS of MAPLE was found to be larger than that of the Eulerian persistence forecasts (Fig. 13). Similar to other verification scores, the FSS also decreased with the increase in the value of Zₜ.

b. MAPLE, Eulerian, and COSMO2 persistence: 12 events

In this section, we compare forecasts from the MAPLE and Eulerian persistence simulations with those from
COSMO2. It may be recalled that COSMO2 forecasts have a spatial resolution of 2.2 km, forecast range of 24 h, and an updating frequency of 3 h. We selected COSMO2 forecasts with a time resolution of 5 min and regridded the fields from 2.2- to 1-km resolution using bilinear interpolation. The initialization to the forecast dissemination time of the COSMO2 model is from about 1 h 30 min to 2 h, which includes the time required to assimilate observations from various sources and solve the governing equations. Therefore, it takes approximately 2 h for the latest COSMO2 run to be available to the user. For each time step within the event, we had to account for this 2-h time lag before selecting the COSMO2 run for a certain lead time. For example, consider the scenario where we are interested in comparing MAPLE and COSMO2 forecasts for a lead time of 3 h from the current time instant of 1600 UTC. Although the last COSMO2 run was initiated at 1500 UTC, the latest forecast available to the user is the run that was initiated at 1200 UTC. Therefore, to compare the performance of COSMO2 and MAPLE

![Fig. 11. Scatterplot of the MAPLE skill (correlation coefficient) against the smoothness of the precipitation field as represented by the major and minor axes of the variogram. The gray scale represents the number of points in each square.](image)
for a lead time of 3 h from 1600 UTC, we selected the 1200 UTC COSMO2 run with 7-h (1200 UTC + 4 h + 3 h) lead time. In short, the evaluation exercise is from a user’s perspective and not from the model developer’s point of view.

For each event we constructed contingency tables for lead times of 5 min to 5 h, and estimated the binary verification scores described in section 5. The 12-event average, minimum, and maximum verification scores were then obtained and compared against the corresponding verification scores from the MAPLE and Eulerian forecasts. Although we estimated all of the skill scores, we present only those of the POD, FAR, and OR in this section. From Fig. 14 it can be seen that the COSMO2 forecasts have lower PODs for shorter lead times. The POD of COSMO2 decreased with the increase in lead time but not as rapidly as the extrapolation forecasts. The 12-event average POD curve of COSMO2 crossed the corresponding average MAPLE POD curve at lead times of 2, 2.2, and 2.4 h for the $Z_t$ values of 10, 25, and 35 dBZ, respectively. From the odds ratio results (Fig. 14, bottom), it can be seen that the skill in the COSMO2 forecasts exceeds the skill in MAPLE at lead times of 3.1, 2.6, and 2.8 h for $Z_t$ values of 10, 25, and 35 dBZ, respectively.

We then estimated the crossover time based on the variation of the correlation and MAE with lead time. Figure 15 shows the comparison of the 12-event average, minimum, and maximum correlation functions obtained from the MAPLE, Eulerian, and COSMO2 forecasts. Comparing the correlations of the MAPLE and COSMO2 forecasts, it can be concluded that the average crossover time is around 2.5 h, with minimum and the maximum values of 2.2 and 3.2 h, respectively (Fig. 15). The 12-event average, minimum, and maximum MAEs between the observed and forecast fields are shown in Fig. 16. The crossover time based on MAE was found to be between 1.7 and 2.7 h with an average value of approximately 2.4 h (Fig. 16).

As stated in the introduction, one of the objectives of this study is to find the crossover lead time beyond which the COSMO2 forecasts have better predictability than the extrapolation forecasts. Based on the binary verification scores presented in Fig. 14, and based on continuous verification skill scores shown in Figs. 15 and 16, we conclude that the crossover lead time is between 2 and 3 h depending on the verification score and rain–no-rain threshold used. Lin et al. (2005) compared the skill...
in hourly accumulations from extrapolation-based and NWP models using U.S. radar-reflectivity composites at 4-km resolution and reported the crossover time to be about 6 h. The shorter crossover time found in the current study can be attributed to the higher space and time resolutions of the forecasts, the better skill of the COSMO2 model in resolving small-scale features, and also the presence of orography, which led to a rapid decay of skill in the extrapolation forecasts with lead time.

7. Orographic effects: Challenges and outlook

The main reason for the rapid decay of skill in the extrapolation techniques is the lack of a mechanism to account for the initiation, growth, and dissipation of storms. As mentioned in the introduction, we aim to incorporate the effects of growth and dissipation into the MAPLE model to obtain improved rainfall nowcasts over the complex orography. We intend to approach the problem in a statistical manner, by training the MAPLE model using a large historical archive of radar-rainfall fields. The key is in arriving at a set of predictors based on an archival dataset, and then using those predictors to incorporate the effect of orography into the output fields from MAPLE. The approach is similar to the analog-based Nowcasting of Orographic Rainfall by Means of Analogues (NORA) model (Panziera et al. 2011) developed at MeteoSwiss and based on the observed relationship between air mass stability and the orographic conditions.

FIG. 14. POD, FAR, and OR of the MAPLE, Eulerian, and COSMO2 forecasts as a function of lead time for three rain–no-rain thresholds.
precipitation in the 8100 km² region to the south of the Alps (Panziera and Germann 2010).

For a larger region such as the one considered in this study, one predictor could be the velocity field from the variational echo-tracking algorithm. That is, any growth or blockage of rainfall patterns by the mountains should lead to a reduction in the speed estimated from the VET algorithm and should also affect the direction of the motion vectors. In the Fig. 16, we show the average velocity field for events 07189 and 10157. These fields were estimated by performing the weighted average of the velocity fields at each time step within an event with the weights estimated as a function of the precipitation intensity. Note that weighting was performed to avoid the effects of interpolated motion vectors on the average velocity field.

The events shown in Fig. 17 represent two extremes in the sense that for event 07189, no effect of orography was
seen in the motion vectors, and for event 10157 the orographic effect is evident from the change in the velocity field above the Alps. The orographic effect can also be seen in Fig. 2, which shows the event-scale accumulations. Therefore, the velocity field can be used as one of the predictors to induce the growth and dissipation of rainfall patterns. Another possible predictor is the time of day. We are investigating the effects of diurnal variability on the performance of MAPLE. The forecast fields from the MAPLE model can therefore be corrected based on the time of day.

8. Summary and conclusions

Very short-term forecasting of precipitation at high spatial and temporal resolutions is critical, particularly for regions dominated by complex orography. To answer the question raised in the title of this paper, we carried out an evaluation of high-resolution forecasts from Eulerian persistence and Lagrangian extrapolation (MAPLE) models using 854 h of summer rainfall events between 2005 and 2010. Rigorous evaluation of 1-km, 5-min forecasts was performed using binary, continuous, and neighborhood verification approaches. The forecasts from the MAPLE model outperformed the Eulerian persistence forecasts for all lead times and for different rain–no-rain thresholds. The lifetime of the MAPLE forecasts as defined by a 1/e decorrelation time was found to be about 3 h on average. The study showed that the skill of the MAPLE forecasts depends on the spatial structure of the initial radar-reflectivity field. Smoother initial radar-reflectivity fields generally lead to better performance. The results showed a relative gain of approximately 1 h in the lifetime of forecasts by adopting a Lagrangian extrapolation approach than a simple Eulerian persistence.

Comparison of MAPLE forecasts with those obtained from the high-resolution NWP COSMO2 model revealed that the extrapolation forecasts have much higher skill for shorter lead times. However, their skill decreased rapidly compared to NWP model forecasts. The crossover lead time beyond which the NWP forecasts have better skill than the Lagrangian extrapolation was found to be around 2.5 h for the study region characterized by the complex Alpine orography. The crossover lead time varied between 2 and 3 h depending on the verification score and the rain–no-rain threshold used. To the best of our knowledge, the crossover times of 2–3 h are the shortest reported in the literature. The shorter crossover time found in this study is due to (but not limited to) the rapid deterioration of the extrapolation forecast skill with lead time, advances in numerical weather prediction and the usage of robust data assimilation schemes, and the assimilation of better quality MeteoSwiss radar-rainfall fields into the COSMO2 model.

Efforts are under way at MeteoSwiss to incorporate the effects of the growth and dissipation of the precipitation patterns into the extrapolation techniques, and to represent the uncertainties in the estimation of storm dynamics (velocity field and time evolution) in the form of ensembles.

Acknowledgments. The authors would like to acknowledge the support of the European Union FP7 project IMPRINTS (226555). We thank our colleagues at MeteoSwiss, Locarno-Monti, and two anonymous reviewers for many helpful suggestions. We also thank Daniel Leuenberger at MeteoSwiss for providing the COSMO2 data.

REFERENCES


