

# SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts

HEINI WERNLI AND MARCUS PAULAT

*Institute for Atmospheric Physics, University of Mainz, Mainz, Germany*

MARTIN HAGEN

*Institut für Physik der Atmosphäre, DLR Oberpfaffenhofen, Germany*

CHRISTOPH FREI

*Federal Office of Meteorology and Climatology (MeteoSwiss), Zürich, Switzerland*

(Manuscript received 12 October 2007, in final form 25 February 2008)

## ABSTRACT

A novel object-based quality measure, which contains three distinct components that consider aspects of the structure ( $S$ ), amplitude ( $A$ ), and location ( $L$ ) of the precipitation field in a prespecified domain (e.g., a river catchment) is introduced for the verification of quantitative precipitation forecasts (QPF). This quality measure is referred to as SAL. The amplitude component  $A$  measures the relative deviation of the domain-averaged QPF from observations. Positive values of  $A$  indicate an overestimation of total precipitation; negative values indicate an underestimation. For the components  $S$  and  $L$ , coherent precipitation objects are separately identified in the forecast and observations; however, no matching is performed of the objects in the two datasets. The location component  $L$  combines information about the displacement of the predicted (compared to the observed) precipitation field's center of mass and about the error in the weighted-average distance of the precipitation objects from the total field's center of mass. The structure component  $S$  is constructed in such a way that positive values occur if precipitation objects are too large and/or too flat, and negative values if the objects are too small and/or too peaked. Perfect QPFs are characterized by zero values for all components of SAL. Examples with both synthetic precipitation fields and real data are shown to illustrate the concept and characteristics of SAL. SAL is applied to 4 yr of daily accumulated QPFs from a global and finer-scale regional model for a German river catchment, and the SAL diagram is introduced as a compact means of visualizing the results. SAL reveals meaningful information about the systematic differences in the performance of the two models. While the median of the  $S$  component is close to zero for the regional model, it is strongly positive for the coarser-scale global model. Consideration is given to the strengths and limitations of the novel quality measure and to possible future applications, in particular, for the verification of QPFs from convection-resolving weather prediction models on short time scales.

## 1. Introduction

Verification of numerical forecasts is an essential part of the numerical weather prediction (NWP) enterprise. On the one hand, it helps identify model shortcomings and systematic errors; on the other hand, it is

key for a quantitative assessment of the improvement with time of current forecasting systems and of their predictability limits. Quality measures like the root-mean-square (RMS) difference or anomaly correlations are simple in terms of implementation and are therefore routinely used to monitor and compare general forecast quality at operational prediction centers (e.g., Simmons and Hollingsworth 2002). The quality of quantitative precipitation forecasts (QPF) is typically measured in terms of categorical verification scores (Jolliffe and Stephenson 2003), a process that requires

---

*Corresponding author address:* Heini Wernli, Institute for Atmospheric Physics, University of Mainz, Becherweg 21, D-55099 Mainz, Germany.  
E-mail: wernli@uni-mainz.de

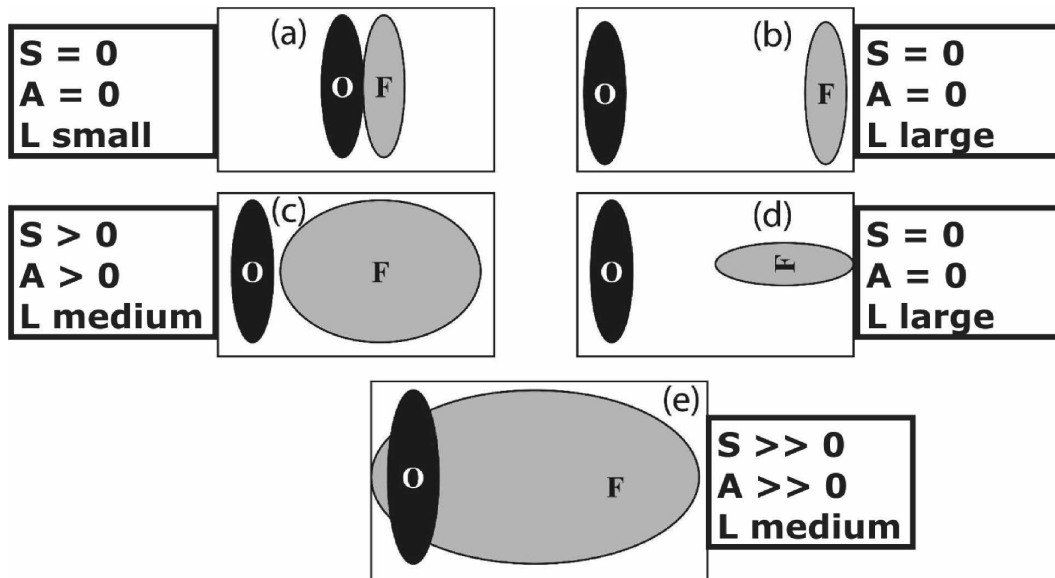


FIG. 1. A schematic example of various forecast and observation combinations, modified from Davis et al. (2006a). For the qualitative application of SAL, it was assumed that precipitation rates are uniform and the same in all objects.

the specification of a precipitation threshold. Examples for this category of QPF verification studies can be found, for instance, in Damrath et al. (2000) for Germany and Ebert et al. (2003) for the United States, Australia, and Germany.

Gridpoint-based error measures are appropriate for the verification of fields dominated by synoptic-scale structures (e.g., the 500-hPa geopotential height field), but for parameters like precipitation, which are characterized by complex structures on scales of less than 100 km, these measures are regarded as problematic and several new approaches have been suggested and developed during the last decade (e.g., Ebert and McBride 2000; Casati et al. 2004; Davis et al. 2006a, and references therein). The classical example to illustrate the limitations of gridpoint-based error measures is the “double penalty problem”: a prediction of a precipitation structure that is correct in terms of amplitude, size, and timing but (maybe only slightly) incorrect concerning position is very poorly rated by categorical error scores and the RMSE. In such a situation, the hit rate of the forecast with the misplaced precipitation structure is as bad as that of a forecast that totally missed the event, and the RMSE is even worse. Also, hit rate and RMSE are equally bad for forecasts that misplaced the event, independent of the degree of the misplacement (see Figs. 1a,b, modified from Davis et al. 2006a), and therefore the verification result does not pinpoint the nature of the error (i.e., the displacement). These issues become even more important with the advent of very-

high-resolution numerical models (with horizontal grid spacings of 1–4 km), which produce precipitation fields that are comparable to radar information in terms of complexity and variety of structures.

The novel approaches of QPF verification try to avoid the double penalty problem and aim to provide useful information about the characteristics and scales of the identified prediction error. They can be categorized into “fuzzy” scores, techniques that focus on spatial scales, and object-based approaches. Different fuzzy scores have been proposed (e.g., Theis et al. 2005; Roberts and Lean 2008) that consider neighboring grid points when comparing simulated and observed fields to account for spatial and temporal uncertainty in the forecast. In the second category, the approach of Casati et al. (2004) using a two-dimensional wavelet decomposition yields useful skill information on different spatial scales. A typical result is that the loss of forecast skill is due to relatively intense events on scales smaller than 40 km. A pioneering study for the object-based category is the one by Ebert and McBride (2000), who decomposed the total mean squared error into components associated with the location, rain volume, and pattern of identified precipitation objects (referred to as “contiguous rain areas”). For the identification of such objects in daily accumulated precipitation fields in Australia, a fixed threshold of  $5 \text{ mm day}^{-1}$  has been used. It was found that the volume error is typically smallest, except for intense events where underestimation of rainfall amounts becomes an issue. Another ob-

ject-based technique has been introduced by Davis et al. (2006a), who used a convolution, smoothing, and thresholding procedure to define meaningful objects. A matching algorithm served to find object pairs in the forecast and observations. For hourly forecasts produced by the Weather Research and Forecasting (WRF) model with a horizontal resolution of 22 km over the United States, one of the interesting results was that the model overestimated the size of the objects, in particular during the later afternoon. Also, it turned out that object matching was increasingly difficult for smaller-scale objects. In a companion study (Davis et al. 2006b), the technique was applied to convection-resolving WRF simulations, again revealing valuable information about the model's QPF performance that could not be obtained with standard verification approaches. An alternative object-based approach has been proposed by Marzban and Sandgathe (2006), based on a cluster analysis technique. Finally, the study by Keil and Craig (2007) is mentioned, who focused on the forecasts' displacement error, which has been calculated with a pyramid matching algorithm, without specifying individual objects.

It is important to note that the current efforts to define alternative error measures are not only motivated by practical and technical issues (i.e., by the fact that gridpoint-based error measures do not provide enough useful information, and that they suffer from the double penalty problem) but they are also rooted in our current understanding of atmospheric predictability. Theoretical and model studies on error propagation (e.g., Zhang et al. 2002, 2003, 2006; Walser et al. 2004; Walser and Schär 2004; Hohenegger et al. 2006) indicate that the predictability limit falls off rapidly toward small scales (1–100 km) mainly due to upscale error propagation associated with individual convective cells. However, several of these studies also emphasize that predictability of QPFs strongly depends on the weather situation and the underlying topography. The presence of convection alone does not necessarily limit predictability, at least in mountainous regions (Walser and Schär 2004), and strongly organized convective systems tend to be characterized by increased predictability (Fritsch and Carbone 2004).

In this study, a novel three-dimensional quality measure is proposed, which separately considers aspects of the structure (*S*), amplitude (*A*), and location (*L*) of a QPF in a certain region of interest (e.g., a major river catchment). This quality measure, referred to as SAL, aims to address the following issues:

1) it measures quantitatively three distinct aspects of the quality of an individual precipitation forecast in

a previously specified area, integrated over time periods ranging from 1 to 24 h;

- 2) it takes into account the “structure” of the precipitation event (e.g., scattered convective cells, convective complex, frontal rain system), which is regarded as a direct fingerprint of the physical nature of the event;
- 3) it does not require a one-to-one matching between the identified objects in the observed and simulated precipitation fields; and
- 4) it is close to a subjective visual judgment of the accuracy of a regional QPF.

To accomplish these tasks, simple measures are specified to characterize the forecast quality in terms of structure, amplitude, and location. For the structure and location components, it will be necessary to identify coherent objects in the observed and predicted precipitation fields. The definition of the three components and some technical details are given in the next section. In section 3, idealized examples are presented to illustrate the functioning of SAL. A first application of SAL to daily accumulated precipitation forecasts for a German river catchment is presented in section 4 for a global and limited-area NWP model, respectively.

## 2. Definition of the three components of SAL

Consider a domain  $\mathcal{D}$  (e.g., a catchment area) represented by a set of  $N$  grid points in both the observational and model datasets. The precipitation field is denoted as  $R$ , and where a distinction between observed and simulated precipitation is necessary, the symbols  $R_{\text{obs}}$  and  $R_{\text{mod}}$  are used (see also Table 1 for an overview on the notation). The order in which the components of SAL are described is guided by their degree of complexity and goes from *A* to *L* and finally to *S*. But first, the issue of the identification of objects is briefly discussed.

### a. The identification of objects

The computation of the location and structure components (as defined later) requires first the identification of individual precipitation objects within the considered domain, separately for the observed and forecast precipitation fields. Several possibilities exist to perform this task, for instance, the method introduced by Davis et al. (2006a). Here we use a simple (and subjective) approach, where a threshold value

$$R^* = fR^{\text{max}} \quad (1)$$

is specified to identify coherent objects enclosed by the threshold contour.  $R^{\text{max}}$  denotes the maximum

TABLE 1. Notation used in this study.

$\mathcal{D}$	Considered domain for verification (set of grid points in domain)
$N$	Number of grid points in domain
$d$	Largest distance of two grid points in domain
$R$	Precipitation field
$R_{ij}$	Precipitation value at grid point $(i, j)$
$\mathbf{x}$	Center of mass of precipitation field in domain
$R^{\max}$	Maximum precipitation value in domain
$R^*$	Threshold value to identify objects
$\mathcal{R}_n$	Precipitation object with index $n$ (set of grid points that belong to object)
$M$	Number of precipitation objects in domain
$R_n^{\max}$	Maximum precipitation value in object $n$
$R_n$	Area-integrated precipitation in object $n$
$\mathbf{x}_n$	Center of mass of precipitation object $n$
$V_n$	Scaled precipitation volume of object $n$
$V$	Weighted average of the scaled precipitation volumes of all objects in the domain
$r$	Weighted averaged distance between individual objects and $\mathbf{x}$

value of precipitation that occurs within the domain  $\mathcal{D}$ . Grid points belonging to an object are selected using an algorithm developed previously for the identification of coherent potential vorticity features (Wernli and Sprenger 2007). Starting from a grid point that corresponds to a local precipitation maximum exceeding the threshold  $R^*$ , neighboring grid points are included in the object as long as the grid point values  $R_{ij}$  are larger than  $R^*$ . The objects are denoted as  $\mathcal{R}_n, n = 1, \dots, M$ , where  $M$  corresponds to the number of objects in  $\mathcal{D}$ .

The choice of the factor  $f$  in Eq. (1) is not based on objective criteria. Our choice used throughout this study ( $f = 1/15$ ) was motivated by the fact that for most considered cases (like the examples shown in Fig. 6), this contour separates features of the precipitation field that correspond reasonably well to distinct objects that can be identified by eye. When discussing the results of SAL in section 4, consideration will be given to their sensitivity to the choice of the threshold factor.

*b. The amplitude component A*

The amplitude component of SAL corresponds to the normalized difference of the domain-averaged precipitation values:

$$A = \frac{D(R_{\text{mod}}) - D(R_{\text{obs}})}{0.5[D(R_{\text{mod}}) + D(R_{\text{obs}})]}. \tag{2}$$

Here,  $D(R)$  denotes the domain average of  $R$ :

$$D(R) = \frac{1}{N} \sum_{(i,j) \in \mathcal{D}} R_{ij}, \tag{3}$$

where  $R_{ij}$  are the gridpoint values. This provides a simple measure of the quantitative accuracy of the total

amount of precipitation in a specified region  $\mathcal{D}$ , ignoring the field's subregional structure. The values of  $A$  are within  $[-2 \dots +2]$  and 0 denotes perfect forecasts in terms of amplitude. The value of  $A = +1$  indicates that the model overestimates the domain-averaged precipitation by a factor of 3; a value of  $A = -1$  goes along with an underestimation by a factor of 3. Overestimations by factors of 1.5 and 2 lead to values of  $A = 0.4$  and 0.67, respectively.

*c. The location component L*

The location component of SAL consists of two parts:  $L = L_1 + L_2$ . The first one measures the normalized distance between the centers of mass of the modeled and observed precipitation fields,

$$L_1 = \frac{|\mathbf{x}(R_{\text{mod}}) - \mathbf{x}(R_{\text{obs}})|}{d}, \tag{4}$$

where  $d$  is the largest distance between two boundary points of the considered domain  $\mathcal{D}$  and  $\mathbf{x}(R)$  denotes the center of mass of the precipitation field  $R$  within  $\mathcal{D}$ . According to Eq. (4), the values of  $L_1$  are in the range  $[0 \dots 1]$ . The term  $L_1$  gives a first-order indication of the accuracy of the precipitation distribution within the domain. In case of  $L_1 = 0$ , the centers of mass of the predicted and observed precipitation fields are identical. However, many different precipitation fields can have the same center of mass, and therefore  $L_1 = 0$  does not necessarily indicate a perfect forecast. For instance, a forecast with two precipitation events on opposite sides in the considered domain can have the same center of mass as an observed precipitation field with one event located in between the two predicted events (see also discussion in section 3a).

The second part,  $L_2$ , aims to distinguish such situations and considers the averaged distance between the center of mass of the total precipitation fields and individual precipitation objects. After identifying the objects separately in the observations and the forecast (as outlined in section 2a), the integrated amount of precipitation is calculated for every object as

$$R_n = \sum_{(i,j) \in \mathcal{R}_n} R_{ij}.$$

The weighted averaged distance between the centers of mass of the individual objects,  $\mathbf{x}_n$ , and the center of mass of the total precipitation field,  $\mathbf{x}$ , is then given by

$$r = \frac{\sum_{n=1}^M R_n |\mathbf{x} - \mathbf{x}_n|}{\sum_{n=1}^M R_n}. \tag{5}$$

The maximum value of  $r$  is  $d/2$  (i.e., half the maximum distance between two grid points in the domain). In the case of a single object in the domain, Eq. (5) yields  $r = 0$ . As an aside, it is noted that the denominator in Eq. (5) is not equal to the sum involved in the computation of  $D(R)$  [see Eq. (3)], because the latter includes all grid points whereas  $\sum_{n=1}^M R_n$  extends only over grid points with  $R_{ij} \geq R^*$ . Now,  $L_2$  can be calculated as the difference of  $r$  calculated for the observed and forecasted precipitation fields:

$$L_2 = 2 \left[ \frac{|r(R_{\text{mod}}) - r(R_{\text{obs}})|}{d} \right]. \quad (6)$$

This quantity can only differ from zero if at least one of the datasets contains more than one object in the considered domain. The factor of 2 is used to scale  $L_2$  to the range  $[0 \dots 1]$  (i.e., the same range as for  $L_1$ ). Hence, the total location component  $L$  can reach values between 0 and 2, and the value of 0 can be obtained only for a forecast, where both the center of mass as well as the averaged distance between the objects and the center of mass agree with the observations. As a caveat, it is mentioned that despite the consideration of  $L_2$ , different situations can still yield the same value of  $L_1 + L_2$ . In particular, the definition of  $L$  is not sensitive to rotation around the center of mass.

#### d. The structure component $S$

Finally, for the structure component  $S$ , the basic idea is to compare the volume of the normalized precipitation objects. As will be shown in several examples, such a measure captures information about the size and shape of precipitation objects. Technically, for every object a “scaled volume”  $V_n$  is calculated as

$$V_n = \sum_{(i,j) \in \mathcal{R}_n} R_{ij}/R_n^{\text{max}} = R_n/R_n^{\text{max}}, \quad (7)$$

where  $R_n^{\text{max}}$  denotes the maximum precipitation value within the object (i.e.,  $R_n^{\text{max}} \leq R^{\text{max}}$ ). The scaling with  $R_n^{\text{max}}$  is necessary to make  $S$  distinct from the amplitude component  $A$  (see examples in next section). The scaled volume  $V_n$  is calculated separately for all objects in the observational and forecast datasets. Then, the weighted mean of all objects’ scaled precipitation volume, referred to as  $V$ , is determined for both datasets. As in Eq. (5), the weights are proportional to the objects’ integrated amount of precipitation  $R_n$ :

$$V(R) = \frac{\sum_{n=1}^M R_n V_n}{\sum_{n=1}^M R_n}. \quad (8)$$

Note that  $V(R)$  is proportional to the second moment of the precipitation field [ $V(R) \propto \sum R_n^2$ ], whereas  $D(R)$  (used for the computation of the  $A$  component) is proportional to the first moment. The component  $S$  is then defined as the normalized difference in  $V$ , analogous to the  $A$  component [cf. Eq. (2)]:

$$S = \frac{V(R_{\text{mod}}) - V(R_{\text{obs}})}{0.5[V(R_{\text{mod}}) + V(R_{\text{obs}})]}. \quad (9)$$

Here,  $S$  becomes large if the model predicts, for instance, widespread precipitation in a situation of small convective events. The possibility to identify these kinds of errors is one of the key characteristics of SAL. Negative values of  $S$  occur for too small precipitation objects, too peaked objects, or a combination of these factors (see examples in sections 3 and 4).

### 3. Idealized examples

To illustrate the characteristics of the precipitation field captured by the SAL components, it is useful to apply their definitions to synthetic precipitation objects with highly idealized, simple shapes. First, a few examples are considered in a qualitative way, and then SAL is applied quantitatively to a set of synthetic fields.

#### a. Qualitative considerations

For simplicity, it is assumed that the observations contain only one object in the considered domain, with a right circular conelike shape (Fig. 2). The left panel shows a contour plot of the object with a maximum value of  $R_{\text{obs}}^{\text{max}}$  and a threshold  $R^*$  defining the border of the object. The center panel provides a section across the center of the object, and the right panel shows the same cross section, after applying the scaling with  $R_{\text{obs}}^{\text{max}}$ . For the calculation of  $V_n$  [Eq. (7)], only the grid points of the circular cone where  $R_{\text{obs}} > R^*$  are considered (see gray shaded area).

SAL is now qualitatively determined for different forecast examples (Figs. 3, 4), which are also characterized by circularly symmetric objects, however differing in amplitude, size, shape, or number. For single-object situations (examples 1–3), the calculation and interpretation of the  $L$  component is straightforward and the discussion therefore focuses on  $A$  and  $S$ .

#### 1) EXAMPLE 1

For the first example (Fig. 3a), the forecast object has the same base area but a reduced amplitude compared to the observed object (Fig. 2; i.e.,  $R_{\text{mod}}^{\text{max}} < R_{\text{obs}}^{\text{max}}$ ). The object identification threshold  $R^*$  does not play a role for the calculation of  $A$ , and therefore  $A$  simply de-

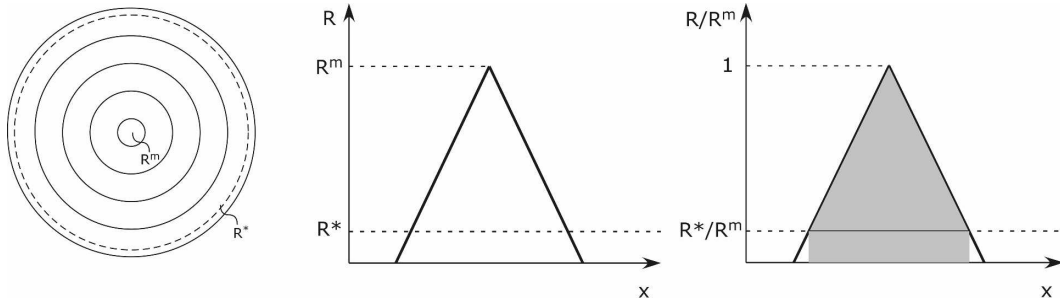


FIG. 2. Idealized precipitation object with the shape of a right circular cone (assumed to represent the observations). (left) Contour plot of the object with a maximum value of  $R^{max}$  (denoted briefly as  $R^m$  in the figure) and a threshold  $R^*$  defining the border of the object. (center) Section across the center of the object. (right): Same cross section after applying the scaling with  $R^{max}$ . The volume  $V$  is marked by gray shading.

depends on the ratio of the maximum values. For the situation shown in Fig. 3a, area-integrated precipitation is underestimated by the model, yielding a negative value for  $A$ . The independent scaling with the maximum value in both datasets leads to two identical ob-

jects (cf. right-hand sides of Figs. 2, 3a), and therefore the structure component  $S$  becomes zero. The interpretation is, according to SAL, that the simulated precipitation field has the correct structure ( $S = 0$ ) while underestimating the total amount of precipitation ( $A < 0$ ).

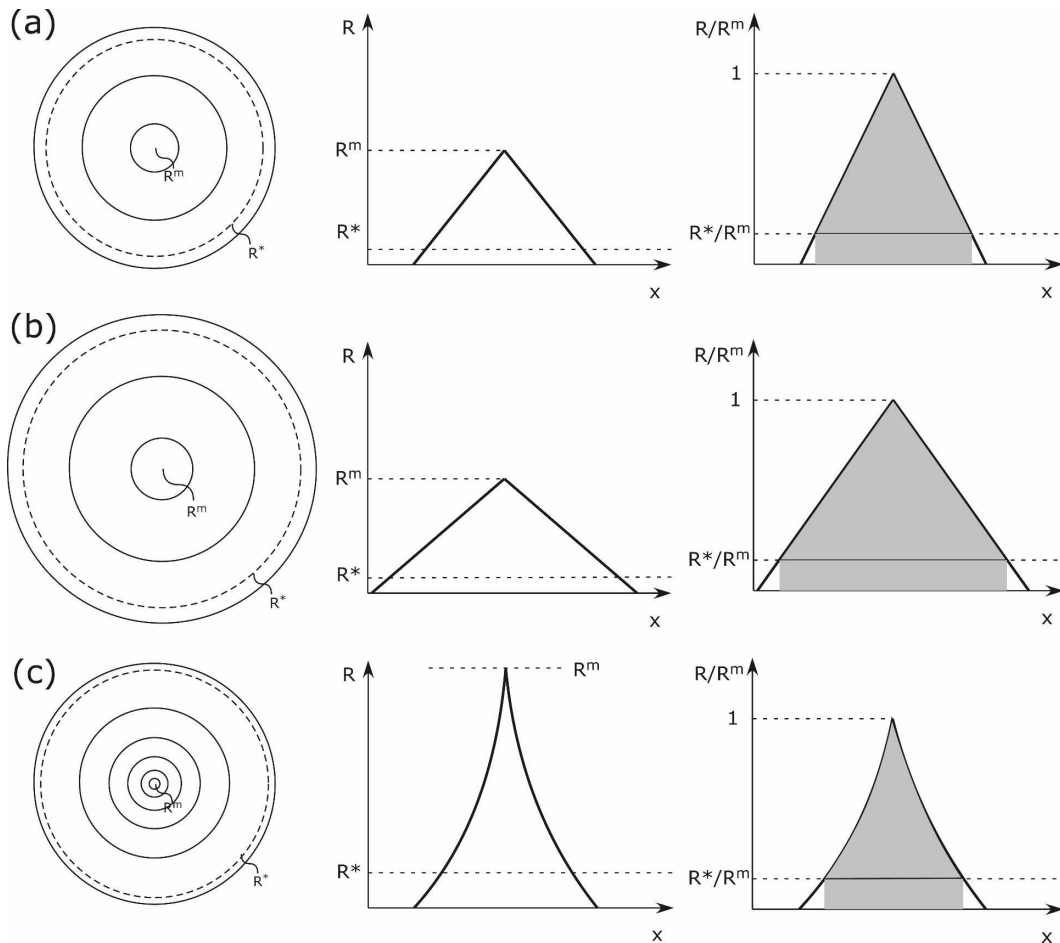


FIG. 3. Same as Fig. 2, but now for three forecast objects: (a) a right circular cone with reduced amplitude (compared to Fig. 2); (b) a right circular cone with reduced amplitude and larger base area (but with the same total precipitation amount as the object in Fig. 2); and (c) a peaked circular cone with the same base area as in Fig. 2.

Note that application of the so-called contiguous rain areas (CRA) technique introduced by Ebert and McBride (2000) would lead to a different result: for the considered example, their error decomposition yields both a volume and a pattern error. Also, the volume error would be positive and not point to the underestimation of the precipitation amplitude in the forecast.

### 2) EXAMPLE 2

We now consider a case (Fig. 3b) in which the errors in the amplitude and base area of the simulated circular conelike object compensate for each other, such that the domain-integrated precipitation value is the same as in the observations (Fig. 2). Consequently, there is no amplitude error ( $A = 0$ ). However, because the base area of the precipitation object is overestimated by the model, the scaling in the calculation of  $S$  [Eq. (7)] leads to a larger scaled precipitation volume in the simulation and to a positive value for  $S$ . In this case, the forecast has no amplitude error but rather a positive structure error due to the too large base area of the object. Again, as for the first example, the CRA error decomposition would lead to both a volume and a pattern error.

From these two examples it becomes obvious that  $A$  and  $S$  are distinct components of SAL—and that they differ significantly from the volume and pattern components of the Ebert and McBride (2000) decomposition of the mean squared error. The scaling involved in the calculation of  $V$  [Eq. (7)] is essential to allow for  $S = 0$  in the presence of an amplitude error (first example) and for identifying a structure error also in case of a correct total precipitation amount (second example). Also note that for these examples, a simpler definition of  $S$  that considers only the objects' base area would lead to the same results as the more complex definition applied here [Eq. (9)]. The next example illustrates the additional distinction that is possible when considering the scaled volume instead of the base area for the calculation of  $S$ .

### 3) EXAMPLE 3

Figure 3c shows a circular object that has the same base area but is more peaked than the right circular cone (Fig. 2). To focus on  $S$ , we can assume that the amplitudes of the two objects are such that they yield the same domain-averaged precipitation values. However, scaling with  $R^{\max}$  leads to a smaller value of  $V$  for the peaked object and therefore to a negative value of  $S$ . Similarly, a flat object with a concave shape would lead to a positive value of  $S$ , if compared with the right circular cone (Fig. 2). This example shows that SAL [with the  $S$  component as defined in Eq. (9)] is able to

distinguish between peaked and flat objects, even if they provide the same total amount of precipitation. The usefulness of this distinction stems from the assumption that widespread stratiform precipitation typically leads to flat objects, whereas in convective situations, objects tend to be much more peaked. It is in this sense that SAL is sensitive to the physical nature of the precipitation event.

Now we consider a few examples in which the simulated precipitation field contains more than one object in the considered domain. For simplicity, we still assume that the observed precipitation is given by the single object shown in Fig. 2.

### 4) EXAMPLE 4

If the simulated field has two objects like the one shown in Fig. 2, then the total amount of precipitation is overestimated by a factor of 2, leading to a value of  $A = 2/3$  [Eq. (2)]. The component  $S$  is zero, because both objects have the correct scaled volume  $V_n$  (recall that for the calculation of  $S$ , the averaged value of all  $V_n$  is considered). The component  $L$  depends on the location of the two simulated objects relative to the observed one. This is discussed in more detail in the next example.

### 5) EXAMPLE 5

If the simulated field has two objects, like the one shown in Fig. 3a, that are right circular cones with half the amplitude compared to the single observed object, then both  $A$  and  $S$  are zero. In the special situation in which the two objects are displaced by the same distance relative to the observed object but exactly in the opposite direction, then the two centers of mass are identical and the component  $L_1$  is zero. It is for this reason that we introduced the second component  $L_2$ , which is positive in this situation and avoids a nonperfect forecast yielding zero values for all components of SAL. Clearly, if the two objects are located in a different way relative to the observed object, then  $L_1$  is also positive leading to a larger location error  $L$ . These considerations are equally valid for example 4.

### 6) EXAMPLE 6

As a last example, consider the situation in which the simulated field has a large object (as shown in Fig. 3b) and a peaked object with a (much) smaller base area (as shown in Fig. 3c). The component  $A$  is most likely positive in this case, unless both objects have a much smaller amplitude than the observed one. As discussed above,  $V_1(R_{\text{mod}})$  (the scaled volume of the large object) is larger, and  $V_2(R_{\text{mod}})$  (the scaled volume of the small

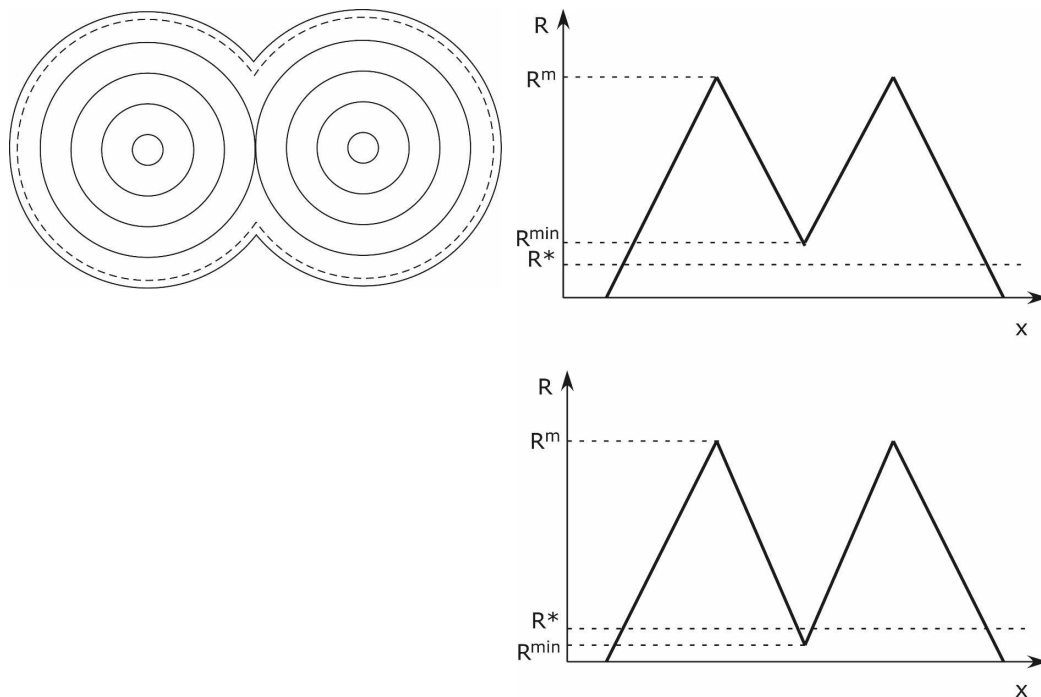


FIG. 4. Example of a precipitation structure with two local maxima to illustrate the camel effect. In the top right situation, one object is identified, whereas two objects are found in the bottom right situation. The two situations depicted on the right differ only in terms of  $R^{\min}$ , the minimum precipitation value along a straight line that connects the two local maxima.

object) is smaller than the scaled volume of the observed object (Fig. 2). Because according to Eq. (8), the resulting  $V(R_{\text{mod}})$  depends on the objects' total precipitation,  $S$  can be either positive or negative. If the peaked object has a small base area and/or amplitude, then the large object dominates the calculation of  $V(R_{\text{mod}})$  and  $S$  will be positive. In contrast, if the large object has a much smaller amplitude than the peaked one, then the latter might dominate (in the sense of having a larger weight  $R_n$ ) and  $S$  turns out to be negative. This example shows that due to the weighting of the objects' scaled volumes  $V_n$  with their contribution  $R_n$  to the total precipitation, the structure component  $S$  yields information primarily about the most relevant objects.

Before we turn to a quantitative application of SAL, an important caveat associated with the choice of threshold  $R^*$  used for the identification of objects should be discussed. In certain situations in which the precipitation field in a given domain contains several local maxima, the identification of objects can be ambiguous, in the sense that a small change of the threshold can lead to a different number and size of objects, and therefore to different values of  $S$  and  $L_2$  (note that  $A$  and  $L_1$  are independent of the object identification). We refer to this effect as the “camel effect” because it

can be illustrated in a simple way with a double-hump precipitation structure (see Fig. 4). Depending on the minimum amplitude  $R^{\min}$  along a line connecting the two maxima, the structure will be identified as a single object ( $R^{\min} > R^*$ ) or as two objects ( $R^{\min} < R^*$ ). This has a large effect on  $S$ : assuming the two humps to be equal in size and amplitude, then  $V$  is larger by about a factor of 2 for the single object. This means that in a situation in which both observations and simulation yield a camel-like object, a relatively large (positive or negative)  $S$  value can occur if the minimum  $R^{\min}$  is (slightly) above the threshold in one of the two fields and (slightly) below in the other. For such precipitation fields, a slight change of the threshold can significantly influence the values of SAL—which is not a desired property of object-oriented error measures. However, such situations are relatively rare and do not influence the results of a *climatological* evaluation of precipitation forecasts with SAL, as further discussed in section 4.

It is also possible to use SAL to reconsider the schematic examples of observed and forecasted precipitation objects discussed by Davis et al. (2006a; see Fig. 1). Assuming uniform precipitation rates, forecasts shown in Figs. 1a,b,d yield no amplitude and no structure error ( $A = S = 0$ ). However, they differ in terms of  $L$ , with



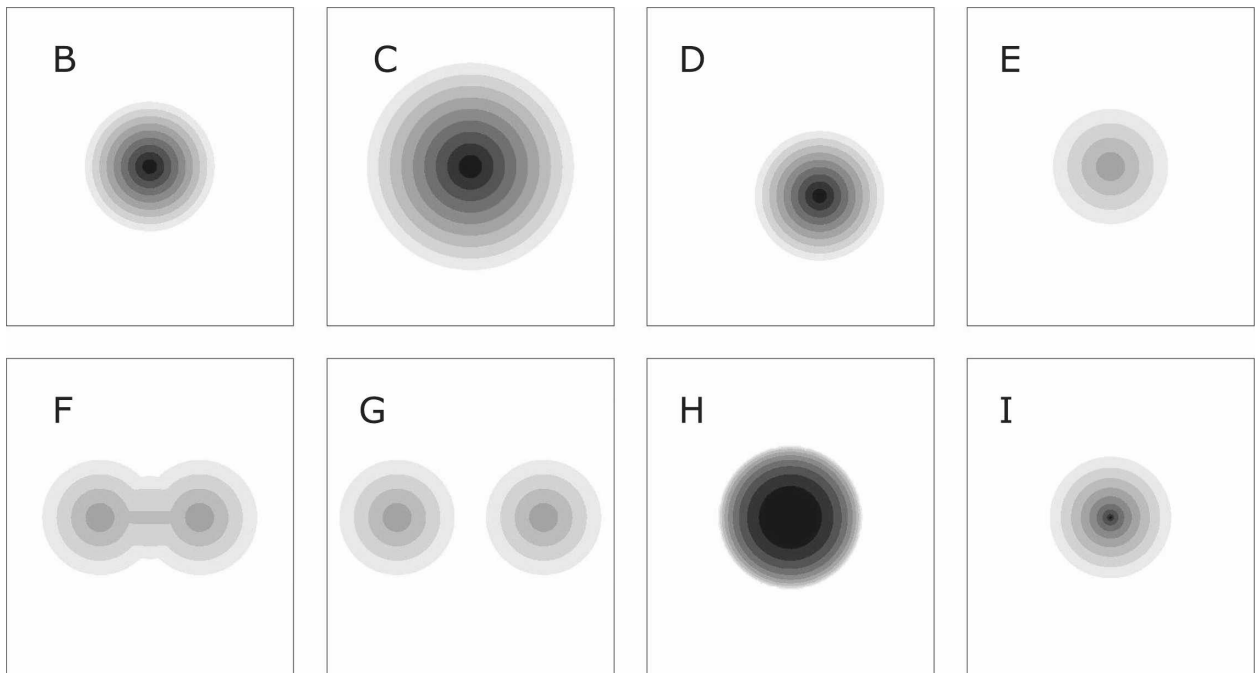


FIG. 5. Contour plots of idealized precipitation objects used for a quantitative evaluation of SAL. The objects are referred to as B, C, D, etc. (from top left to bottom right). The scale is arbitrary, with dark gray denoting more intense precipitation. Compared to the right circular cone (object B), the objects differ as follows: C has a larger base area; D is shifted; E has a reduced amplitude; F and G consist of two right circular cones, each with the same amplitude as E, which overlap in the case of F and don't overlap in the case of G; H is a flat and I a peaked cone.

the smallest location error associated with Fig. 1a. Because SAL does not consider the orientation of objects, it is not able to distinguish between predictions shown in Figs. 1b,d. The too-large precipitation objects in Figs. 1c,e lead to positive values of  $S$  and  $A$ . Note that the example shown in Fig. 1e, which scores best in terms of the hit rate, is regarded as a very poor prediction in terms of SAL, whereas the example in Fig. 1a is regarded as the best.

#### b. A quantitative evaluation

Figure 5 shows eight idealized precipitation fields on a quadratic grid with  $99 \times 99$  grid points. They are labeled as fields B to I (the label A has been omitted to avoid confusion with the amplitude component  $A$ ). Here, B–E are single right circular cones (cf. Figure 2), and when compared to B, C has a larger base area; in D, the object is shifted toward the lower right corner, and E has a reduced amplitude (by a factor of 2). The F and G are precipitation fields with two local maxima (of equal amplitude) that are farther apart in G compared to F. The object in H is convex (or flat) with a large plateau, whereas the object in I is peaked (cf. Figure 3c). A threshold factor of  $f = 1/15$  is used [cf. Eq. (1)] and with this threshold two objects are identified in G and only one in all other situations (including F).

SAL has been applied quantitatively to all possible pairs of these fields, and the results are summarized in Table 2. The entries in the row B and column C, for instance, indicate the SAL values in the situation in which B represents the forecast and C the observations. All diagonal elements are zero, which shows that all components of SAL are zero if the observed and predicted fields are identical. No values are given to the left of the diagonal because the table obviously is anti-symmetric for  $S$  and  $A$ , but symmetric for  $L$ .

Results from the quantitative evaluation (Table 2) agree with the qualitative considerations in the previous subsection. First, situations are discussed where only one of the three components are non zero. BE only yields an amplitude error. BD and BG only yield a location error, however for different reasons: for BD, the component  $L_1$  is positive (D is shifted relative to B), whereas for BG,  $L_2$  is responsible for the location error. Consequently, DG yields a location error that corresponds to the sum of the  $L$  components of BD and BG. BF is the only situation with only a structure error (that arises mainly because the precipitation object F is larger than B).

Now considering situations in which at least two components of SAL are nonzero, there are several examples in which the components  $S$  and  $A$  have the same

TABLE 2. SAL values (format *S/A/L*) for all possible pairs of precipitation structures B–I as shown in Fig. 5. The matrix is antisymmetric for *A* and *S*, but symmetric for *L*. To enhance readability only values above the diagonal are given. Columns denote observations, and rows denote forecasts.

	B	C	D	E	F	G	H	I
B	0/0/0	−0.88/−0.88/0	0/0/0.14	0/0.67/0	−0.67/0/0	0/0/0.18	−0.68/−0.67/0	0.40/0.38/0
C		0/0/0	0.88/0.88/0.14	0.88/1.35/0	0.24/0.88/0	0.88/0.88/0.18	0.23/0.25/0	1.17/1.16/0
D			0/0/0	0/0.67/0.14	−0.67/0/0.14	0/0/0.32	−0.68/−0.67/0.14	0.40/0.38/0.14
E				0/0/0	−0.67/−0.67/0	0/−0.67/0.18	−0.68/−1.20/0	0.40/−0.31/0
F					0/0/0	0.67/0/0.18	−0.01/−0.67/0	1.00/0.38/0
G						0/0/0	−0.68/−0.67/0.18	0.40/0.38/0.18
H							0/0/0	1.01/0.98/0
I								0/0/0

sign (e.g., BC and CE) and only one where the signs differ (EI). This indicates that in most of these idealized situations, an underestimation of the amplitude goes along with a negative structure error (e.g., BC) and an overestimation of the amplitude with a positive value of *S* (e.g., CI). It is important to note that this is not a consequence of the mathematical design of the SAL components but rather due to the chosen examples. Consider, for instance, forecast B and observations C, and increase the amplitude of B continuously: this would not change  $S < 0$ , but it would increase the *A* component until it eventually becomes positive. Interesting is the comparison EI in which the simulated right circular cone underestimates the amplitude of the observed peaked object, along with a positive structure error. This can occur if a forecast misses the high-amplitude localized nature of a convective precipitation event.

Also of interest is FG, in which two seemingly similar precipitation fields are compared. However, in F the two local maxima are close to each other and only one object is identified. Compared to G with two well-separated objects, this yields a positive structure error (the object in F is too large) and a nonzero location error (due to  $L_2$ ).

A caveat of SAL can be noticed for CD and CG: here, the error components are very similar, however the fields D and G, which both score poorly compared to C, are rather different. This shows that SAL might indicate similar errors for differently shaped precipitation fields—a direct consequence of trying to capture the essential aspects of complex precipitation fields with three scalar parameters only.

#### 4. Application to precipitation forecasts for the German part of the Elbe catchment

In this section, operational QPFs from the regional model Consortium for Small-Scale Modeling-Alpine

Model (COSMO-aLMo) and the global European Centre for Medium-Range Weather Forecasts (ECMWF) model will be considered for the summer seasons 2001–04. Results of the application of SAL will be presented, first for four selected examples (section 4a), and then for a climatological analysis of the entire time period (section 4b). Also, to assess these results statistically, they will be compared with SAL calculations for persistence and random forecasts in section 4c.

COSMO-aLMo is a version of the nonhydrostatic limited-area model developed by COSMO (Steppeler et al. 2003) and operated at Meteo Swiss. Its horizontal resolution is 7 km on a rotated stereographic grid. The model has 40 vertical levels, subgrid-scale convection is parameterized with the Tiedtke scheme, and a single-moment bulk microphysical scheme is used that considers cloud water and cloud ice (since September 2003). Advection of precipitating hydrometeors is neglected until the implementation of a so-called prognostic precipitation scheme in November 2004 for the hydrometeor classes of rain and snow. Initial and boundary conditions were provided by the global model of the German Weather Service (GME), until September 2003, and by the ECMWF thereafter. Here, COSMO-aLMo forecasts that were started at 0000 UTC are used, and daily precipitation totals were taken as the accumulated precipitation between forecast times 6 and 30 h.

For the time period considered, operational ECMWF forecasts have a spectral resolution of T511, corresponding to about 0.4° latitude–longitude. To perform the comparison on the same grid, ECMWF forecasts have been interpolated onto the COSMO-aLMo grid with 7-km resolution. Also here, daily totals correspond to accumulated precipitation between forecast steps 6 and 30 h from simulations started at 0000 UTC.

The observational dataset of 24-h accumulated precipitation is based on rain gauge measurements, which are recorded daily at 0630 UTC. About 3500 stations in Germany are operated by the German Weather Ser-

vice, and the average distance between the stations is  $\sim 10$  km. Using the gridding technique of Frei and Schär (1998), the observations have been interpolated to the COSMO-aLMo grid. Further details of this gridded observational dataset for Germany can be found in Paulat (2007).

In summary, all three datasets of daily precipitation used in this study are available during four summer seasons on the same grid with a horizontal resolution of 7 km covering Germany. To illustrate the application of SAL, this study focuses on the summer season, which presents the largest variability in terms of precipitation structures, from small convective cells to widespread stratiform rain. The region considered is the German part of the catchment of the Elbe River, with an area of 97 175 km<sup>2</sup>. The Elbe originates in the Czech Republic, flows across eastern Germany, and has a total length of 1165 km. As a side remark, it is noted that in August 2002 (i.e., within the considered time period) a three-week flooding of the Elbe River saw water levels reach 150-yr highs (Rudolf and Rapp 2002). Large areas were inundated and the resulting insurance claims were in the multimillion Euro range.

#### a. Selected examples

Four days have been selected during summer 2001 for a detailed consideration of the application of SAL. They differ in terms of meteorological conditions and the resulting daily total precipitation patterns.

- Case 1 (2 June 2001): An intense low-pressure system was located over Denmark; warm and cold fronts moved over Germany, leading to widespread precipitation in the Elbe catchment (Fig. 6a). Maximum temperatures in the catchment were below 17°C.
- Case 2 (6 June 2001): Maximum temperatures were again rather low ( $< 18^\circ\text{C}$ ). A developing depression over the North Sea led to scattered showers in the Elbe area (Fig. 6c).
- Case 3 (7 July 2001): A mesoscale cyclone with a pronounced warm sector crossed Germany, leading to very intense precipitation (Fig. 6e). Maximum temperatures were up to 32°C.
- Case 4 (29 July 2001): A large-scale high-pressure system was situated over western and central Europe and maximum temperatures were again up to 32°C. Localized convection occurred in parts of the Elbe catchment (Fig. 6g).

Table 3 presents the SAL values for these examples. In the first example (Figs. 6a,b), the precipitation distribution is fairly homogeneous, and therefore at almost all grid points  $R_{ij}$  exceeds the threshold  $R^*$ . In

both datasets there is just one large object. All components of SAL are fairly small, indicating a high-quality forecast.<sup>1</sup> The largest error occurs in terms of amplitude ( $A = 0.312$ ), which is mainly due to an overestimation of precipitation in the northwest part of the domain. The two centers of mass nearly coincide, and therefore  $L$  is essentially zero.

In the second example (Figs. 6c,d), the precipitation distribution is more variable, in particular in the observations. Four larger (and several very small) objects are found in the observations and one dominant large one in the forecast. The large positive value of  $S$  indicates that the forecast does not capture the localized and rather peaked character of the observed precipitation. Also, there is a general overestimation of the precipitation amount ( $A = 0.88$ ). The location error is much larger than in the first example (but still moderate). The main contribution to  $L$  stems from the second component,  $L_2$ , because the model does not capture the distribution of the objects relative to the center of mass. The latter, however, is very well predicted in the western part of the catchment and hence  $L_1$  is almost zero.

Very intense precipitation is observed and simulated in the third example (Figs. 6e,f). Both observations and forecasts are dominated by one large object. The amplitude component  $A$  is essentially zero. The component  $S$  is negative (but in absolute numbers much smaller than the positive  $S$  values for examples 2 and 4), indicating that the model object is not flat enough. In the forecast, there are too-steep gradients between the heavy rain area and the surroundings, which are affected only by light rain. The location error is almost identical as in the second example, however with reversed importance of the two components. Here, the contribution of  $L_2$  can be neglected (as expected if most of the precipitation occurs in single objects), and the fairly large value of  $L_1$  corresponds to the significant eastward shift of the precipitation area in the forecast.

Finally, the fourth example presents a rather poor forecast as indicated by the large values of  $S$ ,  $A$ , and  $L$ . The model predicts one large object with intense rainfall in the southern part, whereas the observations reveal several small objects and showerlike precipitation in the central part of the catchment. Consequently, all three components of SAL are positive, indicating an overestimation of the total precipitation in the catchment, a failure in capturing the rather small-scale and peaked character of the precipitation objects, and a significant southwestward shift of the rainfall area.

<sup>1</sup> Here, an SAL component is termed “small” if it is much smaller than typical values of a random reference forecast, as discussed in section 4c.

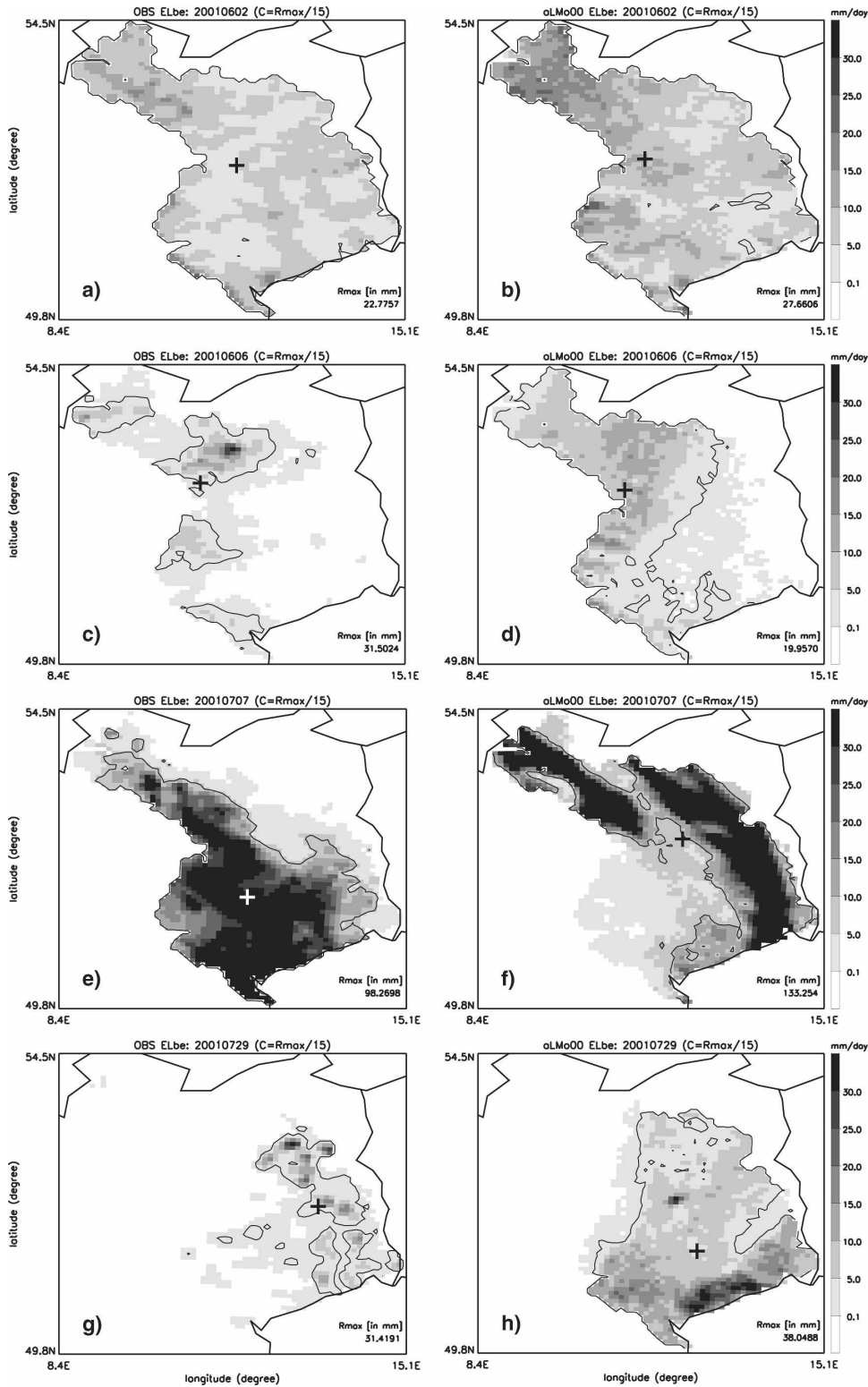


FIG. 6. Examples of daily precipitation fields in the German part of the Elbe catchment. (left) Observations and (right) COSMO-aLMO forecasts. (a),(b) 0600 UTC 2 Jun–0600 UTC 3 Jun 2001; (c),(d) 6–7 Jun 2001; (e),(f) 7–8 Jul 2001; (g),(h) 29–30 Jul 2001. The thin black line denotes the threshold value  $R^*$  used for the identification of the objects. The black cross [white in (e)] denotes  $x$ , the center of mass of the precipitation field in the domain.

TABLE 3. SAL values for the four example cases shown in Fig. 6. Also shown are the two parts  $L_1$  and  $L_2$  that contribute to the  $L$  component. For the definition of objects a threshold of  $R^* = 1/15 \cdot R^{\max}$  has been chosen.

Case	$S$	$A$	$L$	$L_1$	$L_2$
1	0.119	0.312	0.033	0.033	0.000
2	1.597	0.877	0.202	0.026	0.176
3	-0.430	-0.057	0.196	0.191	0.005
4	1.598	1.325	0.303	0.174	0.129

It is instructive to consider the sensitivity of the  $S$  and  $L$  components to the subjective choice of the object identification threshold [Eq. (1)] for these real data examples. In addition to the standard value  $f = 1/15$ , four different threshold factors between  $1/17$  and  $1/13$  have been used for the calculation of SAL. For three examples (1, 2, and 4) the variability of  $S$  and  $L$  is fairly small:  $S$  varies by less than 1.5% (which is negligible), and  $L$  by 2%–3% for examples 2 and 4, and by 10% for example 1 (note that for this example  $L$  is almost zero anyway). For example 3, the values are also almost constant for  $1/14 \leq f \leq 1/17$ . However, the camel effect (see discussion toward the end of section 3a) occurs if the threshold factor is further increased to  $f = 1/13$ :  $S$  jumps from  $-0.430$  to  $-0.830$ , and  $L$  from  $0.196$  to  $0.366$ . The reason is that in this example, similar to the idealized situation shown in Fig. 4, two almost equally large objects are identified in the forecast (Fig. 6f) when using this larger threshold, instead of one object with the slightly lower thresholds. The two objects in the forecast compare even less favorably with the observations than the single object identified with the standard threshold factor  $f = 1/15$ , and therefore  $S$  attains a more negative value and  $L$  increases due to an additional contribution from  $L_2$ . All in all, this brief sensitivity analysis indicates that the SAL values are robust, except for the well-understood situations where the camel effect occurs.

### b. Comparison of global and mesoscale model forecasts

Here we perform a climatological SAL analysis of the QPF capabilities of the global model ECMWF and the limited-area model COSMO-aLMo in the German part of the Elbe catchment for the four summers 2001–04. The main goals are (i) to introduce the compact SAL diagram, (ii) to quantify SAL values of QPFs from state-of-the-art NWP models, and (iii) to identify systematic differences in terms of SAL performance between coarser and finer-scale models. A threshold of  $0.1$  mm (corresponding about to the observational detection limit) is used for the maximum gridpoint value

of precipitation in the domain to distinguish between days with rain (wet days) and without rain (dry days). In case of a dry forecast and/or dry observations, no SAL values can be computed, because, for instance, the center of mass of the precipitation distribution is not defined in such a situation.

Figure 7 shows SAL diagrams for the COSMO-aLMo (Fig. 7a) and ECMWF (Fig. 7b) models, respectively. The small contingency table in the bottom right-hand corner of the SAL diagram provides information about the number of dry and wet days in the observations and forecasts, respectively. Only one day was dry according to both observations and COSMO-aLMo (Fig. 7a). On 18 days, the model missed the precipitation event, and on 13 days, the model produced a false alarm. It is important to consider the number of these cases, because they correspond to particular categories of poor forecasts but are not accessible to the SAL technique. Accordingly, they do not appear in the SAL diagram. During the majority of days (334), both observations and forecasts were characterized by rain, and all these days contribute with one entry to the SAL diagram. Abscissa and ordinate correspond to the  $S$  and  $A$  components, respectively, and the color of the dots represents the  $L$  component (see grayscale in the top left). Excellent forecasts (small values of all three components) are found as white and light gray dots in the center of the diagram. Dashed lines indicate the median values of  $S$  and  $A$ , and the gray-shaded box denotes the 25th and 75th percentiles of the two components. The median and the 25th and 75th percentiles of  $L$  are indicated by the thick and thinner white lines plotted in the grayscale.

For COSMO-aLMo (Fig. 7a), most forecasts are found in the first (top right) and third (bottom left) quadrant of the diagram. In the first quadrant, forecasts overestimate both the amplitude and the structure components of SAL. In the third quadrant, both components are underestimated. The high density of entries along the main diagonal indicates that the model typically tends to overestimate the precipitation amount in the considered area by producing too-large and/or flat precipitation objects. Analogously, underestimations of the amount go typically along with too-small and/or peaked objects. Particularly notable is the cluster of dark gray dots in the top right-hand corner of the diagram. Further analysis shows that in these cases fairly little precipitation was observed, but the model predicted significant precipitation both in terms of amplitude and extension. These cases can also be regarded as false alarms. In comparison, the lower density of dots in the bottom left-hand corner indicates that the model rarely missed a significant precipitation event (values of

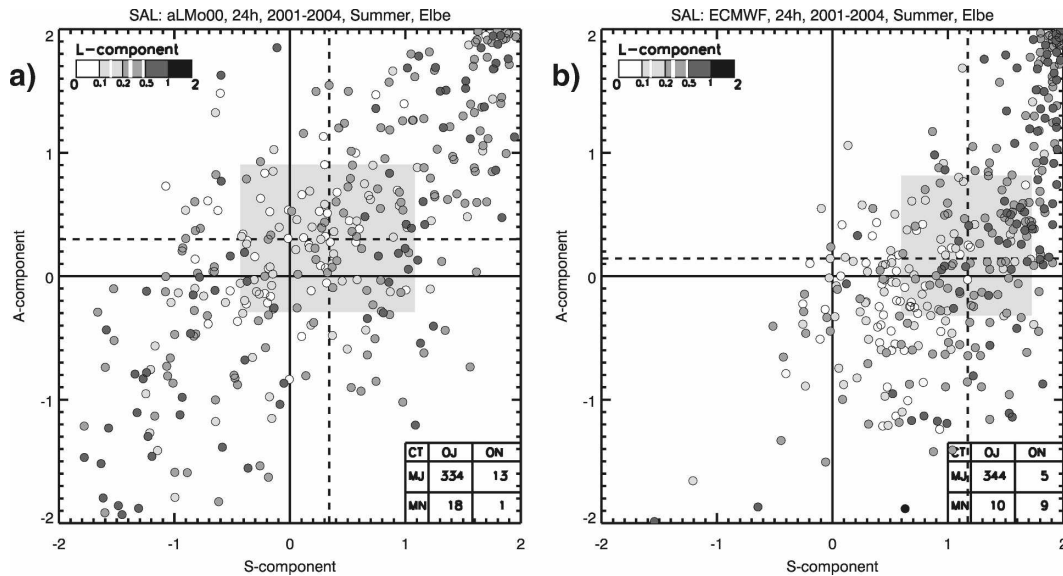


FIG. 7. SAL diagrams for the daily precipitation forecasts of the (a) COSMO-aLMO and (b) ECMWF models during the summer seasons 2001–04 in the German part of the Elbe catchment. Every dot shows the values of the three components of SAL for a particular day. The *L* component is indicated by the color of the dots (see grayscale in top left). Median values for the *S* and *A* components are shown as dashed lines, and the gray box extends from the 25th to the 75th percentile of the distribution of *S* and *A*, respectively. See section 4b for more details.

$A < -1.5$  are relatively rare). The second (top left) and fourth (bottom right) quadrant contain only few SAL entries. Forecasts in the second quadrant produce too much rain, however, with objects that are too small and/or too peaked. This could occur, for instance, if intense showers are predicted in a situation with rather weak stratiform precipitation. Predictions in the fourth quadrant underestimate the amplitude of precipitation and simultaneously produce objects that are too large and/or flat. A possible scenario here is an erroneous forecast of stratiform rain in a situation with intense localized showers. It is notable that no forecasts are situated in the top left-hand and bottom right-hand corners of the diagram, indicating that it is difficult to produce for instance a *strong* overestimation of precipitation amplitude with much too small objects. The *L* component does not show a systematic behavior with the other two components. Light and dark dots (i.e., forecasts with a small and large location error, respectively) occur in all quadrants. A slight concentration of white dots occurs near the center, and darker dots are more frequent in the left and right part of the diagram (i.e., for large absolute values of *S*). The median values of *S* and *A* are positive (both about 0.3), whereas the interquartile distance is about 1.2 for *A* and 1.5 for *S*. These values of the interquartile distances are relatively large and indicate that frequently COSMO-aLMO forecasts score poorly in terms of one of the two or both components.

As for the four examples in section 4a, sensitivity calculations have been performed to assess the frequency of the camel effect. Comparison of *S* and *L* values computed with  $f = 1/13$  and  $1/17$  (recall that our standard value is  $f = 1/15$ ) for the 334 COSMO-aLMO forecasts in the Elbe catchment yielded for *S* an absolute difference of more than 0.1 in 10% and of more than 0.3 in 3% of the cases. For *L*, a difference of more than 0.1 occurred in 5% and of more than 0.3 in 2% of the cases. This indicates that the sensitivity of the SAL values with respect to the threshold factor *f* is typically small and that the camel effect, which is associated with a large sensitivity to *f*, occurs in about 3% of the cases. Note that for a more complete analysis of the uncertainties of the resulting SAL values, it would be important to also consider the uncertainties associated with the observational dataset, for instance, through probabilistic upscaling by ensembles of stochastic simulations conditioned to the available observations (Ahrens and Beck 2008).

Now considering the performance of the global ECMWF model (Fig. 7b), a striking difference occurs. Almost all forecasts are characterized by positive values of *S*. Compared to the results for COSMO-aLMO, the entire distribution is shifted toward the right, indicating that the global model produces too large and/or too flat precipitation objects. This is not surprising, given the coarser model resolution; however, unlike classical error scores, SAL is able to identify and quan-

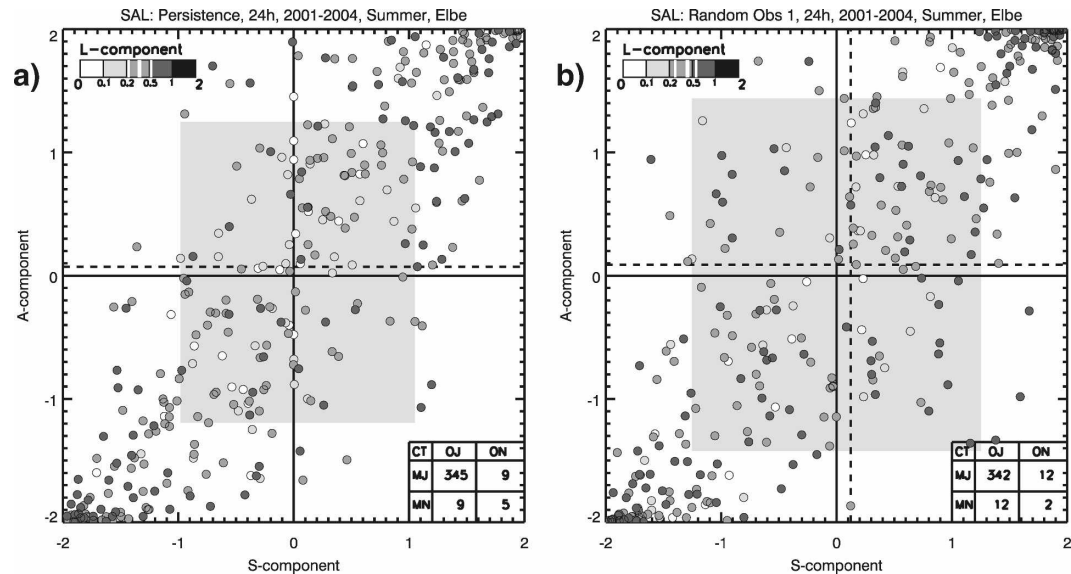


FIG. 8. SAL diagrams for (a) persistence and (b) random forecasts. Plot conventions as in Fig. 7. See section 4c for details.

tify this specific characteristic of the forecasts. Considering the two other aspects ( $A$  and  $L$  components), the two models perform similarly, except that strongly negative values of  $A$  occur less frequently for the global model. Note also that the number of missed events and false alarms is slightly larger for the regional model. In summary, SAL indicates that in the considered area, summertime QPFs from the higher-resolution regional model are superior to the ones from the global model, because they are superior in capturing the structure of the precipitation objects.

### c. Comparison with persistence and random forecasts

For a statistical investigation of the SAL results presented in the previous subsections, the SAL technique has also been applied to sets of persistence and random forecasts, respectively. This is important to assess the quality of NWP model predictions relative to standard reference forecasts. Both reference forecasts are based on the observational dataset described in section 4a and therefore independent of a particular NWP model. For the persistence forecasts, observations from a given day are used as predictions for the next day. For the random forecasts, for every day, a forecast field has been randomly chosen among the set of observed fields, in such a way that every observed field is chosen once as a forecast. In other words, every observed field is considered once as the observations and once as the forecast.

The results of these experiments are shown in Fig. 8. Clearly, in both cases there are much fewer SAL values

in the center of the diagrams. The median values of  $A$  and  $S$  are close to zero, which should be expected because of the symmetry in the construction of the experiments. The median values of  $L$  are about 0.3 for the persistence and almost 0.5 for the random forecasts, respectively. They are both larger than the corresponding values for the NWP model forecasts (see Fig. 7). The interquartile distances of  $A$  and  $S$  (i.e., the gray boxes) are much larger than in Fig. 7, which statistically corroborates the quality of the QPFs from the numerical models, at least for a significant portion of the forecasts. Table 4 provides quantitative information on this issue. The radius  $\rho$  of a sphere in the three-dimensional space spanned by the components of SAL has been calculated, which contains the best 5%, 10%, 20%, and 50% of the forecasts. The values reveal that at least 20% of the COSMO-aLMo forecasts are better than the 5% best random forecasts, and 50% of the COSMO-aLMo forecasts are better than the best about 25% of the random forecasts. Comparing with the SAL values of the examples discussed in section 4a

TABLE 4. Radius  $\rho$  of the sphere in SAL space that contains the best 5%, 10%, 20%, and 50% of the forecasts, for the NWP model COSMO-aLMo and for the persistence and random forecasts.

Forecast	$\rho$ for best forecasts			
	5%	10%	20%	50%
COSMO-aLMo	0.30	0.40	0.55	1.05
Persistence	0.35	0.55	0.85	1.45
Random	0.55	0.70	0.95	1.65

(cf. Table 3) indicates that example 1 belongs to the 5% best COSMO-aLMO forecasts, and example 3 to the best 15%. These forecasts are better than random forecasts with a statistical significance of more than 95%. In contrast, the QPFs shown in examples 2 and 4 score rather poorly ( $S > 1.5$ ) and have a quality that is met also by about 50% of the persistence or random forecasts.

Note that similar to the results for the COSMO-aLMO model (Fig. 7a), the SAL values for the persistence and random forecasts are rarely in the second and fourth quadrants of the diagram (Fig. 8). This indicates that the predominance of SAL values in the first and third quadrants, found for COSMO-aLMO, is not a particular feature of the model but a rather intrinsic characteristic of SAL. It reflects the fact that it is difficult to strongly overestimate the amount of precipitation with too-small objects (and vice versa). In contrast, the shift toward positive values of  $S$  for the ECMWF model (Fig. 7b) points to a systematic deficiency of the coarser-scale global model in realistically capturing the structure of precipitation events.

## 5. Discussion

A novel quality measure, SAL, has been introduced for the verification of QPFs. It can be categorized as an object-oriented verification approach, with the specific characteristics outlined at the end of section 1. The three components of SAL quantify distinct aspects of the quality of a QPF, which are associated with the structure, amplitude, and location of the precipitation field. These three components describe aspects of QPF quality that are directly relevant to forecast users. Consider, for example, the hydrological modeling for a river catchment: the  $A$  component is based on the catchment mean precipitation and hence describes the overall bias in the precipitation input to the hydrological model. Obviously, errors in  $A$  can be expected to result in systematic runoff biases. On the other hand, the  $L$  component describes the accuracy with which precipitation is located/distributed between several subcatchments. Forecasts with nonzero  $L$  would give rise to random errors in the resulting river runoff. Finally, the  $S$  component specifically addresses the effect of QPF errors in connection with the nonlinear processes at the soil surface. The spatial intensity distribution is critical for the repartitioning of precipitation water between surface runoff and infiltration into soils. A nonzero value of  $S$  in the time mean will affect the soil water balance even when the domain mean value is correct (i.e.,  $A = 0$ ). Moreover, it has consequences on the frequency statistics of runoff, unless compensated for by other errors. Altogether, SAL is a quality measure that helps the

user to anticipate effects of QPF limitations in a hydrological application.

The SAL technique has been tested with synthetic fields and applied to forecasts from a regional and global NWP model, as well as to persistence and random forecasts. To this end, it was important to have all datasets (observations and forecasts) available on the same grid. It was shown that for case studies, SAL can provide meaningful and quantitative information about QPF errors. When applied to a large set of QPFs, it pinpointed the generally more realistic structure of precipitation events as one of the major advantages of QPFs from high-resolution models. It was also shown that the COSMO-aLMO model performed significantly better than random forecasts that are not based on a NWP model.

In the following paragraphs, a few aspects will be discussed in more detail, related to the choice of the threshold for the definition of objects, absolute versus relative quality measures, and alternative definitions of the components  $S$  and  $L$ . Also, possibilities for future extensions and applications of SAL are mentioned briefly.

In contrast to other object-oriented verification approaches (e.g., Ebert and McBride 2000; Davis et al. 2006a), no fixed precipitation threshold is used to identify the objects. The advantage of a fixed threshold is that verification can focus on a particular category, for instance, of intense events, and the statistical results are not blurred by (very) weak events that might be of less interest. However, specification of a fixed threshold excludes poor forecasts from an object-oriented verification in situations in which the threshold is not exceeded in either the model ("missed events") or the observations ("false alarms"). This leads to a positive bias in the object-oriented evaluation of a model's QPF performance, because only reasonably good forecasts enter the statistics. It is for this reason that we adopted an alternative approach and used a flexible threshold for the identification of objects, which in general differs in the forecast and observations. With this approach, very few days are excluded from the analysis (only when one of the two datasets contained no precipitation in the entire domain). The possibility still exists to stratify the results according to the observed intensity of the events and thereby to learn more about the QPF performance for weak, medium, and intense events. Such an analysis is documented in Paulat (2007).

Another difference, for instance, to the CRA method by Ebert and McBride (2000) is that the three components of SAL are not absolute but relative (dimensionless) measures. The motivation for the use of relative measures is that they potentially allow a direct com-



parison of the QPF performance during weak and intense precipitation events.

A third and important difference is that our definitions of the three components do not follow from a mathematical decomposition of a well-known error measure (like the mean-squared error in case of the CRA technique). This renders the definition of the components subjective, at least to a certain degree. The advantage, however, is that the components can be tailored such that they become close to a subjective visual judgment. In any case, other definitions would be possible and could be regarded as variants of the SAL technique proposed here. For instance, instead of an absolute displacement component  $L_1$ , a vector location error  $L_1$  would provide additional information about the direction of the displacement. Alternatively, the Hausdorff distance metric could serve as a more sophisticated approach to define a location error component (Venugopal et al. 2005). For the structure component, the volume of the scaled precipitation objects has been used as the key parameter. A simpler possibility would be to use the objects' base area. However, this would lead to a loss of information, because no distinction would be possible between peaked and flat objects. In contrast, a more refined alternative would be to use the surface of the objects instead of their volume. This would allow to additionally distinguish, for instance, between right circular cones and right elliptic cones with the same base area, because these objects have the same volume but not the same surface. Such an extension of  $S$  would be desirable; however, the accurate computation of the surface of complex-shaped precipitation objects is not straightforward and for this reason has not been pursued in this study.

In the future, SAL will be applied to assess the QPF performance of several models on daily and hourly time scales. Currently, several forecasting centers are about to introduce operational short-range forecasts with high-resolution, convection-resolving model versions (e.g., at the German Weather Service, the 21-h COSMO-DE forecasts with a horizontal resolution of 2.8 km). There are considerable expectations that this new model generation can significantly advance QPF quality and overcome some of the inherent problems with the parameterization of deep convection (Ebert et al. 2003; Fritsch and Carbone 2004). Model case studies without parameterized convection (e.g., Steppeler et al. 2003; Done et al. 2004; Trentmann et al. 2007) indicate that this new category of NWP models provides a more accurate depiction of the physics of convective systems (e.g., cold pool formation and the organization of the systems). However, for single cases, this does not necessarily imply an improved QPF performance (Zhang

et al. 2006). First statistical investigations indicate an improved prediction of larger accumulations when using convection-resolving models compared to coarser-scale models with parameterized convection (Mittermaier 2006, using the technique introduced by Casati et al. 2004). Also, as found by Davis et al. (2006b), a convection-resolving version of WRF tends to delay the onset of precipitation systems, which then last too long and are characterized by a too-broad intensity distribution. It will be interesting to compare QPFs from the two categories of NWP models with the SAL technique.

Another application of SAL will be to quantitatively analyze QPF differences in case study sensitivity experiments, where model numerics, physical parameterizations, or the initial and boundary data are varied to assess the importance of this NWP component for QPF accuracy. Here, SAL might be useful to categorize the simulation differences in terms of key aspects of the precipitation field. Along the same lines, forecasts from an ensemble prediction system could be compared to observations and the ensemble spread of the precipitation forecast expressed in terms of the three components of SAL. Also, SAL can be used to compare the characteristics of different climatological precipitation datasets, for instance, provided by regional climate models and satellite retrieval methods (Früh et al. 2007).

*Acknowledgments.* We wish to thank the German Weather Service (DWD) for providing rain gauge measurements and access to ECMWF data, MeteoSwiss for granting access to their operational COSMO-aLMO forecasts, and Caren Marzban and two anonymous reviewers for their constructive and helpful comments. MP acknowledges funding from the German Research Foundation (DFG) priority program on Quantitative Precipitation Forecasts (SPP 1167).

#### REFERENCES

- Ahrens, B., and A. Beck, 2008: On upscaling of rain-gauge data for evaluating numerical weather forecasts. *Meteor. Atmos. Phys.*, **99**, 155–167, doi:10.1007/s00703-007-0261-8.
- Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141–154.
- Damrath, U., G. Doms, D. Frühwald, E. Heise, B. Richter, and J. Steppeler, 2000: Operational quantitative precipitation forecasting at the German Weather Service. *J. Hydrol.*, **239**, 260–285.
- Davis, C. A., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.
- , —, and —, 2006b: Object-based verification of precipi-

- tation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, doi:10.1002/asl.72.
- Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179–202.
- , U. Damrath, W. Wergen, and M. E. Baldwin, 2003: The WGENE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Meteor. Soc.*, **84**, 481–492.
- Frei, C., and C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, **18**, 873–900.
- Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season. *Bull. Amer. Meteor. Soc.*, **85**, 955–965.
- Früh, B., J. Bendix, T. Nauss, M. Paulat, A. Pfeiffer, J. W. Schipper, B. Thies, and H. Wernli, 2007: Verification of precipitation from regional climate simulations and remote-sensing observations with respect to ground-based observations in the upper Danube catchment. *Meteor. Z.*, **16**, 275–293.
- Hohenegger, C., D. Lüthi, and C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Mon. Wea. Rev.*, **134**, 2095–2107.
- Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley and Sons, 240 pp.
- Keil, C., and G. C. Craig, 2007: A displacement-based error measure applied in a regional ensemble forecasting system. *Mon. Wea. Rev.*, **135**, 3248–3259.
- Marzban, C., and S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, **21**, 824–838.
- Mittermaier, M. P., 2006: Using an intensity-scale technique to assess the added benefit of high-resolution model precipitation forecasts. *Atmos. Sci. Lett.*, **7**, 36–42, doi:10.1002/asl.127.
- Paulat, M., 2007: Verifikation der Niederschlagsvorhersage für Deutschland von 2001–2004. Ph.D. thesis, University of Mainz, 155 pp.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97.
- Rudolf, B., and J. Rapp, 2002: *Das Jahrhunderthochwasser der Elbe: Synoptische Wetterentwicklung und klimatologische Aspekte*. (The Elbe flood of the century: Synoptic weather evolution and climatological aspects). Klimastatusbericht 2002, German Weather Service, 173–188.
- Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647–677.
- Steppele, J., G. Doms, U. Schättler, H. W. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, **82**, 75–96.
- Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12**, 257–268.
- Trentmann, J., U. Corsmeier, J. Handwerker, M. Kohler, and H. Wernli, 2007: Evaluation of convection-resolving model simulations with the COSMO-Model in mountainous terrain. *Proc. 29th Int. Conf. on Alpine Meteorology*, Chambéry, France. [Available online at [http://www.cnrm.meteo.fr/icam2007/html/PROCEEDINGS/ICAM2007/index\\_authors.html](http://www.cnrm.meteo.fr/icam2007/html/PROCEEDINGS/ICAM2007/index_authors.html)]
- Venugopal, V., S. Basu, and E. Foufoula-Georgiou, 2005: A new metric for comparing precipitation patterns with an application to ensemble forecasts. *J. Geophys. Res.*, **110**, D08111, doi:10.1029/2004JD005395.
- Walser, A., and C. Schär, 2004: Convection-resolving precipitation forecasting and its predictability in Alpine river catchments. *J. Hydrol.*, **288**, 57–73.
- , D. Lüthi, and C. Schär, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.*, **132**, 560–577.
- Wernli, H., and M. Sprenger, 2007: Identification and ERA-15 climatology of potential vorticity streamers and cutoffs near the extratropical tropopause. *J. Atmos. Sci.*, **64**, 1569–1586.
- Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the surprise snowstorm of 24–25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632.
- , —, and —, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185.
- , A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166.