The skill of convective parameters and indices to predict isolated and severe thunderstorms

M. Kunz

Nat. Hazards Earth Syst. Sci., 7, 327–342, 2007
www.nat-hazards-earth-syst-sci.net/7/327/2007/

Natural Hazards and Earth System Sciences

The skill of convective parameters and indices to predict isolated and severe thunderstorms

M. Kunz

Institut für Meteorologie und Klimaforschung, Universität Karlsruhe/Forschungszentrum Karlsruhe, Karlsruhe, Germany

Received: 6 November 2006 – Revised: 20 March 2007 – Accepted: 20 April 2007 – Published: 3 May 2007

Abstract. The preconvective environment on days with ordinary, widespread, and severe thunderstorms in Southwest Germany was investigated. Various thermodynamic and kinetic parameters calculated from radiosoundings at 12:00 UTC were verified against subsequent thunderstorm observations derived from SYNOP station data, radar data, and damage reports of a building insurance company. The skill of the convective parameters and indices to predict thunderstorms was evaluated by means of probability distribution functions, probabilities of thunderstorms according to an index threshold, and skill scores like the Heidke Skill Score (HSS) that are based on categorical verification.

For the ordinary decision as to whether a thunderstorm day was expected or not, the best results were obtained with the original Lifted Index (80% prediction probability for LI≤ -1.73; HSS=0.57 for LI$\leq$1.76), the Showalter Index, and the modified K-Index. Considering days with isolated compared to widespread thunderstorms, the best performance is reached by the Deep Convective Index. For days with severe thunderstorms that caused damage due to hail, local storms or floods, the best prediction skill is found again for the Lifted Index and the Deep Convective Index, but also for the Potential Instability Index, the Delta-θ_e Index, and a version of the CAPE, where the lifting profile is determined by averaging over the lowest 100 hPa.

1 Introduction

The prediction of thunderstorms is one of the most difficult issues in weather forecasting. Deep convective clouds develop on a rather small spatial and temporal scale in the order of 1–10 km and 1–12 h. So far, operational numerical models with horizontal resolutions of about 10 km have often failed to predict not only the location and the time of convection initiation (Anquetin et al., 2005; Meißner et al., 2007), but also the type and intensity of thunderstorms. Severe thunderstorms are frequently associated with heavy rainfall, hail, or local storms that are a major cause of natural disasters particularly over mountainous terrain that favor the initiation or the triggering of convection (Orville, 1965; Banta, 1990; Barthlott et al., 2006). Improving the prediction of thunderstorms, especially of severe ones, consequently is a challenging task that may help to prevent or mitigate damage.

The general requirements for a preconvective environment are well known and were summarized by many authors, for example by Doswell (1987) or Houze (1993): 1) Any kind of instability over a layer of sufficient depth, 2) a moist layer at lower levels, and 3) a mechanism that triggers the convection. To quantify the first two conditions, various thermodynamic and kinematic parameters have been designed in the past decades. These so-called convective parameters and indices reflect the potential for thunderstorm development according to the prevailing properties of the air mass. In many studies, the efficiency of the various indices derived from the observed vertical profiles for thunderstorm prediction was investigated, for example by Schulz (1989), Lee and Passner (1993), Fuelberg and Biggar (1994), Huntrieser et al. (1997), Haklander and Van Delden (2003), and Manzato (2005). However, little attention has been paid so far to the index-based prediction of severe thunderstorms. The present study aims at evaluating several convective indices with respect to their skills and efficiency to predict thunderstorms of variable severity. The study tries to give some useful insight into the characteristics of the preconvective environments that are decisive for the initiation of thunderstorms.

The area under investigation is located in the northern parts of the federal state of Baden-Württemberg in Southwest Germany (Fig. 1). The terrain exhibits a certain complexity with some rolling terrain and the low mountain ranges of the Black Forest and Swabian Jura. Local wind systems

Correspondence to: M. Kunz
(michael.kunz@imk.uka.de)

Published by Copernicus GmbH on behalf of the European Geosciences Union.

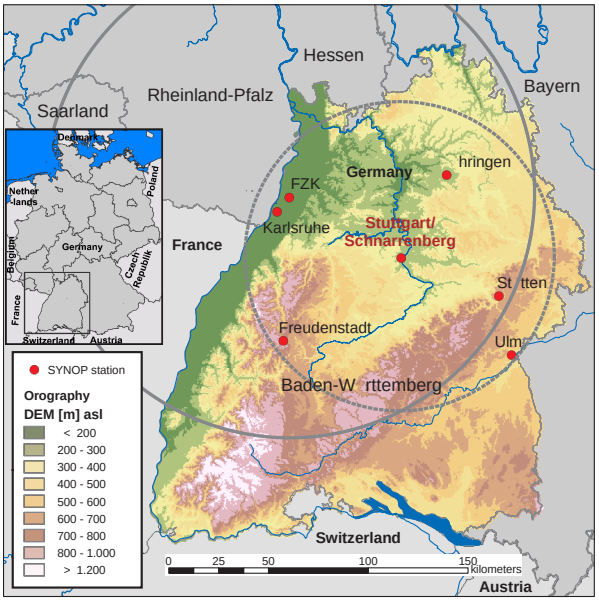

Fig. 1. Area under investigation with the radiosonde station Stuttgart-Schnarrenberg, SYNOP stations, range of the radar (solid circle), and area of the considered insurance data (dotted circle).

often develop at the slopes and in the larger valleys (Koßmann and Fiedler, 2000), which may trigger the initiation of convection. It is well known that this region favors the development of deep convection (Meißner et al., 2007), often associated with damage due to hail, local storms, or floods (Kunz and Kottmeier, 2005). To obtain comprehensive information about both thunderstorm days and the intensity of the thunderstorms, observation data from different networks were used in this study: Data from SYNOP stations, radar data, and damage reports from a building insurance company.

The paper is structured as follows: The data sets and their characteristics as well as the derived convective indices will be specified in Sect. 2. Section 3 will describe the methods applied for the evaluation of the various indices. The skills of the indices to predict thunderstorms in terms of statistical parameters, probabilities of occurrence, and skill scores will be discussed in detail in Sect. 4. The last Sect. 5 will give a summary of the results and some conclusions.

2 Data sets

The prevailing preconvective environment is described by various convective and thunderstorm indices derived from radiosonde observations at 12:00 UTC at the station of Stuttgart-Schnarrenberg (315 m a.s.l.; see Fig. 1) of German Weather Service (Deutsche Wetterdienst (DWD)). Substantial cold air advection on higher levels, often associated with cold front passages, may reduce static stability profoundly. Therefore, days with a temperature decrease of more than 5 K at 850 hPa in the succeeding 00:00 UTC sounding, were excluded from the analysis (66 days between 1986 and 2003, i.e. around 2% of all days). This criterion was tested against other criteria including other levels and checked for synoptic consistency in several cases.

For the assessment of the indices, information about thunderstorm occurrence and their characteristics is necessary with a high spatial coverage and for a sufficiently long term. Since these high requirements are not met by a single data set, data from different kinds of observations were used: SYNOP station data (SY), radar data (RA), and insurance data (SV). Depending on the data set, the days were divided into days without thunderstorms and days with different characteristics of thunderstorms. (see Table 1 for the number of days). Since the analysis is based on the 12:00 UTC soundings thunderstorm occurrence is considered between 12:00 and 23:50 UTC only. This was not taken into account for the insurance data as they have no time indication. All examinations are restricted to a 6-month period between April and September, when severe thunderstorms in Germany occur almost exclusively.

2.1 Synoptic station data (SY)

Data measured by all DWD SYNOP stations with a distance of less than 75 km around the radiosonde station between 1986 and 2003 were used to determine thunderstorm days (red circles in Fig. 1: Stuttgart, Freudenstadt, Stötten, Öhringen, Karlsruhe, and Ulm). Days with more than 50% data set to false, e.g. no observer present, were excluded from the list. Such days totaled 236 days or 7.2%. In the SYNOP report, the present and past weather types at and around a station are hourly encoded by the ww-code, a two-digit number between 00 and 99. The following numbers indicate the occurrence of thunderstorms: ww=13, 17, 29, 91–99. If any of these numbers is recorded at any station, the day is considered a thunderstorm day (SY). Figure 2a shows the monthly means of thunderstorm days according to the observations at the SYNOP stations with a distinct maximum in July and a minimum in November.

2.2 Radar data (RA)

Based on the high-resolution data from the C-band radar of IMK between 1998 and 2001, days with various thunderstorm intensities could be distinguished. The radar situated at the Forschungszentrum Karlsruhe (FZK) covers a range of 120 km in radius (Fig. 1). It has a spatial resolution of 1 km × 1 km and a temporal resolution of 10 min. Days with less than 50 scans, i.e. less than 490 min observing time, were excluded from the study.

To determine intensities and characteristics of the convective cells in the whole area, the tracking algorithm TRACE-3D (Handwerker, 2002) was applied to the raw 3-D radar data. The algorithm identifies convective cells by specific

Table 1. Numbers of thunderstorm days derived from observational data for different definitions: Thunderstorm days according to SYNOP data (SY); days with all kinds of thunderstorms (RA1), widespread (RA2), and severe (RA3) thunderstorms according to the radar data (RA); days with damage due to hail (SV1) and widespread hail (SV2); days with damage due to storm/flood (SV3) and widespread storm/flood (SV4) according to the data of the SV insurance company (SV).

	SY	RA1	RA2	RA3	SV1	SV2	SV3	SV4
Non-thunderstorm days	2.251	461	–	–	3.170	–	3.099	–
Thunderstorm days	807	117	86	60	124	21	195	39

radar signatures and follows them in space and time using successive radar images. For each scan, the algorithm determines the total number of cells, maximum reflectivity, and the volume and spatial extension of the convective area.

A day was classified as a thunderstorm day, if more than 6 scans during the relevant period, 3 of them consecutive, met the following conditions: Maximum reflectivity $\geq$60 dBZ, number of cells $\geq$6, convective area $\geq$60 km^2, and convective volume $\geq$120 km^3 (RA1). If all these criteria were fulfilled for 15 scans with 5 of them consecutive, the day was classified as a day with widespread thunderstorms (RA2). Finally, if the same conditions were valid as for RA1, but with a maximum reflectivity of $\geq$65 dBZ and a convective area of $\geq$100 km^2, the days were categorized as severe thunderstorm days (RA3). The different criteria were adjusted and tested manually by convective signatures in summertime reflectivity images (MaxCappi).

The classification scheme for the radar data did not reveal any thunderstorm day between November and March (Fig. 2a) because of the low vertical extension of convection and the low reflectivity for ice particles. But also for the other months, the number of thunderstorm days according to the radar data is lower compared to that determined from the SYNOP data. Hence, the criteria for defining a thunderstorm day are more stringent for the radar data in comparison to the SYNOP data.

2.3 Insurance data (SV)

Another classification of days with severe thunderstorms was based on loss data of the SV Sparkassen–Versicherung building insurance company (hereinafter referred to as SV) for the period 1986–2003. Between 1960 and 1994, a building insurance against natural hazards – earthquakes, landslides, floods, storms, hail, or avalanches – was obligatory in Baden-Württemberg and offered exclusively by the monopolist Gebäudeversicherung Baden–Württemberg. This results in a good representativeness of the loss data, even though the number of policies, now taken out by the successor SV, had decreased successively since the abolishment of the obligation in 1994.

Separated into five-digit postal code zones, the data comprise the date of each damage event, the number of notification of claims, and the kind of the hazard. The total number

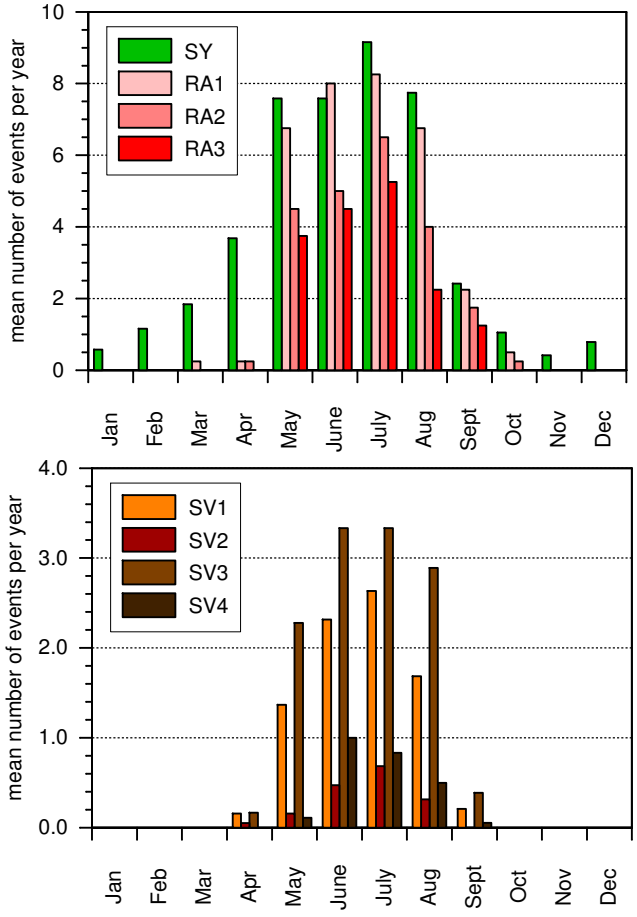

Fig. 2. Monthly mean of thunderstorm days as defined by SYNOP station data and radar data (top) and number of hail or storm/flood days according to the SV insurance data (bottom).

of policies per year and per postal code zone allows for a data correction to account for the yearly variability of the portfolio and especially for the aforementioned decrease after 1994.

In this study, data of loss due to hail or due to floods and storms associated with severe thunderstorms were used for all postal code zones within a region of less than 75 km in radius around the radiosonde station (Fig. 1). A day was classified as a hail day (SV1) or storm/flood day (SV3) when more than 10 claims (corrected) were settled on that day. This

lower threshold was necessary, because some damage notifications were ascribed to a wrong day. To separate local from widespread events, a second threshold was defined by the 80% percentile of the frequency distribution of the number of claims per day. This gives a threshold of 330 claims per day for widespread/strong hail days (SV2) and 165 claims for widespread storm/flood days (SV4).

Damage-causing thunderstorms as reflected by the insurance data occur almost exclusively between May and August (Fig. 2b). The absolute maximum for days with hail is found in July, for storm/flood in June.

2.4 Radiosonde data and convective indices

From the Stuttgart radiosonde observations at 12:00 UTC, various convective and thunderstorm indices that were considered to be predictors for expected thunderstorm development, intensity of thunderstorms, or thunderstorm probability were calculated. The soundings contain profiles of air pressure, geopotential height, temperature and dewpoint temperature, wind speed and wind direction. As the radiosonde data are archived only on constant pressure levels as well as on significant levels, where the gradient of one of the observed variables changes noticeably, all soundings were interpolated into equidistant increments of 10 m.

During the observation period between 1 April 1986 and 30 September 2003, a total of 3256 complete soundings, that is more than 98.9% of all possible soundings in that period, were performed. Some vertical profiles exhibit incomplete humidity or wind data, especially at upper air levels. If it was not possible to calculate a certain convective index from such an incomplete sounding, this index was excluded from further examinations, but not the other indices on that day.

The theoretical concept underlying most of the different indices is to represent conditional and/or latent and/or potential instability. The state of a layer is referred to as conditionally instable when the environmental lapse rate curve is between the dry and the moist adiabatic lapse rate curves on a thermodynamic diagram (Haurwitz, 1941). Parameters that account for this are the Vertical Totals (VT) or the Boyden Index (BOYD). A state is referred to as latent instability when the actual lapse rate above the level of free convection (LFC) is lower than the moist-adiabatic lapse rate. In this concept, conditional instability in a relatively dry environment can be caused by a moist air parcel rising from below the LFC. This is the underlying mechanism for the Convective Available Potential Energy (CAPE), the Lifted Index (LI), Showalter Index (SHOW), and Deep Convective Index (DCI). Finally, a state of an unsaturated column of air where the equivalent potential temperature θ_e decreases with height is called potential instability (Emanuel, 1994). If such a column is lifted entirely until complete saturation, it will become unstable regardless of its initial stratification. This kind of instability is described by the KO Index (KO), the Potential Instability Index (PII), and the Wet Microburst Index or Delta-θ_e Index (DTeI).

Other indices are a combination of the three different concepts: The Total Totals (TT), K-Index (K), S-Index (S), or the Jefferson Index (JEFF). Kinematic information in terms of wind shear, wind speed, or wind direction at different levels is incorporated in the Severe Weather Threat Index (SWEAT), the SWISS Index (SWISS), and the Bulk-Richardson Number (RIB). Information about hail size if expected can be derived from the Wet Bulb Zero height (WBZ). All indices used in this study and the respective equations are listed in Table A1 in the Appendix. A detailed description of most of the indices can also be found in the study of Haklander and Van Delden (2003).

3 Methods

3.1 Probability distributions

A possibility to assess the skill of the various indices and to estimate the range of values for the different categories of days is to compare the probability distribution functions. A predictor is most efficient if the probability distributions of the particular categories are clearly separated, i.e. if the overlap between the distributions is small. To obtain an overview of the indices' efficiency for direct comparison, the distribution functions were also characterized by three different percentiles: The median and the 15.9% and 84.1% percentiles whose distance is twice the standard deviation σ in case of a normal distribution, i.e. approximately 68% of all values lie in between.

3.2 Probability of thunderstorms

The probability of thunderstorms occurring during a 12-h period simply can be determined by counting all days with and without thunderstorm observations for a specific index value. For this, an ordered list from high to low index values with the corresponding thunderstorm observations has to be created at first. From this list, the 1st to the k-th day is taken to calculate a mean index value with a standard deviation and the ratio of thunderstorm days to the total number k of days in this interval. By considering the elements "correct forecast" a and "false alarm" b of the contingency table (see Table 2 and next section) only, this ratio may be regarded the thunderstorm probability. The fixed interval with k days is shifted for about one day and so on, until the end of the list is reached, giving $(n-k+1)$ thunderstorm probabilities and associated mean values for a sample size n. The interval size k for the SY data (n=3.058) was set to 200, for the RA3 data (n=518) to 20, and for the insurance data (n=3.294) to 100. The different interval sizes are due to the varying number of thunderstorm days in the data sets.

To estimate the thunderstorm probability for a comparison of all indices, mean values and standard deviations for

Table 2. Contingency table for a dichotomous categorical verification of forecasts.

		Observation	
		YES	NO
Forecast	YES	a correct event forecast	b false alarms
	NO	c surprise events	d non-events

discrete probabilities P were quantified. They cannot be derived directly from the graphs because of a lacking unambiguous relation between probabilities and index values, i.e. the same probability may be obtained at several thresholds. Hence, the method of Haklander and Van Delden (2003) was applied who estimated the thresholds for a specific thunderstorm probability P by determining the least index value λ_{high} (with standard deviation σ_{high}), where the probability reaches P, and the least index value beyond, λ_{low} (with standard deviation σ_{low}), where the probability stays $\geq P$. Finally, the mean value λ and the corresponding standard deviation σ are given by:

$$\lambda(P) = 0.5\left[(\lambda_{\text{high}} + \sigma_{\text{high}}) + (\lambda_{\text{low}} - \sigma_{\text{low}})\right], \quad (1)$$

$$\sigma(P) = 0.5\left[(\lambda_{\text{high}} + \sigma_{\text{high}}) - (\lambda_{\text{low}} - \sigma_{\text{low}})\right].$$

3.3 Categorical verification and skill scores

An objective method to assess the prediction skill of the various indices and to find appropriate thresholds is provided by the categorical verification. The data sets are entered into a 2×2 contingency table (Table 2), with four elements a to d based on whether an event was observed (YES/NO) and predicted (YES/NO). These methods are widely used for the verification of weather forecasts (Wilks, 1995), but in many studies also for the evaluation of thunderstorm indices, e.g. by Doswell et al. (1990), Lee and Passner (1993), or Huntrieser et al. (1997). Whereas the observations (predictands) match a type of binary scheme – like thunderstorm occurrence vs. no thunderstorm occurrence –, the various parameters as predictors may assume a wide range of values. By defining an appropriate threshold, also the index values are separated into two parts. For indices that associate higher values with a higher thunderstorm potential (e.g. CAPE or DCI), a thunderstorm day is only predicted, when the threshold is reached or exceeded. For indices that associate lower values with a higher thunderstorm potential (e.g. LI or KO) it is just reverse.

The problem then is to find an appropriate threshold, where the correct event forecasts a are maximized and both false alarms b and surprise events c are minimized. Figure 3 (top) illustrates the changes of the numbers in the categories a to d as a function of the Lifted Index LI_{100} (vertical profile averaged over the lowest 100 hPa). For solving this optimization problem, several skill measures were introduced in the past. The index value, at which an appropriate skill score reaches its maximum, is then assigned the optimal threshold for the distinction of thunderstorm and non-thunderstorm days. For the evaluation of thunderstorm indices, both the Heidke Skill Score HSS (Heidke, 1926) and the True Skill Statistic TSS (Hanssen and Kuipers, 1965) are used frequently (see Appendix B for a description of the scores). Both skill scores receive a value of 1 for a perfect forecast, 0 for a totally random forecast, and a negative value for a worst forecast.

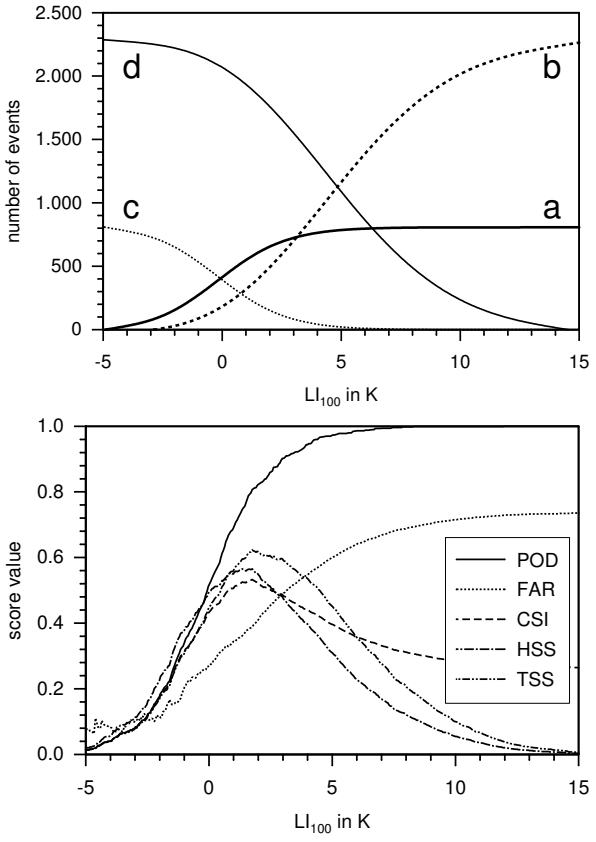

Fig. 3. Number of events for the different elements of the contingency table according to Table 2 (top) and skill scores (bottom) as a function of the Lifted Index LI_{100}, verified using the SYNOP data.

An example of the behavior of several scores as a function of the LI_{100} is given in Fig. 3 (bottom). As the threshold increases, both the Probability of Detection POD and the False Alarm Rate FAR increase more or less monotonically as the correct forecasts a and the false alarms b increase at the same time. In contrast to this, the curves for HSS, TSS, and the Critical Success Index CSI exhibit a distinct maximum for slightly different values of LI_{100}. Hence, the assigned optimal threshold is subject to the applied skill score.

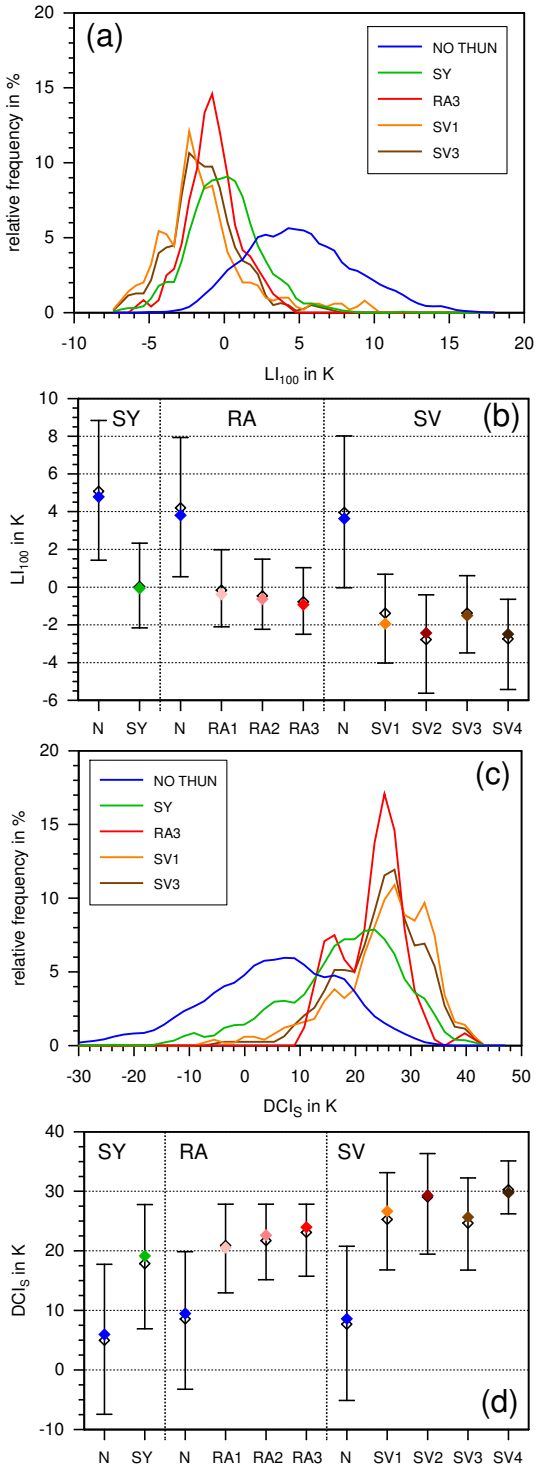

Fig. 4. Relative frequency distribution and range of the values between the 15.9% and 84.1% percentiles for the Lifted Index LI_{100} (**a** and **b**) and the Deep Convective Index DCI_S (**c** and **d**). Blank squares in (b) and (d) indicate the mean values, filled squares the median; the percentiles correspond to twice the distance of the standard deviation for a normal distribution.

The optimal thresholds according to the TSS frequently exhibit a lower thunderstorm potential of the atmosphere than the values for HSS. In most cases, the FAR for the maximum TSS exceeds that of HSS. Following Doswell et al. (1990), who demonstrated that for very rare events (<1%) TSS approaches the Probability of Detection POD that only accounts for the observed events, the HSS was used here for the assessment of the prediction skill of the indices.

4 Assessment of the convective indices

4.1 Probability distributions and mean values

First, the skills of the various indices are assessed by means of probability distribution functions and related percentiles as described in Sect. 3.1. Figures 4a and b show the probability distributions of the Lifted Index LI_{100} and the Deep Convective Index DCI_S (based on the surface Lifted Index LI_S) for non-thunderstorm and thunderstorm days according to SYNOP data (SY), severe thunderstorm days according to radar data (RA3), and days with hail damage (SV1) and storm/flood damage (SV3). It should be noted that the area below the distributions corresponds to 100%. For both indices, the probability distributions for the SYNOP data overlap in a relatively small area only. For LI_{100}, the size of the area that is not jointly covered reaches 62%, for DCI_S 45%. When comparing thunderstorm days and days with severe thunderstorms, e.g. SY with SV1 data, the non-overlapping area is 37% for LI_{100} and 35% for DCI_S.

Almost the same characteristics are obtained by the percentiles of the particular categories of predictands (Figs. 4b and d). Again, the LI_{100} yields the clearest distinction of all categories. For example, the range of values captured by the lower and upper percentiles (i.e. 68% of all days) for days with hail and storm/flood damage lies completely outside the range of non-thunderstorm days according to the SYNOP data. Furthermore, days with widespread and/or severe thunderstorms are also indicated by smaller (for LI_{100}) or higher (for DCI_S) index values – at least on the average for various events. The range of values between the percentiles defined is shown in Fig. 5 for all convective indices used in this study. Most of the indices separate the different days quite well. The worst indices according to the distribution functions are $SWISS_{12}$, HI, and SWEAT.

An estimation of expected hail size, if applicable, is given by the WBZ height. According to the SV1 (and SV3) loss data, the potential for large hail is highest for WBZ heights between 2500 and 3500 m. Below as well as above this range the hail size rapidly diminishes, resulting in an unambiguous relation between thunderstorm potential and WBZ height. Since only in a few cases the WBZ exceeds a height of 3800 m on days without thunderstorms, the upper range can be neglected in the distributions function.

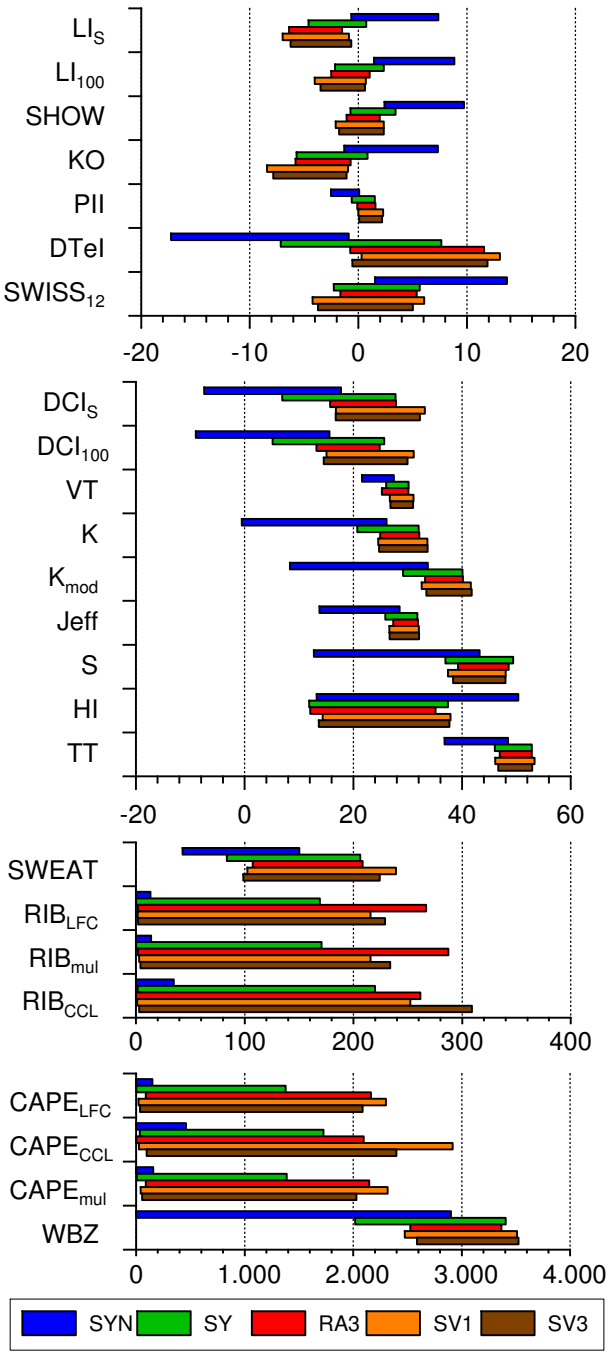

Fig. 5. Range of the index values between the 15.9% and 84.1% percentiles of the distribution function evaluated for days without thunderstorms (SYN), with all kinds of thunderstorms (SY), with severe thunderstorms (RA3), hail damage (SV1), and storm/flood (SV3) damage.

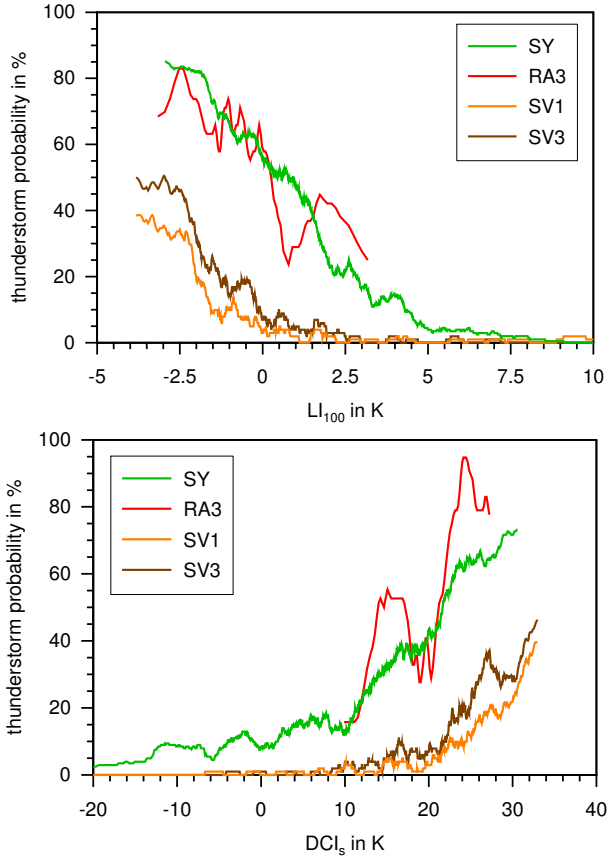

Fig. 6. Probability of thunderstorms as a function of the Lifted Index LI_{100} (top) and the Deep Convective Index DCI_S (bottom) for days of different categories.

4.2 Probability of thunderstorms

Now, the thunderstorm probabilities are derived as a function of the index values λ, as described in Sect. 3.2. By way of example, Fig. 6 shows the probabilities for the Lifted Index LI_{100} and the Deep Convective Index DCI_S. As expected, the probability of all kinds of thunderstorm days increases with decreasing LI_{100} and vice versa for DCI_S. Highest probabilities of up to 80% are reached when distinguishing between thunderstorm and non-thunderstorm days (SY) only. For example, for a mean LI_{100} of -3.0 K, a highest probability of 85% was reached, i.e. 170 of the 200 days in this interval were thunderstorm days (Fig. 6a). When considering the insurance data (SV1 and SV3), however, the computed probabilities of damage due to hail or storm/flood are significantly lower than for the other data sets. Considering a mean value of -3 K for LI_{100}, the probability of damage by storm/flood is 49.5%, by hail 34.7% (Fig. 6a). This is mainly due to the fact that it is only distinguished between days with and without damage, regardless of whether thunderstorms occurred on that day or not. The highest probabilities are reached for the decision between ordinary and severe thunderstorms according to the radar data. Probabilities above 90% are reached, whereas the curves also show a strong variability. This is mainly caused by the fact that

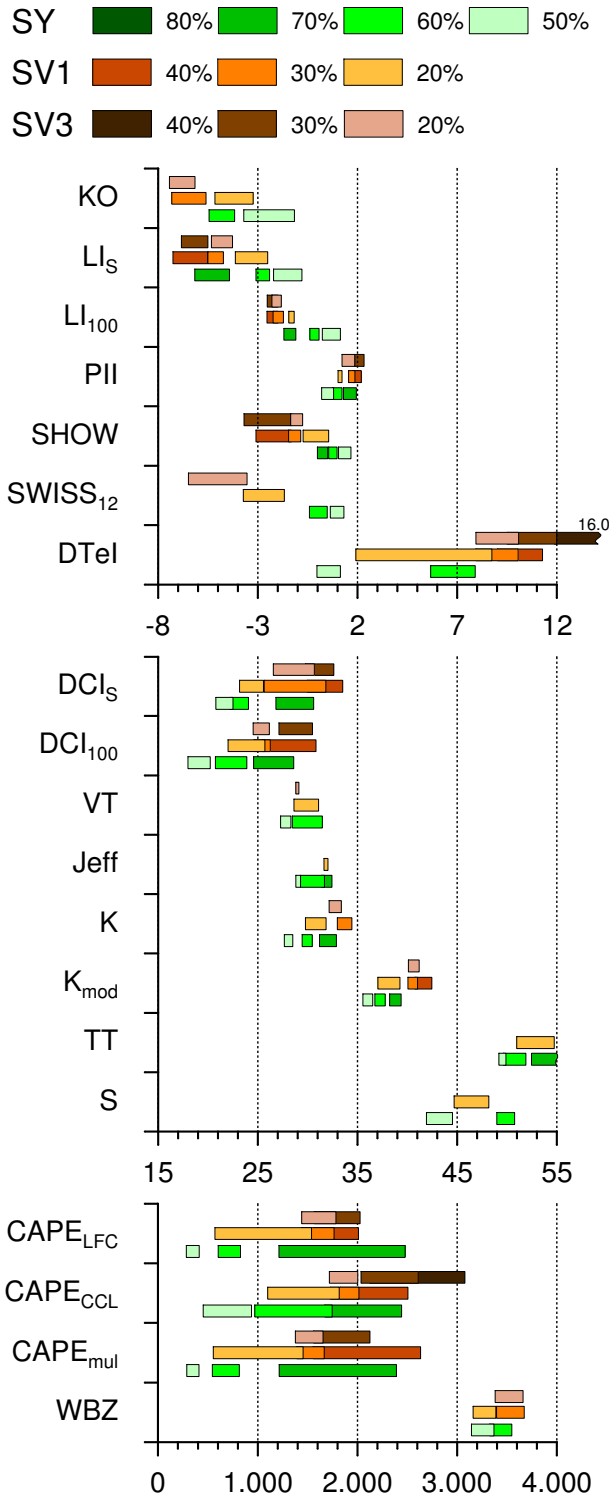

Fig. 7. Probability of thunderstorms and related thresholds (bars are drawn as $\lambda(P)\pm\sigma(P)$) for different convective indices according to days with thunderstorms (SY), hail damage (SV1), and storm/flood damage (SV3). Indices with a maximum probability of less than 60% for the SYNOP data are not shown.

Table 3. Thunderstorm indices with optimal thresholds and different skill scores for the thunderstorm prediction verified against SYNOP data between 1986 and 2003. The table shows the 12 indices with the highest HSS.

Index	Threshold	HSS	TSS	POD	CSI	FAR
SY: thunderstorm vs. non-thunderstorm days						
LI_{100}	$\leq$1.76 K	0.57	0.62	0.81	0.53	0.39
SHOW	$\leq$2.51	0.55	0.58	0.75	0.51	0.38
LI_S	≤-0.22 K	0.49	0.54	0.73	0.47	0.43
K_{mod}	$\geq$33.9 K	0.48	0.49	0.65	0.45	0.40
PII	≥-0.17 K km^{-1}	0.47	0.52	0.72	0.46	0.44
TT	$\geq$48.1 K	0.47	0.49	0.67	0.45	0.42
Jeff	$\geq$27.9 K	0.46	0.50	0.70	0.45	0.44
$CAPE_{mul}$	$\geq$159 J kg^{-1}	0.46	0.47	0.63	0.44	0.41
$CAPE_{CCL}$	$\geq$380 J kg^{-1}	0.46	0.48	0.65	0.44	0.42
K	$\geq$26.1 K	0.46	0.47	0.63	0.44	0.42
KO	≤-0.31 K	0.46	0.50	0.72	0.45	0.45
VT	$\geq$26.9 K	0.46	0.50	0.71	0.45	0.45

severe thunderstorms in the radar data are comparably rare (60 days) due to the comparatively small sample size. Consequently, the results should be interpreted with care.

Discrete thunderstorm probabilities P as a function of the thresholds, as described in Sect. 3.2, are shown in Fig. 7 and additionally listed in Table C1 in the Appendix for the SYNOP data (SV1) and the insurance data (SV1 and SV3). The results for the radar data were omitted because of their lacking of representativeness due to the small sample size of events that is too low for the probability analysis. The highest probabilities are reached when separating between thunderstorm and non-thunderstorm days, regardless of their intensities (SY). Two indices, namely, LI_{100} and K_{mod}, reach a thunderstorm probability of more than 80%; thirteen of the 19 indices reach probabilities of more than 70%. Note that the bar's length in the figure is proportional to the standard deviation $\sigma(P)$ that is determined by the distance between λ_{high} and λ_{low} (see Eq. 1). As expected, the probabilities of the occurrence of damage are significantly lower than those of the occurrence of all kinds of thunderstorms. Only 15 of the 19 indices reach a probability of more than 30% for hail; for storm/flood events, the number even decreases to 10.

4.3 Prediction skills and appropriate thresholds

An objective method to assess the prediction skill of the indices is provided by the categorical verification as described in Sect. 3.3. Both the optimal thresholds and the ranking of the indices regarding their prediction skill are determined by the maximum of the Heidke Skill Score HSS. The results of the categorical verification are listed in Table 3 for the SYNOP data, in Table 4 for the radar data, and in Tables 5 and 6 for the insurance data.

Although the ranking of the indices depends on the observed thunderstorm characteristics and, hence, differs from

Table 4. Same as Table 3, but for radar data between 1998 and 2001.

Index	Threshold	HSS	TSS	POD	CSI	FAR
(a) RA1: thunderstorm vs. non-thunderstorm days						
K_{mod}	$\geq$36.3 K	0.50	0.48	0.57	0.42	0.39
SHOW	$\leq$2.10	0.48	0.55	0.73	0.43	0.49
LI_{100}	$\leq$0.28 K	0.47	0.49	0.62	0.41	0.45
K	$\geq$28.1 K	0.46	0.46	0.56	0.40	0.43
DCI_{100}	$\geq$16.2 K	0.42	0.46	0.62	0.38	0.51
$CAPE_{mul}$	$\geq$520 J kg^{-1}	0.41	0.45	0.62	0.37	0.52
DCI_S	$\geq$18.7 K	0.39	0.45	0.64	0.36	0.54
Jeff	$\geq$29.2 K	0.39	0.41	0.56	0.35	0.51
$CAPE_{LFC}$	$\geq$520 J kg^{-1}	0.39	0.43	0.60	0.36	0.53
PII	≥-0.15 K km^{-1}	0.39	0.49	0.74	0.37	0.57
LI_S	≤-2.42 K	0.38	0.42	0.58	0.35	0.53
DTeI	$\geq$1.34 K	0.37	0.42	0.60	0.34	0.55
(b) RA2: widespread vs. isolated thunderstorm days						
DCI_{100}	$\geq$12.2 K	0.41	0.39	0.87	0.74	0.17
LI_S	≤-1.32 K	0.38	0.40	0.81	0.71	0.16
S	$\geq$41.8 K	0.37	0.39	0.78	0.68	0.15
DCI_S	$\geq$14.1 K	0.36	0.34	0.88	0.74	0.18
WBZ	$\geq$2458 m	0.33	0.31	0.86	0.72	0.19
SHOW	$\leq$2.10	0.33	0.33	0.81	0.69	0.18
LI_{100}	$\leq$0.63 K	0.33	0.35	0.77	0.67	0.17
KO	$\leq$0.08 K	0.31	0.30	0.85	0.71	0.19
K_{mod}	$\geq$33.5 K	0.31	0.31	0.83	0.70	0.18
K	$\geq$26.3 K	0.31	0.33	0.78	0.67	0.17
DTeI	$\geq$0.65 K	0.31	0.35	0.73	0.64	0.16
$CAPE_{LFC}$	$\geq$329 J kg^{-1}	0.30	0.32	0.77	0.66	0.18
(c) RA3: severe vs. non-severe thunderstorm days						
DTeI	$\geq$0.93 K	0.43	0.43	0.83	0.60	0.32
DCI_S	$\geq$22.3 K	0.42	0.42	0.65	0.53	0.25
SHOW	$\leq$0.83	0.40	0.40	0.73	0.56	0.30
LI_S	≤-2.42 K	0.38	0.38	0.77	0.56	0.32
LI_{100}	$\leq$0.33 K	0.36	0.36	0.80	0.57	0.34
PII	$\geq$0.03 K km^{-1}	0.36	0.36	0.80	0.57	0.34
$CAPE_{LFC}$	$\geq$732 J kg^{-1}	0.35	0.35	0.67	0.51	0.31
KO	≤-1.00 K	0.35	0.34	0.83	0.57	0.36
DCI_{100}	$\geq$12.9 K	0.34	0.34	0.90	0.59	0.37
$CAPE_{mul}$	$\geq$922 J kg^{-1}	0.34	0.34	0.57	0.47	0.28
VT	$\geq$27.0 K	0.33	0.33	0.75	0.54	0.35
WBZ	$\geq$2507 m	0.33	0.32	0.88	0.58	0.38

table to table, the highest skill scores are generally achieved with the traditional Lifted Index LI_{100}, the Showalter Index SHOW, and the modified K-Index K_{mod}. The most appropriate indices for the prediction of severe thunderstorms in descending order are: LI_{100}, DCI_S, DCI_{100}, PII, DTeI, $CAPE_{CCL}$, SHOW, and $CAPE_{mul}$. In contrast, lowest skill scores are reached in general by the Humidity Index HI, the Bulk-Richardson number RIB, the Jefferson Index JEFF, the S-Index, and the Boyden Index BOYD (not listed in the tables).

For the prediction of a thunderstorm day (SY in Table 3 and RA1 in Table 4a), the Lifted Index LI_{100}, the Showalter Index SHOW, and the modified K-Index K_{mod}, perform

Table 5. Same as Table 3, but for hail days according to the insurance data between 1986 and 2003.

Index	Threshold	HSS	TSS	POD	CSI	FAR
(a) SV1: hail vs. non-hail days						
LI_{100}	≤-2.07 K	0.39	0.44	0.48	0.27	0.62
$CAPE_{CCL}$	$\geq$1763 J kg^{-1}	0.36	0.41	0.45	0.24	0.66
$CAPE_{mul}$	$\geq$1474 J kg^{-1}	0.36	0.39	0.43	0.24	0.65
$CAPE_{LFC}$	$\geq$1474 J kg^{-1}	0.35	0.39	0.42	0.23	0.66
DTeI	$\geq$10.3 K	0.35	0.34	0.36	0.23	0.62
DCI_{100}	$\geq$25.7 K	0.34	0.40	0.44	0.23	0.68
LI_S	≤-4.19 K	0.34	0.47	0.52	0.23	0.71
DCI_S	$\geq$29.2 K	0.33	0.33	0.36	0.22	0.63
PII	$\geq$1.90 K km^{-1}	0.33	0.30	0.32	0.22	0.59
SHOW	≤-0.85	0.31	0.33	0.36	0.20	0.69
K_{mod}	$\geq$38.9 K	0.27	0.38	0.44	0.18	0.77
KO	≤-6.32 K	0.26	0.29	0.33	0.17	0.74
(b) SV2: widespread vs. isolated hail days						
PII	$\geq$2.71 K km^{-1}	0.31	0.25	0.29	0.24	0.40
SWEAT	$\geq$287	0.31	0.23	0.25	0.23	0.29
DCI_S	$\geq$35.4 K	0.30	0.22	0.24	0.22	0.29
SHOW	≤-3.97	0.30	0.22	0.24	0.22	0.29
LI_{100}	≤-4.21 K	0.29	0.27	0.38	0.25	0.58
$CAPE_{CCL}$	$\geq$2431 J kg^{-1}	0.27	0.30	0.48	0.26	0.64
DCI_{100}	$\geq$30.1 K	0.25	0.27	0.43	0.24	0.64
LI_S	≤-5.67 K	0.21	0.27	0.52	0.23	0.70
K_{mod}	$\geq$42.0 K	0.21	0.19	0.29	0.19	0.63
DTeI	$\geq$11.8 K	0.20	0.23	0.43	0.22	0.69
WBZ	$\geq$3485 m	0.20	0.22	0.38	0.21	0.68
K	$\geq$31.1 K	0.18	0.26	0.57	0.23	0.73

best. The differences between both tables concerning the ranking of the indices and their optimal thresholds are mainly due to the different criteria that were used for the definition of a thunderstorm day. For the radar data, the criterion is more stringent, as already reflected by the lower number of radar-defined thunderstorm days (Fig. 2). Considering the widespread vs. isolated thunderstorm day scheme (RA2 in Table 4b), the ranking of the indices is slightly different with highest scores for the Deep Convective Index DCI_{100}, the surface Lifted Index LI_S, and the S-Index. The defined thresholds indicate a higher thunderstorm potential compared to those discussed above.

For the prediction of severe thunderstorms that are associated with damage due to hail (SV1 and SV2 in Table 5) or storm/flood (SV3 and SV4 in Table 6), the highest skill scores are obtained again for the Lifted Index LI_{100} and the two versions of the Deep Convective Index, DCI_S and DCI_{100}, but also for the CAPE in association with very high thresholds (e.g. $\geq$1763 J kg^{-1} for $CAPE_{CCL}$). The listed thresholds of the indices for hail days reveal the highest convective potential compared to all other data sets. Also the thresholds for widespread damage days (SV2 and SV4) generally exhibit a higher thunderstorm potential compared to the other damage days (SV1 and SV3). Another interesting result is that indices that were designed for the prediction of

Table 6. Same as Table 3, but for storm/flood days according to the insurance data between 1986 and 2003.

Index	Threshold	HSS	TSS	POD	CSI	FAR
(a) SV3: storm/flood vs. none storm/flood days						
LI_{100}	≤ -1.96 K	0.42	0.39	0.42	0.29	0.52
DCI_{100}	≥ 25.3 K	0.40	0.39	0.42	0.28	0.55
DCI_S	≥ 25.4 K	0.39	0.45	0.51	0.27	0.63
$CAPE_{CCL}$	≥ 1763 J kg^{-1}	0.37	0.34	0.37	0.26	0.56
PII	≥ 1.05 K km^{-1}	0.36	0.45	0.52	0.26	0.67
$CAPE_{mul}$	≥ 806 J kg^{-1}	0.35	0.49	0.57	0.25	0.69
K_{mod}	≥ 38.0 K	0.35	0.47	0.55	0.25	0.69
SHOW	≤ 0.54	0.35	0.53	0.63	0.25	0.70
DTeI	≥ 4.34 K	0.35	0.48	0.56	0.25	0.69
$CAPE_{LFC}$	≥ 1495 J kg^{-1}	0.34	0.30	0.33	0.23	0.57
LI_S	≤ -4.12 K	0.34	0.37	0.42	0.24	0.65
KO	≤ -3.39 K	0.32	0.50	0.61	0.23	0.73
(b) SV4: widespread vs. isolated storm/flood days						
DCI_S	≥ 31.8 K	0.38	0.36	0.46	0.33	0.47
LI_{100}	≤ -3.24 K	0.34	0.33	0.44	0.30	0.50
DCI_{100}	≥ 25.7 K	0.34	0.46	0.74	0.35	0.61
KO	≤ -6.01 K	0.31	0.37	0.59	0.31	0.60
$CAPE_{mul}$	≥ 1400 J kg^{-1}	0.30	0.39	0.67	0.32	0.62
DTeI	≥ 7.20 K	0.30	0.41	0.72	0.32	0.63
$CAPE_{CCL}$	≥ 2486 J kg^{-1}	0.30	0.27	0.36	0.26	0.50
LI_S	≤ -4.63 K	0.30	0.38	0.64	0.31	0.62
PII	≥ 1.95 K km^{-1}	0.29	0.31	0.49	0.29	0.59
$CAPE_{LFC}$	≥ 1400 J kg^{-1}	0.29	0.37	0.64	0.31	0.63
SHOW	≤ -2.00	0.29	0.25	0.33	0.25	0.50
WBZ	≥ 3218 m	0.25	0.39	0.80	0.30	0.67

severe events, such as SWISS, SWEAT, or WBZ, exhibit no significantly high scores. Only the SWEAT is listed in second position for widespread hail events.

When ranking the indices based on the TSS instead of the HSS, the results and the optimal thresholds are almost the same for the SYNOP and the radar data (not shown). Major differences between both skill scores are observed for the insurance data only (SV1 to SV4 in Table 4). For all indices, the optimal thresholds determined by the TSS maxima indicate a lower thunderstorm potential of the atmosphere compared to that determined by the HSS maxima. The reason for the differences is that the TSS is more related to a high POD than to a low FAR, while HSS emphasizes a low FAR associated with a higher thunderstorm potential (see Fig. 3). Consequently, the FAR reaches high values of up to 0.8 for the TSS maxima. Only when the number of days with thunderstorms observed almost equals the number of days with predicted thunderstorms, are the scores almost the same for HSS and TSS, as it is the case for RA3 and, partly, RA2.

Regarding the layer that appears to be most relevant to the thunderstorm potential, a systematic relation cannot be found for both isolated and widespread thunderstorms. In case of severe thunderstorms, however, higher scores are obtained for indices that depend on the lowest layers. The two indices with the highest scores in the RA3 data directly depend on temperature and humidity near the surface. For the SV data, highest scores are reached by indices that either are derived from properties near the surface (e.g. CAPE or DCI_S) or that are averaged over the lowest 100 hPa (e.g. LI_{100} or DCI_{100}). The question, over which layer the temperature and humidity profile should be averaged to reach highest scores will be addressed in the next section.

4.4 Sensitivity of skill scores to changes of the lifted profile

A crucial issue for the indices that represent latent instability is their direct dependance on the properties of a particular air parcel that is assumed to be lifted from a certain level. Especially on high radiation days with a strongly superadiabatic stratification and a strong increase of the mixing ratio in the lowest layers, it is questionable whether the values near the surface in particular may be representative of the whole lifting process and for a larger region. On such days, the temperature of an air parcel lifted from the surface to a certain level exceeds that of a parcel lifted from any height between the surface and the condensation level. Hence, indices that are related to the lifting profile, like the CAPE or the LI, are very sensitive to the vertical profiles of temperature and humidity in the lowest layers.

Four convective indices were chosen to examine the sensitivities of the forecasting skill to an averaging of the initial profiles: the Lifted Index LI, the Deep Convective Index DCI, and two versions of CAPE, once determined by the LFC, $CAPE_{LFC}$, otherwise by the CCL, $CAPE_{CCL}$. All four indices are based on the properties of an air parcel that is assumed to be lifted either from the near-surface layer (CAPE, DCI) or from a layer determined by vertical mixing (LI). To test the sensitivities of the indices, the mixing ratio, temperature, and initial level, from where the parcels are assumed to be lifted, were modified by density-weighted averaging from the surface to a level between 10 and 100 hPa above.

The Heidke Skill Score HSS results for different initial values are shown in Fig. 8. As can be seen, a general standard for averaging that gives best results for all indices and all kinds of thunderstorm days cannot be derived. In case of the Lifted Index (Fig. 8a), the best results for SY, RA1, and SV are obtained by averaging the vertical profile over the lowest 100 hPa that corresponds to the traditional LI_{100}. For the prediction of severe thunderstorms only (RA3 and SV1–SV4), the score then increases with the vertical extension of the averaging layer. The lowest score is obtained for the surface-based Lifted Index, LI_S – again except for the two radar data sets.

Even though the Deep Convective Index DCI is directly based on the LI, the variability of the skill scores for different averaging levels is obviously higher than that for the LI. For the decision of thunderstorms vs. no thunderstorms as captured by both SY and RA1, the scores obtained for the various averaging levels of LI differ only marginally. For these categories, DCI_{100} scores best. However, when considering

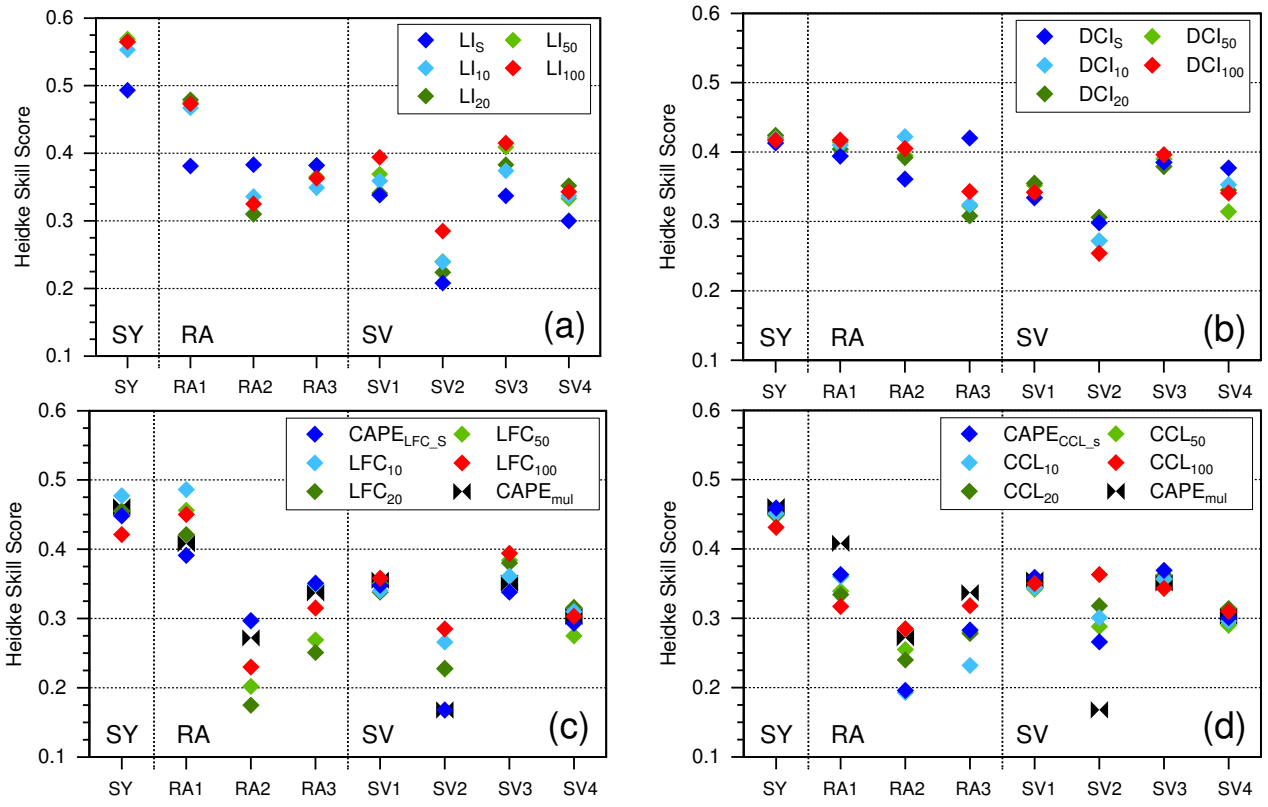

Fig. 8. Heidke Skill Score HSS for Lifted Index LI **(a)**, Deep Convective Index DCI **(b)**, CAPE from the LFC **(c)**, and CAPE from the CCL **(d)**. Initial temperature and mixing ratio of the assumed lifted air parcel are averaged over a certain depth (in hPa) above the surface indicated by the subscription; a subscribed s means no averaging.

the prediction of severe thunderstorms only, the results for RA3, SV2, and SV4 behave in an opposite way, with highest scores for DCI derived from near-surface values without averaging, DCI_S.

No general standard can be derived for the two versions of CAPE as displayed in Figs. 8c and d. In case of severe thunderstorms, best results are obtained for $CAPE_{LFC}$ when the profile is averaged over the lowest 100 hPa. Thus, high CAPE values resulting from a superadiabatic stratification of the layers directly above the surface are reduced. Note that the prediction skill of $CAPE_{CCL}$ for SY and SV – except for SV2 – is not very sensitive regarding the averaging layer. Even though the results for CAPE differ from case to case in this analysis, it is recommended to average the vertical profile over the lowest 100 hPa for the assumed lifted parcel – at least for the prediction of severe thunderstorms.

5 Discussion and conclusions

The aim of this study is to assess the skill of various convective or thunderstorm indices derived from radiosonde observations at 12:00 UTC for the prediction of thunderstorm occurrence. To obtain comprehensive information about both thunderstorm occurrence and intensity, different kinds of observation data sets were used. Data from SYNOP stations allow to determine days without and with thunderstorms for a long term, but with a low spatial coverage. Radar data exhibit additional information about thunderstorm properties in terms of maximum intensity, number of cells, and spatial extension of the convective area, but are available for some years only and cover a very large region of 120 km in radius. Loss data from a building insurance company facilitate the identification of days with extreme thunderstorms associated with hail or storm/flood damage. By bringing together the different data sets, the prediction skill and appropriate thresholds of the various convective indices for thunderstorms with different intensities are assessed comprehensively.

Frequency distributions of the index values and derived percentiles already revealed the efficiency of the indices and yielded an estimate regarding the range of values for the different thunderstorm categories. Most of the indices showed quite a good separation of the ranges of values according to the different categories of days, especially LI_{100}, LI_S, SHOW, DTeI, and PII.

The categorical verification divided the investigated data sets into a 2×2 contingency table to assess the prediction

skill of the different indices. By considering the elements of the table associated with an index-based thunderstorm forecast only, the probability of thunderstorms was derived. To decide on thunderstorm vs. non-thunderstorm days, the LI_{100} and K_{mod} reach a probability of more than 80%. Considering days with damage due to hail or storm/flood only, the probabilities obtained are significantly lower. For the prediction of hail days, probabilities of more than 40% are reached for the two versions of LI and DCI, for all three versions of CAPE, for SHOW, DTeI, PII, and K_{mod}. For the storm/flood days, the same probability is reached by $CAPE_{CCL}$ and DTeI only.

Using all the information of the contingency table, the Heidke Skill Score HSS was used to determine appropriate thresholds for a thunderstorm forecast and to rank the indices according their prediction skill. For the decision between thunderstorm and non-thunderstorm days, LI_{100} and SHOW score best with values of 0.57 and 0.55 for HSS. For the decision on severe vs. non-severe thunderstorms, an HSS value in excess of 0.4 is reached by DTeI, DCI_S, and SHOW only. Considering damage days, the index with the highest score differs from one data set to another. The six indices with the highest skill scores for the loss data sets are in descending order: LI_{100}, DCI_S, DCI_{100}, PII, DTeI, and $CAPE_{CCL}$. Most of the indices that scored best in terms of the maximum HSS also reached a high thunderstorm probability, despite the fact that these two parameters consider different elements of the contingency table.

The study of both the maximum thunderstorm probability and the skill scores reveals that the prediction of thunderstorm days with different characteristics cannot be achieved by a single convective index that fits best to the observations. The skill of the indices rather depends on the thunderstorm intensity and the objective of the investigation. That also applies to the problem of finding an appropriate threshold. When summarizing the results, the indices with the highest skills for thunderstorm prediction based on the 12:00 UTC sounding are LI_{100}, DCI_S, SHOW, DTeI, PII, and CAPE. In contrast to this, the prediction efficiency is found to be poorest for HI, RIB, S, TT, and BOYD.

When assessing the skill of the indices, it must be taken into account that several indices were designed for the prediction of a special kind of thunderstorm. The Boyden Index BOYD originally was designed to assess the thunderstorm probability during frontal passages in the UK. The Bulk-Richardson number RIB was created to estimate the thermodynamic and dynamic forcing of convective development that may help to separate single- or multicell thunderstorms from supercell thunderstorms. This explains the low skills of both indices in our study. Several other indices were designed to forecast severe thunderstorms, such as DCI, SWEAT, or the SWISS Index. However, when considering severe thunderstorms only, a high prediction skill was found for DCI, but not for the two other ones.

Regarding the theoretical concept underlying the different indices, latent instability (represented by LI, DCI, CAPE, SHOW) or potential instability (represented by KO, PII, DteI) are the most important conditions for the onset of convection. As regards the decision between thunderstorm and non-thunderstorm days according to the SYNOP data, also indices that combine the two concepts (TT, K, JEFF, S) reach high probabilities as well as high skill scores. In contrast to this, the combined indices exhibit poor benchmarks when considering days with severe thunderstorms only. Here, higher scores are obtained only by indices representing latent or potential instability and based on temperature and humidity values either from a near-surface level or averaged over the lowest 100 hPa. Indices considering additional dynamic information like the RIB or the SWISS Index exhibit significantly smaller skills for all types of thunderstorm days. It is interesting to note that several other indices are more suitable to predict thunderstorms than the KO Index that was designed and is still used by the DWD.

The study revealed that an index-based prediction of severe thunderstorms that are associated with hail or storm/flood damage is a big challenge. Skill scores for the prediction as well as the maximum probabilities of severe thunderstorms are quite low compared to the prediction of thunderstorm vs. non-thunderstorm days. It must be kept in mind, however, that vulnerability plays a decisive role in the loss data, since, especially in sparsely populated regions, not each severe storm causes the minimum number of ten claims that define a damage day in this study. Besides, hail events without a sufficient size of the hailstones do not cause any damage of buildings and, thus, are not recorded by the data. In the categorical verification scheme, this leads to an overprediction with high FAR and low skill scores in case of severe thunderstorms. Although the scores reached are significantly lower than for all other kinds of thunderstorms, any information about a possible occurrence of severe thunderstorms is very valuable and may help to prevent or reduce damage.

A limiting factor of the sounding-based prediction concept arises from the assumption that the atmospheric conditions are homogeneous for the next 12 h after the sounding. When air masses with different properties are advected, for example, in conjunction with a frontal passage, the sounding cannot be representative of the whole area and the whole period. Another constraint of this method results from the neglect of forced ascent acting on several scales. Local-scale flow effects and evolving convergence zones over topographically structured terrain may often trigger the onset of convection. Synoptic-scale lifting associated with differential vorticity advection, temperature advection, and/or diabatic heat transfer is an important trigger especially for scattered and severe thunderstorms. Besides, the horizontal moisture convergence particularly near the earth surface is decisive for the life cycle of thunderstorms. It it obvious, that all these effects cannot be taken into account by analyzing single radiosonde observations. To overcome these constraints, the convective indices could be combined with appropriate parameters from

a numerical model. In a recent study, Van Zomeren and van Delden (2007) combined different versions of the Lifted Index with the vertically integrated moisture flux derived from weather analysis data from the European Centre for Medium-range Weather Forecasts (ECMWF). They showed in their study that the prediction of severe thunderstorm days and tornado events over Europe improves significantly when employing the moisture flux in the prediction scheme.

State-of-the-art weather forecast tools include several models that consider different spatial and temporal scales of atmospheric disturbances. For the prediction of deep convection, nowcasting models with a lead time of up to 2 h provide information of thunderstorm intensity and expected cell tracks that are extrapolated from observation data of remote sensing systems (e.g. KONRAD of DWD). However, the life cycle of thunderstorms or new cell formation cannot be forecast. Regional high-resolution weather forecast models that explicitly resolve deep convection, have typical lead times around 6–12 h. The quality of convection forecast is high if real-time precipitation data, e.g. from radar data, are assimilated into the system. Weather prediction for lead times of more than 12 h, i.e. short–range (up to 3 days) and medium–range (up to 10 days) weather forecast, is well covered by several models (e.g. COSMO-LME and GME of DWD). However, up to now these models still have problems to predict the life cycle of deep convective clouds due to inadequate convection parameterization schemes, error in the initial conditions, or turbulent closure problem (Hense et al., 2003).

Convective indices derived from radiosonde observations may give additional information about atmospheric stability and conditions for lead times between 1 and 12 h. This time range is of great importance to many users and for issuing warnings of local convective extreme events associated with heavy precipitation, hail, severe downdrafts, or even tornados. In the hierarchy of weather prediction models, this time range is covered by both, nowcasting tools and very short-range forecasts. Hence, convective indices could be employed in both systems. They could be included in nowcasting tools to estimate the thunderstorm probability and/or thunderstorm intensity. And they are still a helpful measure to better estimate the convective situation for weather forecaster since they represent a real state of the atmosphere.

Appendix A

Definition of convective indices

A summary of all convective parameters and indices used in this study is listed in Table A1.

Appendix B

Skill scores

Based on the definition of the contingency table (Table 2), the scores for categorical weather elements used in this study are defined as follows:

Probability of Detection POD

$$\mathrm{POD} = \frac{a}{a+c} \tag{B1}$$

The range of POD is from 0 to 1, with 1 for a perfect forecast; POD increases with overforecasting events; it includes no false alarms.

False Alarm Rate FAR

$$\mathrm{FAR} = \frac{b}{a+b} \tag{B2}$$

The range of FAR is 0 to 1, with 0 for a perfect forecast; FAR increases with underforecasting events.

Critical Success Index CSI

$$\mathrm{CSI} = \frac{a}{a+b+c} \tag{B3}$$

The range of CSI is 0 to 1, with 1 for a perfect forecast; CIS includes both false alarms and surprise events; however, events with different frequencies cannot be compared.

Frequency Bias FBI

$$\mathrm{FBI} = \frac{a+b}{a+c} \tag{B4}$$

The range of FBI is between 0 and ∞, with 1 for a perfect forecast. FBI is the ratio between all events forecasted and all events observed; the FBI does not include the non-event forecasts d.

Heidke Skill Score HSS

$$\mathrm{HSS} = \frac{a+d-R}{a+b+c+d-R} \tag{B5}$$

$$\text{with } R = \frac{(a+b)\times(a+c)+(c+d)\times(b+d)}{a+b+c+d}$$

The range of HSS is from $-\infty$ to 1, with 1 for a perfect forecast; R is the chance. Surprise events and false alarms are included as well as the effect of a reference forecast. The HSS is based on the hit rate as the basic accuracy measure and follows the form of a generic skill score (Wilks, 1995).

True Skill Statistic (Hanssen-Kuiper discriminant): TSS

$$\mathrm{TSS} = \frac{a\times d - b\times c}{(a+c)\times(b+d)} \tag{B6}$$

Table A1. Summary of convective parameters and indices: T and Td are the temperature and dewpoint temperature (°C), θ_e and θ_w are the equivalent potential and wetbulb potential temperature (K), Z is the geopotential height (pgm), and R_d is the gas constant for dry air (J/kg K). The subscript indicates a certain constant pressure level, the character s the surface; an arrow in the subscription indicates the lifting of an air parcel (e.g. $T'_{x\rightarrow y}$ indicates the temperature T of a parcel at the y-level, which was initially lifted dry adiabatically from the x-level to its condensation level and moist adiabatically thereafter).

Index name	Equation	Reference	Comment
A: Indices describing conditional instability			
Vertical Totals	$\mathrm{VT}=T_{850}-T_{500}$	Miller (1972)	
Boyden Index	$\mathrm{BOYD}=0.1(Z_{700}-Z_{1000})-T_{700}-200$	Boyden (1963)	
B: Indices describing latent instability			
Lifted Index	$\mathrm{LI}_{100}=T_{500}-T'_{i\rightarrow 500}$	Galway (1956)	i: p, T and Td averaged over the lowest 100 hPa
Deep Convective Index	$\mathrm{DCI}=(T+Td)_{850}-\mathrm{LI_S}$	Barlow (1993)	$\mathrm{LI_S}$: Surface Lifted Index (no mixing)
Showalter Index	$\mathrm{SI}=T_{500}-T'_{850\rightarrow 500}$	Showalter (1953)	
Convective available potential energy	$\mathrm{CAPE_{LFC}}=R_d\int_{LFC}^{EL}(T'_v-T_v)\,\mathrm{d}\ln p$	Moncrieff and Miller (1976)	T'_v is the virtual temperature of an air parcel lifted from the surface to the level of free convection (LFC) up to the equilibrium level (EL)
	$\mathrm{CAPE_{CCL}}=R_d\int_{CCL}^{EL}(T'_v-T_v)\,\mathrm{d}\ln p$		same as above, but the air parcel is lifted moist adiabatically from the cumulus condensation level (CCL)
	$\mathrm{CAPE_{mul}}=R_d\int_{i}^{EL}(T'^{*}_v-T_v)\,\mathrm{d}\ln p$		defined for a parcel with T, Td, and p at a level where θ_e reaches its highest value in the lowest 250 hPa
C: Indices describing potential instability			
KO Index	$\mathrm{KO}=0.5(\theta_{e500}+\theta_{e700})-0.5(\theta_{e850}+\theta_{e1000})$	Andersson (1989)	since the local pressure is often below 1000 hPa, we used 950 hPa instead;
Delta-θ_e	$\mathrm{DTeI}=\theta_{e\mathrm{S}}-\theta_{e300}$	Atkins and Wakimoto (1991)	designed to assess the potential for wet microbursts
Potential Instability Index	$\mathrm{PII}=(\theta_{e925}-\theta_{e500})/(Z_{500}-Z_{925})$	Van Delden (2001)	
D: Combination of A–C			
Total Totals	$\mathrm{TT}=(T+Td)_{850}-2T_{500}$	Miller (1972)	
K-Index	$\mathrm{K}=(T_{850}-T_{500})+Td_{850}-(T-Td)_{700}$	George (1960)	developed for forecasting air mass thunderstorms
modified K-Index	$\mathrm{K_{mod}}=(T^{*}-T_{500})+Td^{*}-(T-Td)_{700}$	Charba (1977)	T^{*} and Td^{*} calculated by averaging between the surface and the 850 hPa level
Humidity Index	$\mathrm{HI}=(T-Td)_{850}+(T-Td)_{700}+(T-Td)_{500}$	Litynska et al. (1976)	
Jefferson Index	$\mathrm{JEFF}=1.6\theta_{w850}-T_{500}-0.5(T_{700}-Td_{500})-8$	Jefferson (1963)	
S-Index	$\mathrm{S}=\mathrm{TT}-(T-Td)_{700}-\Phi$		where Φ=0 for VT$\geq$25, Φ=2 for 25>VT>22, and Φ=6 for VT$\leq$22; designed to indicate the thunderstorm potential from April to September
Wet Bulb Zero Height	WBZ	Miller (1972)	hight where the wetbulb profile transitions from a positive to a negative temperature; indicates the potential for hail
E: Indices considering kinematic properties			
SWISS Index	$\mathrm{SWISS_{12}}=LI_{\mathrm{S}}-0.1\mathrm{WSh}_{0-3}+0.1(T-Td)_{650}$	Huntrieser et al. (1997)	where WSh_{0-3} is the wind sheer in the lowest 3 km agl; the index was designed for the 12:00 UTC sounding in Switzerland
Severe Weather Threat Index	$\mathrm{SWEAT}=12Td_{850}+20(\mathrm{TT}-49)+2f_{850}+f_{500}+125[\sin(d_{500}-d_{850})]+0.2$	Miller (1972)	where f and d are wind speed in knots and direction in (0–360°) on the indicated levels; the first two terms must be greater or equal than zero; the last term is set to zero if any of the conditions are not met: $130^\circ\leq d_{850}\leq 250^\circ$, $210^\circ\leq d_{500}\leq 310^\circ$, $d_{500}>d_{850}$, and both f_{850} and $f_{5000}\geq 15$ knots. SWEAT was designed for the prediction of severe thunderstorms

The range of TSS is the same as for HSS. It equally emphasizes the yes/no events. TSS approaches the POD for very rare events. It is formulated similarly to the HSS, except for the reference hit rate in the denominator being that for random forecasts that are constrained to be unbiased. Hence, if the Frequency Bias FBI is near unity, then TSS≈HSS.

Appendix C

Climatological means and standard deviations of the indices

Table C1. Mean and standard deviation of the indices for specific probabilities of thunderstorm occurrence, derived for thunderstorm days (SY), days with hail damage (SV1), and days with storm/flood damage (SV3).

Index	Thunderstorm days SY				Hail days SV1			Days with storm/flood SV3		
	80%	70%	60%	50%	40%	30%	20%	40%	30%	20%
VT (K)			30.0	27.8			29.9			29.0
			±1.50	±0.52			±1.23			±0.16
BOYD			99.0	97.9		100	98.9			
			±0.63	±0.35		±0.44	±0.58			
LI_S (K)		−5.28	−2.74	−1.48	−6.38	−5.11	−3.91		−6.13	−4.76
		±0.87	±0.34	±0.71	±0.87	±0.39	±0.25		±0.66	±0.53
LI_{100} (K)	−1.73	−1.38	−0.14	0.70	−2.26	−1.97	−1.31		−2.22	−2.02
	±0.38	±0.30	±0.23	±0.20	±0.28	±0.26	±0.14		±0.27	±0.24
DCI_S (K)		28.7	23.2	21.7	31.8	28.7	24.4		31.2	28.6
		±1.90	±0.93	±0.86	±1.76	±3.10	±1.19		±1.43	±2.06
DCI_{100} (K)		26.6	22.4	19.1	28.1	25.4	23.9		28.8	25.4
		±2.01	±1.58	±1.11	±2.69	±0.82	±1.85		±1.68	±0.81
$CAPE_{LFC}$ ($J\,kg^{-1}$)		1836	706	339	1786	1601	1055		1808	1625
		±632	±112	±64.9	±225	±163	±485		±229	±171
$CAPE_{CCL}$		2049	1350	685	2166	1872	1459	2613	2335	1872
		±385	±388	±242	±339	±142	±359	±478	±288	±142
$CAPE_{mul}$ ($J\,kg^{-1}$)		1793	670	340	2095	1529	1005		1855	1529
		±589	±135	±60.3	±539	±139	±450		±283	±139
SHOW		0.26	0.78	1.37	−2.20	−1.15	−0.09		−2.47	−1.01
		±0.26	±0.21	±0.31	±0.89	±0.31	±0.64		±1.18	±0.30
KO (K)			−4.79	−2.42		−6.45	−4.19			−6.75
			±0.64	±1.27		±0.86	±0.96			±0.64
DTeI (K)			6.80	0.57	10.15	9.00	5.33	13.62	10.72	9.04
			±1.12	±0.59	±1.13	±1.06	±3.41	±2.49	±1.24	±1.07
PII ($K\,km^{-1}$)		1.63	1.00	0.51	1.98	1.71	1.12		2.13	1.58
		±0.32	±0.24	±0.30	±0.21	±0.16	±0.10		±0.24	±0.32
TT (K)		53.8	50.8	49.6			52.9			
		±1.35	±1.12	±0.35			±1.88			
K (K)		32.1	30.0	28.1		33.7	30.8			32.8
		±0.84	±0.51	±0.42		±0.73	±1.04			±0.61
K_{mod} (K)	40.6	38.8	37.3	36.1	41.6	40.6	38.2			40.7
	±1.08	±0.58	±0.52	±0.49	±0.84	±0.48	±1.11			±0.54
Jeff (K)		32.0	30.5	29.1			31.8			
		±0.52	±1.20	±0.23			±0.21			
S (K)			49.9	43.2			46.4			
			±0.89	±1.30			±1.74			
WBZ (m)			3427	3247		3534	3274			3533
			±111	±113		±139	±114			±140
$SWISS_{12}$			0.05	0.99			-2.70			-4.97
			±0.45	±0.34			±1.03			±1.47

Acknowledgements. The author thanks the SV Sparkassen-Versicherung for the provision of loss data, the German weather service DWD for the provision of the SYNOP data, J. Handwerker for the provision of the radar data and the application of the cell tracking algorithm Trace-3D, and O. Weber and H. Konow for editing the SV data. Thanks to the comments of two reviewer.

Edited by: A. Mugnai
Reviewed by: A. van Delden and another referee

References

Andersson, T. A. M. J. C. N. S.: Thermodynamic indices for forecasting thunderstorms in southern Sweden, Meteorol. Mag., 116, 141–146, 1989.

Anquetin, S., Yates, E., Ducrocq, V., Samouillan, S., Chancibault, K., Davolio, S., Accadia, C., Casaioli, M., Mariani, S., Ficca, G., Gozzini, B., Pasi, F., Pasqui, M., Garcia, A., Martorell, M., Romero, R., and Chessa, P.: The 8 and 9 September 2002 flash flood event in France: a model intercomparison, Nat. Hazards Earth Syst. Sci., 5, 741–754, 2005, http://www.nat-hazards-earth-syst-sci.net/5/741/2005/.

Atkins, N. and Wakimoto, R.: Wet microburst activity over the Southeastern United States: Implications for forecasting, Wea. Forecasting, 6, 470–482, 1991.

Banta, R.: The role of mountain flows in making clouds, Meteorol. Monographs., Amer. Meteorol. Soc., 1990.

Barlow, W.: A new index for the prediction of deep convection, in: Preprints, 17th Conf. on Severe Local Storms, St. Louis, MO, pp. 129–132, Amer. Meteorol. Soc., 1993.

Barthlott, C., Corsmeier, U., Meißner, C., Braun, F., and Kottmeier, C.: The influence of mesoscale circulation systems on triggering convective cells over complex terrain, Atmos. Res., 81, 150–175, 2006.

Boyden, C.: A simple instability index for use as a synoptic parameter, Meteorol. Mag., 92, 198–210, 1963.

Charba, J.: Operational system for predicting thunderstorms two to six hours in advance, Tech. rep., NOAA. NWS TDL-64. [Techniques Development Laboratory, National Weather Service, Silver Spring, MD], 1977.

Doswell, C. I.: The distinction between large-scale and mesoscale contributions to severe convection: A case study example, Wea. Forecasting, 2, 3–16, 1987.

Doswell, C. I., Davies-Jones, R., and Keller, D.: On summary measures of skill in rare event forecasting based on contigency tables, Wea. Forecasting, 5, 576–585, 1990.

Emanuel, K.: Atmospheric Convection, Oxford Univ. Press, 1994.

Fuelberg, H. E. and Biggar, D. G.: The preconvective environment of summer thunderstorms over the Florida Panhandle, Wea. Forecasting, 9, 316–326, 1994.

Galway, J.: The lifted index as a predictor of latent instability, Bull. Amer. Meteorol. Soc., 37, 528–529, 1956.

George, J.: Weather Forecasting for Aeronautics, Academic Press, New York, 1960.

Haklander, A. J. and Van Delden, A.: Thunderstorm predictors and their forecast skill for the Netherlands, Atmos. Res., 67–68, 273–299, 2003.

Handwerker, J.: Cell Tracking with TRACE3D, – a New Algorithm, Atmos. Res., 61, 15–34, 2002.

Hanssen, A. and Kuipers, W.: On the relationship between the frequency of rain and various meteorological parameters, Meded. Verh., 81, 2–15, 1965.

Haurwitz, B.: Dynamic Meteorology, McGraw-Hill Book Co., 1941.

Heidke, P.: Berechnung des Erfolges und der Güte der Windstärkenvorhersage im Sturmwarnungsdienst, Geogr. Ann., 8, 310–349, 1926.

Hense, A., Adrian, G., Kottmeier, C., Simmer, C., and Wulfmeyer, V.: Priority Program of the German Research Foundation: Quantitative Precipitation Forecast (DFG-SPP 1167), 2003.

Houze, R. A.: Cloud Dynamics, no. 53 in International Geophysics Series, Academic Press, San Diego, 1993.

Huntrieser, H., Schiesser, H., Schmid, W., and Waldvogl, A.: Comparison of traditional and newly developed thunderstorm indices for Switzerland, Wea. Forecasting, 12, 108–125, 1997.

Jefferson, G.: A further development of the stability index, Meteorol. Mag., 92, 313–316, 1963.

Koßmann, M. and Fiedler, F.: Diurnal momentum budget analysis of thermally induced slope winds, Meteorol. Atmos. Phys., 75, 195–215, 2000.

Kunz, M. and Kottmeier, C.: Meteorologische Ereignisse mit großem Schadenpotenzial, in: KLARA – Klimawandel, Auswirkungen, Risiken, Anpassung, edited by: Stock, M., PIK Report Nr. 99, 161–172, 2005.

Lee, R. R. and Passner, J. E.: The development and verification of TIPS: An expert system to forecast thunderstorm occurrence, Wea. Forecasting, 8, 271–280, 1993.

Litynska, Z., Parfiniewicz, J., and Pinkowski, H.: The prediction of airmass thunderstorms and hails, 450, 128–130, 1976.

Manzato, A.: The use of sounding-derived indices for a neural network short-term thunderstorm forecast, Wea. Forecasting, 20, 896–917, 2005.

Meißner, C., Kalthoff, N., Michael, and Adrian, G.: Initiation of shallow convection in the Black Forest Mountains, Atmos. Res., in press, doi:10.1016/j.atmosres.2007.03.003, 2007.

Miller, R.: Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central, Tech. Rep. 200 (Rev.), AWS, U.S. Air Force. 102 pp., (Headquarters, AWS, Scott AFB, IL 62225), 1972.

Moncrieff, M. and Miller, M.: The dynamics and simulation of tropical cumulonimbus and squall lines, Quart. J. Roy. Meteorol. Soc., 102, 373–394, 1976.

Orville, H.: A numerical study of the initiation of cumulus clouds over mountainous terrain, J. Atmos. Sci., 22, 684–699, 1965.

Schulz, P.: Relationships of several stability indices to convective weather events in northeast Colorado, Wea. Forecasting, 4, 73–80, 1989.

Showalter, A.: A stability index for thunderstorm forecasting, Bull. Amer. Meteorol. Soc., 34, 250–252, 1953.

Van Delden, A.: The synoptic setting of thunderstorms in Western Europe, Atmos. Res., 56, 89–110, 2001.

Van Zomeren, J. and van Delden, A.: Vertically integrated moisture flux convergence as a predictor of thunderstorms, Atmos. Res., 83, 435–445, 2007.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, San Diego, 1995.