VALIDATION OF METEOSAT STORM DETECTION AND NOWCASTING BASED ON LIGHTNING NETWORK DATA

Tobias Zinner[1,2], Hans D. Betz [3,4]

[1]Deutsches Zentrum für Luft- und Raumfahrt, Wessling, Germany
[2]Meteorologisches Institut, Ludwig-Maximilians-Universität München, Germany
[3]Department für Physik, Ludwig-Maximilians-Universität München, Germany
[4]Nowcast GmbH, München, Germany

Abstract

Validation of satellite based storm nowcasting is a topic of highest interest as a consequence of increasing capabilities of nowcasting tools and thus increased use of the products. The value of pre-convective instability indices, of detections of convective initiation or developed storms, and eventually of nowcasting products have to be quantified. Only this way they can be used and – from a developer's point of view – improved.

Here we show a validation of detections and nowcasting products generated by the current version of the DLR Meteosat thunderstorm TRacking And Monitoring algorithm Cb-TRAM against NOWCAST ground-based LIghtning NETwork data (LINET). The systematic comparison of objects detected and nowcasted as developed thunderstorms by Cb-TRAM and with lightning detections clustered in time and space (defining lightning storm cells) allows for the specification of probability of detection and false alarm rate for the satellite products related to thunderstorms with different lightning intensity thresholds.

1. INTRODUCTION

Today a range of possibilities for thunderstorm nowcasting is provided based on satellite data, especially from a geostationary perspective like Meteosat SEVIRI. Instability indices are derived for cloud free areas before even first cloud development occurs, e.g., demonstrated for METEOSAT data in the EUMETSAT global instability index product (ATBD GII, 2007, Fig 1). The indices are based on an estimate of the vertical atmospheric temperature and moisture profile derived from the water vapour and infra-red window channels of SEVIRI and a first-guess atmospheric profile from numerical weather models. The next stage of convective development is covered by detection schemes for the first appearance of clouds (convective initiation), e.g., described in Mecikalski and Bedka (2006). Using a series of threshold test (instantaneous and time trends) they indentify the cloudy pixels which are most likely showing substantial convectively induced cloud growth. A similar but simpler CI detection scheme is also implemented in Cb-TRAM as well (see section 2). The third step, the detection of existing thunderstorms and monitoring of their life cycles is covered by schemes like the Rapid Development Thunderstorms tool (RDT) of MeteoFrance and Nowcasting SAF (ATBD RDT, 2007) or Cb-TRAM (Zinner et al, 2008). All of these techniques have in common that users and developers need quantitative characterization of their capabilities. Users base important, and not unlikely quantitative, decisions on their output; developers need to check the impact of changes to the algorithms systematically.

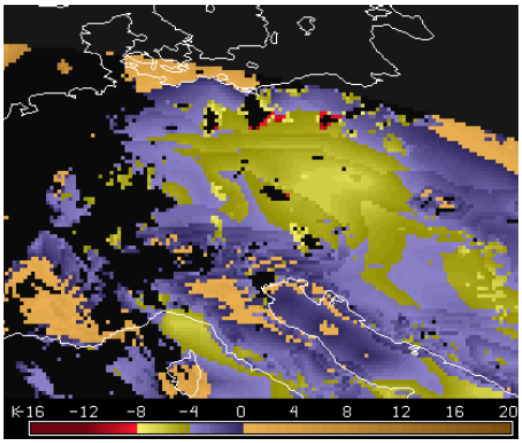

Figure 1: (Figure 6 from EUMETSAT GII ATBD, 2007) Lifted Index for 05 June 2003, 0900 UTC for Central Europe. Increasing negative values,i.e. increasing instability, are shown in blue to yellow to red, while brown denotes stable air. Black areas are clouds.

Figure 2 shows an example of detection scheme output which illustrates the issues arising once the decision to compare to independent data is reached. The image shows Cb-TRAM output on a high resolution visible (HRV) channel SEVIRI image. Contours show current detections of convective initiation (yellow), rapid development (orange) and mature thunderstorm (red), related tracks show past positions and development stages (see next section for more detail). Pink crosses depict the position of flash reports through the LINET network within 5 minutes around the SEVIRI data acquisition time (for Central Europe). For each of the Cb-TRAM development stages an own validation strategy should be used. Areas suspected to be in a CI stage do not necessarily show any current independent signal of convective activity, no lightning can be expected, most likely no precipitation has formed yet. CI detections can only be validated by their future development. But what should be the size of the space and time window to be checked for convective activity? Similarly the rapid development stage is probably not accompanied by any precipitation which is already of unmistakably convective intensity. Are these early detections correct, if there is any storm activity at any arbitrary future point in time in this area? Or for this particular detection's future track? Mature thunderstorm detections should display strong precipitation signals, e.g. in radar data, and lightning activity, but what are the intensity thresholds to look for? Is one flash already a sign of an intense storm?

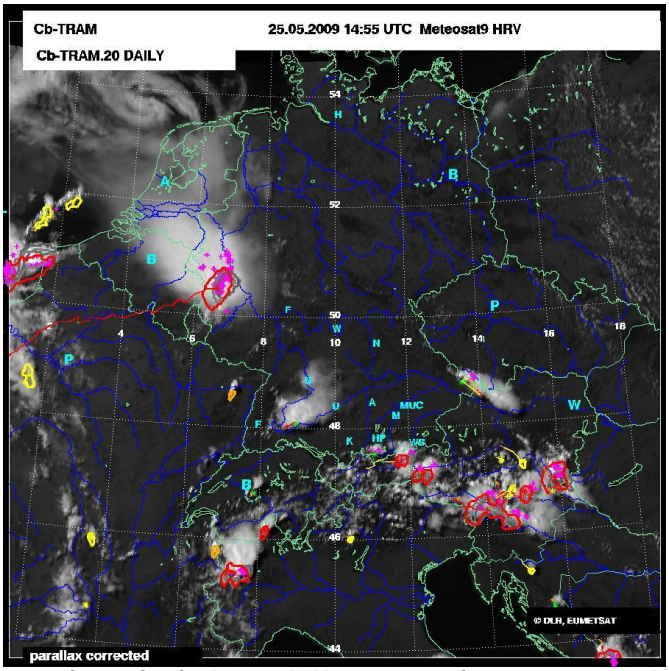

Figure 2: Meteosat-8 SEVIRI HRV image for the 25 May 2009, over central Europe, 14:55 UTC (time of data acquisition). Overlaid are (1) Cb-TRAM output (yellow, orange, red contours and tracks) and (2) LINET flash reports (pink crosses).

These different levels of significance have to be carefully characterized by means of independent data wherever available and the gaps in data availability, e.g. early developments, might be filled with (dependent) data, e.g. from the Cb-TRAM tool itself. There are first comparisons of Cb-TRAM detections with severe weather reports from the European Severe Weather Database (ESWD, e.g., Forster and Dotzek, 2009), but systematic validation against ground truth data of coverage, at least, comparable to Meteosat data is missing. In the following, we show a first approach to characterize the Cb-TRAM detection quality for the stage of mature thunderstorms with different thresholds of lightning intensity. A validation of the other two detection stages for earlier development is planned to follow.

2. THE METEOSAT THUNDERSTORM TRACKING MONITORING ALGORITHM CB-TRAM: RECENT IMPROVEMENTS

Cb-TRAM is documented in Zinner et al. (2008) summing up work which was going on at DLR for already more than 10 years. The core algorithms of Cb-TRAM, namely the image matching and motion vector derivation were used for different purposes before: Contrail detection (Mannstein et al., 1999), stereo imagery (Muller et al. 2007), but also convective storm studies (Mannstein et al. 2002). Once the tool was established for day-to-day detection and tracking of convective cells on a project basis (EU projects RiskAware, 2004-2006; FLYSAFE, 2006-2009, Tafferner et al., 2008; ongoing DLR project Wetter&Fliegen) a rapid evolution of the detection schemes themselves was initiated driven by weaknesses appearing during regular operation. The current setup is described in the following.

At the center of the Cb-TRAM still stands the image matching technique which analyses the motion field, or more precisely the transformation field, that describes the change from one image to the next. A continuous field of vectors is obtained from all features visible in the image regardless of its physical nature (be it albedo features or different level clouds). The image is analysed stepwise from large scale to small scale features ("pyramidal" matching). This vector field can be utilized to generate intermediate or extrapolated synthetic images. The extrapolations are used throughout Cb-TRAM for several purposes. First it is used in the tracking scheme to match cloud objects identified in the detection scheme at one time with its alter ego at the next time step. This feature improves the standard cloud object overlap matching technique by accounting for cloud motion. Especially the matching over long time periods or of small objects is improved. In a similar way the influence of cloud element motion is removed from the analysis of visible growing tendencies (CI detection) and strong IR cooling trends (rapid development detection). Finally extrapolation in time is used to generate simple nowcasts of cloud object positions. More detail is given in Zinner et al. (2008). Three separate detection schemes are employed. For all stages a minimum size requirement of 3 connected pixels (8-connectivity) is implemented to avoid numerous spurious and fluctuating detections.

Stage 1 "convective initiation" identifies cloud objects which show any change in horizontal cloud extent in the HRV channel. An object consists of all connected pixels which show an increase in reflectivity which is accompanied by any IR 10.8 cooling. Difference images for both channels are analysed for that purpose and horizontal advection is accounted for by forming the difference between the current and a synthetic image obtained from the past image by a one-timestep forward extrapolation.

Stage 2 "rapid development" identifies cloud objects which show a rapid cooling (evaluated analogous to stage 1 trends) of more than 1K/ 15 min in the WV 6.2 channel. Thereby, parts of cloud tops are detected which grow rapidly at heights at or close to the WV tropospheric background temperature. This is a usual sign of clouds growing close to strong inversions in the middle troposphere or at the tropopause level. Their absolute temperature is not yet cold enough to be detected by stage 3.

Originally ECMWF tropopause data was used for the stage 3 "mature thunderstorms" detection scheme. Although this was already better than using a fixed temperature threshold, detection failures occurred for low-capped thunderstorms and for the application in tropical environment (with a much less distinct cold point tropopause). This leads to changes compared to the version presented in Zinner et al. (2008). This scheme is now composed of two main criteria: (1) as a new temperature criterion the channel difference T_6.2-T_10.8 is introduced which, as in the original version, is complemented (2) by a HRV texture information during daylight hours. Both are combined in a way

that a close miss of the storm threshold in one criterion can be compensated by a clear signal in the second. This leads to much more consistent detections over a storm life cycle than the isolated use of only an arbitrary fixed threshold for the temperature criterion. In addition, the use of HRV texture improves the separation of large areas of storm anvils and high cloud top frontal systems from the small cores of convective activity we are most interested in. The temperature difference alone was found to be too insensitive to make this separation and, at the same time, very sensitive to the exact value of the actually chosen temperature difference threshold.

First the WV6.2-IR10.8 difference is evaluated. Wherever it is positive, cloud tops are suspected to reach or overshoot the tropospheric background which is a clear sign of strong convective activity (Schmetz et al., 1997). As mentioned above, looking for a positive difference of these two channels alone leads to miss-detections of large cloud areas, especially in frontal systems. Raising this detection threshold to positive values causes missed detections. These and similar reasons in the original setup are the main motivation to mix different detectable signs of storm activity in a weighted non binary sense.

During daylight hours (defined as local solar zenith angle SZA>75°) the "local standard deviation" is used as a texture measure for the HRV image. This standard deviation is obtained by using a Gaussian weighting kernel centred on the pixel of interest to find a neighbourhood typical value and derive the weighted standard deviation from this value (Zinner et al., 2008). If this standard deviation is larger than the standard deviation found for 65% of all thunderstorm (Cb-TRAM detection without texture criterion), the temperature difference is weighted with this standard deviation in a way that increases the likelihood for a detection. Technically the detection threshold could be lowered by up to 10 K this way: even a difference of -10 K could still be detected as mature storm, if the local standard deviation is large enough. Practically the most extreme values of local standard deviation observed lead to the detection of storms which show a negative WV6.2-IR10.8 difference of -3 K. Areas which do not show a clear texture signal of a turbulent thunderstorm cloud top, on the other hand, are less likely to be detected due to this combination of criteria. This excludes large cloud areas especially in situations of passages of fronts. The dependence of the texture signal on solar zenith angle is correctly considered, thereby becoming independent from time of day and season.

During night time the HRV texture is replaced by an analogous WV6.2 texture signal. Although we cannot use the high-resolution information provided by the HRV, there still is lower resolution information on the variability of the cloud top in the IR channels. To provide comparable detection sensitivity between day and night, originally the temperature criterion had to be adjusted to be more conservative to compensate the missing HRV contribution (Zinner et al., 2008). Now we facilitate this by adding a similar contribution, although less effective in identification of most active cells. A local standard deviation exceeding the value which 75% of all thunderstorms show (detected without texture criterion) increases the detection likelihood. Apart from the change of channels, this relative threshold value for the standard deviation is the only parameter changed compared to the daytime algorithm (in a way to provide comparable detection sensitivity).

3. LIGHTNING NETWORK DATA – LINET

Lightning detection can be performed by means of quite different techniques, but in many countries fully automated networks are most common, which utilize a number of antennae for the measurement of electric and/or magnetic fields emitted during lightning discharges. The sensor data are transmitted to a central processor, where lightning locating is performed. LINET exploits the VLF/LF regime and combines the measurement of cloud-to-ground (CG) and inter-cloud (IC) strokes within a single technology, employing baselines of 200–250 km for an adequate coverage in the central parts of the network (Betz et al. 2008). Presently, in many border areas with the inclusion of the Mediterranean Sea the baselines between stations are larger; consequently, the detection efficiency is reduced, i.e. weak IC and CG signals are not located. Nevertheless, data comparisons have revealed that LINET is very sensitive and reports more events than other networks with comparable sensor geometry. Figure 4 shows the sensor locations as of April 2008.

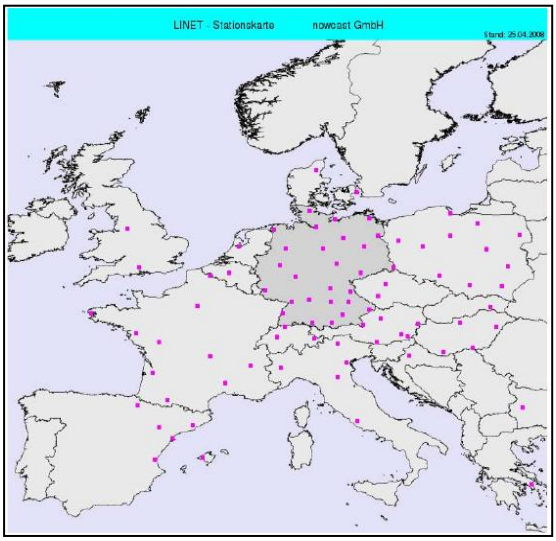

Figure 3: Location of 90 sensor sites of the lightning detection network LINET, as of April 2008 (taken from

4. DEFINITION OF LIGHTNING CELLS

For a comparison of this lightning data to Cb-TRAM detected objects, the lightning reports of LINET are grouped to form lightning cells. This way storm objects on the basis of satellite data can be validated against storm objects based on lightning data. Figure 4 illustrates this process. The lightning data is not divided into CG and IC strokes; multiple detections of a single event are filtered out by the requirement of a minimum time and space separation.

As pointed out in the introduction, several definitions of what a “good” storm detection has to identify are imaginable. Here this is in part already set by our definition of lightning cells. It is implied that severe thunderstorms, which have to be detected reliably by our mature stage scheme, are always accompanied by strong lightning activity confined to strong updrafts. At the same time we use a minimum size requirement of three connected pixels (8-connectivity) analogous to the Cb-TRAM object definition. Following the literature (Oettinger et al. 2001, Betz et al. 2008) and personal communication (Alain Delannoy, ONERA, FLYSAFE project) a series of possible lightning density threshold are defined as a sign of convective activity: 0, 1, 5, 10, 20 or 40 flash reports within 3 km radius and within 5 minutes. These values roughly translate into densities of 0, 2, 10, 20, 40, and 80 flash reports per SEVIRI pixel per 15 minute time interval for Central Europe. Cells which are smaller than three connected SEVIRI pixels or containing a lower lightning activity than one of these thresholds will not be regarded a miss in a respective validation. In the following, results for the activity levels “> 0 = any activity” and “> 10 = intense storm cell” will be inspected closer. After filtering out activity below the threshold and cells too small, the remaining cell size is extended by adding additional edge pixels following a morphologic dilation operation. This way, small gaps between small areas of current lightning activity are combined to obtain larger cells. At the same time, of course, the requirement for exact positioning of SEVIRI and LINET measurements onto the common grid is relaxed this way (compare to Fig 4d). Yet again analogous steps are used to obtain the Cb-TRAM objects; thus a demand for more detailed detection capabilities would not be fair either. All this said, we move on to the comparison between lightning cells and satellite detected mature objects.

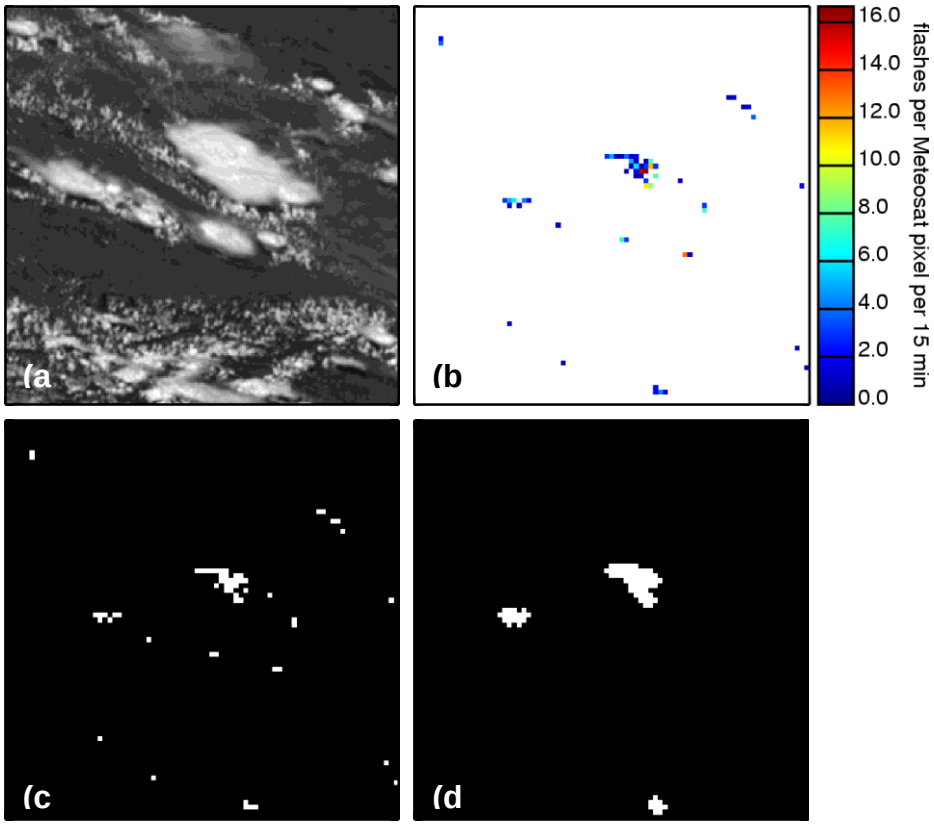

Figure 4: (a) shows an example of a cloud situation in the HRV. (b) is the density of related LINET flash reports mapped on the Meteosat SEVIRI normal resolution pixel grid (time of Meteosat image +/- 7.5 minutes). (c) shows a masked version of (b) highlighting each single SEVIRI pixel containing any lightning activity, (d) shows another more restrictive possible mask of intense lighting cells (> approx. 10 flashes per pixel, added edge pixels,

5. VALIDATION OF CB-TRAM DETECTIONS AGAINST LINET CELLS

In the following we present values of the two most used scoring parameters, the probability of detection (POD) and false alarm rate (FAR). Comparing the positions of two types of objects which are very different in the underlying data generates further need for definitions. POD is defined with respect to lightning cells, while FAR is related to Cb-TRAM objects.

A CORRECT detection is a Cb-TRAM object which has any overlap with a lightning cell of the required threshold activity. A FALSE detection is a Cb-TRAM object without any such overlap. A MISS is a lightning cell which has no overlap with a Cb-TRAM object; a HIT is a case with overlapping Cb-TRAM object. POD and FAR are defined as follows. The subscript *lt* emphasizes the dependence on the lightning activity threshold selected.

$$POD_{lt} = \frac{HIT_{lt}}{HIT_{lt} + MISS_{lt}} \qquad \text{and} \qquad FAR_{lt} = \frac{FALSE_{lt}}{CORRECT_{lt} + FALSE_{lt}}$$

Our validation data base consists of 92 consecutive days of data for an area between 42 and 54°N and -5 and 16°E for a time period throughout the main thunderstorm season in summer 2008 (June, July, and August). Figure 5 shows the results day by day and the average values for $POD_{>10}$ (approx. lightning density >10 flashes within the area of a Meteosat pixel), $FAR_{>10}$, and $FAR_{>0}$ as 11-day moving average and JJA average.

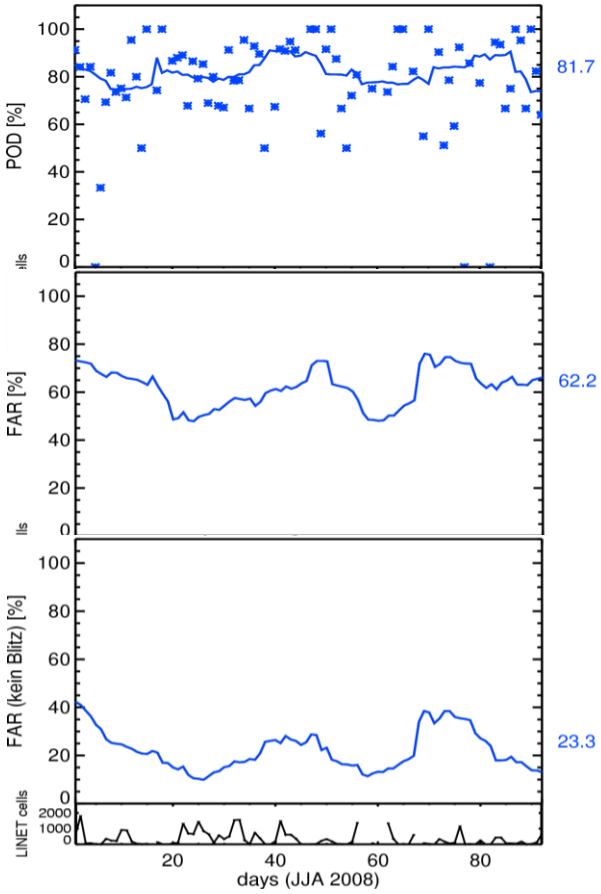

Figure 5: (top) $POD_{>10}$, (center) $FAR_{>10}$, and (bottom) $FAR_{>0}$. In (top) the $POD_{>10}$ values for each single day in June/July/August 2008 are given (stars). The line shows the average over all storm cases within a moving 11 day period. The single value to the right is the average value in percent over all cases within the analysed 92 day period. (b) shows the $FAR_{>10}$, and (c) $FAR_{>0}$.The convective activity for each day is illustrated by the total number of lightning cells (*lt* > 10) in the bottom of (c).

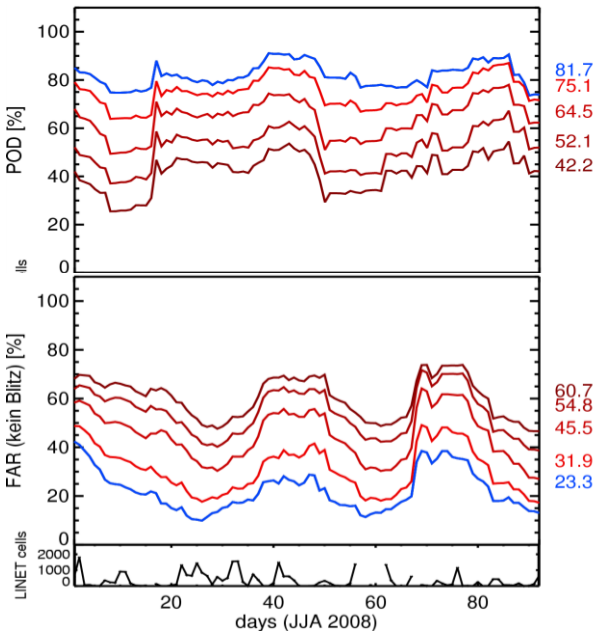

Figure 5: 11-day moving average of $POD_{>10}$ (top) and $FAR_{>0}$ (bottom) for detections (compare to Figure 4, blue lines) and for the nowcasts +15, +30, +45, and +60 minutes (orange to dark red lines). At the far right of the image the 92 day average values are given.

A lightning cell with an activity above 10 flashes within a Meteosat pixel area (approx. 20 km^2) within 15 minutes is detected with a probability of 82% over the whole period. Single day values jump mostly between 50 and 90% with the most extreme values appearing on days with very low activity (the only cell of the day is either detected or not). This detection rate for thunderstorm is rather high, but this is accompanied by a high $FAR_{>10}$ of 62% over the validation period. That means 62% of all mature stage Cb-TRAM objects do not show a lightning intensity which above 10 flashes per 15 min and 20 km^2. However, if the lighting intensity requirement is released to a value of *lt* > 0, it is revealed that still 77% of all objects identified as mature thunderstorm objects contain three connected pixels displaying lightning activity. Only 23% of the objects are un-disputable false alarms in the sense that they do not show any lightning activity and thus should definitely not be identified as mature convective storm.

6. VALIDATION OF CB-TRAM NOWCASTS AGAINST LINET CELLS

In figure 6 the analogous validation of the nowcast capabilities of Cb-TRAM is presented. Once a Cb-TRAM object has reached the mature stage, the nowcast of this object's position up to 60 minutes into the future is investigated. In this case POD and FAR could well be defined in their logical correct sense. That means, both quantities could be defined with respect to Cb-TRAM objects: detected objects and extrapolated objects from an earlier time step. Nonetheless for reasons of continuity and clarity, we prefer to keep the above definitions, i.e., the analysis against the ground truth from lightning data.

The preliminary analysis shows interesting information and clear limits of our extrapolation technique. While the $POD_{>10}$ is still about 65% for a 30 minute nowcast, it drops down to around 40% for the 60 minute nowcast. At the same time even the tolerant $FAR_{>0}$ reaches values of 46 (30 minutes) and 61% (60 minutes).

On the one hand this is probably owed to the life cycle of convection which is ignored in our extrapolation algorithm. Even mature thunderstorms covering the required minimum area which are detected at one time can easily decay within 60 minutes. On the other hand, the results are a clear sign of technical characteristics of our extrapolation algorithm. The motion or transformation vector fields which we derive are obtained from matching small scale brightness values in the context of a larger scale analysis step (pyramidal matcher). That can lead to sharp gradients in the vector field when small isolated features (clouds) move over large scale stationary background (surface). Vector fields extracted this way are well suited for the use in extrapolation for one or two time steps, as long as the motion still takes place in a similar surrounding motion regime (the next larger scale in the matching scheme). That means, e.g., thunderstorms embedded in larger scale cloud systems or situations of broken cloud fields covering some area allow for better extrapolation results than small isolated convective cells. For general reliable nowcasts of more than 30 minutes improvements are necessary.

7. DISCUSSION

There are important limits to the method we apply to compare objects and derive POD and FAR from object counts. These numbers are readily understandable and the definitions seem to be obvious, but an important property neglected here is the area of the detected Cb-TRAM object. Usually the POD would be expected to increase when a detection threshold is interpreted less and less strict, detecting more and more possible hits. At the same time the FAR would go up as well, as the chance for false detections rises. Conducting an object based comparison, this situation changes depending on a certain (large) size of detected objects. An object covering the whole analysis domain can have a POD of 100% and at the same time a FAR of 0%, the optimal result, although the detection is useless: Each (lightning) thunderstorm occurring would be a detected storm (a hit) as overlap is always given and, at least during summer season, the chance is high that somewhere in the domain a single lightning cell would make the full domain detection a correct one at the same time. That means a full analysis had to include a measure preventing this unintentional result. A solution would be the definition of POD and FAR on a pixel by pixel basis. The above mentioned full domain detection would then have a 100% POD, but also a FAR close to 100% which would signal its low quality. Levels of POD and FAR would in this respect be solid measures, but at the same time the very low values obtained for POD and the very large values for FAR do not give a user the correct impression. One cannot expect a satellite detection to find the exact pixels of lightning activity, already for cloud physical reasons. This example again emphasizes the ambiguity of most validation methods.

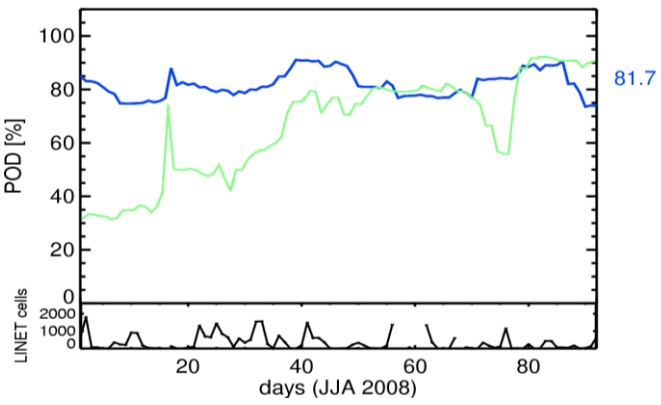

Figure 6: Development of the "operational" Cb-TRAM version throughout the summer 2008. The green line documents the improvements of the Cb-TRAM detection schemes with respect to the lightning validation.

To account for the discussed problem, additional pixel based quantities have to be defined. In a later more detailed analysis it is planned to use an area based "fraction of the Cb-TRAM object showing lightning". For example here it is found that for a Cb-TRAM mature stage object about 12% of the pixels show lightning activity.

The above chapter already points out an important benefit of the systematic validation data which we plan to exploit more systematically. Further developments of our nowcasting and our detection scheme will be based on this JJA2008 lightning validation data set. The impact of changes to the nowcasting scheme can

easily be tested. Figure 7 illustrates this possibility in a similar context. Shown is the 11-day moving average of the Cb-TRAM stage 3 detection alongside the adjustments made throughout summer 2008 with positive and negative impact on $POD_{>10}$. E.g., in the beginning of June there were several days with extremely weak detection probability, because the original tropopause based detection scheme was weak in situations of low capped thunderstorms occurring at that time.

REFERENCES

ATBD GII (2007), The Global Instability Indices Product, Algorithm Theoretical Basis Document, EUMETSAT report, EUM/MET/REP/07/0164.

ATBD RDT (2007), Algorithm Theoretical Basis Document for “Rapid Development Thunderstorms”, Nowcasting Satellite Application Facility (NWC-SAF) Report, SAF/NWC/CDOP/MFT/SCI/ATBD/11, Issue 1, Rev.3.

Betz, H. D., K. Schmidt, W. P. Oettinger, and B. Montag (2008), Cell-tracking with lightning data from LINET, Adv. Geosci., 17, 55–61.

Forster, C. and N. Dotzek (2009), Quantitative intercomparison of Meteosat-based thunderstorm detection and nowcasting with in situ reports from the European Severe Weather Database (ESWD), Proc. 2009 EUMETSAT Meteorological Satellite Conference, Bath, 21-25 September 2009, 10 pp.

Mannstein, H., Meyer R., Wendling P. (1999), Operational detection of contrails from NOAA-AVHRR data. *International Journal of Remote Sensing*, 20, pp. 1641-1660.

Mannstein, H., H. Huntrieser, and S. Wimmer (2002), Determination of the mass flux in convective cells over Europe, Proceedings EUMETSAT Meteorological Satellite Conference 2002, Dublin, Ireland, EUMETSAT, 264-269.

Muller, J.-P., M. A. Denis, R. Dundas, K. L. M. Mitchell, C. M. Naud, and H. M. Mannstein (2007), Stereo cloud-top height and amount retrieval from ATSR2. International Journal of Remote Sensing, 28, 1921-1938.

Mecikalski, J. R. and Bedka, K. M. (2006), Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery, Monthly Weather Review, 134, 1, pp.49-78.

Oettinger, W. P., B. Eisert, H.-D. Betz, A. Gerl (2001), Automatische Erkennung extremer Wettersituationen im Nowcasting-Bereich, Proceedings DACH Tagung, Wien.

Schmetz, J.; Tjemkes, S. A.; Gube, M.; van de Berg, L. (1997), Monitoring deep convection and convective overshooting with METEOSAT, Advances in Space Research, Volume 19, Issue 3, p. 433-441.

Tafferner, A., C. Forster , S. Senesi, Y. Guillou, P. Laroche, A. Delannoy, B. Lunnon, D. Turb, T. Hauf, and D. Markovic (2008), Nowcasting thunderstorm hazards for flight operations: the Cb-WIMS approach in FLYSAFE, Proc. 26th Int. Congress of the Aeronautical Sci. (ICAS), Anchorage, Alaska, 14 -19 September 2008.

Zinner, T., Mannstein, H, and Tafferner, A. (2008), Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data**,** *Meteorology and Atmospheric Physics*, 101, pp 191-210, DOI 10.1007/s00703-008-290-y.