Report
on the testing of EUMETSAT NWC-SAF products
at the ESSL Testbed 2017

This work was carried out according to the contract of 29 November 2016 for the Associated Scientist for the activity "Perform the second (European) edition of the NWC-SAF Testbed Facility", including Annex I, the Visiting Scientist Proposal "NWC/CDOP2/SAF/AEMET/SCI/VSP/10" of 24 August 2016.

Alois M. Holzer (Visiting Scientist) and Pieter Groenemeijer

Wiener Neustadt, 29 September 2017

Table of content

I. Introduction

EUMETSAT Nowcasting SAF products were tested during the ESSL Testbed 2017 on a European scale, putting products at test within a realistic forecaster environment – including other data sources and time pressure.

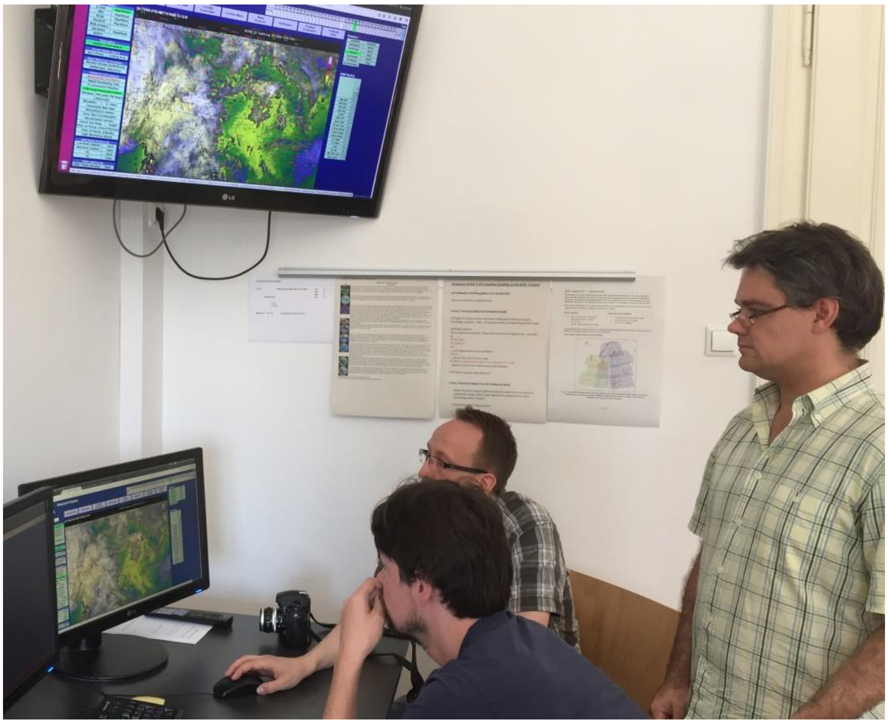
Preparation, participation and post-processing phases covered the period from 1 November 2016 to 31 October 2017.

During the 5-week participation phase in June and July 2017, 52 participants from 17 different countries attended the ESSL Testbed and worked with the NWC-SAF products.

Germany	12	Hungary	1
Spain	8	The Netherlands	1
Croatia	8	France	1
Austria	5	Czechia	1
USA	4	Latvia	1
Finland	2	Switzerland	1
Slovakia	2	Portugal	1
Greece	2	Romania	1
		Italy	1

Table 1: Number of participants per country at the ESSL Testbed 2017

For further details regarding the ESSL Testbed venue, concept and tested products we kindly refer to the attached ESSL Testbed Operations Plan 2017.



Participants from Spain, Finland and Slovakia discussing NWC-SAF products at the ESSL Testbed 2017

II. Performed Tasks

1. Necessary software and hardware were installed at the ESSL Testbed site in time to have all the defined NWC-SAF products available for the Testbed.
2. Testing of the defined NWC-SAF products was performed during the regular annual ESSL Testbed. Details on the results, including the collected participants' responses, are presented in this report.
3. Post-processing of then responses collected at the testbed related to NWC-SAF was done after the participation phase. A synthesis of the results is presented in this report.

Task 2, the evaluation of NWC-SAF products, was carried out in two ways, as foreseen:

Assessment of the product usability for a forecaster (manual severe storms forecasting)

The listed NWC-SAF products were assessed for their usability in human forecasting processes via a questionnaire. As anticipated in the Visiting Scientist Proposal, some products are of high value for manual forecasting, nowcasting and warning processes, while some products are of lesser value for manual techniques, but of high value for other applications, for example as an important input or basis for subsequent automatic processes.

Engrossed evaluation of selected NWC-SAF products in the focus of forecasting deep moist convection

The products "Instability index", "Precipitable water content in clear air" and "Convection initiation and rapid development thunderstorms" underwent a more detailed evaluation. Visualization and accuracy of these products were evaluated in group discussions and extended questionnaires.

Potential areas of improvement of these products will be outlined and interesting cases illustrated in this report.

III. Results

The following NWC-SAF products were evaluated by testbed participants via online questionnaires (sample size N=45):

RDT Rapid Developing Thunderstorms
CI Convective Initiation
TPC Total Precipitable Water
PW Layer Precipitable Water (low, mid, high)
Lifted Index, Showalter Index, K-Index
Convective Rain Rate
Convective Rain 1 h Accumulation
Cloud Top Temperature
Cloud Top Height
PoP Probability of Precipitation (cloud physics)
Probability of Precipitation (classic)
High Resolution Winds

Every testbed participant was asked to fill out questionnaires at the end of the respective testbed week. One questionnaire was focused on the added value of the single NWC-SAF products for the process of manual weather forecasting and warning, given that standard RGB satellite imagery is available too. The second questionnaire focused on the visualization of the different NWC-SAF products, while the third questionnaire asked the respondents to name the three most favourite products from their personal point of view. In addition there was room for written comments. Results from these per-single-person questionnaires are presented in the following section III.i.

Engrossed evaluation was performed in small working groups of typically 3 persons each. These small working groups jointly filled out feedback forms during the second half of the respective testbed week. Results from the engrossed evaluation are presented in section III.ii. for NWC-SAF product groups of Instability Indices, Precipitable Water Content in clear air, Convection Initiation and Rapid Development Thunderstorms.

i. Assessment of the product usability for a forecaster

a) Added Value

Question 1: Compared to the standard IR and VIS imagery and its standard RGB composites: How much added value do you see for the following NWC SAF products in the manual forecasting/warning process?

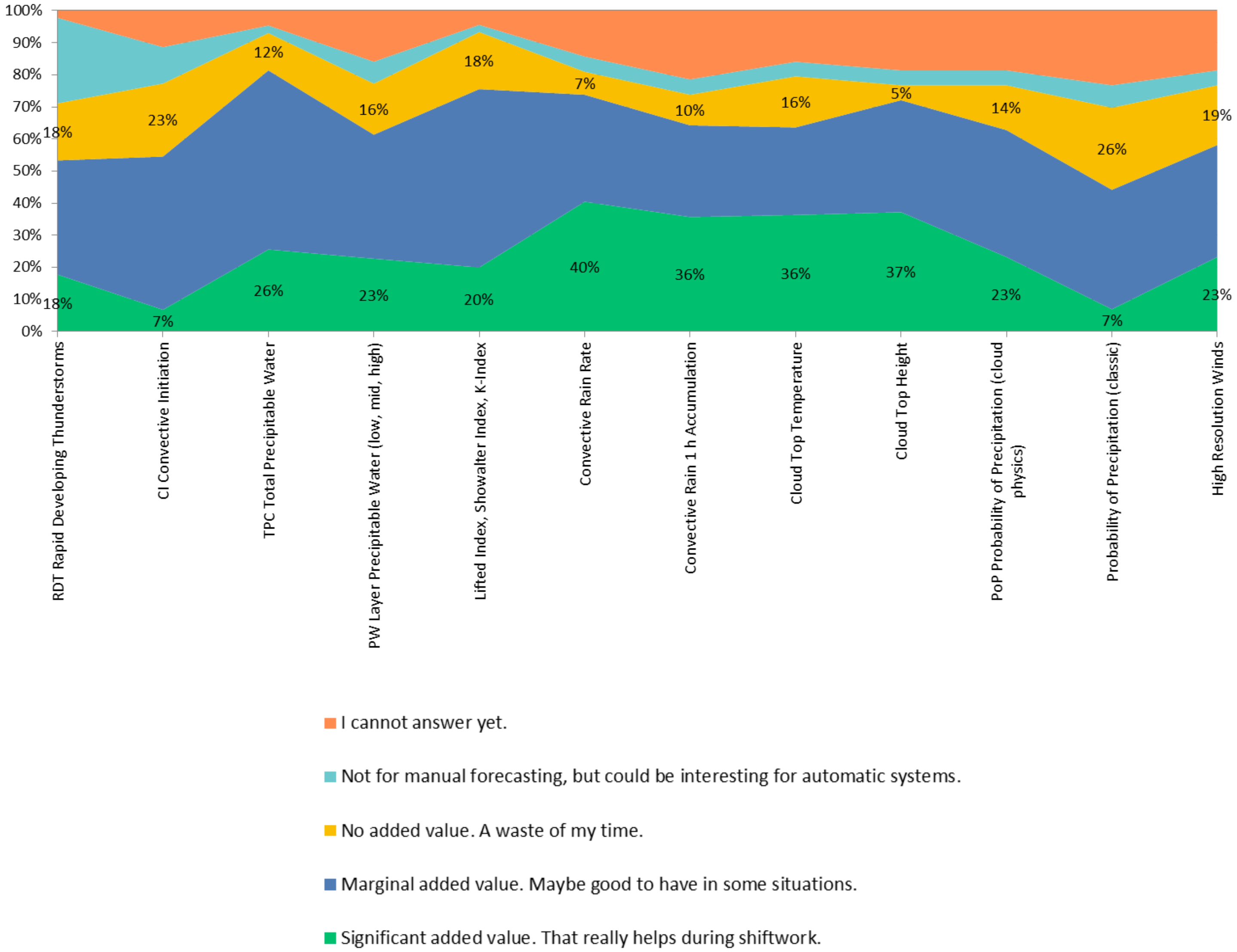


Figure 1: Added Value of NWC-SAF products, percentage of respondents (N=45)

The question to indicate the Added Value of NWC-SAF products was specifically designed to hear the testbed participants' opinion on the added value they see in a forecasting or warning situation, given that regular RGB satellite imagery is available too. For most NWC-SAF products, the category *"Marginal added value. Maybe good to have in some situations."* was selected most frequently. However, for a number of products, i. e. the convective rain rate products and the cloud top temperature and height products, significant added value relative to standard RGB satellite imagery was seen by a relative majority (36 to 40 %) of respondents. The strongest negative feedback was given for the product probability of precipitation (classical version) and convective initiation, to which 26 and 23 % of respondents to the online questionnaire answered *“No added value. A waste of my time.”*.

About 25 % of respondents see potential for automatic systems in the RDT product – by far the highest value of all products, while the added value for manual forecasting is third lowest for the same product.

A quotient of the respondents answering for *“Significant added value. That really helps during shift work.”* and for *“No added value. A waste of my time.”* provides a combined rating for the different products, while *“Marginal added value. Maybe good to have in some situations.”* is not taken into account for this quotient. The “Added Value Quotient” (AVQ) becomes smaller than 1 for products with more negative than positive answers, it becomes larger than 1 for products with more positive than negative answers. Quotients >>1 indicate that Testbed participants show a strong tendency to agree on substantial added value by the respective products.

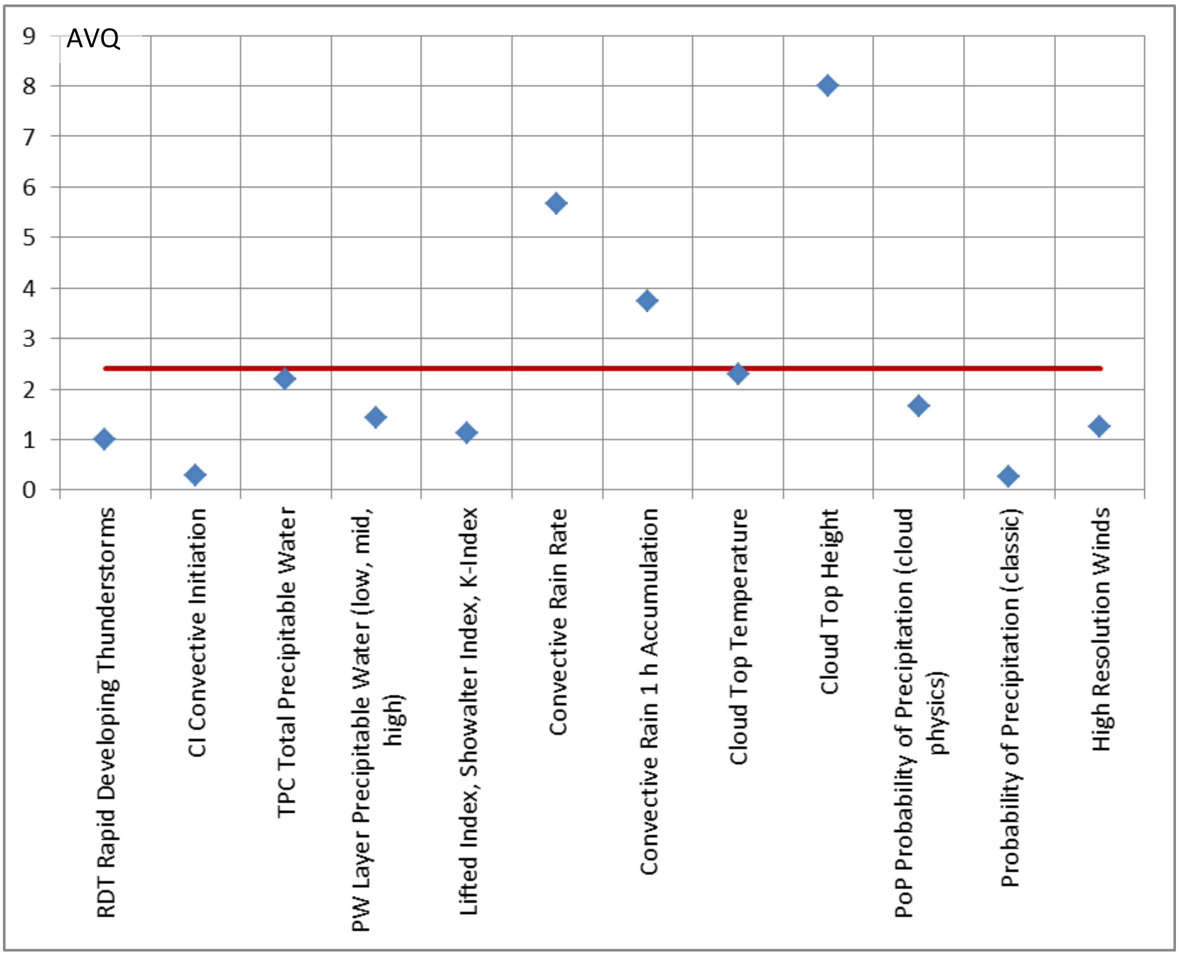


Figure 2: Added Value Quotient (AVQ) for NWC-SAF products, red line indicates mean value of all products

Best rated product from this point of view is “Cloud Top Height” (AVQ of 8), followed by “Convective Rain Rate” (AVQ of nearly 6) and the related accumulation product (AVQ nearly 4), while products CI and classic PoP are rated below 1.

b) Visualization

Question 2: We ask for your feedback in terms of visualization of the following NWC SAF products

This question was designed to collect feedback on the visualization of the single products. While most products were displayed in the proposed way, visualization for "High Resolution Winds" was not determined beforehand and ESSL Testbed visualization there might have been suboptimal.

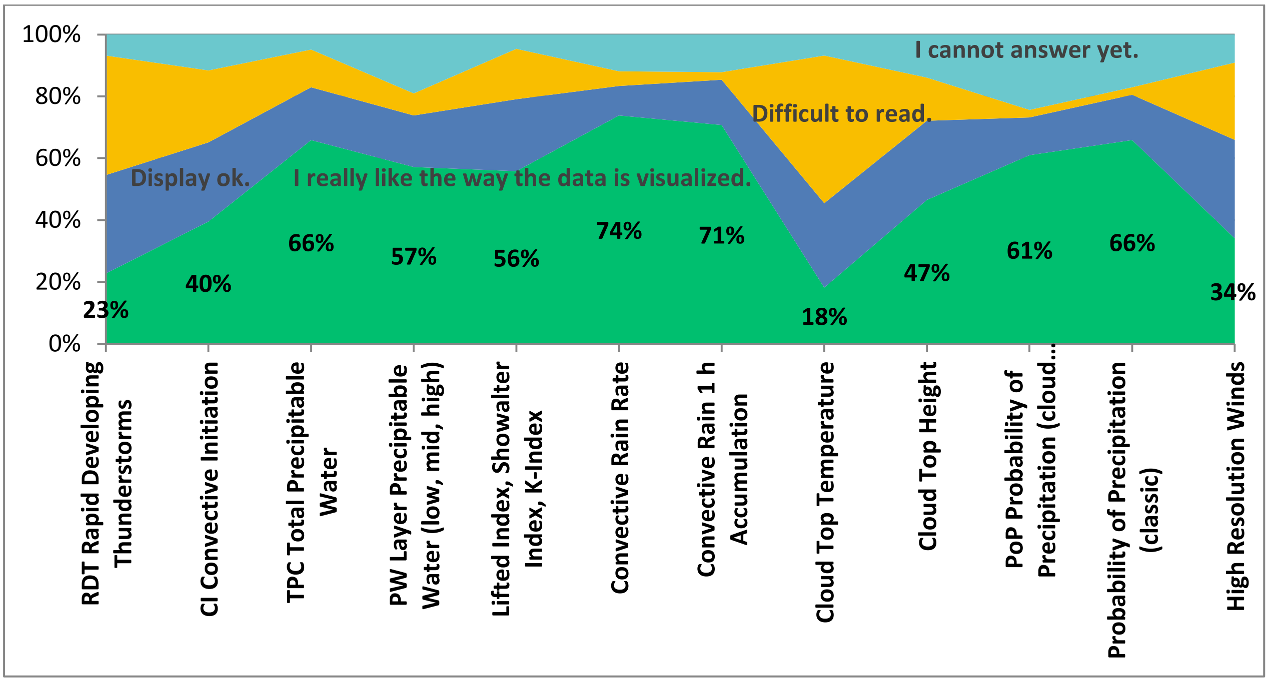


Figure 3: Visualization feedback, percentage of respondents (N=45)

Testbed participants liked most of the NWC-SAF displays and visualizations. The best rating (74 %) is expressed for the Convective Rain Rate product. A number of other products seem to have a well-accepted visualization with over 50 % share for the best display rating.

Ratings for the RDT product were mediocre with about 54 % of all participants answering either with *"Display ok."* or with *"I really like the way the data is visualized"*, while this product received about 38 % negative answers ("*Difficult to read*").

Only 18 % of all respondents liked the visualization of the "Cloud Top Temperature" product, while at the same time it received the largest number of negative ratings (47 %).

c) Subjective Favourite Products

Question 3: Indicate up to three products that you like the most

Here respondents were asked to vote for their most preferred products. This feedback was not aimed to point specifically to added value or to visualization of the single products, but to collect data on the subjective overall favourites of the ESSL Testbed participants, most of them being forecasters or even lead forecasters at their home institutes.

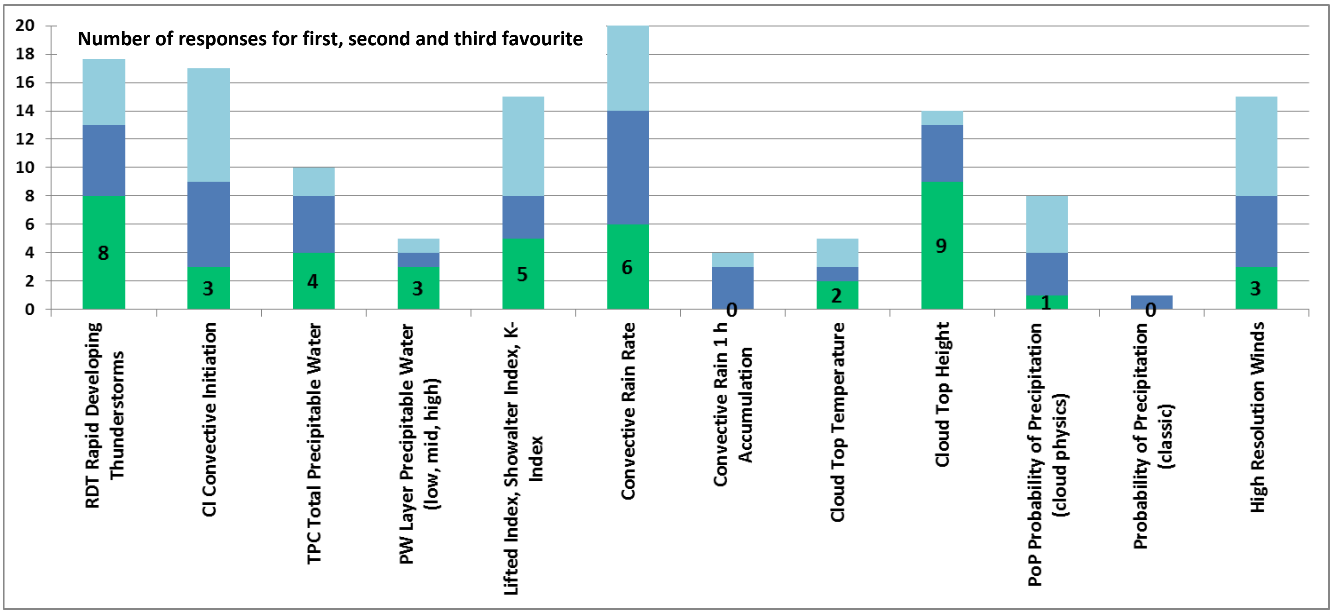


Figure 4: Subjective favourite NWC-SAF product of testbed participants, absolute number of respondents (green … first favourite, dark blue … second favourite, light blue … third favourite product)

The product “Cloud Top Height” received most votes for the absolute favourite product (9), while the accumulated ratings for first, second and third favourite product are highest for “Convective Rain Rate”. Least votes were collected for the classic Probability of Precipitation product, only one second favourite.

A possible interpretation for the relative high number of votes for the RDT product, while added value and visualization feedback was mediocre for the same product, could be that forecasters would like to have such a product with improved value and visualization.

ii. Engrossed evaluation of selected NWC-SAF products

The products “Instability Indices”, “Precipitable water content in clear air (total, low, medium layer)” and “Convection initiation and rapid development thunderstorms” underwent an engrossed evaluation. The usefulness within a scientific forecasting framework was assessed. Visualization and accuracy of these products were evaluated in group discussions and extended questionnaires were filled out for these products by the participants of the ESSL Testbed.

For the Instability Products 11 working groups at the ESSL Testbed provided consensus feedback, for the other products the number of working groups providing feedback was 8. Each group typically consisted of 3 participants.

a) Instability Indices Products

Instability products deliver convective indices for regions with clear air: LI, KI, SHW. Their main value lies in the assessment of the pre-convective environment. The instability products combine model information with satellite data. Difference products (model – satellite data) were not evaluated at this ESSL Testbed.

Some of the respondents saw the availability of data only in cloud-free areas as a major drawback. In addition, some noted pollution by spurious data close to neighbouring clouds.

Participants mainly suggested the following improvements (raw answers to be found in section A.2.1):

1) Enhance the colour scales in order to allow for better distinction in value ranges critical for convective forecasting, i.e. around 0 K for the Lifted Index.

2) To introduce a broader buffer area between clouds and cloud-free zones, to prevent the display of less reliable data.

3) To display departures from model fields instead of combined model and satellite data fields. ESSL would be keen to test such departure from model products in a future edition of the ESSL Testbed.

4) If possible, calculate CAPE in addition to the available indices.

Overall assessment:

The tested instability products have a high potential value, especially in the determination of departures from model fields. Such information cannot be calculated from other sources in comparable temporal and spatial resolution and therefore is highly useful for nowcasting. Future satellite sounder data is expected to deliver vertically higher resolved and more precise data, indicating clear future potential for further improvement.

Example:

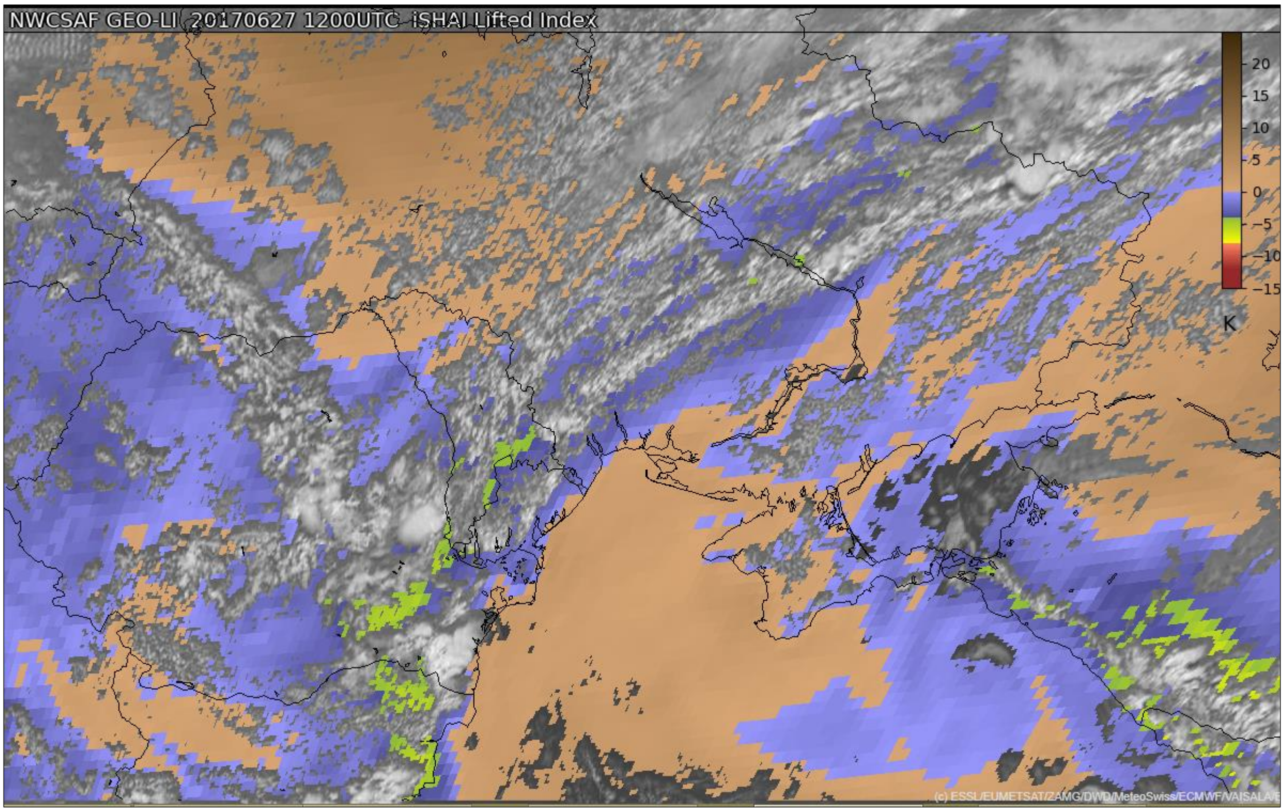


Figure 5: Lifted Index product (LI) for south east Europe. Strong gradients in LI close to cloudy areas led to interesting discussions about the optimum setting for masking out areas adjacent to clouds. ESSL Testbed display screenshot of 27 June 2017, 12 UTC.

b) Precipitable water content in clear air

Precipitable water products for the total column and for three individual layers are available in regions with clear air. Their main value lies in the assessment of the pre-convective environment. The precipitable water products combine model information with satellite data. Difference products (model – satellite data) were not tested at this ESSL Testbed.

Similar to the instability products, respondents saw the availability of data only in cloud-free areas as a drawback. In addition, the fact that pixels close to neighbouring clouds seem to show contamination by unreliable data (leading towards higher water content) was noted negatively.

The availability of such products, particularly low-level humidity, was appreciated, because it is a basic ingredient for deep moist convection. In addition, total column precipitable water is a good predictor for the likelihood of severe convective rain events.

Participants suggested mainly the following improvements (raw answers to be found in section A.2.2):

1) To introduce a broader buffer area between clouds and cloud-free zones, where data should not be displayed because of reliability issues.

2) To display departures from model fields instead of combined model and satellite data fields. ESSL would be keen to test such departure from model products in a future edition of the ESSL Testbed.

3) Modify the colour scale in a way that is more natural for moisture.

Overall assessment:

The tested precipitable water products have a high potential value, especially in the determination of departures from model fields. Such difference information cannot be calculated from other sources in comparable temporal and spatial resolution, and therefore is highly valuable for nowcasting timescales. Future satellite sounder data is expected to deliver vertically higher resolved and more precise data, indicating clear future potential for further improvement.

The main application in forecasting severe convective storms is the satellite versus model comparison of low level moisture and the tracking of atmospheric rivers. High total column precipitable water is important for the anticipation of an elevated risk for excessive convective precipitation.

Example:

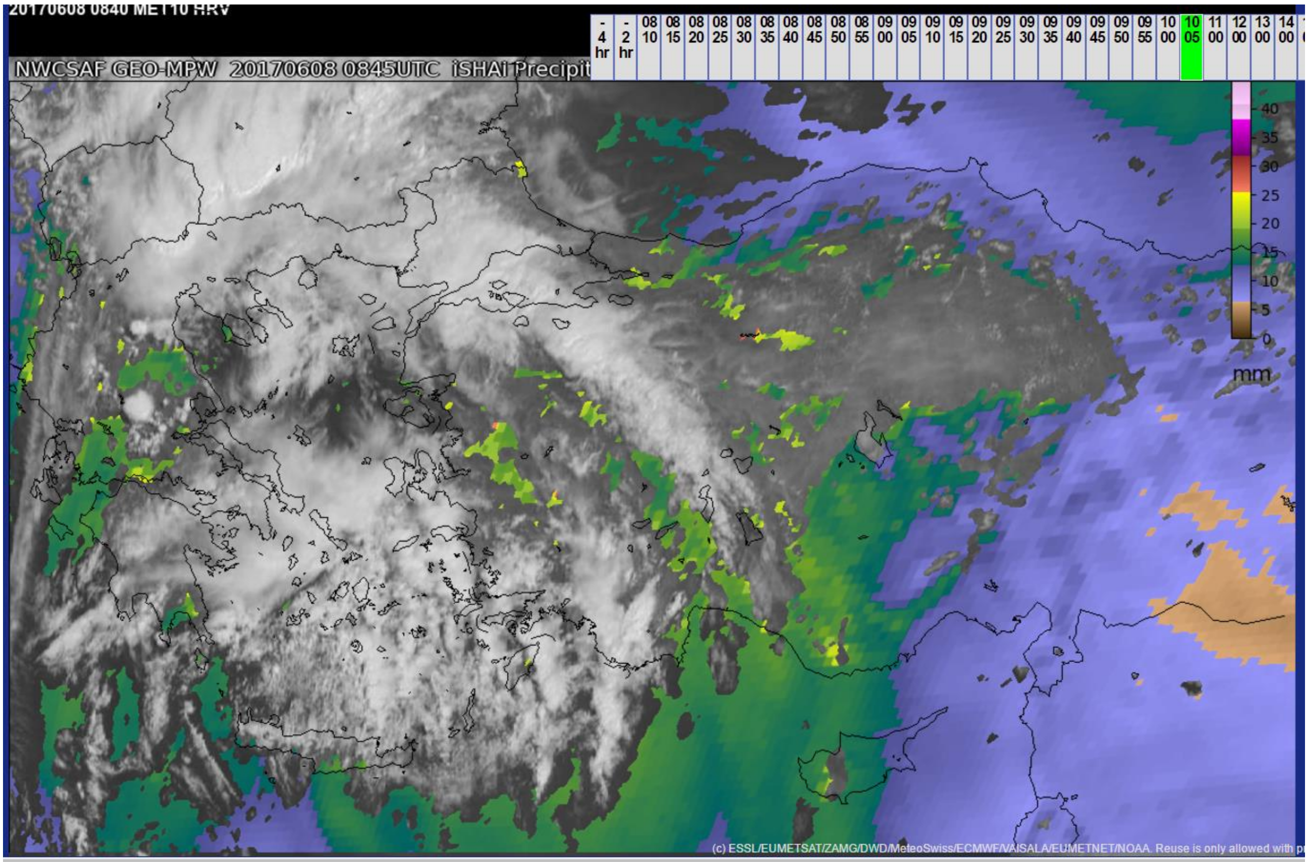


Figure 6: High values of low level layer precipitable water over Turkey confirm that low level moisture is present well inland and not only in a shallow layer over the sea and at the coastline. ESSL Testbed nowcasting display screenshot of 8 June 2017 8:45 UTC.

c) Convection Initiation

The Convection Initiation (CI) products are designed to highlight growing and cooling cumulus cloud areas. At the ESSL Testbed, the CI product with a 30-minute lead time was evaluated. The CI product is classified as experimental by the NWC-SAF because of known deficiencies. This view was shared by most participants of the ESSL Testbed, although it was stated that such a product would be useful to have.

Participants noted mainly the following deficiencies:

1) A high number of false alarms, i.e. clouds that would not develop into convective storms were detected.

2) A high number of misses, i.e. many clouds that turned into deep convective clouds were not detected.

3) Displacements of detections in highly sheared situations

4) Spurious signals near the edges of large anvil clouds.

Only 7 % of all respondents see significant added value compared to standard RGB satellite imagery for this product, the lowest value of all NWC-SAF products (see Table 2).

Respondents (see section A.2.3 for raw answers) saw best performance in situations where high level clouds are absent and low vertical wind shear is present (weak upper level winds). Some participants asked for an extension of the forecast period from 30 to 60 minutes. Although such a products is available, it was not evaluated at the ESSL Testbed, as we wanted to first establish the quality of the 30-minute product.

Two suggestions for improvement were made by testbed participants:

1) Design a product that displays the cloud top cooling rate as a difference product between two consecutive images and display such values above a certain threshold in a colour scale without indicating a probability for initiation of deep moist convection.

2) Cooperate with John Mecikalski and/or Kristopher Bedka, who have long experience in developing CI products. http://journals.ametsoc.org/doi/abs/10.1175/2008MWR2352.1

John Mecikalski
University of Alabama in Huntsville, 301 Sparkman Drive, NSSTC, Room 3040, Huntsville, AL 35899.
E-mail: johnm@nsstc.uah.edu

Kristopher Bedka
NASA Langley Research Center, Mail Stop 420, Hampton, VA 23681.
E-mail: kristopher.m.bedka@nasa.gov

Bedka, who took part in earlier editions of the ESSL Testbed, developed a CI product in the USA that seems to have better handling of CI. Recent related publications can be found on the webpage of the EUMETSAT-ESSL Convection Working Group: https://www.essl.org/cwg/?page_id=620

Overall assessment:

Forecasters are interested to work with a satellite based CI product. The present CI product tested is seen as experimental by the NWC-SAF. This assessment is supported by the feedback from participants of the ESSL Testbed. Suggestions for improvement and cooperation are provided to the developers.

As an alternative to the given CI product, displays of cloud top cooling rates might be of interest, especially if cloud top cooling rates above a certain threshold could be combined with other products, like with cell tracking based on convective rain rate cores and overshooting top information from the RDT product suite.

Example:

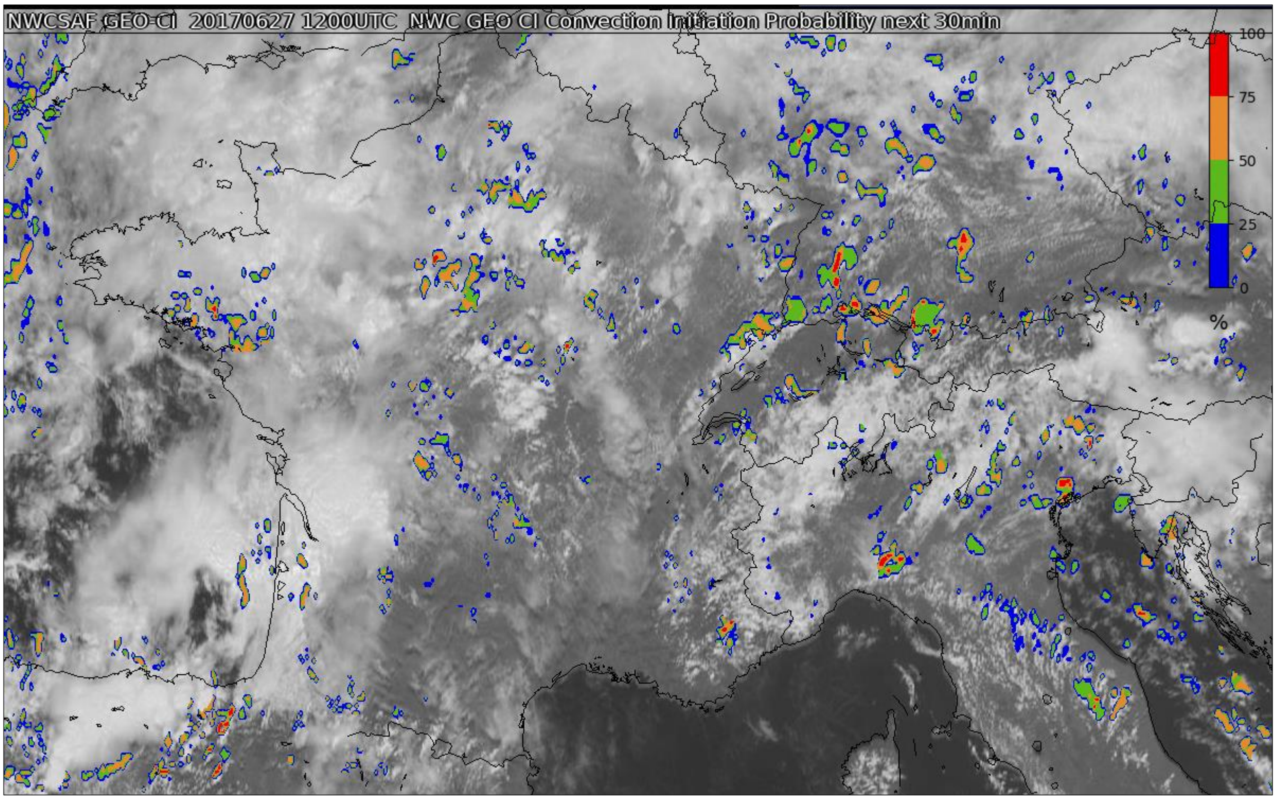


Figure 7: CI is most reliable in regions without high level cloudiness and without strong upper level winds and complex cell clusters. ESSL Testbed nowcasting display screenshot for western Europe, 27 June 2017, 12:00 UTC.

d) Rapid Developing Thunderstorms

The Rapid Developing Thunderstorms (RDT) product is designed to provide information on clouds related to significant convective systems, from meso-alpha scale down to smaller scales (few pixels).

Objectives:

- Identification, monitoring and tracking of intense convective system clouds
- Detection of rapidly developing convective cells
- Forecast of the convective cells

Testbed participants valued that (as already the name of the product suggests) rapid developments in otherwise cloud free environments are best detected. There the product offers good signals especially in the early stage of the convective process.

Because of the condensed data, applications are seen especially where full imagery cannot be derived in real time, i. e. for aviation en-route. The detection of overshooting tops is one of the best features of the product.

The question *"Compared to the standard IR and VIS imagery and its standard RGB composites: How much added value do you see for the following NWC SAF products in the manual forecasting/warning process?"* was answered with *"Not for manual forecasting, but could be interesting for automatic systems."* by 27 % of all respondents (N=45) for the RDT product. Within all NWC-SAF products this is by far the highest response rate for this answer (see Table 2).

Together with the answer *"No added value. A waste of my time."* (18 %) nearly half of all respondents think that the RDT product in its present stage is not suitable for manual forecasting of severe convective storms. Only 18 % of all testbed participants see significant added value for manual forecasting in the RDT product (third lowest after "CI" and "classic probability of precipitation").

Added value to standard RGB imagery?	Significant added value	Marginal added value	No added value	For automatic systems	I cannot answer yet
RDT Rapid Developing Thunderstorms	18%	36%	18%	27%	2%
CI Convective Initiation	7%	48%	23%	11%	11%
TPC Total Precipitable Water	26%	56%	12%	2%	5%
PW Layer Precipitable Water (low, mid, high)	23%	39%	16%	7%	16%
Lifted Index, Showalter Index, K-Index	20%	56%	18%	2%	4%
Convective Rain Rate	40%	33%	7%	5%	14%
Convective Rain 1 h Accumulation	36%	29%	10%	5%	21%
Cloud Top Temperature	36%	27%	16%	5%	16%
Cloud Top Height	37%	35%	5%	5%	19%
PoP Probability of Precipitation (cloud physics)	23%	40%	14%	5%	19%
Probability of Precipitation (classic)	7%	37%	26%	7%	23%
High Resolution Winds	23%	35%	19%	5%	19%

Table 2: Percentages of respondent's answers to the question of how much the different NWC-SAF products add value compared to standard RGB imagery in the process of manual (human) forecasting or warning.

Testbed participants noted mainly the following deficiencies:

1) False detections (mainly large cirrus shields)

2) Misleading placement of the active core of the convective system (too far downstream)

3) Questionable assessments of the main hazard and of the intensity level

4) Unreliable and spurious behaviour of forecast motion vectors

Respondents suggest combining NWC-SAF products differently in order to obtain a more reliable product:

1) Cell tracking and severity levels could be based on microphysics version of convective rain rate (during daytime) by using the cell cores from this product. This should help to smoothen the cell tracking and extrapolation. Furthermore, the most active parts should be captured in a more consistent way.

2) Information on cloud top cooling and overshooting tops should be retained. Latest research suggests that the area covered by an overshooting top can be a good indicator for the severity of a convective storm. Therefore information on the OT size (for example in km²) or diameter could be helpful and could be provided together with the cloud top temperature development.

Overall assessment:

The RDT product in its present form is difficult to use in manual nowcasting and warning because of issues within the placement of the active cell core, the erratic behaviour of the past and forecast motion vectors, and the size of the cells, that often included inactive parts of a large blown-off downstream cirrus shield.

Elements of the RDT product still have significant potential for the nowcasting of severe convective storms, namely the cloud top temperatures and the overshooting top detections. In combination with improved cell tracking - during daytime the new microphysics convective rain rate output can be a good candidate to extract the most active cores from - these features could be combined into a more powerful product.

Current RDT performs reasonably well in an environment of little high clouds and fresh cell developments without strong shear. The larger a system grows, the more cells are merging, and the more high-level clouds are present, the less reliable the output is.

Example:

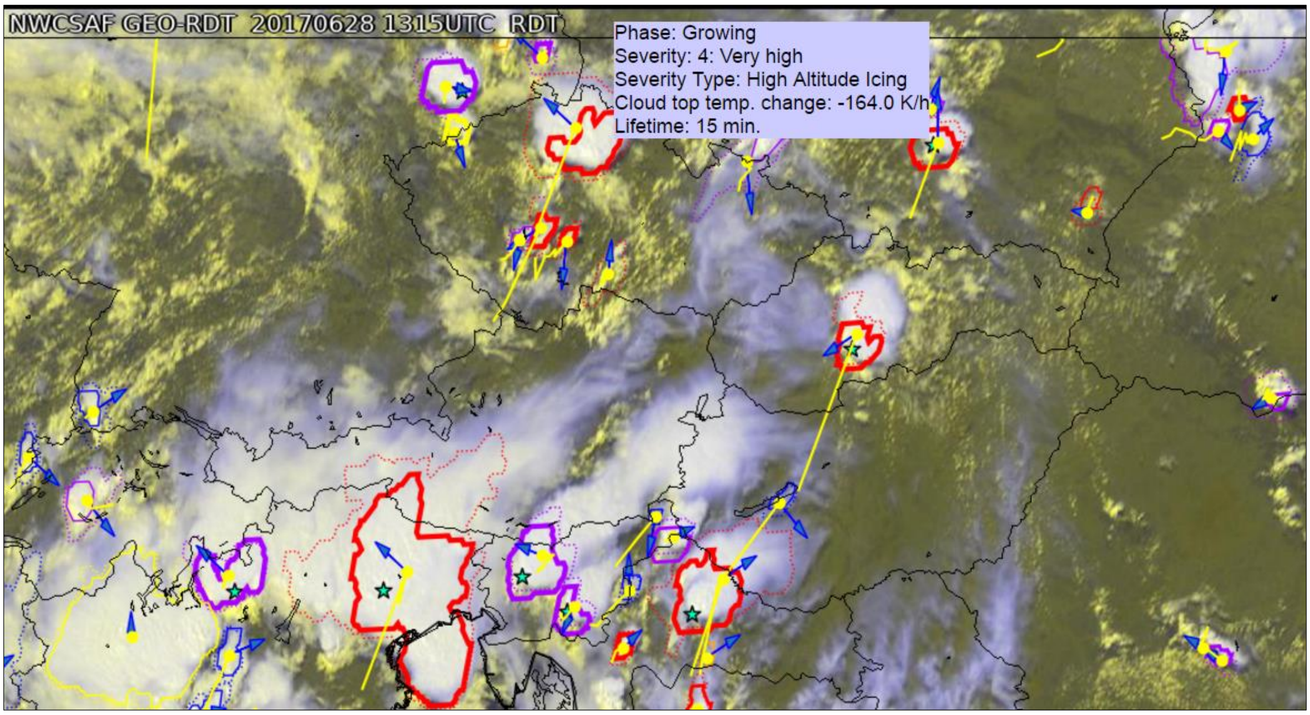


Figure 8: The RDT product best performs in situations with fresh and rapidly growing cells in an environment of little high level cloudiness. ESSL Testbed nowcasting display screenshot with mouse-over function of 28 June 2017 13:15 UTC.

IV. Summary

The **microphysics version of the convective precipitation rate** (daytime product) was strongly appreciated during the discussions at the ESSL Testbed, and this is also reflected by high ratings in the questionnaires. The product comes surprisingly close to radar signatures of convective cell structures.

High potential is also seen for the **products that cover the pre-convective environment in clear air**. Future sensor improvements are expected to further enhance the usefulness of these products (Instability Indices and Precipitable Water Products) because of higher resolution and precision of input data. At this testbed only combined products of model and satellite data were tested. In future testbed editions difference products between model and satellite data would be interesting to evaluate.

It can be confirmed that the **convection initiation** product (CI) still is in an experimental stage, as stressed by the NWC-SAF developers. Suggestions for possible improvements were made.

The **rapid development thunderstorm** product (RDT) includes valuable features like detection of overshooting tops and tracking of cloud top temperatures. In combination with different cell tracking options this product has potential to be of value for manual nowcasting, while the tested version is mainly interesting for some users with limited data access.

Cloud top temperature and cloud top height products are of high interest for nowcasting, foremost in the aviation sector. Adapted displays and colour scales could be tested at future editions of the ESSL Testbed.

High resolution winds were not optimally displayed in the current version of the ESSL Testbed Nowcasting Display. In future, this could be improved based on best practice examples. As a derivate of this product difference vectors to model data could be of interest in the future assessment of vertical wind shear.

The **probability of precipitation** product is of minor interest for forecasting severe convective storms but can be of value in different applications.

The interest in NWC-SAF products at the ESSL Testbed was very high. A number of forecasters present at the testbed had never been exposed to them before this activity. The atmosphere during the participant phase of the testbed was very constructive, leading to honest positive and negative feedback, which is expected to help the developers in the continued improvement of their products.

Acknowledgements

Special acknowledgement goes to the colleagues from AEMet for the good collaboration during the whole project, namely to Pilar Rípodas Agudo and to Xavier Calbet Álvarez, as well as to Zsofia Kocsis (OMSZ) and Ján Kaňák (SHMU) for their support in the hardware and software preparation phase. We are thankful to our colleagues within ESSL for all of their help, especially to Tomáš Púčik, who intensely worked with the testbed participants to collect high quality feedback.

V. Attachments

Attachment 1 – Raw Feedback for Product Usability

This section contains all raw written feedback that was collected regarding product usability for a forecaster (manual severe storms forecasting), including visualization, behaviour and content of the single NWC-SAF products. No syntax correction or text editing was applied to the feedback.

Participants were asked to express their thoughts freely, as this is the most valuable feedback for the developers. Based on the answers below it can be said that direct and honest feedback was collected.

A1.1 RDT

The RDT is missing a legend. The interpretation of thicknesses and colors is anyway difficult. The meaning of the green stars must be explained.

RDT: The arrows and vectors associated to this product are not very useful because we could not interprete them. It seemed to be an error with the yellow lines and the blue arrows.

with RDT you can see very well the cells since the first moment of convection until of their dead, but you can not see if they produce severe storm/

The yellow lines on the RDT product can be very confusing.

RDT: Shading the area for the solid-lined areas could be helpful, possibly with a level of transparency.

RTD - at times, the tracks did some strange things. For example, sometimes the last point would be somewhere far away from all the other points.

Choice of colors for the different stages in the RDT could be more intuitive.

RDT: Storm movement should be measured from the cell cores.

For RDT and CI the legend was not available on the screen.

Very inconsistent from timestep to timestep - growing/ mature/ decay/ mature/ decay/ growing/ mature etc.

There are some inconsistencies between the developing phases of a certain cell. Also direction arrows seem to be wrong sometimes

The display is very difficult to read.

good, but very difficult to read under time pressure

In general it catches the rapid developments. The attributes can be helpful to see the current state of the cell development. Sometimes additional care is needed to discriminate between convection related to pure orographical cumulus clouds and possible severe convection.

Interesting product, may be very useful for nowcasting or to recognize new developments immediately.

Very interensting product and this week had some good responses

Didn't really get the sense of the dashed line.

too many colour or different storm types (at least do not distinguish between splitting cells and triggering cells?)

The display is difficult to read.

Oftentimes this would label storms incorrectly.

The vectors of propagation often do not fit... maybe a combination with the high resolution winds is possible to get a better estimation for the propagation

not able to follow one storm

It could very very useful for nowcasters. I observed some temporal inconsistency (up and down behaviour).

The indication of overshooting tops with in the RDT was really useful, however, the RDT itself was a bit confusing due to the very jumpy appearance o the cells from one time step to the next, and the strange past motion and motion vectors.

I have no idea why one would look at this product

There should be a clearer indication about how display the data quickly and efficiently. It includes many parameters, and the novice user does not have much idea of what can be more interesting there.

Very difficult to use. Much changes between time frames. Hardly any added value in some situations.

inconsistency is the major problem

from one timestep to the next the arrows can jump 180 degree, so it is misleading. just focus on the rapid developing cells, and leave the rest (ie. mature, decaying)

too noisy

Quite a lot of information in a single product.

A1.2 CI

CI - I acknowledge that smoothing gives the product a bit more of a "cellular" appearance, but when objects are smaller, the CI product looks very noisy. Has an unsmoothed version of this product been evaluated? If so and it did worse, disregard comment.

CI - the display is fine, but this product gives MANY false alarms Cloud Top Temp - the black and white color scale is very difficult to read.

CI: too patchy with too much pixels, should highlight more the really significant cells

In the CI product, it was quite confusing that the marked cells had very sharp edges and that the product often only showed filaments. (This might be due to the masking a certain areas.?)

CI: The output is extremely noisy, developing C/TCU clouds can be better seen from ordinary HRV satellite products

CI-fields could be smoother to reduce the noisiness

It does not work well, but it would be a great product if it worked well.

Looks decent with some decent lead-times in certain spots of today's convection but misses or false alarms in others. Has a rigorous climatology been done?

It could not help too much due to its current experimental phase.

No experience with that. But how can the satellite know if a cumulus is growing to an CB or if there is an inverison? Does the product produce many false alarms?

too much flickering for me to interpretate the image. Probably show only the maximum or the strongest storms respectively?

Many false alarms - maybe some value in extending the time period over which to use this product

CI: simply too many false alarms, this product should not be considered operational yet. see example of 22 June 2017 at 21:30 UTC: Munich storm missed, but many false alarms to the W.

no logic, especially if high clouds are there

High potential, but I have not checked it enough.

Seems to be easily fooled by mobile cirrus clouds

too many false alarms

Absolutely too noisy, no confidence.

Many false alarms, sometimes also misses.

too much false alarms

A1.3 TPW and layer PW

For PW, ans the instability indices better colorscales can be used.

TPW - the colors should be more transparent because they obscure the coast lines and the borders.

PW: for me total column is enough

I like how this product looks.

It is very good for air masses developments triggered by diurnal heating in a cloud-free area. It is really useful to see differences with the model and anticipate the coming convection. For zones with pre-formation of cloud shields or over pixels in between cloudy areas it seems to be contaminated. I would suggest to exclude these areas of the algorithm because it could lead to confusion.

usefull to compare TPW data from the model, even with is limitation (cloud area) it allows us to see what is ahed of the area of interest

not very useful at the boarder to cloud shields (maybe add some space around clouds without calculating TPC?)

The display is difficult to read.

ok

This is great!

I really cannot trust these products which are using satellite based temperature and moisture profiles.. They tend to be bad at BL, which is the most important one f.ex. severe thunderstorm forecasting

Can be good in cloud-free areas, but just very little added value in areas that contain many clouds.

good to know enviroment where the cell is moving to

nearly no additional information to model data

Having a product that shows the difference with the models is useful.

It helps to see which levels contain the humidity. The low one could be useful to see low-level jets that bring moisture

maybe will be useful

for all clear air products: see if you could mask out a little more of the edges near the clouds. there seem to be 1 or 2 pixels that typially are not ok near to the cloud edges (maybe just add a certain distance from the cloud mask edge)

A1.4 Instability Indices

Especially for the LI, it is important to have bigger contrasts just above and below 0.

higher resolution of lifted index!

LI, Showalter, K-Index: more focus of scale on most interesting values

These products are relatively basic. Could more mathematically rigorous products be evaluated?

too many fails

partly interesting for a short owerview

It is necessary to know how the lifted index is computed to make a realistic comparation with the predicted soundings. In general it is useful to anticipate potential unstable areas.

In my opinion difficult to interpretate, because you don't know how much model input there is.

good product, but if you don`t have any clouds coverage is not so good

Only use the LI, pixel contamination and clouds are the product limitations

I do not use that.

I think Showalter and K-index are useless, but LI is really nice. It would help to fill in the cloudy data voids (perhaps with a model background/first guess field)

Gives a broad overview. Added value possible, but very limited in partly cloudy situations.

In my opinion these indices are the most useful, but it would be nice to expand them somehow to cloudy areas as well...

nearly no additional information to model data

A1.5 Convective Rain Rate and Accumulation

The colorbar for some of the precipitation products was not appropriate. Very often the cells reached the maximum and the cells had an empty spot in the middle. (Remark: This was a bug in the ESSL Testbed display that was fixed after the first days of the testbed.)

Convective Rain rate: other color scale

ConvectiveRainRate: Visualization is already quite good, but needs more resolution in the high rain rate values and maybe some smoothing to make the visualization nicer.

Didn't really use when radar data are available

only important in areas without radar

Very useful to see where are the most intense precipitating systems. Reasonable agreement with radar signal.

No experience with that, but seems as it could be very useful when there is no radar.

Good addition when radar not available, overestimate prep.

It has a good looking. I like it, especially in areas of poor radar coverage.

The new microphysics version really looks good. This is a true alternative to radar in remote areas. The older CRR version was not as good, and many cases there go off-scale, like on 8 June 2017 at 11:50 UTC over Turkey and SE Europe.

Particular useful in the areas where radar coverage is limited. The product depicts the cell cores quite well. Some interesting procuts like "2h maximum rain rate track" could be very useful for detecting the movement of CB cores.

Good addition when radar not available.

usable when no radar data available

Good visualization

The microphysics version of this product was very nice - where radar wasnt available this was an extremely useful product to understand storm cores and storm mode. The color scale could be improved to delineate between categorizations.

The new microphysics version really looks good. This is a true alternative to radar in remote areas. The older CRR version was not as good.

A1.6 Cloud Top Temperature

CTT colorscale is useles,s because one cannot read the values.

Cloud Top Temperature - the visualisation should be in colors, and not black-white. Maybe adding a certain threshold under which black-white display is used.

Cloud Top Temperature - the Black/Gray color scheme is terrible to diagnose subtle differences in cloud top temperatures. A color scheme would work better.

coloured cloud top temperature, isolines?!

Concerning the CTT the scale colours do not help too much to see the temperature differences

Cloud Top Temp: Think about the gray tones and the steps. Could be helpful to go a bit coarser.

Cloud top temperature could be coloured for better optical differentiation of the temepratures

Cloud top : use a specific scale for Cb with a coloured scale (the same as for satellite IR-convection)

CTT/CTH: using sandwich product instead is quite more efficient!

CTT: greys scale should be larger.

Cloud top temperature: greyscale is useless. Use enhancements from sandwich product by Setvak.

Cloud TOP products: The visualization should have much more resolution on the cold CB tops. Otherwise almost the whole CB anvils will be over-saturated giving absolutely no information.

Cloud top temperature: Change color scale (more colors)

For CTT the colours palett could be improved.

Make color scale easier to read for cloud top products.

I am not sure about the accuracy of this product, but if it were accurate ot would be of great help in my work.

Poor color scheme made this product unusable.

visuallization not good, parameter sencefull

Not very useful because of the scale. It makes difficult to appreciate differences.

Is very very helpful, but on the ESSL Nowcast Display the vizualisation is not good.

The Top Temperature scale should be clearer and more visible for be useful: not the grey scale now.

I think a mouse-over readout function for the temperature values would be best for operational forecasting. The product has a big potential, but the current colour scale is terrible. You need to be able to especially distinguish temperatures in the range between -20 and -70 C. You could use the BT enhancements used by Martin Setvak from CHMI.

With a better visualization very good product

Good.

usable additional information but hard to read colour (grey) scale

A1.7 Cloud Top Heights

CTH colorscale should have better contrast around tropopause height, the range where it matters for aviation forecasting.

Cloud Top Height - when you hover over a point with your cursor, a pop-up window should appear, indicating the clodu top temperature and height at that point.

The scale/legend for Cloud Top Temperature and Height is much too imprecisely.

Cloud Top Height: Might be helpful only to have negative values coloured in red.

Cloud Top Height - perhaps only use one color with different shades for this plot?

Cloud top height: the most interesting heights for intense thunderstorms are not distinguishable (light grey)

change colour scales for many products - more clear distinction between the levels highlight most important levels more

I found it odd that an area with an overshooting top had a lower cloud height that the anvil. It made me initially very weary of the product.

operively very usefull for aviation

the best saf product

Good to see the evolution of tops but the pixel size is too big for single storms.

check example of 23 June 2017 at 0:30 UTC: were the clouds really not getting higher than 13 km near Munich?

The Top Height scale should differentiate better (with more different colours) layers above 10000m, to better detect overshooting tops and maximum height of cumulonimbus.

Make the heights for the range from 10 to 18 km much more distinguishable.

With a better visualization very good product

Good. usable additional information

A1.8 Probability of Precipitation

good

only important in areas without radar

Maybe useful if radar is not working.

Coud be a very useful product. But does it recognize drizzle?

It seems to be quite OK.
Good addition when radar not available.

not much additional information

PoP (cloud physics) seemed to provide a better look/correlation to radar reflectivity than this classic version.

A1.9 High Resolution Winds

HRW barbs should not be plotted for each single barb. The barbs were sometimes overlapping in such a way that they became unreadable.

In some images the arrows are so close that is difficult to read them.

High Resolution Winds - sometimes the wind bars are so clustered together that you can't read them.

High Resolution Winds: The legend should be the other way round, means bottom of atmosphere down and vice versa

High res winds: only plot 400-550 hPa and 850-1000 hPa winds, so that you can compare with 0-6 km bulk shear from models.

HRW: In cases the wind is strong is difficult to see clearly.

Highreswinds: use more colours to better distinguish the levels.

High Resolution Winds: Potentially very useful tool for forecasting. Visualization is too crowded.. much less vector should be visible.

High res winds: Lessen density of displayed arrows

Wind field would be easier to read if contoured or smoothed.

It could be useful to generate an additional product about vertical wind shear, in cases there are several winds in the same vertical. Considering this, the most interesting data would be: 0-6 km wind shear, 0-3 km wind shear, 0/1 km wind shear. It would also be good to present the wind data in an additional way, as a function of mean values (in a region) and outliers.

With a better visualization very good product

Nice tool, good display when number of arrows is reduced. Can give good information about jet streaks to compare with soundings in model data.

usable additional information

Maybe fewer wind barbs would make maps more easy to read.

Really liked this product - could be visualized better.

Very helpful to detect divergence zones in upper-air levels and imagine which motion is following the storm (especially if there is strong wind shear). In some cases the low-level fluxes can be distinguished.

yes it can be calculated but why? you can see direction of movement from animation, and this information is quite messy

for nowcasting not really sencefull

Often impossible to recognize wind speed because of too many wind barbs.

A1.10 General Feedback

How would a pop-up sounding look derived from these NWC SAF data? How would it compare to real soundings? There may be potential here for future development.

the legend should be more described, on the first and second view, it`t not possible to get to know at which border it seems to be a severe thunderstorm or not.  at some index it is too much coloured   please describe the single products, it`s hard to get a feeling what we see

The SAF products compared to radar information or high resolution models are worsted. In regions without this data, this products can be helpful.

Especially Cloud Top Temperature and Height are very useful and I use them nearly every day when I am at work. RDT product is sometimes confusing for example with the movement of the thunderstorms.

Lots of the products are quite more covered and estimated much better by local radar products only in case of technical problems on that I would shift to those satellite derived products

Five days is a short time to check this products

Some of these products have a high potential but they are not very known among forecasters.

From my point of view it is very difficult to estimate the product in such a short period (less than five days). It would be useful for the forecaster to choose at least two product before testbed and to follow their 'behaviour' for a month . After that we could make a better evaluation of the products.

The operations plan document should contain much more information, however in a condensed way.

Looking forward for further improvements.

Some products could be more useful not directly visualized, but processed in a way that the bias of the model predictions could be evaluated.

legends are missing for many parameters it would be more interesting to have a difference (model - nwc saf) map

Most of the products work only in cloud free areas, which is usually not an area of interest. Therefore not that useful on operational work except in a few cases.

Testbed participants in week 1

Attachment 2 – Raw Feedback for Engrossed Evaluation Products

A2.1 Instability Products

Strengths of instability products

In theory the LI is quite useful.

Algorithm reveals cloud free areas with calculated index

Today The Alps is the area where this product of clear-air can be checked because is not completely cloudy. The LI and Showalter indexes are well represented in the areas where lihtning storms have developed giving a signal between 0 and -4 C.

In general it catches the pre-convective enviroments.

We have been using just the LI product. Strengths on Wednesday: values did seem to become higher in locations where convection developed. This seemed to agree with where the warmer moister surface air was in place.

It provides independent information from model forecasts, which is important in nowcasting.

07.6.2017 N Italy LI gave an indication of unstable stratification exactly in the area, where the only storm of the day developed.

Case: 8.6.2017, Region: Turkey LI: Initiating storms correspond well with the negative LI, implying unstable atmosphere. Showalter Index: The same as above, even though index suggests only slightly unstable atmosphere, despite vigorous initiation. K: Captures very well that thunderstorm formation should take place over the Turkish plateau instead of the Mediterranean Sea. STRENGTH: 1. Seems to work 2. Provides information about the instability in realtime, supplements sounding information very well.

Appear to be useful for environments with no storm development and limited cloud (thus a pre-storm clear sky scenario), however the coverage is frustrating obscured which limits utility.

Can be useful for first guess in cloud free situations - LI had good guess for Cb, but K index was good for isolated convection where CAPE and shear did not (Hungary and Croatia)

Weaknesses of instability products

The K-index is an ancient index that has no clear physical meaning. I cannot work with such indices. Why isn't there a CAPE field: that would be most straightforward. The contribution of the satellite to the estimates of LI is unclear: likely most of the signal comes from the ECMWF model, given the limitations of MSG. The colorscale for LI should show better contrasts just above and just below zero. It is now essentially only possible to say that LI was either above 0, between 0 and -4, between -4 and -8 or below -8. There is little contrasts between a value of, say, -0.1 and -3.9, which is however a big and important difference for convective forecasting.

Not visible in cloudy areas limits of defined indexes

K-index in the same areas than the previous ones appears to be more noisy and less clear in their values.

A small layer of cirrus can cause that the products do not give any signal.

Product is missing in any areas with cloud cover. Unfortunately, where storms are most likely there is often widespread cumulus and cirrus cloudiness, so we had no information. Also, we find that LI is hard to use without information about CIN.

No weakness, except that it can only be used in cloud free areas, but this is ok and understandable from the physical limitations. It would be good to have CAPE in addition as a product.

Showalter and K-index did not indicate any instability in this case.

Interesting areas often are cloudy, not much additional information to model information

Information only in the clear skies. Should be used only in addition to other data, such as the sfc observations and soundings.

Concerns about how representative profiles derived from atmospheric sounders are for the pre-storm environment - does not appear to give any information that is superior to operational models at least in the available cases. Given lowest half of the troposphere is of the greatest importance for severe thunderstorm development, it is unclear if these products are providing any forecast usable information.

It's working only in cloud free situations K and LI index didn't catch severe convection in N Italy Lots of false alarms, especially SH

Other comments and suggestions for improvement

Improve colorscale. Show deltas w.r.t. to model, in order to estimate the influence of the satellite on the instability estimates.

K: stronger difference between K blue and gray clouds

The instability indexes can be also infered from the models and other satellite observations. They do not deliver so much information as other nowcasting products like the RDT. It would be interesting to see how big are the differences between models and satellite observations.

If a model is being used as a first guess field, perhaps the model values could be shown in areas where satellite data are missing.

Make difference products available to model forecasts. Provide CAPE output. Concentrate on the most interesting values for convection forecasting regarding the colour scale. Make values more distinguishable where it counts most.

Difference maps (model to saf) more interesting for detecting areas where something may go "wrong" with the model prediction

Change in the color scale: It would be worthwhile to highlight values of LI already around 2K to suggest that there may be instability below. Color scale would also need more steps, to highlight the changes in the LI. The most important range of LI, to which most color should be concentrated between +5 and -5 K.

It should be less signals, less detections because it produces too much false alarms. Don't see enough added value at all :)

If satellite based profiles can be verified better for these products in near convective environments, then this might be a useful tool for between observed rawinsonde profiles, however, the issues with even thin cirriform layers obscuring any sort of sounding certainly limits utility when there is appreciable warm advection or large scale dynamic forcing. Maybe better for isolated thunderstorms during the summer?

A2.2 Precipitable Water Content in Clear Air

Strengths of precipitable water products

The low-level humidity and mid-level humidity have significance in convective forecasting. total column can say "something" about heavy rain risk with storms. So it is important that satellite-based estimates of these quantitites are used by forecasters, provided they are reasonably accurate and add value to NWP guidance.

Revealing of very dry areas good for finding qualitative area difference for convective ingridients.

We have found a high accordance between values of total PW from NWC SAF and measurements from soundings in different parts of Europe. We were not able to check the Spanish soundings because they are not available in the Wyoming site. We have no way to check PW in the different levels but is a very valuable information in convection forecasting.

Useful to see how a system probably is going to develop

Really additional information that cannot be derived from other sources independent from model data. Very important in the pre-convective environment where the vertical extent of a mois layer is of importance and could otherwise only be analyzed from surface data.

Good product to see moisture in clear air, especially how is distributed vertically

Good idea to differentiate the different levels and its water content.

Better then nothing (if you have no model and no observation data) most recent information about the state of the atmosphere

Weaknesses of precipitable water products

It is questionable how much added value the product has over NWP guidance, since SEVIRI is not able to assigen the detected moisture to a particular layer of air (low, mid or high). For low-level humidity, it may be that we are essentially looking at the corresponding values from ECMWF model output, without much input from satellite. To find this out, it would be more useful to plot differences w.r.t. to NWP output.   For the polar-orbiting sounders and the sounder on MTG, this product may be much more useful than it currently is, as these sounders are/will be much better.

Not defines where PW is mostly released

We have no found discrepancies between real data and the NOWSAF products. Clearly, there is no information under cloudy areas.

Doesn't work over clouds Pixel contamination can lead to some wrong conclusions

The limitations of the product are clear, but I see no weakness.

none

Since thunderstorms usually develop in very moist airmass, there is usually some cloudiness in the region. But the product works only for clear air pixels.

Not much additional information to the model data (values don't change so quickly) only available at clear sky

Other comments and suggestions for improvement

Can be used to compare with the ECMWF TPW product.

The direct difference product compared to model data would be most useful for manual forecasting, so that you can immediately see where the satellite data deviates from model forecasts. Please continue to develop this product, I would very much like to have this in my weather service.

The color scale should be revised (green and blue colors mean more moisture than yellow).

Better: have difference (modelled PW to satellite PW) chart for signal of possible failure in the model data.

A2.3 Convection Initiation (CI)

Strengths of CI products

If the CI product would work, it could help detect glaciating tCu early and point forecasters attention to that. That would be really useful.

Initiation of CB in areas free of other clouds.

Did quite well in a low shear environment where anvil contamination did not mask downwind development.

In nearly cloudless conditions, product gives a good first hint on convective systems, but may be more successful if longer time frames are considered.

In the current version there is no strenght, sorry to say. We were told at the Testbed, that this is experimental, so we were warned, which is ok.

N/A

For first overview: detects areas where convection may happen

Good guess for 08:45 in Romania - TS at 09:15

Weaknesses of CI products

The product simply doesn't work yet. Most detections are false, and most initiating storms are not detected.

Area with large anvils will not be good recognised. It cannot recognize CB that are very fast growing within 30 min.

One product should be generated each time a new satellite product comes in. Rapidly developing convection is missed completely or geographically misclassified in strongly sheared events such as today in south Austria.

There seem to be many false alarms. Also, sometimes for very rapidly developing systems there is little to no signal. There also seem to be problems - especially for mid-high level clouds with false alarms and/or the algorithm turning on and off. Sometimes in these situations storms did eventually develop, but after 30 minutes.

It has far too many false alarms. And it does not highlight the really most active areas. It seems to have big problems with the edges of anvil clouds and with areas of debris clouds or mid- and high-level clouds. I have seen a much better product form Chris Bedka from the USA, maybe you can try to cooperate.   Examples: 8 June 2017 over W Turkey and Aegean Sea: many false alarms, initiation before hailstorm (around 10:45 UTC in ESWD data) not really highlighted.

It appears to have little application for operational forecasting, and the detections that it has seem to be missing the actual development of large thunderstorms. Extremely noisy with many false alarms, the vast majority of which do not show any large convective development. It also seems to produce CI signals in close proximity to large isolated updrafts with expanding anvils, and this is accompanied by subsiding mass flux response around these storms, and so hence whatever it is doing is not physically realistic.

Too many false alarms too ambitious to detect every cell - better to show an wider area

Lots of false alarm areas (Alps, S Germay 06.06.2017. 07:30-09:30, Bosnia, Croatia 6. 6. 2017.8:30-9:45) Not consistent - changing sometimes every 15 minutes Didn't catch some big convection in Slovakia at 12 UTC

Other comments and suggestions for improvement

I don't know, but that a much better working version has been develped in the USA, so in principle this must be possible.

Maybe to plot dashed area where there is let say 90% of possibility that significant growing will not be or where it CBs are decaying

A product update every time a new satellite image comes in.

Better calibrate - lessen false alarms - extend forecast lead time.

This products needs a basic and major overhaul. CI approach from Chris Bedka (I am not sure at which US agency he is working at) could be a better way to deal with it.

This needs a lot of work to make it useful for operational context and considerable filtering to improve the POD/FAR ratio. Perhaps for CI - the rapidly developing (delta T cloud top) might be a better product to consider.

Probability for next hour better than next 30 minutes (delay of satellite image)

Maybe CI should be forecasting for a little bit longer period, 60 min or something like that

A2.4 Rapid Development Thunderstorms (RDT)

Strengths of RDT product

Recognizing cells

Rapid developments are very good detected, the product offers good signals especially in the early stage of the convective process.

Useful in initial stages in storms and their systems. Useful as a back up product in case there is no radar data.

Can be of use for aviation: if pilot does not get realtime satellite imagery the extent of the Cb anvils can be of value during flight on track, e.g. over the ocean.

Detecting of overshooting tops and tracking the storm. Prediction of new position and size of the cloud.

Picking out the very first moments of a growing cell is good.

Better than nothing specially when having no radar data useful for one single point / small area first hint for area of interest.

Weaknesses of RDT product

I have the feeling I better assess severity than RDT - RDT has many false detections (cirrus), as well as misses - RDT changes from one phase to another all the time, e.g. mature-> growing-> decaying->mature, etc. - RDT claims to assess a severity, but I lack confidence in the accuracy of this assessment - RDT claims to assess the primary hazard, but again, I have no confidence that this is done accurately - None from NWCSAF staff attending the testbed was able to explain clearly how this product assesses severity and the type of primary hazard, in order to build confidence - the use of the outer level contour of each detected feature was unclear.

Not intuitive, arrows in not good directions.

The signals during the lifetime seem to be less reliable, we got for example some signals of mature or decaying stage, although the storm was growing.

Some overestimation cases/false positives (especially in ongoing situations). Some problems in detecting core of a storm in cases there are high cirrus clouds.

Mostly useless for manual forecasting. Severity and severity type are often not available in the most important situations when very intense cells are present. If available: "mature" says little about what to expect. Examples: cases of 6 June 2017 over N Italy. The active parts of the cells are rarely captured, instead a large cloud shield is included over Austria causing a strange cell center and tracking behaviour, e.g. at 14:00 UTC. 8 June 2017, 12 UTC: W Turkey: product catches more the anvil shields than the active precipitation cores. New developments are hardly catched. There was no triggering from a split, as the product suggests at one area. 2 June 2017 over the Alps: some cells are nicely captured, other very small ones not (compare with lightning data), for example at 11:40 UTC. Quasi stationary cells over mountains seem to produce a difficult tracking.

Inconsistency of the severity and stage of development, also for direction of the movement. Examples: 8.6.2017.
Greece 11:15 very high severity and growing phase 11:20 phase decaying, severity zero
Romania 11:20 growing, very high severity 11:25 decaying

Too much information, and vectors are pointing in every directions. How beneficial this product could be, when the cell is already mature?

Hard to follow cells in a larger area (too much information in one image) cells are sometimes too big (cirrus shield) - where is the severe weather?

Other comments and suggestions for improvement

In absence of radar, RDT would be potentially useful if... 1. it is stripped from all the unnecessary and unclear classifications 2. its detection becomes better, e.g. a detection of objects in the CRR field.

Some signals are given for low-/mid-layer cumulus that are not growing further, a treshold for the absolute value of the cloud-top-temperature seems to be a solution for that problem. Question: If the first signal in a region is a decaying one, is the information then helpful? The movement-arrows of the system/cells could be better configurated as the directions are jumping even in systems that have obviously a clear movement direction.

It takes time to know the product, it has a lot of information, a bit confusing at times - especially the tracking of cells (yellow lines and blue arrows).

Try out the following: Use the highest intensity core from the cloud microphysics version of the convective rain rate for the cell tracking (which should lead to a considerably smoother track). Then look for the lowest brightness temperature in the environment and plot the curve of the BT over the past 2 hours next to the actual reading of the latest IR BT minimum temperature. Showing the borders of the anvil clouds is useless for manual forecasting (because you can directly and much better see this on the regular imagery), this might only be of interest for pilots on flight that do not have the full imagery available.

Need legend and explanation of the symbols (stars, etc).

The goal of this product should be focus on the appearance of stong, new cells instead of already existing cells.

Put a legend to it. 6 June hail cell southern Czech republic cell signal disappears at one time step. Then reappears 10 minutes before severe hail (13:15). Question: why did the gap happen?

Online briefing supported by EUMETCAL, demonstrating NWC-SAF products to remote participants around Europe.

Attachment 3 - ESSL Testbed 2017 Operations Plan

See separate document.

Vivid discussions about NWC-SAF products as part of the comprehensive suite of products available at the ESSL Testbed, putting products at test within a realistic forecaster environment.