African Satellite Nowcasting Intercomparison Project (ASNIP) Final Report

Authors: Alexander Roberts[1], Jakub Lewandowski[2]

1 National Centre for Atmospheric Science (NCAS), University of Leeds **2** School of Earth and Environment, University of Leeds

1. Purpose

This document outlines the work undertaken as part of the African Satellite Nowcasting Intercomparison Project (ASNIP). ASNIP is a small project aimed at providing a comparison between two storm nowcasting products. The first of these products is the EUMSAT NoWCasting Satellite Application Facility (NWC SAF) Rapidly Developing Thunderstorms (RDT). The second is the newly developed Thunderstorms product (DWD-TS) from the German National Meteorological service, Deutscher Wetterdienst (DWD). Both products aim to make use of EUMETSAT geostationary satellite data (Meteosat Second Generation; MSG) to identify the location of active deep convective storms. However, they use differing approaches to reach the same goal. Therefore an understanding of how these two products differ from one another provides insight into how they could and, indeed, should be used operationally. In addition to this, the development of techniques for such comparisons will likely prove useful in the future due to the proliferation of satellite based nowcasting products potentially available to forecasters in Africa. This proliferation is expected due to the expansion of machine learning approaches into the domain of African nowcasting and the dawn of the Meteosat Third Generation (MTG) era, where the quantity and quality of near real time satellite data is rapidly improving.

2. Introduction

There is a desperate need across the whole of the African continent for good quality weather information. Especially information that can be used to make decisions to mitigate against the effect of high impact weather. Numerical Weather Prediction (NWP) is known to have poor predictive skill in the tropics. There is also a heavy reliance across the continent on global models (with relatively coarse grid spacing and parameterised convection, EUMETSAT, 2024). By their nature simulations of this type are inherently poor at representing deep convective storms meaning that there is little capability within African national meteorological services to predict and provide warnings for high impact weather associated with deep convective storms Vogel et al., (2020). Such storms represent a major risk factor across most of the continent due to the effects of heavy and persistent rainfall, lightning and strong winds. This is especially important due to the highly vulnerable nature of many African communities to such weather (WMO 2022). Despite satellite nowcasting products being available over more than a decade (generated by NWC SAF software) there has been a slow uptake in nowcasting across the continent (Roberts et al., 2022).

The Global Challenges Research Fund (GCRF) African SWIFT (Science for Weather Information and Forecasting Techniques; Parker et al., 2022) and the Weather and climate Information SERvices (WISER) Early Warnings for Southern Africa (EWSA) projects have sought to highlight the value of the satellite nowcasting approach as well as identify barriers to African nowcasting (Roberts et al. 2022). Much work has been focussed on enabling African meteorological services to operationalise nowcasting including the production of guidelines for satellite based nowcasting in Africa by the World Meteorological Organization (WMO). A successful implementation relies on a holistic, end to end approach including the: (1) acquisition and processing of satellite data, (2) production of nowcasting products, (3) examination of products, (4) warning production and dissemination, (5) understanding and

communicating uncertainty and (6) understanding the needs of users and their capacity to understand warnings and act accordingly (WMO 2023). All these steps should be combined to create a nowcasting system that can meet the needs of users and allow for warnings to be generated and issued in a timely manner.

3. Rationale

The work undertaken in ASNIP focuses on just a small part of this endeavour, the use of extant nowcasting products. It is likely that within a robust and successful implementation of a satellite nowcasting system for Africa that a variety of nowcasting products would be used (with some products being circumstance, time or location specific). The use of products generated from a diverse range of approaches (even if they largely utilise the same input data) can provide nowcasters with additional insight into present and near-future storm behaviour. With this in mind it is important to make objective comparisons between products (especially with a lack of sufficient observations to act as a ground truth) to help identify behavioural, regional and temporal differences between them. Knowing how such products differ (and how they are similar) is useful for nowcasters and will inform the subsequent steps in the nowcasting chain. Here we are comparing the well established RDT product and the new DWD-TS product. These both aim to identify the location of deep convective storms and make short term predictions about their position in the future.

In order to see how such comparisons might be of use, the main target period for the ASNIP work coincides with the rainy season in Southern Africa including the 2 week long WISER-EWSA testbed based in Lusaka, Zambia. Therefore, ASNIP daily comparisons have been made for the periods preceding and over the testbed (26th Jan - 2nd Feb 2024). This approach meant that there could be objective comparisons between the main 2 products, but also there could be less formal, subjective comparisons with other nowcasting products being used within the testbed.

4. Data

4.1 NWC SAF RDT-CW

The NWC SAF (NoWCasting Satellite Application Facility) is one of a number of EUMETSAT funded SAFs. These are distributed centres designed to generate tools and products for the better use of EUMETSAT satellite data. The NWC SAF nowcasting software (NWCSAF-GEO) comprises a number of different product generating elements (PGEs) and an overarching task manager that are used to produce near-real-time nowcasting products from a combination of geostationary satellite and NWP data. Of the products that are available from the NWCSAF-GEO software one of the most useful (and most commonly used) is the Rapidly Developing Thunderstorms - Convection Warning (RDT-CW) product commonly referred to as RDT (https://www.nwcsaf.org/rdt_description).

The purpose of the RDT product is to identify convective cells, detect which of these are likely to be associated with deep convection (currently and in the near future) and produce short term predictions of cell movement. The algorithm identifies cells using an adaptive thresholding technique applied to brightness temperatures from the 10.8 micron satellite channel. This approach is useful as the application of a single threshold is problematic. If the threshold is set too cold, then cells are not detected until they are very well developed, whereas a warm threshold produces cells that are caught earlier but quickly merge into other nearby cells and provide little useful information on the behaviour of areas

of interest. Each identified cell in the RDT product is defined by its own specific threshold. This threshold value has to be between two set thresholds (a warm and a cold threshold) and cells must meet a number of additional criteria to be classed as a cells of interest. This approach allows for the identification of not only mature and well developed cells but also those that are rapidly developing.

Once cells have been identified they are tracked in time. This part of the algorithm is based on a cascading area overlap approach. The algorithm also handles the splitting and merging of cloud cells previously identified. The aim of this part of the algorithm is to make links where possible between cells identified in the latest image and cells from the previous image. The first step attempts to make use of the estimated speeds of cells from the previous images. Cells from the previous timestep are forward projected and the area overlap between this prediction and the latest image is assessed. If there is sufficient area overlap a connection between cells over time is made and the latest cell motion is added to the previously identified trajectory. If this approach is unsuccessful for some cells then a backward projection of the cells identified from the latest image is made (using the group velocity of all connected cells locally). Once again if the area overlap threshold is met for particular cells over the timestep they are linked, else a third attempt is made. This time the cells are backward advected and enlarged to increase the chance of overlap. If no links can be found, cells are identified as being newly triggered and are considered the start of a new trajectory.

The cell detection algorithm in RDT can yield very large numbers of cells. This necessitates a discrimination process that needs to be applied to the detected cells so that only cells that are of interest are included in the final product. While this discrimination process can make use of lightning data (if it is available) the RDT data used in this study does not do so. This approach aims to vastly reduce the number of false alarms associated with the RDT product. Discrimination will often yield only around 1% of the identified cells. The approach used is a statistical method trained using lightning data over Western Europe combined with an NWP derived convective mask which limits the region of interest to parts of the image that have sufficiently high instability. The second aim of the discrimination process is to allow for the identification of a cell's storm life phase. Based on the characteristics of a storm (cloud top temperature, size of cell, cooling rate) the cells are grouped into different categories; mature, dissipating and growing. If a cell has not previously been identified then it is classified as triggering.

The final RDT file structure generated is a netcdf that contains polygons representing cell outlines along additional cell metadata such as storm category, intensity and previous storm track. Similarly structured files are generated that provide a forward projection of the cells identified in the latest image. This is a simplistic method whereby the storm motion vector is used to advect the cells into the future, but without any deformation of cell size or shape.

4.2 DWD TS

In comparison to the RDT product there is far less documentation on the methods used to produce the Deutscher Wetterdienst ThunderStorms (DWD-TS) product (https://www.dwd.de/EN/ourservices/fernerkund_thunderstorm_detection/fernerkund_thunderstorm_detection.html). Therefore the description below has less detail on the methods used than the preceding RDT section. However, the goal of the DWD-TS product appears to be

similar to RDT as it makes use of EUMETSAT geostationary satellite data (as well as global lightning retrievals) to indicate the position of deep convective storms.

Thunderstorms are initially detected by making use of the Vaisala GLD360 lightning product. The GLD360 is a global near-real-time lightning product that uses a distributed network of ground based lightning sensors (passive radio wave sensors) to provide information on the location and timing of lightning strikes. The radio waves that are generated by lightning can travel thousands of kilometres. By detecting the same radio signal at multiple locations and with sensors that can accurately record the time that the signal reaches them (precision GPS enabled timing) the location and time from which the radio waves started (the lightning flash) can be ascertained. To make GLD360 a global lightning detection product, the range over which the radio waves can be detected is increased by making use of the reflectance of radio waves by the ionosphere. This increases the detection range to approximately 10,000 km but reduces the accuracy of detection. However the accuracy quoted for the GLD360 product is within 1km of the location of the strike, this is significantly smaller than the pixel size of the current MSG data. For the global product over 80 % of cloud to ground strikes are detected as well as a significant number of cloud to cloud lightning, this suggests that for the purposes of identifying storms which are sufficiently convectively active to be generating lightning the GLD360 is more than capable.

Due to the global nature of lightning detection in the GLD360 product and the very low latency (approximately 30 s) it can be used to identify which storms visible in satellite imagery are electrically active (and therefore are likely to pose a greater risk of intense weather). Thunderstorm intensity is also based on the brightness temperature of the clouds and differences in the water vapour channels (neither of which are daylight dependent). This methodology is designed to ignore storms that do not meet the criteria of being sufficiently electrically active. The format of the output data differs significantly from RDT. DWD-TS is not broken down into discrete cells like the RDT product but is instead a gridded product (0.05° x 0.05°) of thunderstorm intensity (delivered as netcdfs). This intensity value is not related to any specific ground conditions (such as a rainfall rate) but is a normalised value between 0 and 1 that provides information on the severity of an observed storm.

The forward propagation technique also makes use of a different approach to RDT. Instead of attempting to predict motion through generating storm specific motion vectors (from the area overlap approach mentioned above) atmospheric motion vectors are produced using an optical flow technique. This approach is similar to other nowcasting forward propagation techniques such as the NWCSAF-GEO EXIM (https://www.nwcsaf.org/exim_description) and pysteps (Pulkkinen et al., 2019) techniques. Consecutive recent water vapour images are used to produce atmospheric motion vectors using a Farneback optical flow algorithm out to a 105 minute prediction. This motion vector field is then used to forward propagate the storm field. As with all nowcasting approaches that seek to simply extrapolate current conditions forward in time there can be no capacity to represent more complex behaviour (e.g. triggering and dissipation). Therefore, nowcast skill reduces over the timeframe of the extrapolation due to increasing numbers of false alarms and missed cells for longer extrapolation periods.

5. Methods

5.1 Pre-processing

As one product is object based and describes storm cells by use of polygons (RDT) while the other is a gridded product there is a conversion required to allow the products to be compared with one another. This conversion takes the RDT polygons that describe the location of storms and maps them to the same 2 dimensional latitude longitude grid that the DWD-TS product is supplied on. Any grid points that are within the RDT polygons are given a value of 1 while those outside the polygons are designated as having a value of 0. As such, a binary map of RDT storm locations is created. In a similar process the 2 dimensional thunderstorm intensity value given by the DWD-TS product is thresholded so that values over 0 are given a value of 1 and zero values are maintained. It was envisioned that there might be some need to tune the threshold level chosen for DWD-TS to produce similar detection levels between RDT and DWD-TS. However, even with the threshold set to 0 the coverage of the storms detected by the DWD-TS product was significantly lower than that from the processed RDT binary maps so it was decided to keep the detection level for both RDT and DWD-TS at the maximum possible. It is deemed a fair method of comparison as both RDT and DWD-TS include all areas of storms that were identified in their respective algorithms.

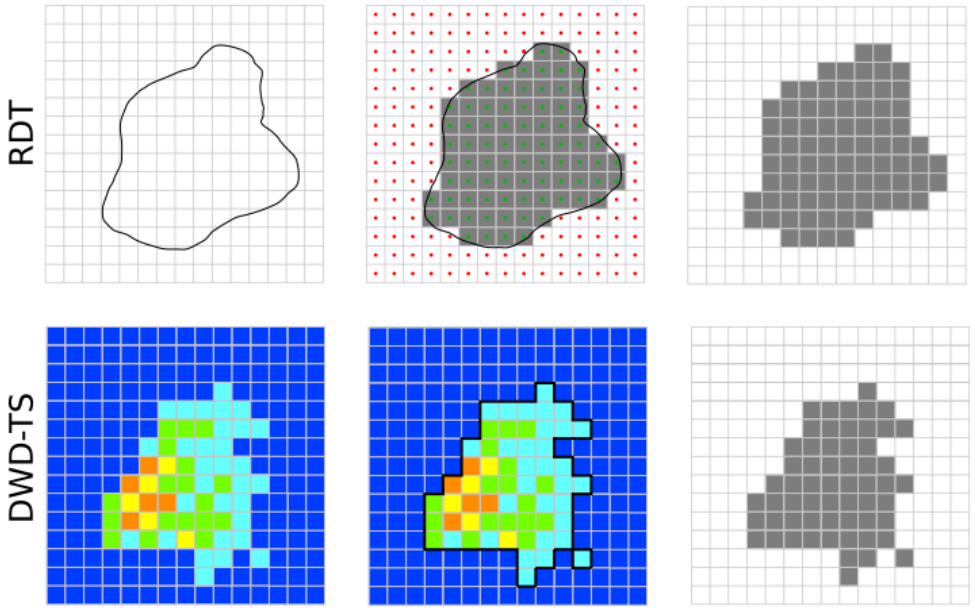


Figure 1. An example of how a single storm represented by an RDT polygon and the DWD-TS thunderstorm intensity product would be reduced to simple binary images ready for comparison with each other. RDT polygons are projected onto a grid (shared with the DWD-TS product) and a binary image created. Similarly the DWD-TS product (which is a dimensionless value between 0 and 1) is thresholded at a value of 0 and a comparable binary image is produced.

5.2 Study domain.

The region being studied for this project has been limited to a domain that covers the WISER-EWSA region of interest (Southern Africa). This is for 2 main reasons. Firstly, the work conducted as part of the ASNIP project was concurrent with the preparation and execution of the WISER-EWSA testbed. Secondly, despite both products being pan-Africa in coverage the period of time for which data was available (November 2023 - March 2024) is the rainy season for Southern Africa. Limiting the analysis to the region where high impact weather (thunderstorms) were present allows for a better comparison of the two products while reducing the overall computational burden.

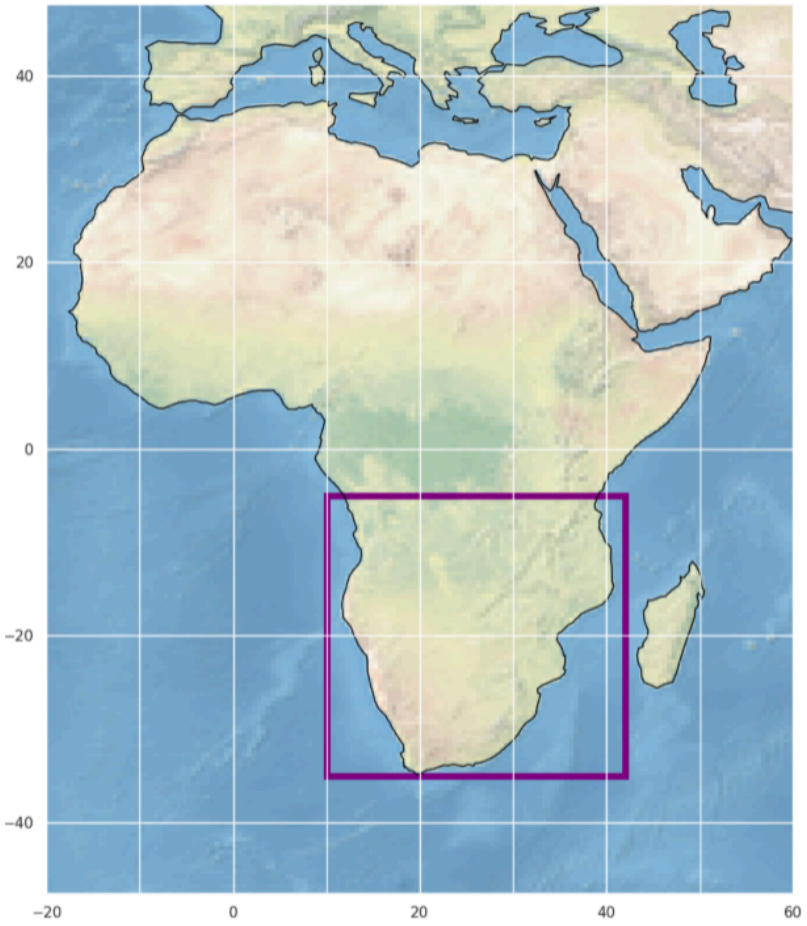


Figure 2. Extent of the WISER-EWSA focussed domain. Analysis has all been conducted inside this domain due to the fact that work was conducted to support the WISER-EWSSA testbed and to focus on a region of Africa during its rainy season.

5.3 Comparison methods.

5.3.1 Simple comparison of storm activation areas

Three simple comparison metrics have been used to help identify the differences in the way that the two products are identifying storms across the southern Africa domain shown above. Only times when both RDT and DWD-TS products are present (files successfully generated/downloaded) have been compared with one another. These are; (1) the total detection area for each product, (2) the intersection of detections between both products, where the storm detections agree with one another and (3) the total additional

detection area for each product. This final metric indicates the number of grid points for each product that have an activation while the corresponding grid point at the same time in the other product does not indicate the presence of a storm. Prior to any analysis there was an expectation that one product would identify storms more readily than the other, or have storms that are bigger and therefore have a greater coverage. This would have led to a situation where there would be a high percentage of intersections between the two products and one product would then show a significant additional detection area while the other showed almost no additional detection (see figure 3). However, it appears that instead of seeing this type of behaviour there are instead significant differences in storm detection between the two products as illustrated in figure 4 (discussed in detail in section 6).

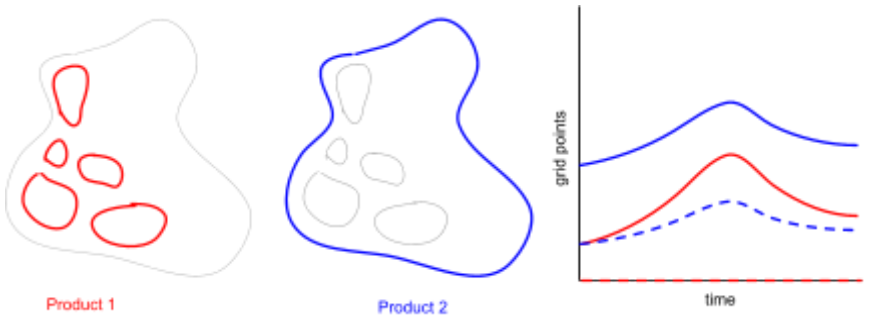


Figure 3. Illustration indicating the expected difference between storm detection algorithms where one product (product 2 here) is more sensitive and so detects storms over a larger area (for longer) than product 1. This would lead to a greater level of storm grid points being detected (solid lines on graph) and a correspondingly low additional detection in product 1 (dashed lines on graph).

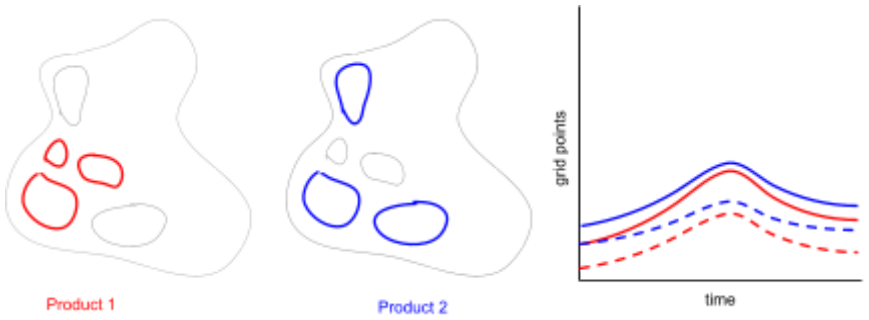


Figure 4. Illustration of what is more common when comparing the two storm identification products.There is no clear pattern of one product detecting larger areas than the other but that instead different regions are often activated. There is a high level of additional detection in both products (dashed lines on graph).

5.3.2 Fractions Skill Score

In addition to the simple comparison metrics, the binary maps created lend themselves to comparison with each other using the Fractions Skill Score (FSS; Roberts and Lean 2008). This methodology compares thresholded values (essentially binary images) against each other over a range of scales. This has the advantage of removing double penalties associated with offsets in storm position and allows for the fraction of a neighbourhood above a threshold to be compared with the corresponding neighbourhood from either a deterministic forecast or another comparable product.

The scales are dictated by the size of neighbourhoods that include an increasing large number of grid boxes. The FSS approach is most commonly applied to forecasts being compared with “truth” for the purposes of verification (this approach is used in this study to compare forward extrapolations of both products with their own retrievals). However, it can also be used when there is no explicit verification data to compare how favourably 2 separate products agree with one another. If two products have high FSS values we can assume that despite the differences in algorithmic methods they are convergent on the same answer. While this type of agreement is encouraging, it should not be confused with evaluation of whether the products give the correct answer, merely that they show (at particular scales) agreement with one another.

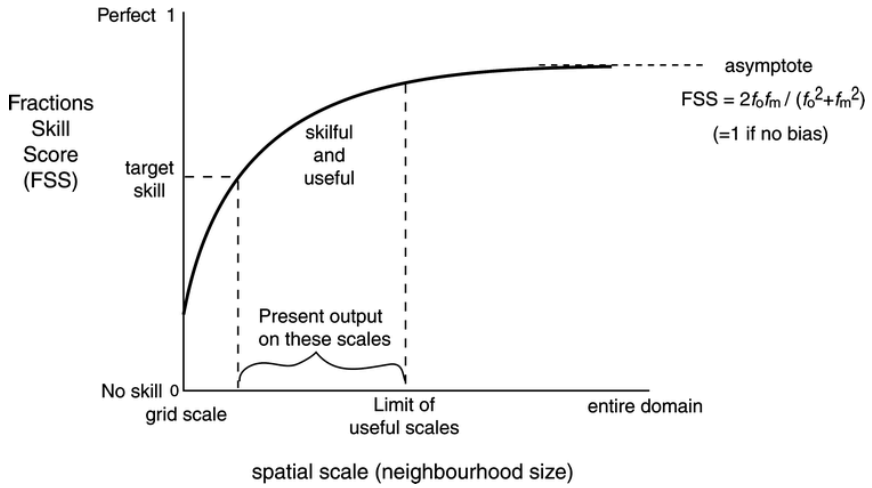


Figure 5 Schematic of fractions skill score against spatial scale illustrating the identification of scales that are both skillful and useful (From Roberts and Lean 2008). In the context of evaluation of predictions against retrievals this can help to identify scales where predictions are sufficiently skillful and still on useful scales.

An advantage of the FSS approach for comparing predictions with retrievals is that it allows for the identification of a range of scales where the predictions are both skillful (or at least in agreement with one another) and useful. At small scales it is likely that skill is below a target FSS value and at large scales there will be little improvement in FSS score with increasing neighbourhood size. In many cases an FSS value in excess of 0.5 is used to identify scales that are skillful/show good agreement.

5.5 Support of WISER-EWSA nowcasting testbed

In order to provide information that would be useful during the WISER-EWSA testbed (held in Lusaka, Zambia Jan 29th to 9th Feb 2024) visualisations of both the RDT and DWD-TS products were made available in near real time (available on the NCAS WISER-EWSA catalogue; https://science.ncas.ac.uk/wiserewsa). Figure 6 provides an example of the types of images generated and provided for use in the testbed. As well as images showing the whole of the southern Africa region smaller domains focused on South Africa, Zambia and Mozambique were produced. These products were used as part of the daily nowcasting activities which aimed to provide very short term predictions of high impact weather for the participatory countries of the the testbed (Zambia, Mozambique and South Africa). As can be seen in Figure 6 the detection of storms in the two products is broadly similar with the same regions being identified as being stormy. However, on finer (storm) scales there are significant differences between the two products. In particular there are: (1) locations where both products indicate the presence of convective storms but the cells highlighted are different (e.g. the intense storms highlighted in the DWD-TS product over the southern Democratic Republic of the Congo), (2) size and intensity variations in DWD-TS that are not mirrored in cell size in the RDT product and (3) regions where once product detects a storm but the other does not (e.g. southern Namibia and across Tanzania in figure 6).

Within the testbed a wide range of products for nowcasting were being used. Some were more familiar to participants than others and so there was not a consistent use of particular products every nowcasting shift. However, built into the structure of the testbeds (Fletcher et al., 2023) is the idea of continuous validation/evaluation with a dedicated group of participants looking at how nowcasting products compared with each other, available observations and forecast products. The comparison of RDT and DWD-TS formed part of this evaluation process. Daily summary figures were created and made available in a directory of the WISER-EWSA testbed google drive. Images were compiled into a standard pdf “report” format with a clear structure and a guidance document to facilitate interpretation (https://drive.google.com/drive/folders/1eVQpBiRa7xq5qWtfExsNUzFdm2IzlY-E?usp=sharing).

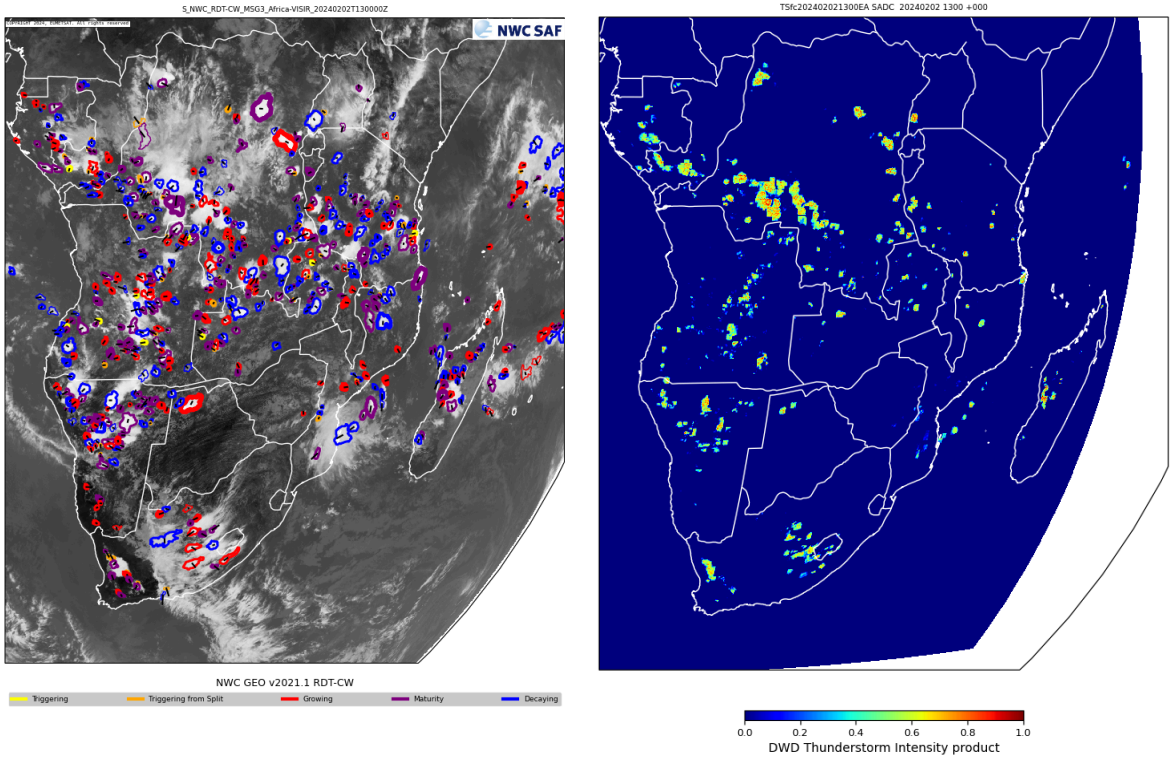


Figure 6 Examples of the RDT and DWD-TS products valid for 2nd Feb 2024 1300 UTC as generated for the WISER-EWSA testbed for a region covering Southern Africa. Overall pattern of storms identification are regionally similar but the location of individual storm regions varies significantly on storm scale.

6. Results

6.1 Product Inter-comparison

6.1.1 Activation maps

The difference in storm detections between the two products is visible by comparing maps of the average activations per day (Figure 7). Figure 7 highlights that while the general pattern of activation from Nov 2023 to Mar 2024 is similar, there is a much higher activation rate in the RDT product compared to the DWD-TS product. This is more pronounced in the more tropical regions of the domain which indicate a much more frequent activation rate. While there are some developments over the season this pattern holds true when comparing maps of product activation for shorter periods of time (monthly figures are provided in appendix 1). Another notable difference is that the RDT product has a stronger response to regular activation regions. For example the hotspot (present in both products) between 10

and 15 degrees south along the west coast of Angola is significantly stronger in the RDT storm detection algorithm. This is particularly clear in the months of Nov, Feb and Mar. The reason for this is not immediately clear, however it is possible that this sea breeze triggering zone commonly reaches the required temperature thresholds and discrimination criteria in RDT but did not become electrically active (required for DWD-TS detection) as frequently. Similarly, the pattern of more frequent RDT activation in the tropical parts of the domain can be explained in the same way. The RDT discrimination criteria contains a focus on instability, which is far more likely to be reached in the tropics as well as convective clouds having higher cloud tops. Both factors could allow for RDT detection to be triggered (and cells to pass the discrimination step) more readily than in the more temperate parts of the studied domain, whereas electrification of the storms (used by DWD-TS) is not necessarily latitudinally controlled in the same way.

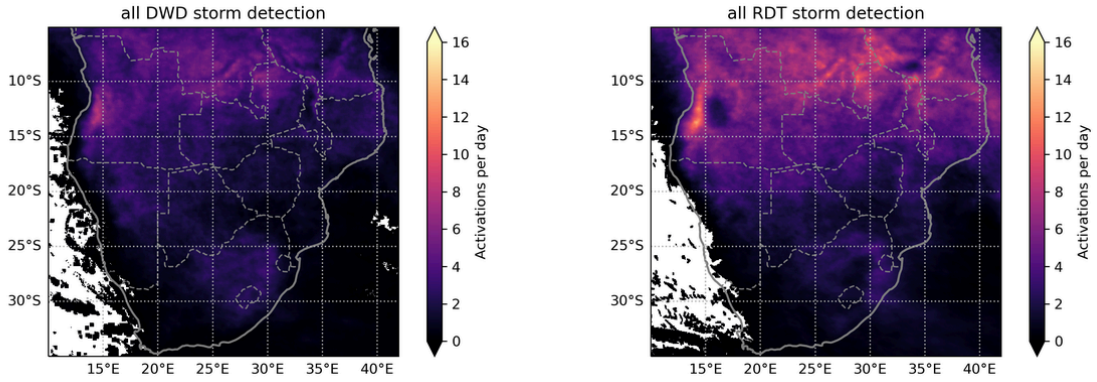


Figure 7. Maps of activation frequency for the 2 studied products (DWD-TS and RDT). Not the higher activation levels in the RDT product compared to DWD-TS especially int he northern half of the domain.

To highlight the agreement and differences of the products Figure 8 shows the intersection of the two products (where they agree on the timing and location of storms) as well as the additional detection (where one product identifies a storm but the other doesn't) plotted on maps in the same style of maps as Figure 7. The relative intensity of the RDT DWD-TS intersection to the two additional detections shows that the additional detections that DWD-TS makes are of a similar magnitude as the intersecting detections, while the RDT additional detection is much more frequent. This indicates that the amount of times that products are in agreement with one another is approximately half of all DWD-TS detections and significantly less than half of all RDT detections. This suggests the behaviour shown in the schematic (Figure 4) where cells that are being identified are sometimes the same but often completely different. Another possible cause of such high additional detection in the RDT product is that storms are detected earlier and persist for longer. The same storms being detected but being bigger in RDT compared to DWD-TS would not lead to such a high level of DWD-TS additional detection, ruling out the behaviour shown in the schematic (Figure 3).

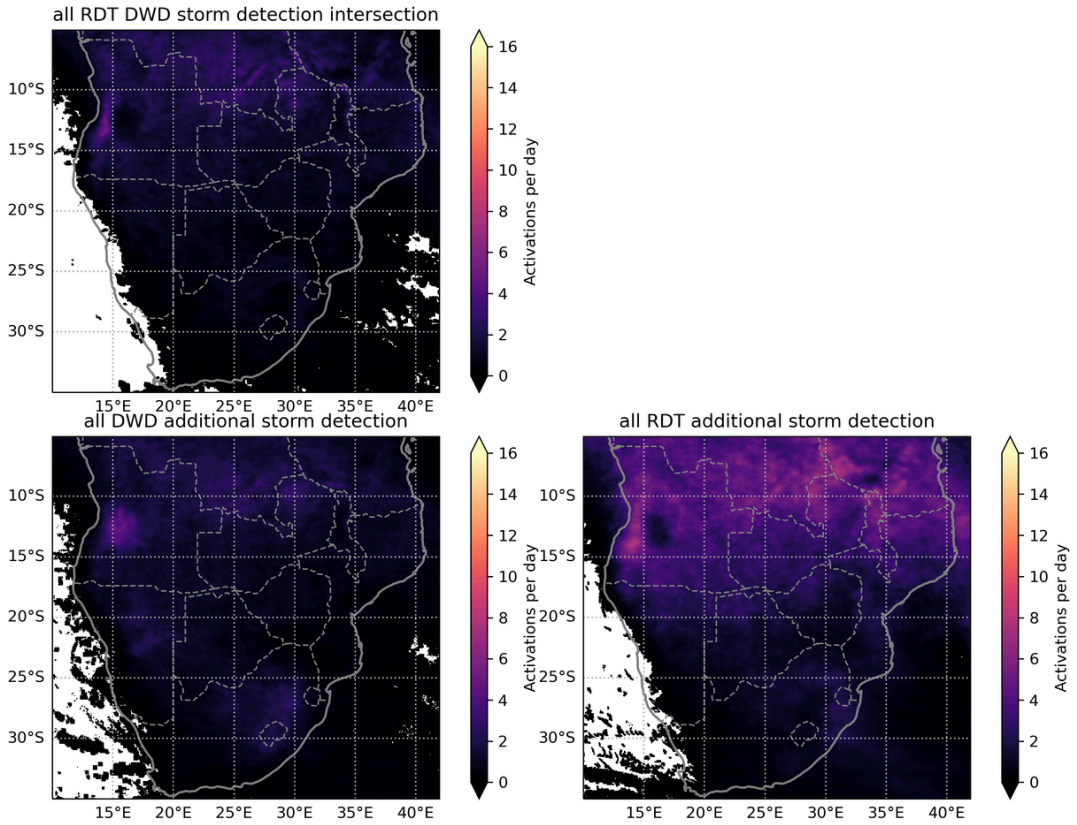


Figure 8. Maps showing the RDT and DWD-TS intersection frequency (top panel) and the additional detection for each product (bottom panels).

6.1.2 Diurnal cycle of detection

To identify whether the additional detection from the RDT product can be explained by differences in the diurnal cycle of detection (i.e. storms being detected earlier in the daily cycle and lasting for longer) Figure 9 shows the diurnal cycle of grid point activations (shaded areas showing the interquartile range). These have been split into storm detection (as shown in Figure 7; left panel of Figure 9) and storm intersection and additional detection (as shown in Figure 8; right panel of Figure 9). The total storm area covered highlights that the RDT product detects much greater storm areas than the DWD-TS product at all times of day. The diurnal cycles are very similar and show a pattern of overnight reduction in storm cover before an increase starting at approximately 0900 UTC and peaking at approximately 1400 UTC. The increases at the peak are larger in RDT compared to DWD-TS but not in a way that suggests that storms detected by the different methods are very different in their longevity. Therefore it appears that the differences between the products are largely driven by the detection of a larger number of storms in the RDT product. Another feature of the diurnal cycles is that the RDT product appears to show an unusual behaviour at approximately 1600 UTC, with a sharp drop in the area coverage of storms. This is, however, an artefact from the imperfect satellite retrievals of MSG data used to generate the RDT product at the University of Leeds. Figure 10 indicates the way in which missing channel data from the MSG multicast can impact two of the most useful NWCSAF-GEO products (CRR and RDT). For an unknown reason the frequency of dropped data files was higher at certain times of day, such dropouts did not impact the ability to generate data in other parts of the domain and therefore led to the creation of valid netcdf output files and

subsequent inclusion in this analysis. An oversight in subsetting for missing data in this way has led to a discrepancy between the RDT and DWD-TS files (visible in Figure 9). However, given that even with missing data RDT still detects significantly higher storm coverage than DWD-TS and only a small drop in the product intersection it is deemed unlikely to impact the interpretation of the findings presented here.

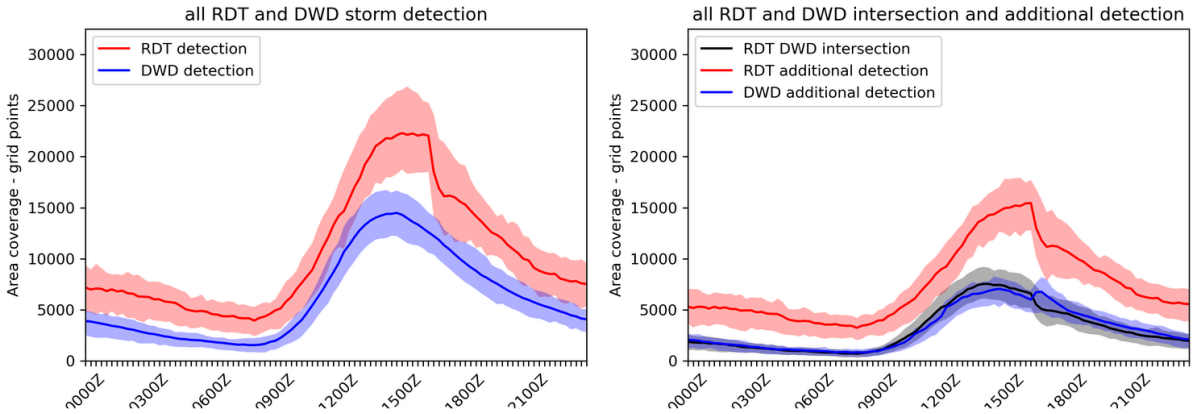


Figure 9. Line graphs of diurnal cycle of total storm coverage for the DWD-TS and RDT products (left panel) and the DWD-TS-RDT intersection and DWD-TS and RDT additional detections (right panel). Shaded regions show the interquartile range and lines indicate the median value across the whole study period (Nov 2023 - Mar 2024).

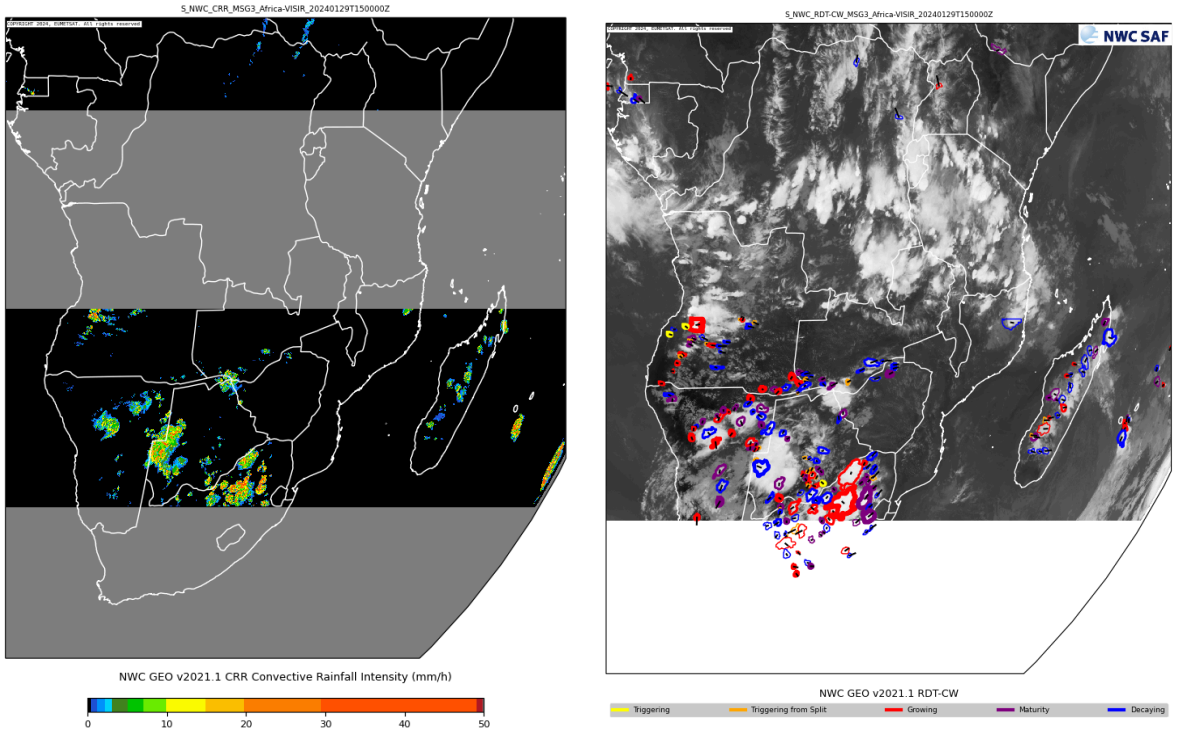


Figure 10. Example images of NWCSAF-GEo products CRR and RDT at 1500 UTC 29th Jan 2024. Missing channel data has prevented the generation of the CRR product in 2 regions (central and southern Africa) while RDT has only been affected in central Africa) meaning that at this time and for those locations storms were missed.

6.1.3 Agreement across spatial scales

Given that we know that there are different regions of activation between the two products with the intersection being approximately half the total DWD-TS detection and a third of the RDT detection, but also that the large scale spatial pattern of detection (Figure 7) is quite similar it is likely that the grid scale (0.05° x 0.05°) is too fine and that at larger scales the two products will be in better agreement. As such, it would be useful to identify the scales at which the different storm detection algorithms agree with one another as this might provide greater confidence in the identification of areas at risk from deep convective storms. As long as the scale at which a high FSS score is reached is still useful then it could be a way of moving toward a probabilistic method of storm identification which focuses on scales where we know algorithms using different methodologies compare well with one another. In the absence of good ground truth data this approach (but used with multiple different algorithms, all attempting to reach the same answer) could allow for the generation of nowcasting metrics which can take into account the agreement of disparate products over a range of spatial scales.

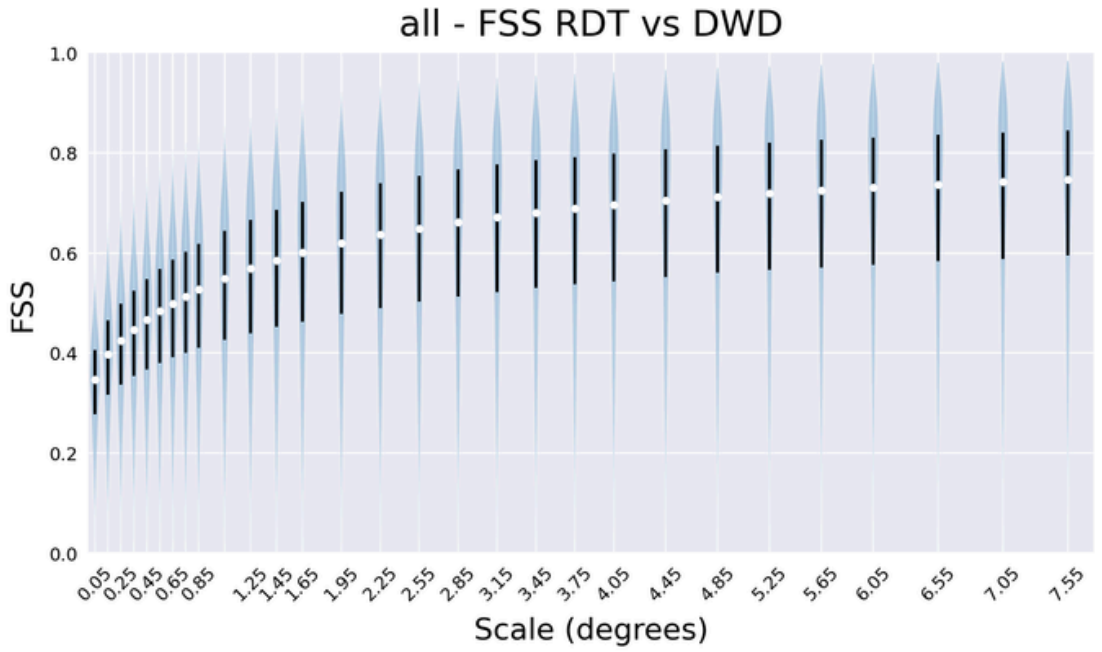


Figure 11 Violin plot of fraction skill score calculated by comparing DWD-TS and RDT storm identification algorithms across a range of spatial scales. White dots indicate median values, black lines show the interquartile range and blue shading shows the distribution. Calculated for the whole study period (Nov 2023 - Mar 2024)

Figure 11 shows a violin plot of FSS values comparing RDT and DWD-TS across a range of spatial scales. At the finest scales used in this study (0.05° x 0.05°) the FSS between the two products is largely below the 0.5 value often used to denote useful skill/agreement. This increases across the scales and levels out at approximately 5 degrees. The value at which the FSS values asymptote is significantly below 1. This is to be expected as however large a neighbourhood is used there will always be a bias towards the RDT detection as it is significantly more frequent than the DWD-TS. That said, the 0.5 FSS value is exceeded by the median at a neighbourhood scale of 0.65° and exceeds 0.6 at 1.65°.

This result indicates that the scales at which there is good agreement is significantly larger than what would be desirable operationally. Reducing the resolution of products to a 1° would put them within the scale range that makes them have good agreement and is still a useful scale. However, it precludes their use for identifying storm regions on a city/neighbourhood scale. Whether the advent of higher resolution satellite data (MTG) means that the agreement between such products is improved on smaller spatial scales is yet to be seen. However, it seems unlikely that improving the satellite resolution will have any impact on the mechanism of the algorithms and so it is likely that even using better input data these products will have significant differences with respect to identifying storm regions.

6.1.4 Diurnal cycle of agreement across spatial scales

In order to test whether the scales of agreement vary significantly across the diurnal cycle, a heatmap of the mean FSS values across times of day and spatial scales has been generated (Figure 12). Figure 12 shows that there is a large diurnal variation in the spatial scale at which the mean FSS is greater than 0.5. Specifically there is a period of much worse agreement overnight, which worsens until approximately 0800 UTC. This coincides with the minima of storm coverage in both products (Figure 9). However when new storms are beginning to be generated the agreement between the products rapidly improves across spatial scales and from 1100 UTC to 1600 UTC the scale at which the mean FSS is above 0.5 is below 0.5°. This spatial scale is much closer to what would be most useful operationally and suggests that the best use of such storm detection products might be made during the part of the day when storm growth and development is occurring.

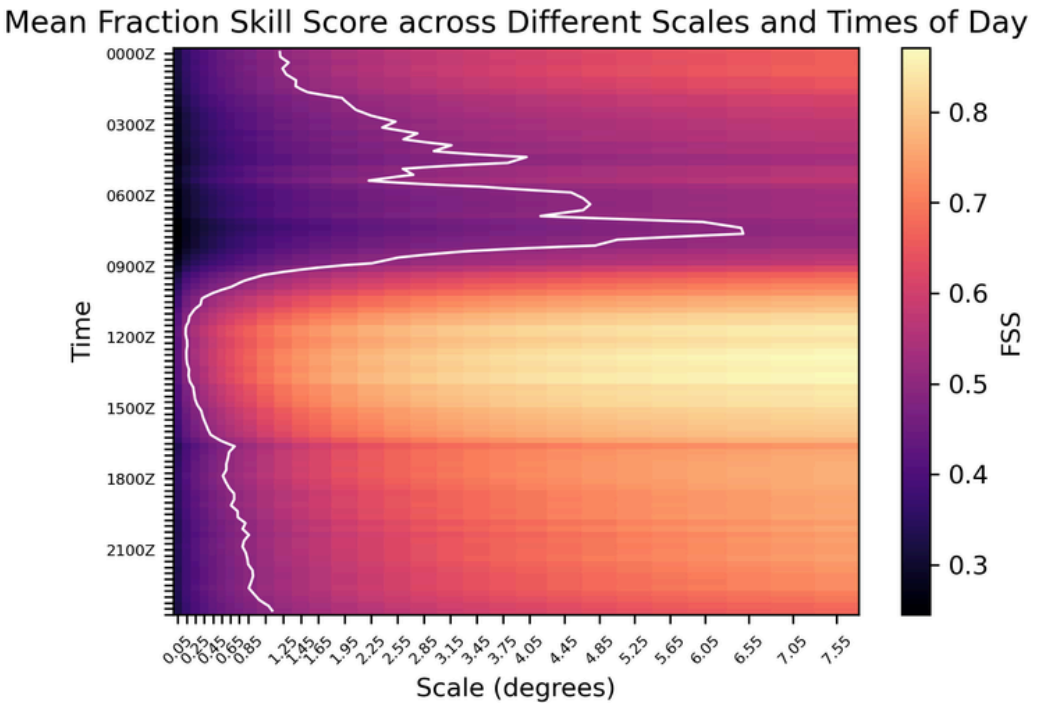


Figure 12 Heatmap indicating the mean FSS value of RDT compared with DWD-TS at different times of day for a variety of spatial scales. The better the agreement the higher the FSS value. The white line marked indicates the scale at which an FSS value exceeds 0.5.

6.2 Forward propagation skill

Nowcasting doesn't only relate to the rapid processing of retrieval data into usable products, it is also concerned with the forward prediction (via extrapolation) of weather in time. The RDT and DWD-TS products both provide forward extrapolations of the storms that have been identified via different methodologies (see section 4). In this section we compare these forward extrapolations using the fraction skill score to see which of the two products compares most favourably when compared with retrievals from the same product. This way the skill of the extrapolation is tested independently of the differences in the storm detection algorithms. The RDT product has a standard forward extrapolation of 60 minutes (in 15 minute steps) While the DWD-TS product extrapolates storm positions to 105 minutes (in 15 minute steps).

Figure 13 (part 1) shows violin plots of FSS from both DWD and RDT for the 15, 30, 45 and 60 minute extrapolations. As expected for both products there is a trend that for longer forward predictions that the skill drops. At the smallest scale (0.05° x 0.05°) the median FSS for DWD-TS drops from 0.72 to 0.44 and 0.73 to 0.43 for RDT. This indicates that on this spatial scale (and over this shared prediction timeframe) the forward extrapolation techniques are very similar and have comparable skill. However, when the increasing neighbourhood sizes are taken into account it is clear that across all the forward projections that the DWD-TS product produces higher FSS values and reaches a stable FSS score at smaller spatial scales and has a smaller distribution of FSS values. This means that predictions that are provided on coarser spatial grids (something that should be considered given the results in section 6.1.3) will be better when using the Farneback optical flow methodology employed by the DWD-TS product than the more simplistic multi-step area overlap technique to generate storm motion vectors. This does not come as a surprise to the authors as optical flow methods are currently the preferred method for nowcasting forward extrapolations (NWCSAF-GEO EXIM and pysteps are examples). However, in the future it would be useful to re-perform similar analysis on emerging techniques for satellite image prediction, such as the use of convolutional neural networks and diffusion models to assess the skill of such techniques for very short term prediction of storms (Capelle 2023).
The additional forward prediction timesteps available for the DWD-TS product have a similar reduction in skill the further forward predictions are made. However, even for the 105 minute extrapolation the median FSS at the 1° scale is still 0.63 suggesting that if a coarsened version of the product were to be used operationally then predictions (on that scale) would have skill in excess of the 30 minute prediction on the finest spatial scale. As discussed earlier, while such a coarsening of the product means that it loses much of the fine detail, much of this detail (in terms of the forward propagations) is not skillful and so there is a danger of a user being overconfident in the predictive capability of the extrapolation. In this sense it might be useful to encourage users to use a coarsened spatial grid and interpret the predictions in a more probabilistic way. While this might make nowcasting operations more complicated it is important (for operational nowcasters and for nowcast consumers) to be aware of the inherent uncertainty of predictions and make sure information and warnings are generated in line with this approach. This method naturally feeds into the likelihood vs impact risk matrix methodology that is already in use for high impact weather across the African continent.

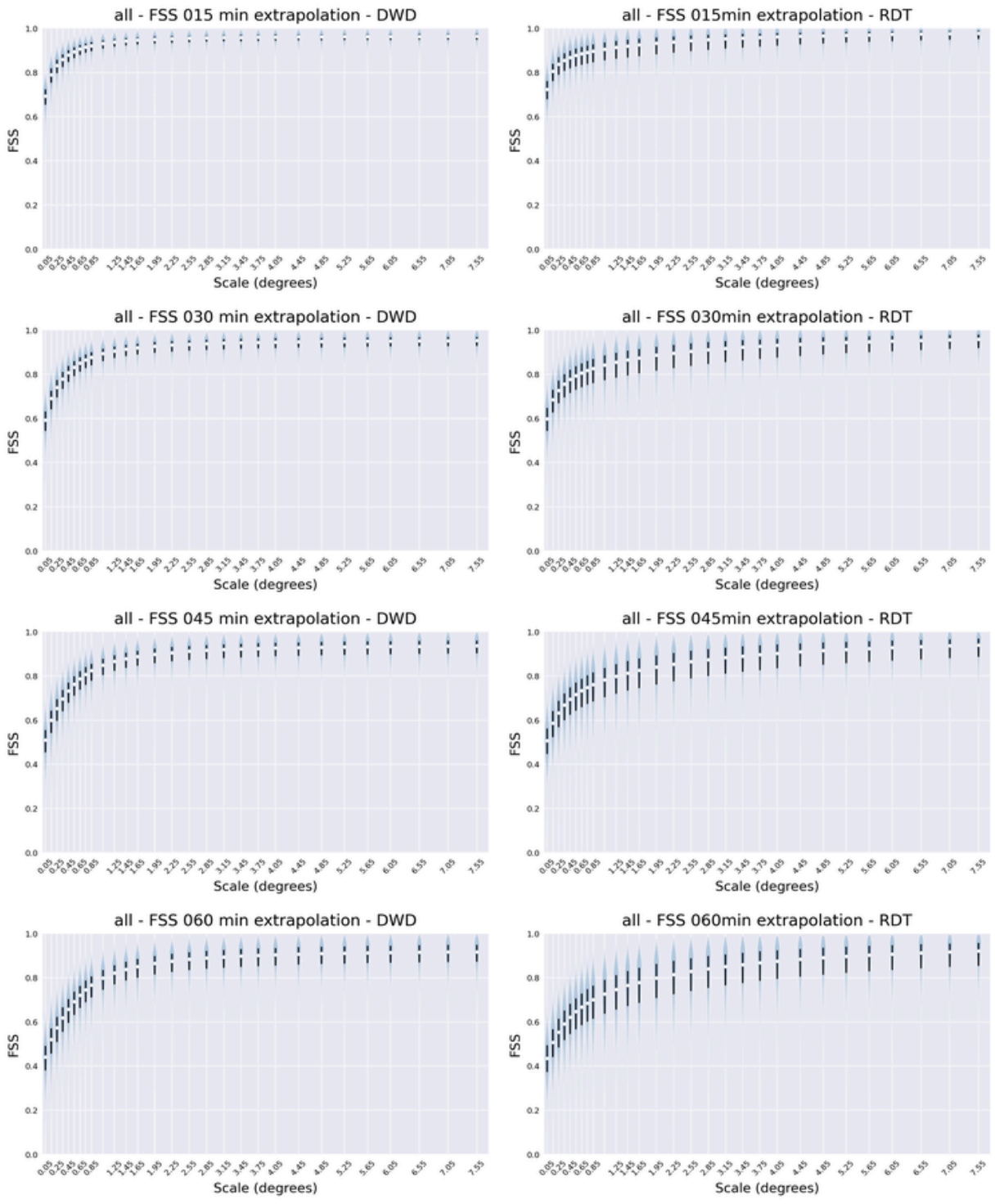


Figure 13 (part 1) Violin plot of fraction skill score calculated by comparing DWD-TS forward extrapolations to retrievals (left column) and RDT forward extrapolations to retrievals (right column) across a range of spatial scales. White dots indicate median values, black lines show the interquartile range and blue shading shows the distribution. Calculated for the whole study period (Nov 2023 - Mar 2024).

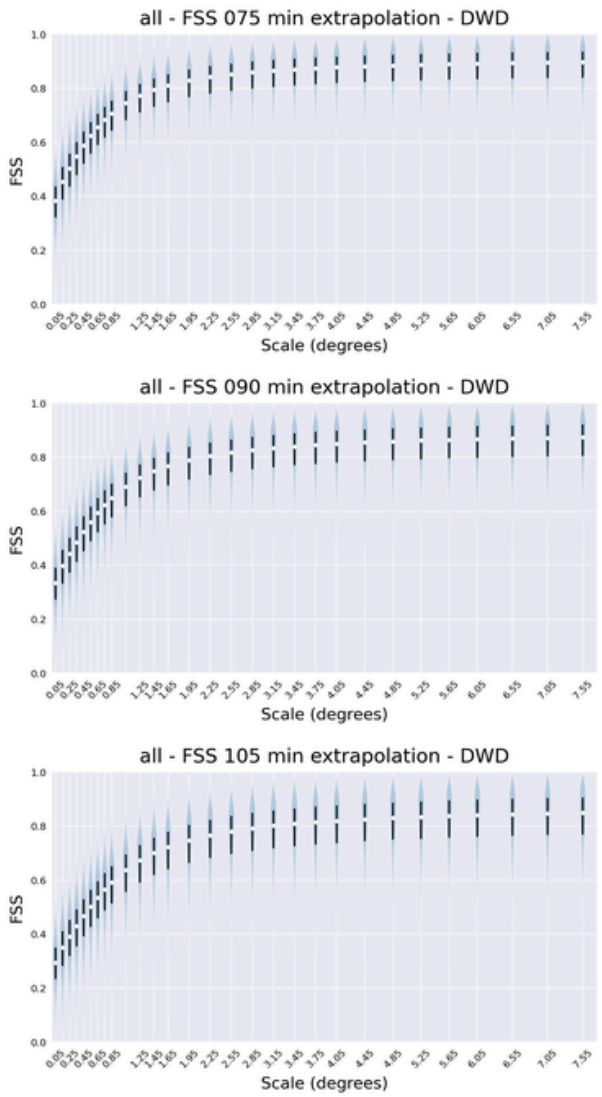


Figure 13 (part 2) as Figure 13 part 1 but only showing the 3 additional extrapolation steps produced by the DWD-TS method.

7. Conclusions

It is evident from the work presented here that the two compared products (RDT and DWD-TS) generally have similar spatial distributions over the study regions and period but have some important differences. Notably, that RDT detects more storms and has a greater tendency to detect storms over triggering hotspots and in the more tropical part of the domain. The intersection between products is relatively low, representing ~50 % of all DWD-TS detections and ~33% of all RDT detections. The diurnal cycles of detection are very similar indicating that differences are mostly driven by different storms being detected rather than one product detecting the same storms earlier and for longer.
Even when comparing products across spatial scales agreement is suppressed due to RDT bias. Agreement between products does increase with larger spatial scales and is also heavily influenced by time of day. The best agreement occurs during the initialisation and growth period of convective cells. It is possible that limiting the use of such storm identification products to such times would be beneficial. This is because this period of time

is when such products have the greatest utility, but also better agreement on smaller scales allows their use on more useful scales (closer to the city scales required for operationalisation).
Forward prediction methods are comparable when considering the finest scales, however DWD-TS has a longer standard extrapolation (105 minutes compared to 60 minutes in RDT). Also, when compared on a larger spatial scales DWD-TS shows improvements over RDT, reaching higher FSS values and stabilising at lower spatial scales. It also has smaller distributions in FSS values across lead times and spatial scales. This indicates that there is some advantage to the DWD-TS optical flow technique when the product is presented on a coarser grid.
Recommendations for future work are listed below:

- Further comparisons to be made with new storm detection methodologies (new algorithms and improved data availability).
- Future comparisons should be made which compare the forward extrapolations techniques currently used for nowcasting with new satellite image prediction techniques such as convolutional neural networks and diffusion models.
- Studies comparing satellite nowcasting methods with high quality verification data, specifically rainfall radar data (South African Weather Service radar network).
- Use of storm detection algorithms on scales tailored to multi-method agreement or if possible scales that show skill against verification data.
- Move toward probabilistic approaches which represent the inherent uncertainty of detection algorithms.
- Encourage use of new and existing products in a probabilistic way to assist in continued generation of warnings using extant risk matrices.

References

Cappelle, T., (2023). *Diffusion on the Clouds: Short-term solar energy forecasting with Diffusion Models.* Available at: https://wandb.ai/capecape/ddpm_clouds/reports/Diffusion-on-the-Clouds-Short-term-solar-energy-forecasting-with-Diffusion-Models--VmlldzozNDMxNTg5 Accessed (10-07-2024).

EUMETSAT (2024). *EUMETCast Africa data catalogue* Available at: https://user.eumetsat.int/data-access/eumetcast-africa/data. Accessed (11-06-2024).

Fletcher, J. K., Diop, C.A., Adefisan, E., Ahiataku, M.A., Ansah, S.O., Birch, C.E., Burns, H.L., Clarke, S.J., Gacheru, J., James, T.D., Ngetich Tuikong, C.K., Koros, D., Indasi, V.S., Lamptey, B., Lawal, K.A., Parker, D.J., Roberts, A.J., Stein, T.H.M., Visman, E., Warner, J., Woodhams, B.J., Youds, L.H., Ajayi, V.O., Bosire, E.N., Cafaro, C., Camara, C.A.T., Chanzu, B., Dione, C., Gitau, W., Groves, D., Groves, J., Hill, P.G., Ishiyaku, I., Klein, C.M., Marsham, J.H., Mutai, B.K., Ndiaye, P.N., Osei, M., Popoola, T.I., Talib, J., Taylor, C.M., and Walker, D. (2023). *Tropical Africa's First Testbed for High-Impact Weather Forecasting and Nowcasting.* Bull. Amer. Meteor. Soc., 104, E1409–E1425, https://doi.org/10.1175/BAMS-D-21-0156.1.

Parker, D. J., Blyth, A.M., Woolnough, S.J., Dougill, A.J., Bain, C.L., de Coning, E., Diop-Kane, M., Foamouhoue, A.K., Lamptey, B., Ndiaye, O., Ruti, P., Adefisan, E.A., Amekudzi, L.K., Antwi-Agyei, P., Birch, C.E., Cafaro, C., Carr, H., Chanzu, B., Clarke, S.J., Coskeran, H., Danuor, S.K., de Andrade, F.M., Diakaria, K., Dione, C., Diop, C.A., Fletcher, J.K., Gaye, A.T., Groves, J.L., Gudoshava, M., Hartley, A.J., Hirons, L.C., Ibrahim, I., James, T.D., Lawal, K.A., Marsham, J.H., Mutemi, J.N., Okogbue, E.C., Olaniyan, E., Omotosho, J.B., Portuphy, J., Roberts, A.J., Schwendike, J., Segele, Z.T., Stein, T.H.M., Taylor, A.L., Taylor, C.M., Warnaars, T.A., Webster, S., Woodhams, B.J. and Youds, L. (2022): *The African SWIFT Project: Growing Science Capability to Bring about a Revolution in Weather Prediction.* Bull. Amer. Meteor. Soc., 103, E349–E369, https://doi.org/10.1175/BAMS-D-20-0047.1.

Pulkkinen, S., Nerini, D., Hortal, A.A.P., Velasco-Forero, C., Seed, A., Germann, U. and Foresti, L. (2019). *Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0).* Geoscientific Model Development, 12, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019.

Roberts, N. M., and H. W. Lean, (2008). *Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events.* Mon. Wea. Rev., 136, 78–97, https://doi.org/10.1175/2007MWR2123.1.

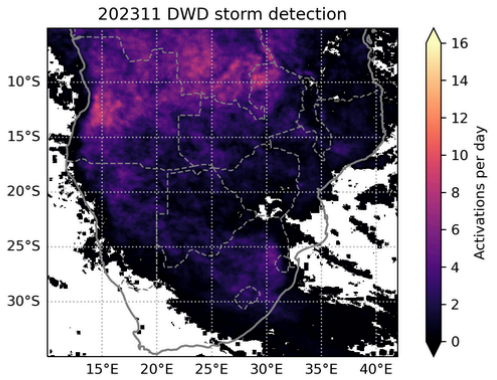
Roberts, A.J., Fletcher, J.K., Groves, J., Marsham, J.H., Parker, D.J., Blyth, A.M., Adefisan, E.A., Ajayi, V.O., Barrette, R., de Coning, E., Dione, C., Diop, A., Foamouhoue, A.K., Gijben, M., Hill, P.G., Lawal, K.A., Mutemi, J., Padi, M., Popoola, T.I., Rípodas, P., Stein, T.H.M. and Woodhams, B.J. (2022), *Nowcasting for Africa: advances, potential and value.* Weather, 77: 250-256. https://doi.org/10.1002/wea.3936.

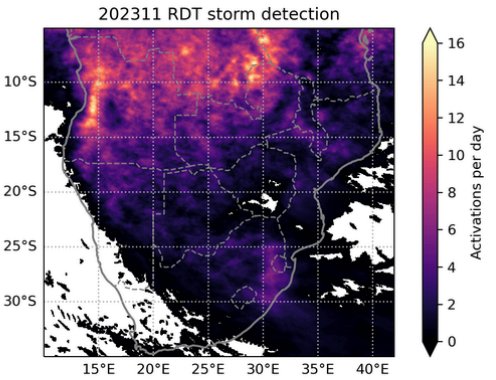
Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting, (2020). *Skill of Global Raw and Post Processed Ensemble Predictions of Rainfall in the Tropics.* Wea. Forecasting, 35, 2367–2385, https://doi.org/10.1175/WAF-D-20-0082.1.

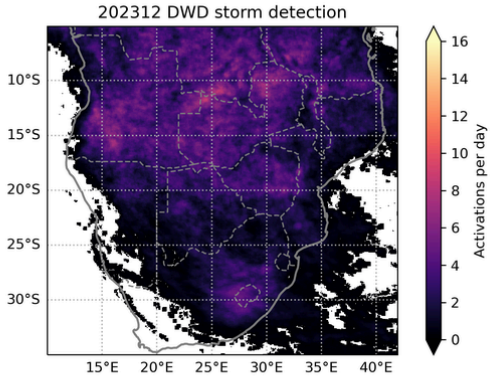
WMO (2022). *Early Warnings for All: Executive Action Plan 2023-2027.* WMO, Geneva, Switzerland. https://library.wmo.int/idurl/4/58209.

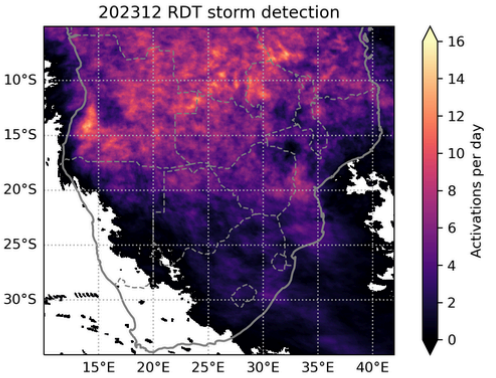
WMO (2023). *Guidelines for Satellite-based Nowcasting in Africa.* WMO, Geneva, Switzerland. https://library.wmo.int/idurl/4/58348.

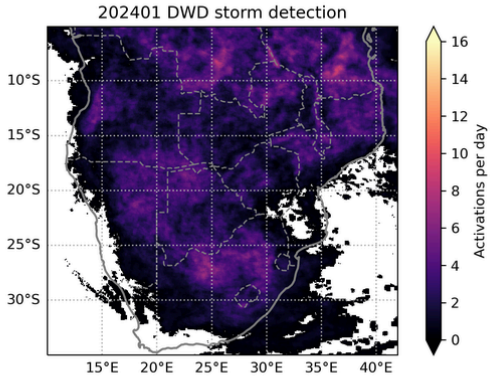
Appendices

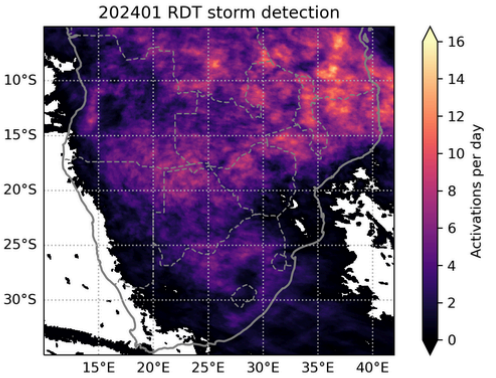

202311 DWD storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

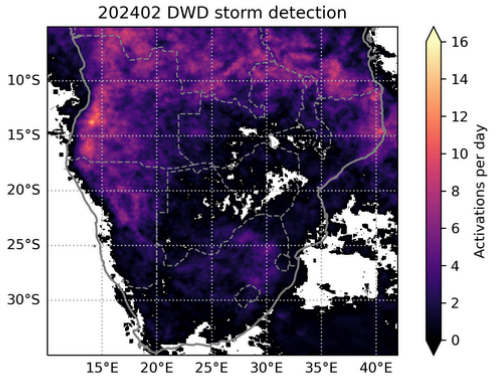

202311 RDT storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

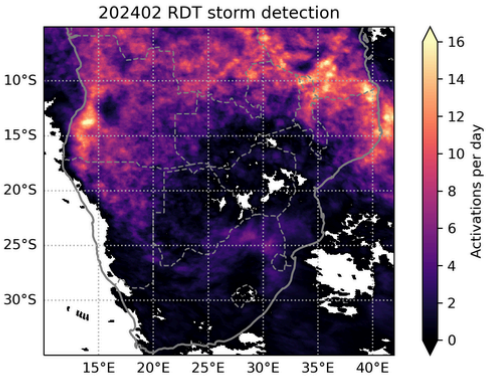

202312 DWD storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

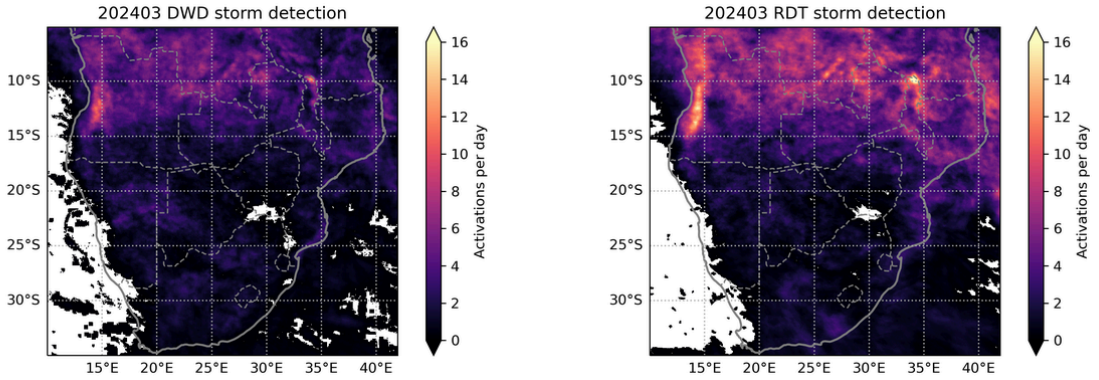

202312 RDT storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day


202401 DWD storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

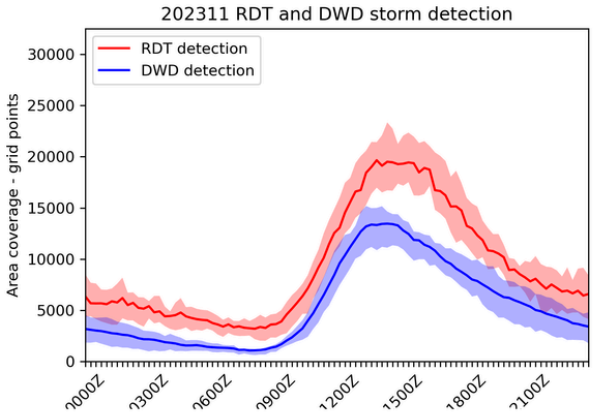

202401 RDT storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

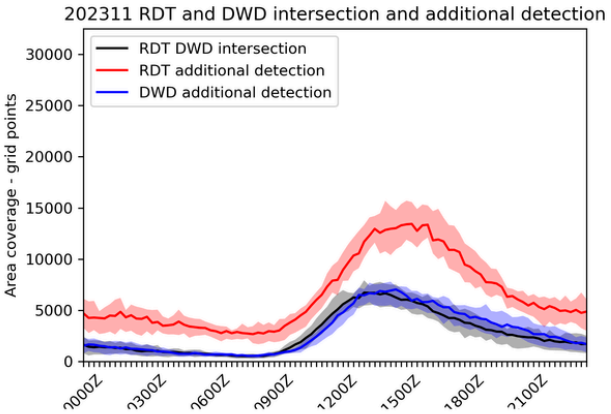

202402 DWD storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

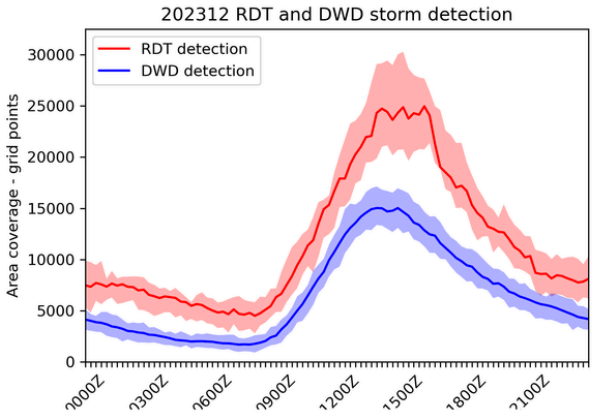

202402 RDT storm detection
10°S
15°S
20°S
25°S
30°S
15°E
20°E
25°E
30°E
35°E
40°E
0
2
4
6
8
10
12
14
16
Activations per day

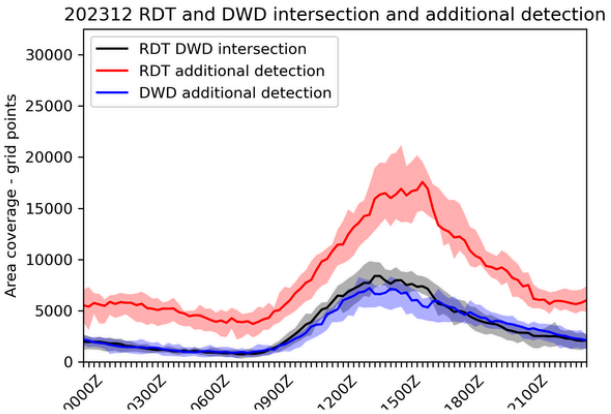


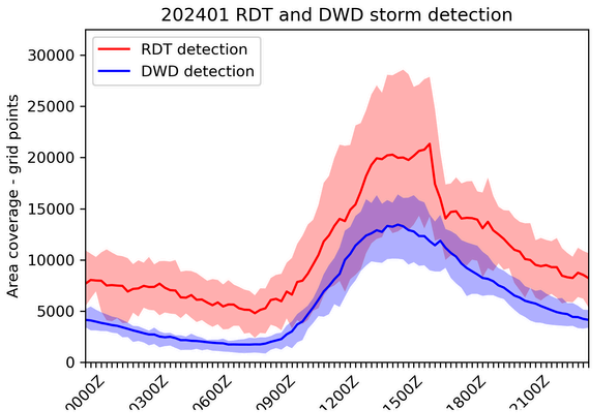
Appendix 1 Monthly storm detection given in activations per day in the DWD-TS product (left column) and the RDT product (right column), after Figure 6.

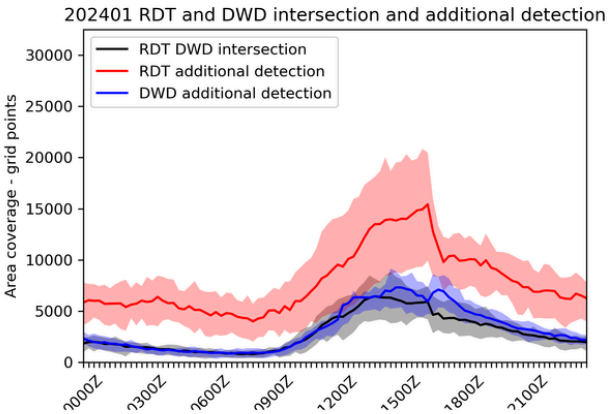

202311 RDT and DWD storm detection
RDT detection
DWD detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z

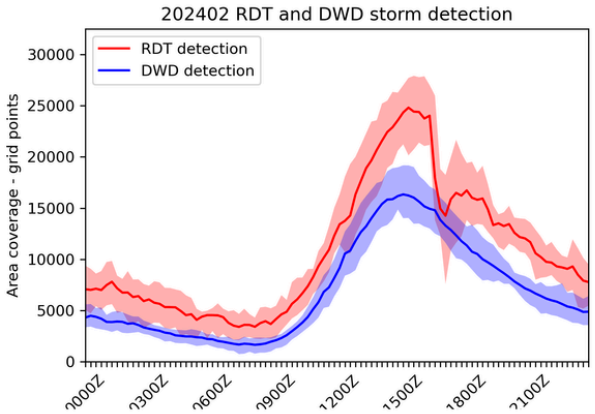

202311 RDT and DWD intersection and additional detection
RDT DWD intersection
RDT additional detection
DWD additional detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z

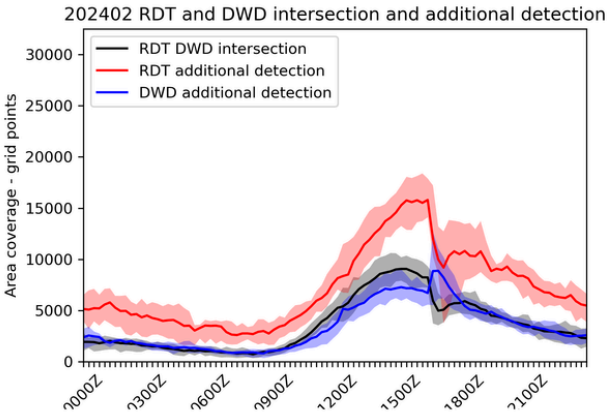

202312 RDT and DWD storm detection
RDT detection
DWD detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z

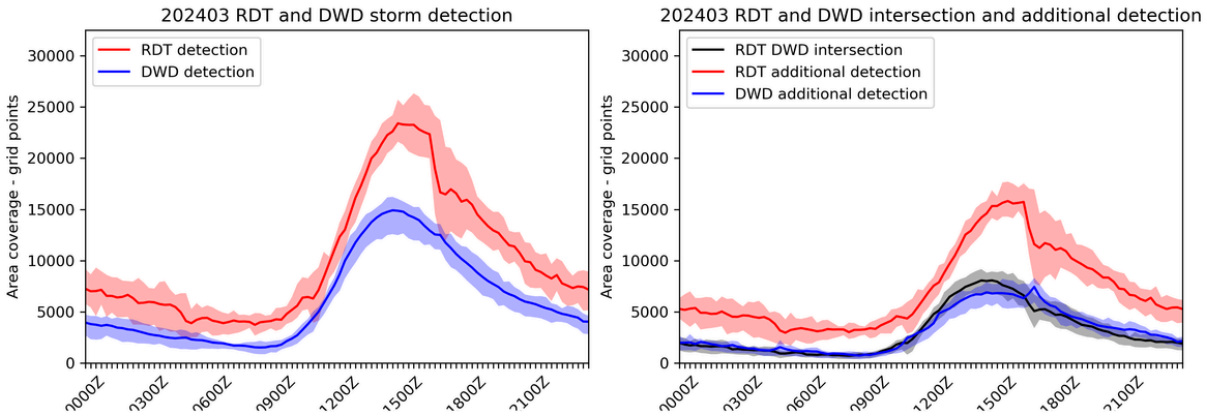

202312 RDT and DWD intersection and additional detection
RDT DWD intersection
RDT additional detection
DWD additional detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z


202401 RDT and DWD storm detection
RDT detection
DWD detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z

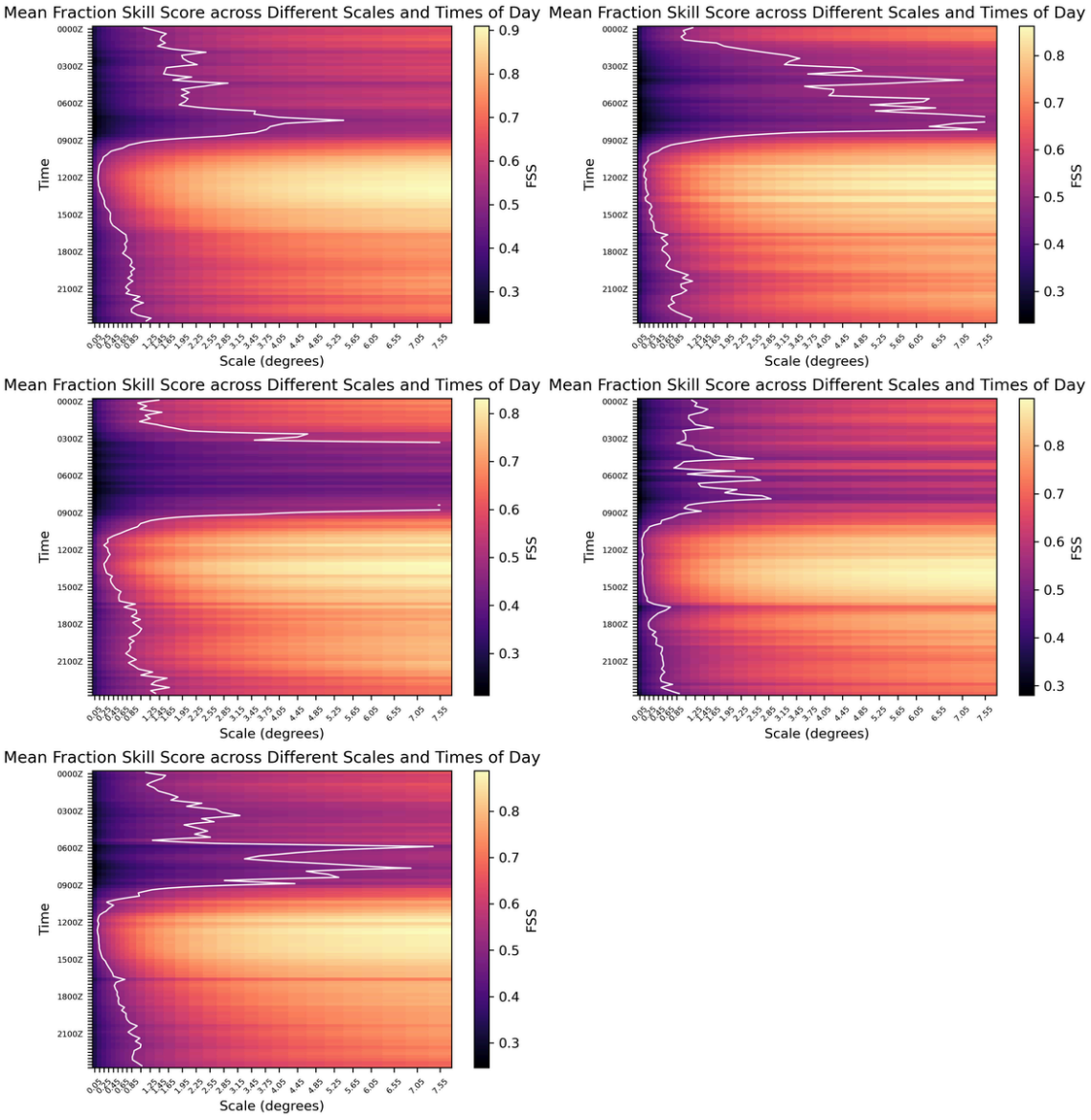

202401 RDT and DWD intersection and additional detection
RDT DWD intersection
RDT additional detection
DWD additional detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z


202402 RDT and DWD storm detection
RDT detection
DWD detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z


202402 RDT and DWD intersection and additional detection
RDT DWD intersection
RDT additional detection
DWD additional detection
Area coverage - grid points
0
5000
10000
15000
20000
25000
30000
0000Z
0300Z
0600Z
0900Z
1200Z
1500Z
1800Z
2100Z



Appendix 2 Monthly storm detection (left column) and intersection and additional detection (right column) for RDT (red) and DWD-TS (blue) after Figure 8.



Appendix 3 Monthly heatmaps of mean FSS split by time of day and scale after Figure 11.