



Application and Evaluation of the WRF Model over North-Africa: Precipitation and Cloud Fraction

R. Moustabchir^{1*}, H. Charifi¹, M. Amchghal^{1,2}, H. Amghar^{1,3},
A. Chirmata^{1,4}, A. Ihlal¹, K. Iraoui¹, M. Ouattab¹

¹Environmental Modeling and Application Research Team (EMART), LMER, Physics department, Faculty of Science, Ibn Zohr University, Agadir, Morocco

²Direction Régionale de la Météorologie, Agadir, Morocco

³Direction Générale de la Météorologie, Casablanca, Morocco

⁴Environment Department, Wilaya of Souss Massa Region, Agadir, Morocco

Received 11 March 2021,

Revised 16 July 2021,

Accepted 17 July 2021

Keywords

- ✓ WRF,
- ✓ Numerical Weather Prediction,
- ✓ Model Evaluation Tool,
- ✓ CMORPH,
- ✓ IMERG,
- ✓ MODIS.

* Corresponding author:

r.moustabchir@uiz.ac.ma

Phone: +212661927251;

Abstract

In this work, we present an assessment of a short-term weather prediction system based on the Weather Research and Forecasting Model (WRF). This study is the second of a two-part WRF model evaluation over the north Africa. In the first part, we examined aspects of the verification process for continuous weather parameters using in-situ surface and upper air observations. In this study we focus on precipitation and cloud fraction evaluation using categorical metrics; We produced model validation statistics using Model Evaluation Tools (MET) to compute different metrics statistics which compared the performance of WRF forecasts with satellite remote sensing products. Satellite-based products are available at very high spatial and temporal resolutions, which has accelerated their use in numerous applications especially over the studied region with the limitations associated with the scarcity of available surface rain and cloud observation data. Fuzzy verification technique was applied to the assessment of the precipitation and cloud fraction gridded forecasts with respect to satellite products. Rather than comparing a single forecast to a single observation at each grid point, these methods compare the observations and forecasts in a neighborhood surrounding the point of interest. Given the strengths and weaknesses of each metric, the results are discussed with multiple scores and methods in order to assess the overall quality of the forecasts and to determine potential biases and outliers.

1. Introduction

Aiming to design a forecast system, it is necessary to perform a quantitative verification to assess model performance based on the weather systems that are intended to predict. The assessment of model accuracy relies on the ability to generate verification statistics that compare the model output to actual observations. This is especially difficult for high-resolution model verification that requires time, as well as spatial forecast verification. Systematic forecast verification can assist in overcoming difficulties in predicting the weather by allowing the quality of the forecasts to be assessed and to track and identify errors and document improvements in the system [1].

Previous papers [2,3] present a comprehensive evaluation of the used forecasting system for surface and upper level continuous observations. In this paper, the results from WRF are evaluated and compared with those of satellite data to assess the impact of the previously used model representations on precipitation and cloud fraction predictions. Radar and rain gauge networks have been the primary tools for precipitation and cloud analyses, but satellite products can be very useful when forecasting for a number of different weather events, particularly in areas with poor coverage from radar and rain gauge data [4]. Because the quantitative accuracy of satellite remote sensing products is affected by various

factors related to uncertainties associated with satellite data, atmospheric and terrain properties, validating numerical weather-prediction model (NWP) using different satellite products over different regions and weather conditions is needed. This study investigated the performance of the developed forecasting system using different satellite-based products for a challenging winter precipitation forecast event over north Africa and Morocco, on December 3-4, 2019. There are many methods available to evaluate and validate precipitation and cloud fractions, among which the categorical metrics method is the most popular. Categorical metrics computed from newly modified neighborhood verification approach [5,6] are a bit superior to continuous metrics [7] but give credit only to the close forecasts. Categorical metrics include scores such as probability of detection (POD), critical success index (CSI), success ratio (SR), bias, accuracy (ACC) and Heidke skill score (HSS)... They are derived from the analysis of a contingency table (hits, misses, false alarms, correct rejections). Like Taylor diagrams [8] for continuous metrics, Roebber performance diagrams [9] provides a way to visualize a number of measures of forecast quality including POD, false alarm ratio, contingency bias, and CSI in a single diagram and give more in-depth information about the model performance. Given the strengths and weaknesses of each metric, it is often desirable to utilize multiple scores in order to assess the overall quality of the forecasts.

In this paper, we produced model validation statistics using the Model Evaluation Tools (MET) tool [10] which includes a set of verification tools to compute multiple scores in order to assess and evaluate the performance of the forecast. In this case, different evaluation methods were combined, and different multivariable diagrams were used which offer a more comprehensive approach to assess the ability of the model to predict weather parameters and events. The technical design of the numerical weather prediction system used in the experiments and the results exploring synoptic charts and the major changes related to the passage of the cold fronts associated with the case study used for the forecast model's evaluation were already explained in previous works [2,3]. The next section gives a quick summary of the different model configurations used and explains the technique selected for the model evaluation. The neighborhood technique deployed to the assessment of the precipitation and cloud fraction gridded forecasts with respect to satellite products are presented in section 2 and 3, respectively. The summary and conclusions are given in section 4.

1. Data and Methodology

1.1. Model System Design

The Weather Research and Forecasting Model (WRF) is a numerical weather-prediction model that has been used for many applications including operational forecasting and research purposes [11-16]. A configuration of the real-time NWP system was setup to support the daily forecasting operations, using the Unified Environmental Modeling System (UEMS) to manage and to produce daily runs on a regional grid over northern Africa and Morocco. A prototype weather forecast system has been developed and applied with careful configuration. In the current set up the model runs one time per day for a period of 72 hours and is initialized with GFS forecasts. The simulated region is given in Figure 1. We use a two-layer nesting scheme: the mother domain (D_1) covers north Africa with a resolution of 12 km, and the inner domain (D_2) includes the Moroccan region with a horizontal resolution of 4 km. As weather forecasting system requires a big amount of computing resources and the simultaneous use of many processors, the model is running on the High Performance Computing facility of the CNRST. Detailed weather domain research parameters and forecasting are given in Table 1. The model configurations and evaluation protocols for continuous variables are described in detail in previous papers [2,3]. Two different configurations were statistically evaluated over the region: (1) the Mellor–Yamada–Nakanishi–

Niino (MYNN) scheme (gfs1 and GVF1 in Table 1) [17-19], and (2) the Yonsei University scheme (YSU; gfs2 and GVF2 in Table 1) [20,21].



Figure 1. The two WRF model domains used in the study.

Table 1. The different configurations used to run the UEMS system over Morocco and north Africa.

Configuration Detail	gfs1 and GVF1 specifications	gfs2 and GVF2 specifications
Model grid: Domain D1	365x281 with 12-km spacing	365x281 with 12-km spacing
Model grid: Domain D2	427x445 with 4-km spacing; 1-way nesting	427x445 with 4-km spacing; 1-way nesting
Vertical sigma levels	42 levels; pressure top of 30 hPa	42 levels; pressure top of 30 hPa
Domain D1 time step	72 seconds	72 seconds
Short/Long wave radiation parameterization	RRTMG	RRTMG
Microphysics scheme	Lin 5-class	Lin 5-class
Convection parameterization	Multi-scale Kain-Fritsch (D ₁ only)	Multi-scale Kain-Fritsch (D ₁ only)
Planetary boundary layer	Mellor-Yamada Nakanishi Niino (MYNN2)	Yonsei University (YU)
Land surface model	Noah	Noah
Initialization and integration	0000 UTC, 72-hour forecasts, once daily	0000 UTC, 72-hour forecasts, once daily
Initial/Boundary conditions	NCEP GFS model 0-72-h forecasts in 3-h intervals	NCEP GFS model 0-72-h forecasts in 3-h intervals
land use categories datasets	gfs1: MODIS data GVF1: daily real-time global NESDIS VIIRS GVF product	gfs2: MODIS data GVF2: daily real-time global NESDIS VIIRS GVF product
Sea surface temperature (SST)	gfs1: NCEP Real-Time Global; fixed for simulation GVF1: SPoRT Sea Surface Temperature data	gfs2: NCEP Real-Time Global; fixed for simulation GVF2: SPoRT Sea Surface Temperature data
Land surface initialization	GFS 0-h soil temperature/moisture	GFS 0-h soil temperature/moisture

The influences of different green vegetation fraction (GVF) input data and Sea Surface Temperature data on the WRF forecast fields were studied using NESDIS/VIIRS green vegetation fraction (GVF) data and SPoRT Sea Surface Temperature (SST) data. For gfs1 and gfs2 (Table 1) surface properties and SSTs are imported from MODIS data with good spatial resolution and NCEP fixed Sea surface temperature. GVF1 and GVF2 (Table 1) correspond to the incorporation of daily global NESDIS/VIIRS

green vegetation fraction (GVF) data, which are generated in real-time by NOAA [22,23] and daily SPoRT Sea Surface Temperature data [24].

Daily accumulated precipitation and mean cloud fraction forecasts from a simulation using the Advanced Research WRF (ARW) dynamics is examined and evaluated. These forecasts were initialized at 0000 UTC and integrated over north Africa (D_1) during a most intense prevailing synoptic weather condition associated with cold fronts passage which took place between 03 December 2019 and 04 December 2019. More information about results exploring synoptic charts, the major changes related to this event and the evaluation of the used forecasting system for surface and upper level continuous observations were explained in previous works [2,3].

1.2. Evaluation Method

Over some parts of Africa, a major concern for the verification of numerical weather prediction systems, is the scarcity of in-situ, surface, upper level and rain gauge observations, which adversely affects the model forecast skill. Fortunately, satellites cover almost all of Earth's surface, and are employed to derive high-quality observations to help fill these data voids. The precipitation outputs of the WRF model were compared with precipitation data estimated by CMORPH [25] and IMERG [26] satellite products, and cloud data produced from MODIS imagery [27-29] were used to evaluate the model skill in representing the spatial pattern and timing of cloud fraction forecasts.

In this study, 24-hourly accumulated precipitations and mean cloud fractions forecasts are compared to those from satellites using Grid-Stat tool, by applying the neighborhood-based technique and thresholds to the grids to compute contingency-table statistics for the entire domain. The satellite products have been interpolated to model resolution from their native resolution and accumulation period also have been matched with the forecast. The evaluation of the WRF precipitation and cloud cover was carried out through two approaches: (1) quantification of the accuracy or discrepancy between WRF estimates and satellite products; and (2) analysis of the capacity to detect precipitation and cloud patterns.

In the first evaluation approach, precipitation and cloud fraction data were considered as continuous variables and three indicators, through Taylor diagram, were used to measure the accuracy or discrepancy between the forecasts and the satellite estimates: the correlation, the normalized root-mean-square (RMS) difference, and the standard deviation. For the second approach, precipitation and cloud fraction observations were considered categorical events and the following indicators will be used: false alarm rate (FAR), probability of detection (POD), critical success index (CSI), and frequency bias (bias) (details can be found in the textbooks of Wilks [30]). The recommended verification measures for forecast categories are: (i) forecasts of the event occurrence meeting or exceeding specific thresholds, (ii) forecasts of event amount, (iii) probability forecasts of event meeting or exceeding specific thresholds, (iv) verification of ensemble probability distribution. In this work, quantitative precipitation and cloud fractions forecasts verification analysis has been conducted using the four grid scores of the (i) category.

For fuzzy or neighborhood verification, Grid-Stat compares the forecasts and observations at grid points in a neighborhood surrounding the point of interest rather than comparing a single point from both fields. With these verification techniques, the user chooses a distance within which the forecast event can fall from the observed event and still be considered a hit [10]. In MET tools neighborhood verification is implemented by defining a square search window around each grid point. Within the search window, the number of observed events is compared to the number of forecasted events [10]. For this study various spatial scales were chosen. The size of the radius of influence considered varied from 12-km (1 grid point) up to 180-km (15 grid points).

The metrics and diagnostics used for scoring arise by defining an event from both the forecast and the observation grids. The event is defined by the use of a category or a threshold that serves as the basis for determining “hits” or “misses”, which follows the established theoretical framework for evaluating deterministic binary forecasts. This framework computes the skill of the forecast by counting the number of times the event was observed (or not) and forecasted (or not) in a contingency table.

To carry out the calculation of the category indicators, a 2x2 contingency table was used as a tool, since it allows a summary view of the frequency in which the event is detected (precipitation or cloud cover threshold $X \geq X_0$) or not ($X < X_0$) by the model or the satellite product. There are 4 possible combinations: (1) hit (a), both the model and the satellite detect the event; (2) false alarm (b), the model detects the event, but the satellite does not; (3) miss (c), the satellite detects the event, but the model not; and (4) correct rejection (d), neither the model nor the satellite detect the event. Table 2 illustrates a typical 2x2 contingency table along with associated definitions.

Table 2. The 2x2 contingency table containing general classification terms.

		Satellite Observation	
		Event detected	Yes ($X \geq X_0$)
Forecast	Yes ($X \geq X_0$)	Hits (a)	False alarms (b)
	No ($X < X_0$)	Misses (c)	Correct rejections (d)

The POD, also known as hit rate, is the fraction of precipitation events detected by the model. This indicator is sensitive to hits but ignores false alarms, so it must be used in conjunction with the FAR. The FAR is the fraction of forecasted events that did not occur, which is sensitive to false alarms, but ignores misses. The CSI is the fraction of correctly forecasted events without considering the correct rejections. The bias measures the ratio of the frequency of forecast events to frequency of observed events, and reveals whether the model over forecasts or under forecasts. A value of one indicates a perfect situation with no bias, a score over one represents over forecasting, and less than one represents undertreating, and it must be used in conjunction with the other three indicators. Table 3 presents the indicators, the equations used for their calculation, their range and their optimal value; a is the total of hits, b is the total of false alarms, and c is the total of misses.

Table 3. Forecast metrics calculated in this study.

Indicator	Equation	Range	Perfect Score
Probability of detection (POD)	$POD = a/(a+c)$	[0, 1]	1
False alarm rate (FAR)	$FAR = b/(a+b)$	[0, 1]	0
Critical success index (CSI)	$CSI = a/(a+b+c)$	[0, 1]	1
Frequency bias index (FBI)	$FBI = (a+b)/(a+c)$	[0, ∞]	1

The overall performance of precipitation and cloud fraction forecasts were tracked on a daily basis as well as cumulatively throughout the entire experiment (domain D_1). Since none of the calculated scores alone is capable to completely assess the quality of a forecasting system, the four defined performance indexes were computed and visualized in a single diagram using Roebber Performance Diagram [9].

2. Evaluation of Precipitation

2.1. Satellite precipitation estimates

Forecasters view and analyze precipitation from multiple sensors, including radars, gauge networks, and satellites. While daily scale precipitation data is available from multiple sources, very few reliable sub-daily rainfall estimates are available over the African region. Advances in satellite tools for

precipitation analysis are showing great promise in such areas with poor coverage from weather radars and rain gauge networks. Satellite-based products become crucial to provide appropriate temporal and spatial resolution in regions with the scarcity in availability of surface weather observations. In this case the merged satellite guidance provides useful details about the movement and evolution of heavy rainfall systems [4,25,26]. This study investigated the performance of our model forecast using two high-resolution global satellite-based precipitation products: the climate prediction center MORPHing technique (CMORPH) and the latest version of the Integrated Multi-SatellitE Retrievals for the Global Precipitation Mission (GPM) algorithm (IMERG-V06), over the north Africa (D₁).

Climate Prediction Center Morphing Technique (CMORPH) produces global precipitation data at very high temporal and spatial resolution [25]; The data are gridded at 0.07277 degrees' resolution (8 km at the equator) with a temporal resolution of 30 minutes. The technique uses precipitation estimates derived from low orbiter satellite microwave observations, and whose features are transported via spatial propagation information that is obtained entirely from geostationary satellite IR data. CMORPH binary files are available at the NCEP website (https://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V0.x/RAW/8km-30min/).

The American National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) launched the Global Precipitation Measurement (GPM) Core Observatory satellite on 2014 [31]. IMERG is the level 3 products of the GPM mission and is available in three product types: Early, Late, and Final; the first two types are near real time with a latency time of four and 12 h after observation, respectively. The final run with latency of up to two months after observation requires gauge analysis correlation and produces a product that is expected to have the best accuracy and quality [26]. This study utilized the IMERG Final run with 0.1°x0.1° spatial resolutions and 30 min temporal resolution; It includes TRMM-era data going back to June 2000, and can be downloaded from <https://gpm.nasa.gov/data/directory>. The IMERG algorithm inter-calibrates, merges, and interpolates all satellite-based microwave data together with microwave-calibrated IR satellite data [26]. Version 6 of IMERG is the latest version of the IMERG algorithm since its inception in 2014. The product includes different new features, such as full inter-calibration to the GPM combined instrument dataset, increasing the maximum precipitation threshold from 50 to 200 mm/h, the use of an updated rain retrieval algorithm, and incorporation of the Advanced Technology Microwave Sounder (ATMS) dataset.

Prior to any comparison, significant data processing was needed for the two used satellite products due to the need for reformatting to raster datasets over the study area, to be regridded to unified grids and aggregated to daily temporal resolution.

Figure 2 shows the 24-h accumulated precipitation of the different products in their original spatial resolution over north Africa and southern Europe. The image is divided as a table, In the first column CMORPH images were placed, the second column denotes the images referring to IMERG images and the third to forecasted precipitations. All the images in the first line refer to 03 December 2019, and the second line denotes the images referring to 04 December 2019. The comparisons considered whether the simulations were able to reproduce the shape, position and intensity of precipitation. Note that only the GVF1 configuration for WRF forecasts were presented in Figure 2. For 03 December 2019, it is noted that the WRF simulations managed to detect the presence and location of the precipitation range, however it showed more intense precipitation compared to satellite products over the highest terrain areas of the Middle Atlas Mountains. In the following day, the model overestimates precipitation over the highest terrain areas of Middle Atlas Mountains, and underestimates it over the northeast Tunisia. In general, the satellite rainfall products and the forecasted one show some similarities about the extent of rainfall systems, shape, and position although the magnitude of the rainfall may be different. Also,

IMERG data showed large bands of low precipitations and more intense rainfall compared to CMORPH and forecasted precipitations.

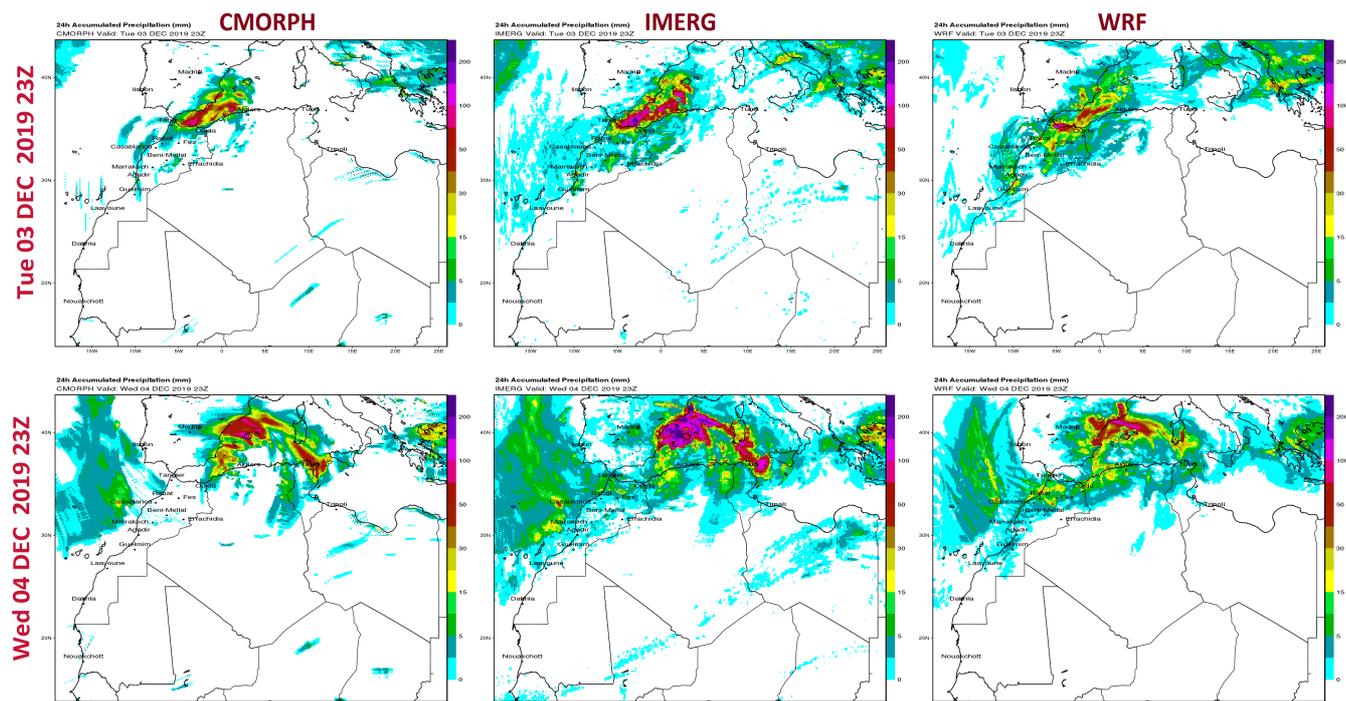


Figure 2. 24-h accumulated precipitation of domain D_1 from 0000 UTC 3 December 2019 to 2300 UTC 4 December 2019 for each precipitation product at their original spatial resolutions.

2.2. Taylor diagrams

The Taylor diagrams in Figure 3 have been used to analyze how closer or how far the four configurations used (gfs1, GVF1, gfs2, and GVF2) match the satellite-based rainfall products over the parent domain D_1 . These diagrams visualize the relative skill of the different model configurations in terms of normalized standard deviation (σ), normalized RMS and Pearson's correlation coefficient (r). The black cross symbol (OBS) represents the skill of the model that perfectly reproduces the observations, where r is 1, normalized RMS is 0, and the normalized standard deviation is 1. The green-scale background represents the skill score which provides further guidance for interpreting the model performances.

The variability is represented by the normalized standard deviation (σ) of the observed and forecasted values. The blue arc line in Figure 3 corresponds to normalized standard deviation value of 1; If $\sigma < 1$ the forecast has less variability than the observation, whereas if $\sigma > 1$ the forecast has more variability than the observation. The red dashed lines highlight the centered RMS values which measure the differences between the predicted and the observed values. Yellow dashed lines are correlation coefficients values (r), which measure the linear relationship between forecasted and observed variables.

For 03 December 2019 (forecast leading time 24), the verification against IMERG data (left panel of the Figure 3) for the four forecast configurations show generally poor predictive skill for 24-h accumulated precipitation, with the best performing configuration, GVF2, having a correlation coefficient of 0.6 and normalized RMS error around 0.8. All model's configurations underestimate the diurnal variability of observed IMERG precipitation by 45-50%. In particular, the variability from the gfs1 and GVF1 configurations is only 50% of the observed value. In the following day (forecast leading time 48), the different forecast configurations used get even worse skills; diurnal variability of observed IMERG precipitation is underestimated by 60-65%.

The validation of the four-configuration using CMORPH data (right panel of the Figure 3), shows that the model performs better than the case when comparing with IMERG data. The scores indicate strong skill in the diurnal variability ($\sigma \sim 0.9-1$) but nearly the same correlation coefficients as those found when comparing with IMERG ($r \sim 0.5-0.6$). Again the best performing configuration is GVF2 and the forecast skill decreases a little bit in the following day. In this case the diurnal variability of observed CMORPH precipitation is underestimated by 10-20%.

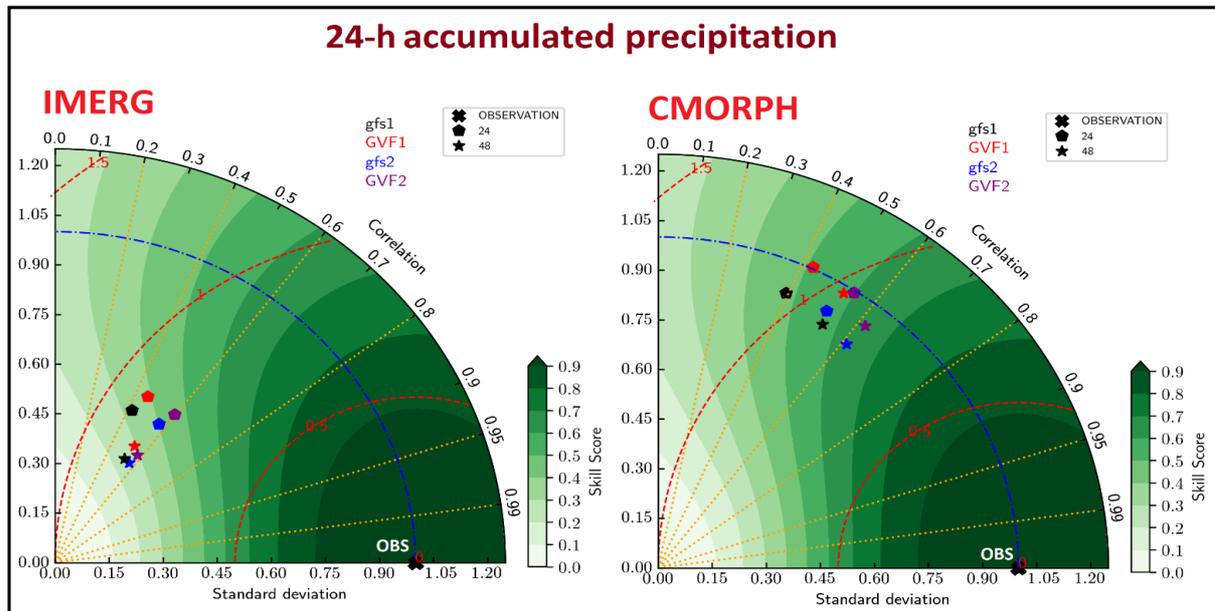


Figure 3. Taylor diagrams comparing the 24-h accumulated rainfall for the four used forecast configurations with: IMERG observations versus forecast lead time (left panel); CMORPH observations versus forecast lead time (right panel). Observation is presented by the black cross symbol (x) on the x-axis (OBS).

As a conclusion, IMERG and CMORPH precipitation products present similar patterns, but have differences in diurnal variability. The underestimation of the diurnal variability of observed CMORPH precipitation at the forecast leading time 48 is consistent with the underforecast of accumulated precipitation over the northeast Tunisia observed on 04 December 2019 (in Figure 2).

2.3. Performance diagrams

By analogy to Taylor [8], who exploited the geometric relationship between three measures of model performance (correlation, normalized root-mean-square difference, and standard deviation) to represent them in a single diagram, Roebber proceeded in a similar fashion with dichotomous (Yes/No) forecasts [9]. He showed that the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and bias are geometrically related and consequently can be plotted in a single diagram. For good forecasts, success ratio ($SR=1-FAR$), POD, CSI, and bias approach unity, such that a perfect forecast lies in the upper right of the diagram (the dark green zone in diagrams of Figures 4 and 5). Deviations in a particular direction will show the relative differences in SR and POD, and consequently CSI and bias. This visual representation is preferable to simple tables, and immediate visualization of differences in performance is obtained.

The overall performance of selected configurations was tracked on a daily basis as well as cumulatively throughout the entire experiment domain (D_1) using Roebber Performance Diagrams, pictured in Figures 4 and 5. Dashed labeled lines represent bias scores, while green contours are CSI. Note that unbiased forecasts will be represented by any point on the 45° line. Figure 4 shows the

performance diagrams of 24-h accumulated precipitation forecast versus precipitation thresholds (1, 5, 10, 25 and 50mm) with respect to IMERG observations (first column), and CMORPH data (second column) for the four used configurations.

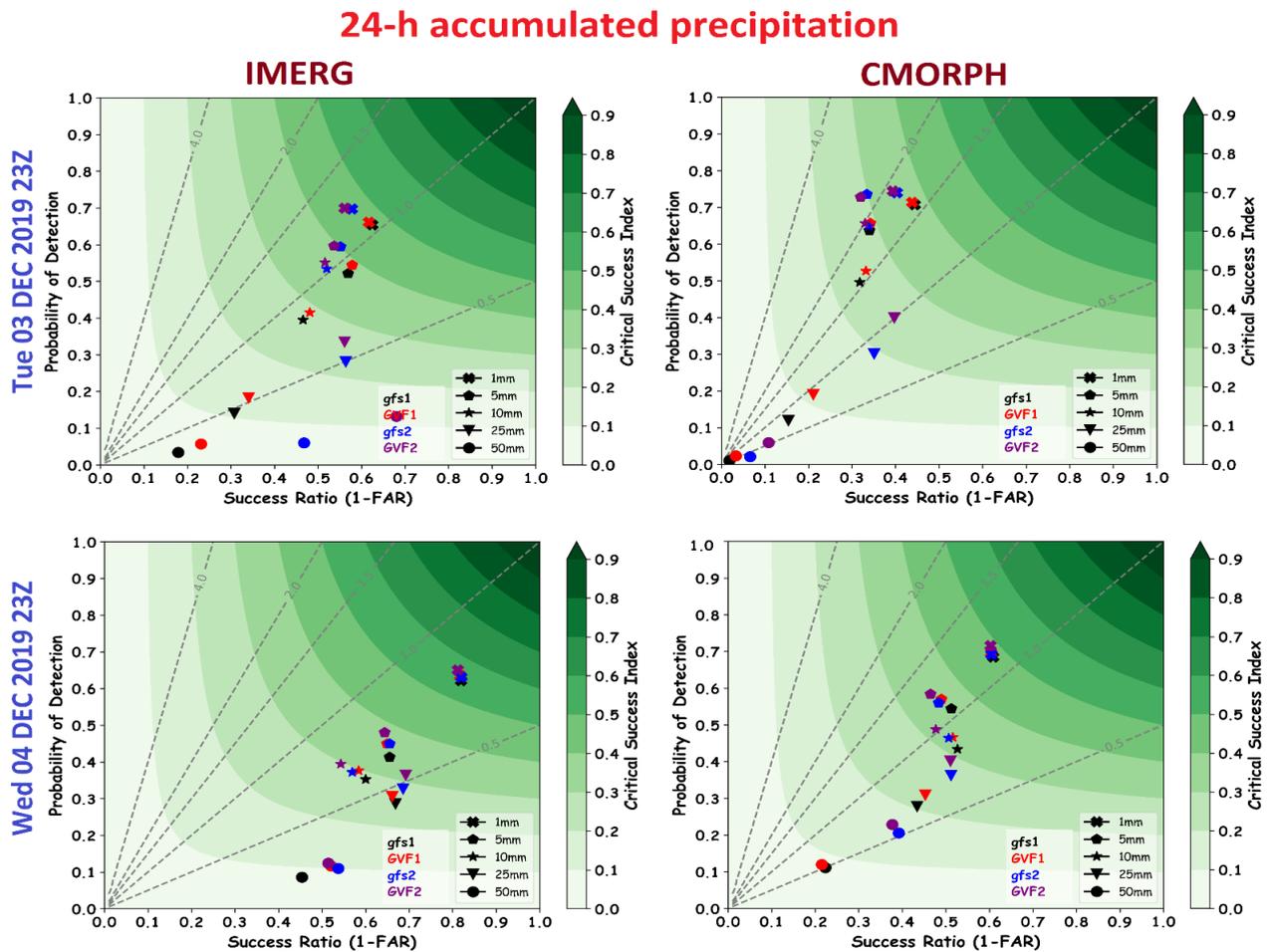


Figure 4. Performance diagrams summarizing the POD, SR, CSI and bias. Shown are daily accumulated precipitation forecast performances versus precipitation thresholds (1, 5, 10, 25 and 50mm) with respect to IMERG observations (first column), and CMORPH data (second column) for the four used configurations. Green curved contours represent CSI and dashed diagonal lines represent the bias.

All the images in the first line refer to 03 December 2019, and the second line denotes the images referring to 04 December 2019. The width of the neighborhood grids over which verification is performed is 1, so only grid point to grid point matching was done in this case. For 03 December 2019, the objective verification against IMERG data (left top panel of the Figure 4) supports that at lighter thresholds of 1 and 5mm, gfs2 and GVF2 forecast configurations have a noticeable wet bias (bias > 1) and at the higher threshold of 25 and 50mm, they have a dry bias (bias < 1). In this case, forecasts are slightly unbiased at thresholds of 10mm. Except, for 1mm threshold where precipitation forecasts are relatively unbiased, gfs1 and GVF1 have a noticeable dry bias at higher thresholds. In the following day (left bottom panel), the forecast skill decreases and there is a large under forecasting bias for all types of forecasts. The validation of the four configurations using CMORPH data (right top panel), shows that the model overestimates observed precipitations at lighter thresholds of 1, 5 and 10mm, and underestimate precipitations at the higher threshold of 25 and 50mm. For the next day (right bottom panel), forecasts have a wet bias at lighter thresholds of 1 and 5mm and they have a dry bias at higher thresholds of 25 and 50mm.

The data demonstrate the meteorological challenge of forecasting heavy precipitation, the performance decreases with rainfall intensity. Also, the gfs2 and GVF2 show higher values of POD statistics, if compared to the simulations with the gfs1 and GVF1 configurations, but are also biased high at thresholds of 1, 5 and 10mm. The high values of POD statistics may be due to the increase of areas covered with rainfall, if the Yonsei University scheme is used. Another discovery of this study was finding a threshold that marks the steep drop of forecast scores which is 25mm. So there is an apparent limitation in the model's ability to simulate heavy precipitation events.

Figure 5 shows performance diagrams valid for the two days of the experiment for GVF2 configuration at two different thresholds: 1mm (first column) and 10mm (second column). In this case computed skill metrics with respect to IMERG and CMORPH observations are compared for different neighborhood sizes (1, 5, 11 and 15 grids) over which neighborhood statistics are accumulated. These figure supports that forecast performance improves with increasing spatial scale, and as rain threshold increased the forecast performance worsened at all scales.

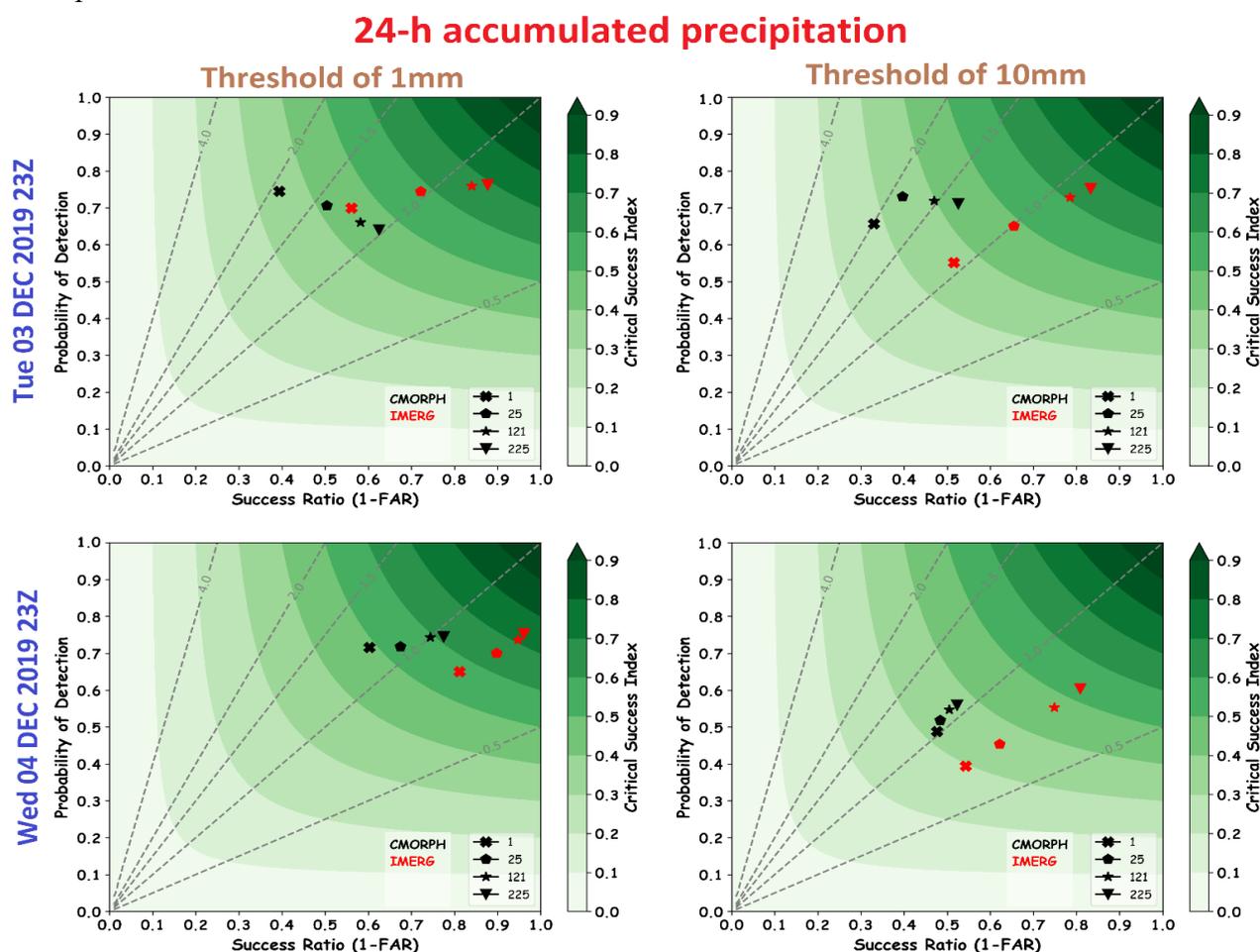


Figure 5. Performance Diagrams summarizing the SR, POD, bias, and CSI for daily accumulated precipitation forecast performances versus number of grid boxes over which statistics are accumulated (1 , 5^2 , 11^2 and 15^2) for 1mm (first column) and 10mm (second column) thresholds.

3. Evaluation of Cloud Fraction

3.1. Satellite cloud fraction estimates

For the evaluation of cloud cover, we use the Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6.1 cloud fraction product from both Terra (MOD06_L2) and Aqua (MYD06_L2) [31,32], over the northern Africa and southern Europe. MODIS is a 36-channel spectroradiometric sensor that was installed on board both Terra and Aqua sun-synchronous, near-polar satellite platforms, which

were launched in December 1999 and May 2001, respectively. The different orbital paths of the two platforms gives different overpass times at morning (Terra descending node 1030 local time) and afternoon (Aqua ascending node 1330 local time) over the equator, which allows MODIS to observe the earth two times during the day.

MODIS has a 1 km² field of view mapping to a swath of approximately 2330 km with global data archived every day. It observed the earth from a 700-km altitude (± 55 degrees' view scan). So it had approximately global coverage every 1–2 days on a daily basis with a repeat cycle of 16 days (<http://modis.gsfc.nasa.gov>). The MODIS data are organized into 5-minute sections called granules. Each granule contains 2030 lines of data, and each line is composed of 1350 pixels. The MODIS cloud product contains both physical and radiative cloud properties, including cloud mask, cloud particle phase (clouds versus snow, ice versus water) mask, cloud top temperature/pressure/height, cloud optical depth and effective cloud particle radius. We leveraged the availability of collocated MODIS (MOD06_L2 and MYD06_L2) data over the experiment domain from both Terra and Aqua to calculate 24-h mean cloud fraction. Figure 6 presents example of 24-h mean cloud cover fraction images from the satellite product and WRF GVF1 simulations for the first day (first line) and the second day (second line) of the case study. The locations of forecasted cloudy regions correspond to the satellite-derived image, but there is some discrepancy in cloud fractions in some regions. Cloud fraction bias is relatively large over southern Tunisia during the first day, but relatively small over Tunisia during the first day and the southern of Morocco during the two days of forecast. Note that, in general the two satellite 24-h average cloud fraction spatial distribution maps show similar patterns.

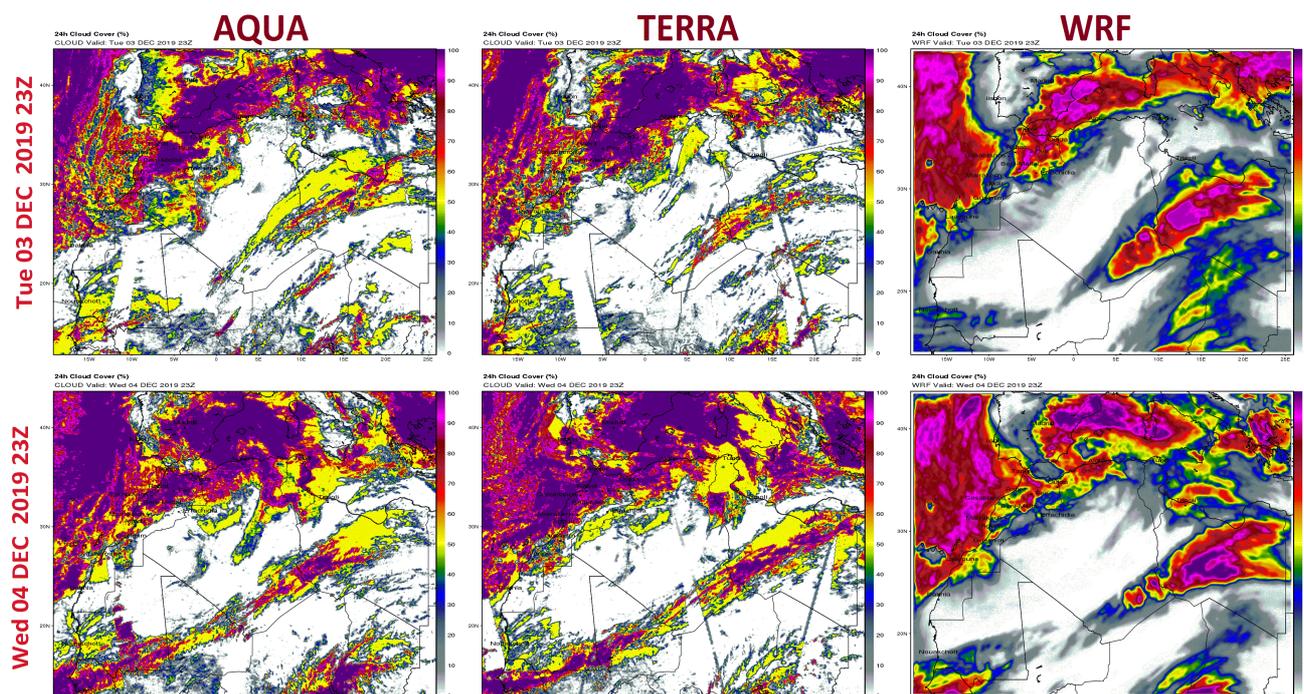


Figure 6. RegridDED 24-h mean cloud fraction from 0000 UTC 3 December 2019 to 2300 UTC 4 December 2019 for each cloud product.

3.2. Taylor diagrams

The Taylor diagrams in Figure 7 have been used to analyze the skill of the four used configurations (gfs1, GVF1, gfs2, and GVF2) over the parent domain D_1 . In general, the four configurations show similar skills for the two forecast days and light differences can be seen when comparing skills with respect to the two used satellite products. All model's configurations underestimate the diurnal variability

of observed satellites cloud fraction by about 25% (normalized standard deviation ~ 0.75), and have a correlation coefficients values of 0.5-0.55 and a normalized RMS error around 0.85.

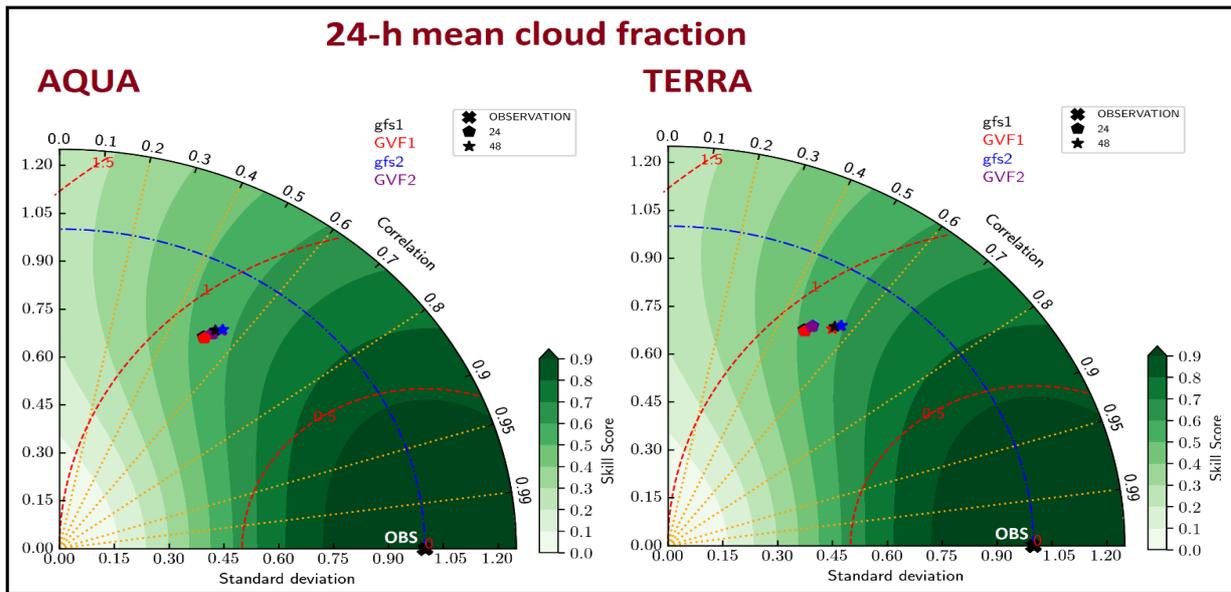


Figure 7. Taylor diagrams comparing the 24-h mean cloud fraction for the four used forecast configurations with: AQUA observations versus forecast lead time (left panel); TERRA observations versus forecast lead time (right panel)

3.3. Performance diagrams

Figure 8 shows the performance diagrams of 24-h mean cloud fraction forecast versus cloud fraction thresholds (20, 40, 60, and 80%) with respect to Aqua (first column), and Terra observations (second column) for the four used configurations.

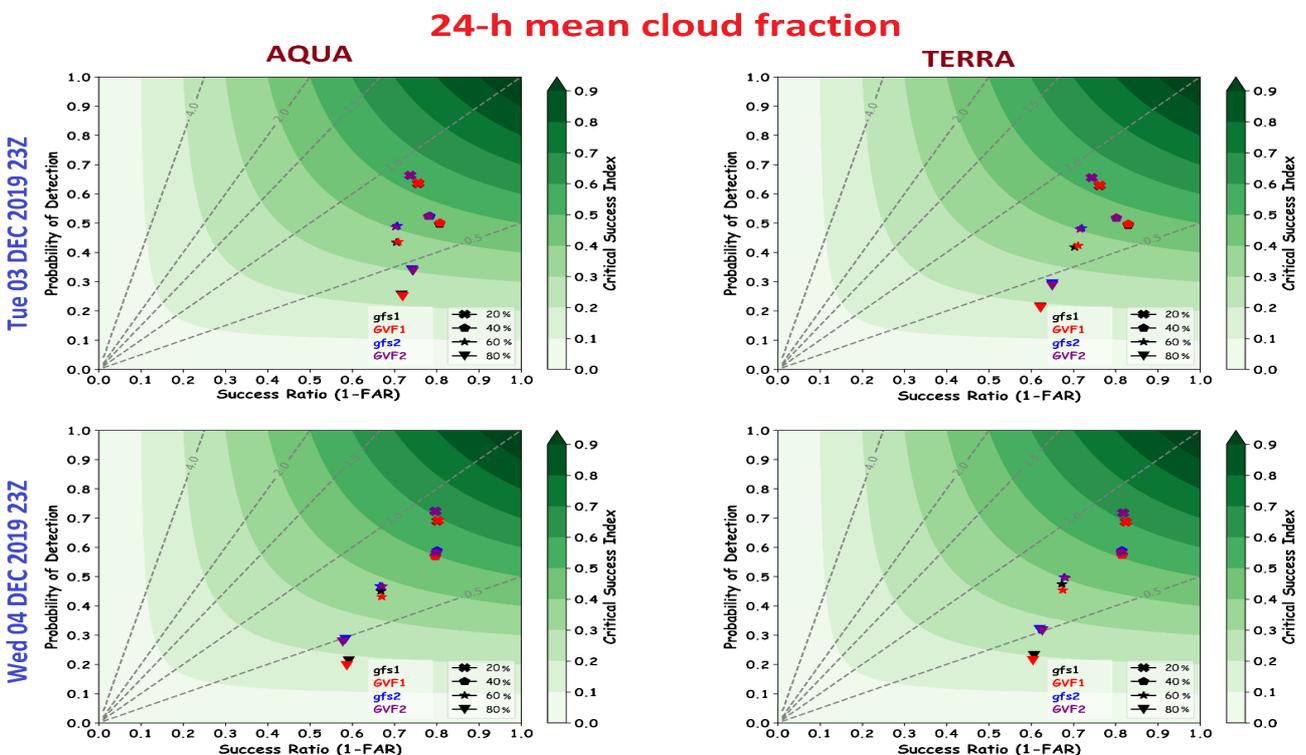


Figure 8. Performance diagrams summarizing the POD, SR, bias, and CSI. Shown are daily mean cloud fraction forecast performances versus cloud fraction thresholds (20, 40, 60, and 80%) with respect to Aqua observations (first column), and Terra data (second column) for the four configurations used. Dashed diagonal lines represent the bias, and green curved contours represent CSI.

Images in the first line refer to 03 December 2019, and those in the second line refer to 04 December 2019. Only grid point to grid point matching was done in this case.

The 24-h average cloud fraction skills with respect to the two satellites show similar results, and the validation of the four configurations shows similar results with an underestimation of observed cloud fractions for all thresholds (bias<1). Again, the performance decreases with cloud fraction intensity; the loss of skill is largely owing to a decreased POD, while SR slightly decreases from 0.85 to 0.65. As for precipitation, we can find a threshold that marks the steep drop of forecast scores. So at threshold higher than 80%, there is a clear limitation in the model's ability to simulate cloud cover fractions.

Conclusions and Future Direction

We developed a NWP system based on the WRF model to predict weather over North Africa and Morocco. We selected a case study day from a challenging winter precipitation forecast event over north Africa and Morocco, on December 3-4, 2019. The WRF forecasts were performed for two domains with spatial resolution of 12-km x 12-km (domain D₁) and 4-km x 4-km grid (domain D₂), and temporal resolution of 1 hour. Forecast calculations were carried out using the 55-km GFS forecast data as input data with different WRF parameter settings. The influences of different green vegetation fraction (GVF) input data and Sea Surface Temperature data on the WRF forecast fields were studied using NESDIS/VIIRS green vegetation fraction (GVF) data and SPoRT Sea Surface Temperature data. For the present study, only the forecast fields for domain D₁ were compared with the observed data.

Grid-Stat tool developed by the National Center for Atmospheric Research called Model Evaluation Tools was used to compute traditional error statistics and verification metrics to compare the performance of the forecast model with observed data for the entire domain. In this study, 24-hourly accumulated precipitations and mean cloud fractions forecasts were compared to those from satellites using the neighborhood-based technique. The evaluation was carried out through two approaches. In the first evaluation approach, precipitation and cloud fraction data were considered as continuous variables and three indicators, through Taylor diagram, were used to measure the accuracy of forecasts: the correlation, the normalized root-mean-square (RMS) difference, and the standard deviation. For the second approach, precipitation and cloud fraction data were considered as categorical events and four indicators (POD, FAR, CSI, and bias) were used and visualized in a single diagram using Roebber Performance Diagram. We believe that combining these different evaluation methods, and using multivariable diagrams offer a more comprehensive approach to assess the ability of the model to predict weather events. From this study, the following conclusions can be inferred:

1. In general, the satellite products and the forecasted one show some similarities about the extent, shape, and position of rainfall and cloud cover fraction systems although the magnitude may be different.
2. IMERG data showed large bands of low precipitations and more intense rainfall compared to CMORPH and forecasted precipitations. Also, IMERG and CMORPH precipitation products present similar patterns, but have differences in diurnal variability.
3. The validation of forecasted rainfall using CMORPH data, shows that the model performs better than the case when using IMERG data. The model overestimates CMORPH observed precipitations at lighter thresholds of 1, and 5mm, and underestimate precipitations at the higher threshold of 25 and 50mm.
4. From the performance diagrams we found a threshold that marks the steep drop of forecast scores (25mm for 24-h accumulated precipitations and 80% for 24-h mean cloud fractions). So there is an apparent limitation in the model's ability to simulate these events.

5. Model forecasts show an underestimation of cloud fraction for all thresholds, and as for 24-h accumulated precipitations the performance decreases with the increase of intensity threshold.

To the authors' knowledge, this is the first NWP system over this region that uses the WRF model at high spatial and temporal resolution for weather forecasting. So there is a further need for more complex evaluation of the forecasts for this study area. Spatially-based techniques could present an alternative for the high resolution rainfall and cloud fraction forecasts evaluation enabling a better representativeness of the uncertainty of these highly variable fields [33]. The MET has proven to be a powerful means assessing the accuracy of the WRF model. Consideration must be given to the use of other verification methods like the Method for Object-Based Diagnostic Evaluation (MODE) tool which is one of the main components of MET [10]. MODE represents a category of spatial verification methods whose purpose is to identify localized features and compare them to identify which features best correspond to each other, rather than evaluating hits and misses at a point/neighborhood as is done with the contingency table. Errors in precipitation or cloud fractions forecast could occur due to prediction of inaccurate event amount or faulty timing of the event or even errors and uncertainties associated with satellite data. In fact, satellite products have been released recently, more comprehensive evaluations are still essential to better understand error characteristics of these products. Therefore, a more critical analysis and evaluation of satellite products is vital to understand their capabilities and their strengths and weaknesses before any application in a specific region.

Acknowledgements: We would like to show our gratitude to the CNRST (Centre National pour la Recherche Scientifique et Technique) for allowing us to use their HPC facility to run our weather forecasting system.

References

1. E. Ebert, and L. Wilson, A. Weigel, M. Mittermaier, P. Nurmi, P. Gill, M. Göber, S. Joslyn, B. Brown, T. Fowler, and A. Watkins. Progress and challenges in forecast verification. *Meteor. Appl.*, 20 (2013) 130-139. <https://doi.org/10.1002/met.1392>
2. R. Moustabchir, and H. Charifi. Real-time NWP simulations over north-Africa and Morocco using the Unified Environmental Modeling System (UEMS). *Materials Today: Proceedings* 27 (2020) 3071–3081. <https://doi.org/10.1016/j.matpr.2020.03.564>
3. R. Moustabchir, H. Charifi, M. Amchghal, H. Amghar, A. Chirmata, A. Ihlal, K. Iraoui, and M. Ouattab. Application and Evaluation of the WRF Model over North-Africa: Continuous Observations. Submitted to *Moroccan Journal of Chemistry* (2021).
4. Using Merged Satellite Precipitation (CMORPH and IMERG) Guidance, COMET MetEd. Available online: https://www.meted.ucar.edu/training_module.php?id=1605#.XmwsSmb7l2w (accessed on 14 September 2020).
5. E. E. Ebert. Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework, *Meteorol. Appl.*, 15 (2008) 51–64. <https://doi.org/10.1002/met.25>
6. N. M. Roberts, and H. W. Lean. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, *Mon. Wea. Rev.*, 136 (2008) 78–97. <https://doi.org/10.1175/2007MWR2123.1>
7. S. E. Theis, A. Hense and U. Damrath. Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach, *Meteorol. Appl.*, 12 (2005) 257–268. <https://doi.org/10.1017/S1350482705001763>

8. K. E. Taylor. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, 106 (2001) 7183-7192 (also see PCMDI Report 55, <http://www-pcmdi.llnl.gov/publications/ab55.html>). <https://doi.org/10.1029/2000JD900719>
9. P. J. Roebber. Visualizing Multiple Measures of Forecast Quality. *Wea. Forecasting*, 24 (2009) 601–608, <https://doi.org/10.1175/2008WAF2222159.1>
10. B. Brown, R. Bullock, T. Fowler, J. Halley Gotway, K. Newman, and T. Jensen. The MET Version 9.1.1 User's Guide. Developmental Testbed Center. Available at: <https://github.com/dtcenter/MET/releases> (accessed on 22 November 2020)..
11. W. C. Skamarock, and J. B. Klemp. The stability of time split numerical methods for the hydrostatic and the non-hydrostatic elastic equations. *Mon. Weather Rev.* 120 (1992) 2109–2127. [http://dx.doi.org/10.1175/1520-0493\(1992\)120<2109:TSOTSN>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<2109:TSOTSN>2.0.CO;2)
12. J. B. Klemp, and W. C. Skamarock, and J. Dudhia. Conservative Split-Explicit Time Integration Methods for the Compressible Nonhydrostatic Equations. *Monthly Weather Review.*, 135 (2007) 2897–2913. <https://doi.org/10.1175/MWR3440.1>
13. W. C. Skamarock, J. B. Klemp, and B. Joseph. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.* 227 (2008) 3465– 3485. <https://doi.org/10.1016/j.jcp.2007.01.037>
14. W. C. Skamarock, J. B. Klemp, J. Dudhia, and D. O. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers. A description of the Advanced Research WRF Version 3. Report TN475+STR. NCAR: Boulder, CO (2008). <http://dx.doi.org/10.5065/D68S4MVH>
15. W. C. Skamarock, and M.L. Weisman. The impact of positive-definite moisture transport on NWP precipitation forecasts. *Mon. Weather Rev.* 137 (2009) 488–494. <https://doi.org/10.1175/2008MWR2583.1>
16. W. Wang, C. Bruyère, M. Duda, J. Dudhia, D. Gill, M. Kavulich, K. Keene, M. Chen, H. C. Lin, J. Michalakes, S. Rizvi, X. Zhang, J. Berner, S. Ha and K. Fossell. WRF-ARW Version 3 Modeling System User's Guide (United States: National Center for Atmospheric Research) (2017).
17. M. Nakanishi, H. Niino. An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorol* 119 (2006) 397–407. <https://doi.org/10.1007/S10546-005-9030-8>
18. M. Nakanishi, and H. Niino. Development of an improved turbulence closure model for the atmospheric boundary layer. *J Meteorol Soc Jpn* 87 (2009) 895–912. <https://doi.org/10.2151/jmsj.87.895>
19. Y. Kitamura. Modifications to the Mellor–Yamada–Nakanishi–Niino (MYNN) Model for the stable stratification case. *J. Meteorol. Soc. Jpn.* 88 (2010) 857–864. <https://doi.org/10.2151/jmsj.2010-506>
20. S.-Y. Hong, Y. Noh, and J. Dudhia. A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Weather Rev.*, 134(9) (2006) 2318-2341. <https://doi.org/10.1175/MWR3199.1>
21. S.-Y. Hong. A new stable boundary-layer mixing scheme and its impact on the simulated East Asia summer monsoon. *Q. J. R. Meteorol. Soc.*, 136(651) (2010) 1481-1496. <https://doi.org/10.1002/qj.665>
22. J. L. Case, Z. Jiang, and M. Vargas. Real-time Suomi-NPP green vegetation fraction for improving numerical weather prediction and situational awareness. 40th National Weather Association annual meeting, Oklahoma City, OK, *Natl. Wea. Assoc.* (2015) AP-10.

23. M. Z. Vargas, J. Ju Jiang, and I. A. Csiszar. Real-time daily rolling weekly Green Vegetation Fraction (GVF) derived from the Visible Imaging Radiometer Suite (VIIRS) sensor onboard the SNPP satellite. Preprints, 20th Conf. Satellite Meteorology and Oceanography, Phoenix, AZ, *Amer. Meteor. Soc.* (2015) P210.
24. B. T. Zavodsky, F. J. LaFontaine, E. Berndt, P. Meyer, and G. J. Jedlovec. Satellite data product and data dissemination updates for the SPoRT Sea Surface Temperature composite product. 13th Annual Symp. New Generation Operational Environmental Satellite Systems, Seattle, WA, *Amer. Meteor. Soc.* (2017) P255.
25. R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie. CMORPH. A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, 5 (2004) 487-503. [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2)
26. G. J. Huffman, D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, P. Xie, S.H. Yoo. Algorithm Theoretical Basis Document (ATBD) Version 4.5: Nasa Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG), NASA/GSFC, Greenbelt, MD, USA (2015).
27. S. Platnick, M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey. The MODIS cloud products: Algorithms and examples from Terra. *IEEE Trans. Geosci. Remote Sens.*, 41 (2003) 459–473, <https://doi.org/10.1109/TGRS.2002.808301>
28. J. Li, H. Huang, C. Liu, P. Yang, T. J. Schmit, H. Wei, E. Weisz, L. Guan, and P. Menzel. Retrieval of cloud microphysical properties from MODIS and AIRS. *J. Appl. Meteor.*, 44 (2005) 1526–1543, <https://doi.org/10.1175/JAM2281.1>
29. E. Weisz, J. Li, W. P. Menzel, A. K. Heidinger, B. H. Kahn, and C.-Y. Liu. Comparison of AIRS, MODIS, CloudSat and CALIPSO cloud top height retrievals. *Geophys. Res. Lett.*, 34 (2007) L17811, <https://doi.org/10.1029/2007GL030676>
30. D. S. Wilks. Statistical methods in the Atmospheric Sciences, International Geophysics Series, Academic Press, 91 (2006) 627p.
31. NASA. Global Precipitation Measurement (GPM) Mission Overview. Available online: <https://gpm.nasa.gov/missions/GPM> (accessed on 1 October 2020).
32. S. A. Ackerman, R. E. Holz, R. Frey, E. W. Eloranta, B. C. Maddux, and M. McGill. Cloud detection with MODIS Part II: Validation, *J. Atmos. Oceanic Technol.*, 25 (2008) 1073–1086, <https://doi.org/10.1175/2007JTECHA1053.1>
33. Ian T. Jolliffe and David B. Stephenson. Forecast Verification: A Practitioner's Guide in Atmospheric Science. Wiley (2011).

(2021) ; <http://www.jmaterenvirosci.com>