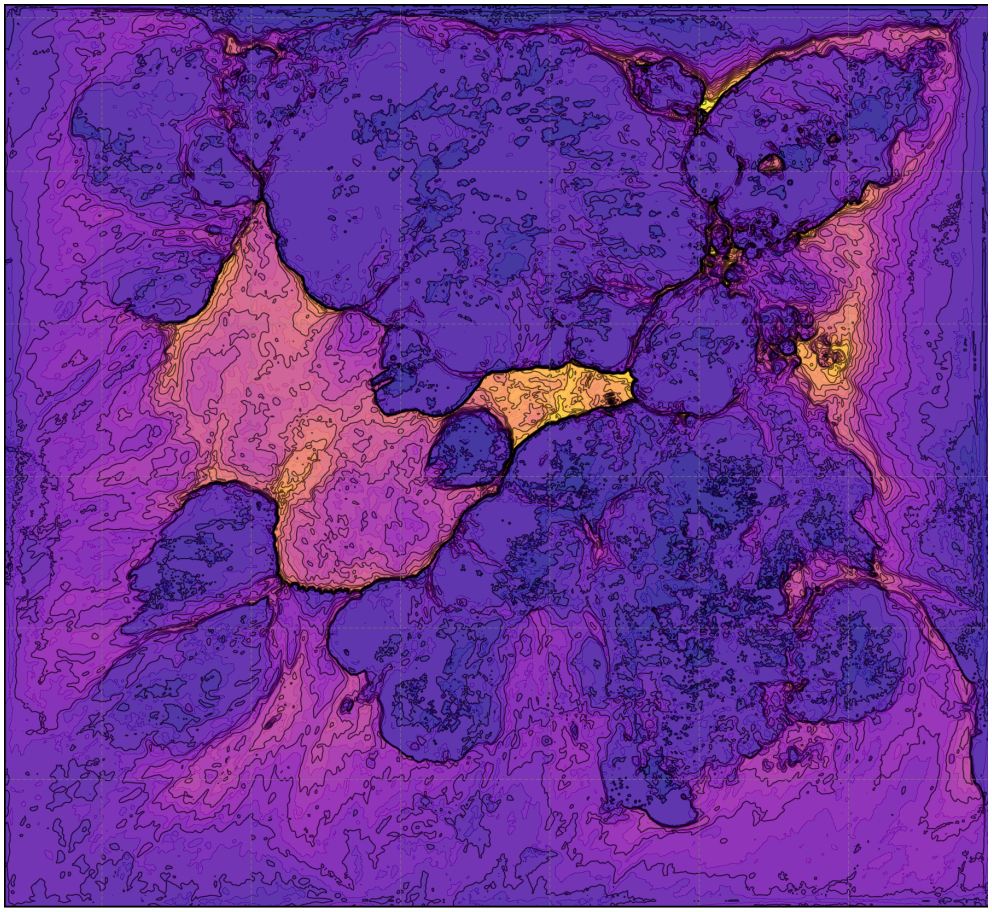
Sensitivity of Mesoscale Convective System forecasting in the Sahel to synthetic sub-daily soil moisture observations using WRF



Nefeli Boscolo

Sensitivity of Mesoscale Convective System forecasting in the Sahel to synthetic sub-daily soil moisture observations using WRF

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

APPLIED EARTH SCIENCES

by

Nefeli Boscolo

Department of Geosciences & Remote Sensing
Faculty CEG, Delft University of Technology
Delft, the Netherlands

Cover picture: Snapshot of boundary layer height from model output.

Sensitivity of Mesoscale Convective System forecasting in the Sahel to synthetic sub-daily soil moisture observations using WRF

Author: Nefeli Boscolo
Student id: 6056520

Abstract

In semi-arid regions such as the Sahel surface soil moisture (SSM) strongly influences the partitioning between sensible and latent heat flux, with SSM anomalies on scales larger than 10 km favoring convection initiation. Mesoscale Convective Systems (MCSs) can be triggered as a result, which numerical weather prediction cannot accurately forecast with current available observational capabilities. This study investigates the sensitivity of MCS forecasting in the Sahel to synthetic sub-daily soil moisture observations using the Weather Research and Forecasting (WRF) model, in the context of the Sub-daily Land-Atmosphere INTEractions (SLAINTE) satellite mission concept. A high-resolution idealized case study was performed to assess the impact of SSM perturbations and data assimilation at different observation spatial resolutions (1 km, 5 km, 12 km) and observation times. Results indicate that synthetic observations three times per day at 5 km resolution would improve forecasting of precipitation intensity and timing. Finer resolution observations could improve forecasts only if observation noise is reduced, while 12 km resolution observations –representing the Advanced SCATterometer (ASCAT) satellite– tend to further disrupt them. The experiments highlight that at least one observation before 12:00 and an observation at 18:00 local time are necessary to constrain the forecasts sufficiently. These findings provide insight into the role of soil moisture observations for convective forecasting in semi-arid regions and contribute to defining requirements for future satellite missions such as SLAINTE.

Thesis Committee:

Chair:	Prof.dr.ir. Susan Steele-Dunne, Faculty CITG, TU Delft
University supervisor:	Dr. Martina Messmer, Faculty CITG, TU Delft
External supervisor:	Dr. Christopher Taylor, UK Centre for Hydrology and Meteorology
Committee Member:	Dr. R. Tri (Rajashree) Datta, Faculty CITG, TU Delft

Preface

I must begin by thanking my wonderful supervisors. Susan and Martina, I could not have wished for a more supportive and motivating supervisory team. Thank you Susan, for reaching out to me all those months ago, for your guidance, for keeping me motivated and for grounding me when I derailed in wild sheep chases. Thank you Martina for your patience, your advice and the countless hours spent discussing results and finding bugs. Thank you both for your kindness and support. Thank you to Chris, for the feedback along the way and the opportunity to work on a topic so fascinating that it inspired me to continue further in this field of research. It has been incredibly valuable to work with all of you and I am lucky to have had the chance to do so.

Thank you to my friends, who got me through these last few months. Thank you Oak, for dragging me all the way to Munich to study and for your patience when I hijacked our study sessions with endless yapping. Thank you Suus, for everything I cannot begin to put into words. Thank you to Haris, for sharing and enduring these two years of studies together. And thank you to Milo, my unexpected thesis buddy.

And finally, thank you to my family. Most of what I have done I owe to your support. Thank you for always motivating me to pursue new fields and experiences, and for sending me care packages when I missed home.

Nefeli Boscolo
Delft, the Netherlands
September 25, 2025

Contents

List of Figures

Chapter 1

Introduction

In this chapter, the context and motivation of this research as well as the research questions that guide it are introduced.

1.1 Background and relevance of the project

This study focuses on the requirements for satellite observation of soil moisture to improve the predictability of the initiation of Mesoscale Convective Systems (MCS) and convective activity in regions where the soil moisture conditions exert a strong control on atmospheric processes. Many studies have been conducted on the influence of soil moisture on convective activity (W. Liu et al., 2022). When soils dry, evaporation can become water limited (Seneviratne et al., 2010), and sensible heat warms and deepens the planetary boundary layer (PBL). Surface heat flux increases under a larger horizontal soil moisture gradient, creating secondary mesoscale circulations similar to sea-breeze flows (W. Liu et al., 2022). Once this soil moisture gradient reaches 10 km on the horizontal scale, a significant heat flux gradient will drive the development of local circulation patterns (Pielke Sr., 2001), and variations in soil moisture on length scales of 10–40 km has been shown to exert a strong control on convective activity and storm initiation (Taylor et al., 2011). Once convective instability has been released, organized clusters of thunderstorms referred to as MCSs can be triggered, which expand and propagate over hundreds of kilometers. The new soil-wetness patterns produced by the precipitation of a MCS in turn increases the probability of another MCS in the following day or two (Taylor et al., 2011).

A region where these phenomena are prominent is the Sahel, a semi-arid and largely flat region bounded to the north by the Sahara desert and to the south by closed canopy tropical forest (Taylor et al., 2011). The Sahel provides a "natural laboratory" for studying soil moisture–MCS interactions and is the focus of systematic analysis of soil moisture controls on convection within mature MCSs anywhere in the world (Koster et al., 2004). During boreal summer, the Sahel is a global hotspot in intense MCS activity (Zipser et al., 2006) and the vast majority of precipitation events in the region are due to MCSs (Mathon et al., 2002). Flash floods often cause damage in infrastructure and affect significantly the local population (Di Baldassarre et al., 2010); in the context of climate change, this issue

is worsening as the region has already seen a rapid increase in the frequency of the most intense MCSs (Taylor et al., 2017), making improved forecasts urgent and crucial.

In the Sahel, convection reaches its diurnal peak in intensity and frequency during the afternoon–evening transition (Taylor et al., 2012); the MCSs that are potentially generated have a lifetime of 6 to 18 hours (Taylor et al., 2022) and the footprint of the rainfall patterns on the soil have a rapidly decreasing effect on consequent convective activity in the next 3 to 4 days (Taylor et al., 2024).

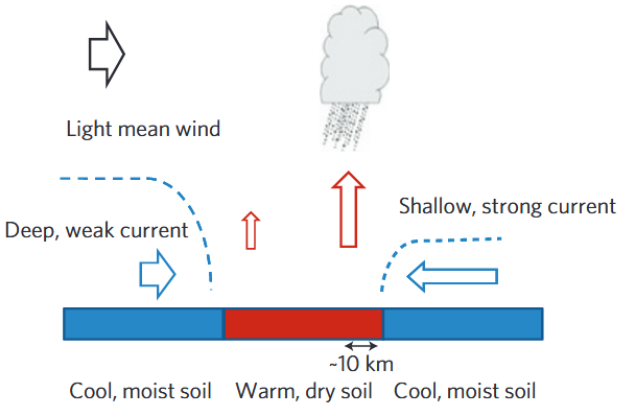


Figure 1.1: Schematic diagram showing how heterogeneity in soil moisture induces convection. Idealized soil-moisture-induced flows (blue arrows) under light synoptic winds (black arrow) create an ascent region (large red arrow) where the shallow, strong current opposes the mean wind. The preferred location for convective initiation coincides with the ascent region induced by the heating gradient at the downwind edge of the dry patch. Additional convergence over the dry patch is provided by a deep, weaker current at its upwind edge, and cross-wind gradients in soil moisture (Taylor et al., 2011)

The sensitivity of mesoscale convective system life-cycles to soil conditions in West Africa has been explored in modelling studies. Gantner and Kalthoff, 2010 used the Consortium for Small-scale Modeling (COSMO) model run at 2.8 km resolution and initialized with European Centre for Medium-range Weather Forecasts (ECMWF) analysis data to investigate convective activity in different soil scenarios, such as observed surface soil moisture (SSM), homogeneous SSM, and homogeneous soil type and SSM. They observed convection initiation in all experiments regardless of the initial soil moisture distribution but found evidence that soil moisture inhomogeneities and drier soil favor the triggering of convection. In an idealized case study using the Colorado State University (CSU) Regional Atmospheric Modeling System (RAMS) at 500 m resolution, Chen and Avissar, 1994 showed that SSM significantly affects the timing of onset clouds as well as the intensity and distribution of precipitation. Modelling studies by Garcia-Carreras et al., 2011, using the Met Office Large Eddy Model run at 250 m resolution and supported with observational evidence by Taylor et al., 2011, have shown that convection is favoured over dry soils and close to negative upwind gradients in sensible heat flux, as convergence is maximized where

large-scale winds opposes shallow surface-induced flow (see Figure 1.1). Soil moisture - precipitation feedbacks have also been studied using the WRF model, indicating that soil moisture heterogeneity can be more important for surface-atmosphere interactions than the actual soil moisture content (Graf et al., 2021) and that both positive and negative soil moisture-precipitation feedbacks are observed depending on the region of interest (Alessi et al., 2022; Graf et al., 2021). These studies demonstrate that these models are sensitive to soil conditions and can reproduce thermally-driven circulations that are observed in reality.

However, over sub-Saharan Africa numerical weather prediction (NWP) forecasting performs poorly (Vogel et al., 2018) due to a combination of weak synoptic forcing, limited observations for data assimilation, and the rapid development of MCSs, which dominate the growth of model errors (Baumgart et al., 2019). Early warning systems in West Africa rely on nowcasting methods through satellite observation, as there are currently no operational ground-based radars in West Africa and thus extrapolation of storm trajectories in combination with NWP or deep learning – employed in other regions such as Tornado Valley in the US – cannot be applied (Taylor et al., 2022). One such nowcasting approach, developed by Anderson et al., 2024, combines land surface temperature anomaly (LSTA) as a proxy for soil moisture and cloud top temperature observation to detect convectively active regions most likely to result in heavy precipitation events (convective cores).

Nonetheless, previous research indicates that better representation of the land surface with satellite-derived soil moisture can improve the forecasting skills of NWP (Drusch, 2007; Gao et al., 2006; Kerr et al., 2010; Lin et al., 2017; Santanello Jr. et al., 2019). The direct use of the Soil Moisture Active Passive (SMAP) soil moisture data in NASA Unified WRF (NU-WRF) modeling system's inital conditions was found to outperform other soil moisture products when used for short-term forecasting over the US southern Great Plains (Santanello Jr. et al., 2019). Similarly, incorporating a 6 hour nudging scheme with TMI (TRMM Microwave Imager, Gao et al., 2006) derived soil moisture in data assimilation experiments with the Integrated Forecast System (IFS) of the European Center for Medium-Range Weather Forecasts (ECMWF) improved local weather parameters, including planetary boundary layer height and cloud coverage (Drusch, 2007). Lin et al., 2017 showed that simultaneous assimilation of Tropical Rainfall Measuring Mission(TRMM) (Huffman et al., 2007) satellite precipitation and Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2010) soil moisture observations into the coupled Weather Research and Forecasting (WRF) and Noah land surface model effectively improve the forecast skills of precipitation, top 10-cm soil moisture, and 2-m temperature and specific humidity. These studies indicate that integrating SSM observations into NWP models such as WRF can improve model performance.

Currently, ESA's C-band Advanced SCATterometer (ASCAT) (H SAF, 2021) is the longest available operational SSM product and has finer spatial and temporal resolution than passive microwave products counterparts, such as from the SMOS and SMAP. Specifically, it has a return period of 1.5 days and provides SSM observations daily at a maximum resolution of 12.5 km. While it has finer temporal resolution than alternative soil moisture products from passive microwave remote sensing, its temporal resolution is still not sufficient to resolve the fast evolution of weather systems in the Sahel. Therefore, in the context of nowcasting, it is primarily used for validation of soil moisture proxies such as LST rather

than directly integrated into forecasts (Taylor et al., 2012).

A recently proposed interferometric L-band Synthetic Aperture Radar (SAR) mission concept with sub-daily revisits could provide an opportunity to assimilate sub-daily soil moisture to improve forecasts. The Sub-daily Land Atmosphere INTEractions (SLAINTE) mission concept aims to provide sub-daily, $\leq$1 km scale observations related to ecosystem water status, including vegetation water content and surface soil moisture over key regions of scientific, ecological, societal and economic interest. It would do so through a constellation of several identical SAR satellites which would provide observations of the same site multiple times on the same day with an overall return period of 3 days (Matar et al., 2024; Steele-Dunne et al., 2024).

Here, we aim to quantify the mission and L2 product (SSM) requirements of the SLAINTE mission concept to improve MCS forecasting through a series of synthetic experiments using the WRF model version 4.6.0 (Skamarock et al., 2021).

1.2 Research Questions

First, we will reproduce a case study of MCS development in the Sahel by initializing the model with ERA5 and ERA-Land products and run the model at the highest computationally possible resolution (1 km) – this will be the "nature" run. It can be considered as our synthetic truth or reality that we will try to capture. The soil moisture output of the nature run will be further used by upscaling it to a coarser resolution and adding gaussian noise to simulate a noisy observational dataset. These will be our synthetic observations. Then we will create perturbed and assimilation experiment runs, in which the perturbed run is created similarly to the nature run but with coarsened and perturbed soil moisture initial conditions, and the assimilation run in which soil moisture synthetic observations are directly inserted at various time intervals starting from the perturbed run to simulate the effect of assimilating either ASCAT or SLAINTE measurements. The effect of these different methods on the simulations will be compared to quantify the impact of the spatial and temporal resolution of SSM observations on the forecasts.

These experiments will allow us to evaluate how sub-daily soil moisture measurements with the Sub-daily Land-Atmosphere INTEractions (SLAINTE) satellite could improve the forecasting of Mesoscale Convective Systems (MCS) initiated due to Soil Moisture Anomalies (SMA) in the Sahel, West Africa. Specifically, we will address the following questions:

- How are precipitation and cloud formation affected by the soil moisture assimilation at different times and different resolutions?
- What is the impact of the spatial resolution of the assimilated soil moisture on the forecasts?
- What is the impact of assimilating SSM (synthetic) observations twice or three times a day on the forecasts?
- What are the optimal observation times?

Chapter 2

Methodology

In this chapter the methodology used in this research project is described. First, we describe the process of selecting a case study and its characteristics, followed by the model and model setup that was used to reproduce it. Then, we define regions where convection is triggered based on the model output variables. Finally, the experiment set up and evaluation methods are presented.

2.1 Case study selection and description

2.1.1 Nowcasting Portal

Convective activity in the Sahel region can be visualized through the UK Center for Ecology and Hydrology Nowcasting Portal for Sub-Saharan Africa (https://africa-hydrology.ceh.ac.uk/nowcasting/) (Anderson et al., 2024).
The portal is designed to help forecasters in Sub-Saharan Africa predict severe convective storms over the next few hours using near-real time satellite data — a process known as nowcasting. The available datasets that can be visualized through the portal include cloud top temperature (CTT), land surface temperature anomaly (LSTA) and convective cores, further described below.

SEVIRI Cloud Top Temperature

This product identifies where high clouds are present based on thermal infrared brightness temperature imaging (10.8 μm, channel 9) from the Spinning Enhanced Visible and Infrared Imager of the geostationary Meteosat Second Generation (MSG) satellites. The data are freely available from the EUMETSAT website (http://www.eumetsat.int) every 15 min at an approximate spatial resolution of 3 km over West Africa. This product is a useful proxy for precipitation, as deep convective clouds, responsible for the vast majority of intense rain across tropical Africa, have cloud-top temperatures typically well below -40 °C (Klein et al., 2018). However, the presence of cold cloud-tops does not always imply convective rainfall.

To help identify the most convectively-active parts of the cloud shield, the convective core product is introduced.

Convective cores

This product identifies areas within large storm clouds where cloud top temperatures are particularly cold (<-70° C), distinguishing the most convectively active parts of a cloud system (the cores) from slightly warmer regions which are less likely to produce extreme precipitation. It is based on the study of Klein et al., 2018, which correlates extreme precipitation with cloud top temperature in the Sahel, defining a threshold of cloud cover area and cloud top temperature value most likely to result in extreme precipitation.

Land Surface Temperature Anomaly

The Land Surface Temperature anomaly (LSTA) product is a daily product derived from daytime, clear sky Meteosat Second Generation (MSG) observations. Every 15 minutes, the Land Surface Analysis Satellite Applications Facility (LSA SAF) produce an LST image from MSG channels 9 and 10 in the infra-red. For each 15-minute image, additional cloud-screening is applied and compared with a climatology based on data for the same pixel and month from the period 2004–2015. A daytime mean LSTA is computed using all cloud-screened data between the hours of 0700 and the current time or 1700 (local time), whichever is the earlier. It is a useful proxy for SSM and as such provides important information on convective activity (Taylor et al., 2011).

Specifically, areas with positive LST anomalies (drier soil) and strong LST gradients (strong SSM gradient) are more likely to experience deep convection.

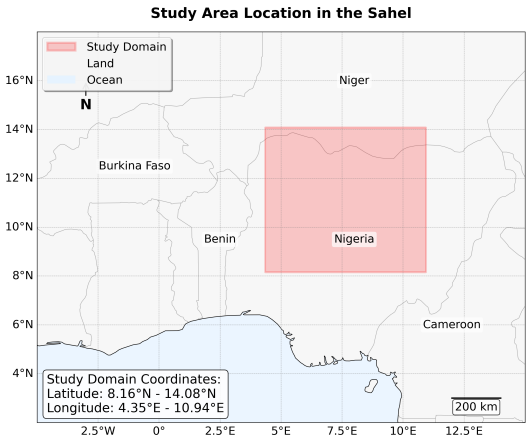


Figure 2.1: WRF simulation domain area in the Sahel indicated by the red box, enclosing an area of approximately 660×720 km^2 at 1 km grid.

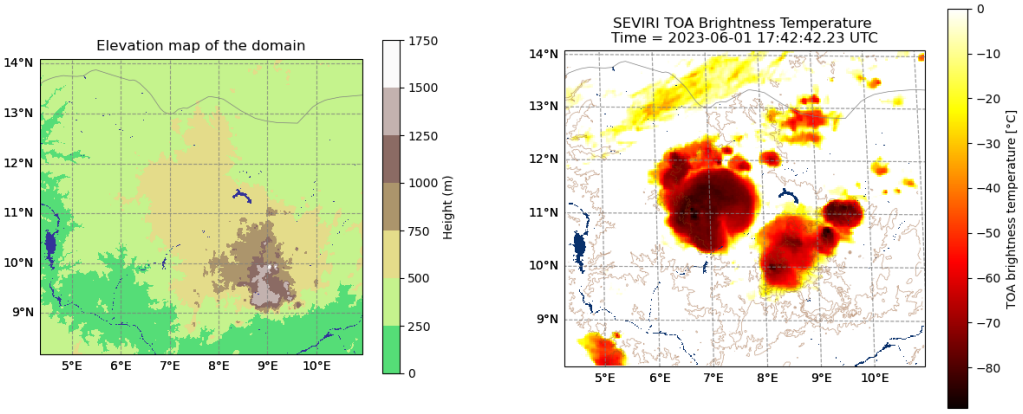
2.1.2 Case study selection

Research case studies were selected by observing convective activity in the Sahel through the Nowcasting Portal (section 2.1.1). In particular, we considered days when convective cores appeared between the hours 12 and 21 (local time) in areas characterized by small-scale scale variability in land surface temperature anomaly. The afternoon hours are a time of peak convective activity in the region when the soil exerts strongest control on atmospheric processes (Taylor et al., 2012)), making it more likely that cloud formation was triggered by soil conditions. The search was also limited to the first two weeks of June of the years 2022 and

2023 (the years available on the nowcasting portal), as early June marks the beginning of the wet season in West Africa and vegetation is still scarce in the Sahel at that time, allowing for a stronger coupling between soil moisture and the atmosphere.

Initially, two case studies were considered: 2022-06-11 over Burkina Faso and 2023-06-01 over northern Nigeria. On these days, the LSTA showed significant small-scale variability, and mesoscale convective systems developed in the afternoon period before dissipating after 21 UTC. The clouds developed over areas that are flat, where no water bodies are present and that are sparsely vegetated, making it more likely that the convection initiation was due to soil moisture anomalies.

The description of the area over Burkina Faso is omitted as it was eventually only used for calibration of the model parameterizations, and the simulations were carried out only in the region in northern Nigeria as further explained in section 2.4.



(a) (b)

Figure 2.2: (a) Elevation map of the model domain of the case study in northern Nigeria (U.S. Geological Survey, 2010). (b) SEVIRI TOA brightness temperature of the case study event over the model domain, showing the cloud formation on 2023-06-01 in the afternoon. The water bodies are identified in blue in both figures.

The location of the case study in northern Nigeria can be seen in Fig. 2.1, and in the elevation map of the region in Fig. 2.2a. The domain encloses the region with coordinates [8.16–14.08 °N, 4.35–10.94 °E] (approximately 660×720 km^2). Of note is the mountainous terrain present in the south-eastern part of the domain, as well as the water bodies such as rivers and lakes present throughout.

The SEVIRI top of the atmosphere brightness temperature in Fig. 2.1.1 shows the two separate mesoscale convective systems that developed in the afternoon of 2023-06-01: a MCS due to orographic lifting (the cloud in the south-west) and a MCS over the flat region near the center of the domain. The latter convective system developed in an area where no topography or water bodies are present and is the one of interest which prompted the selection of this case study.

The local time in Nigeria is UTC+1. However, since we are interested in the timing of the solar cycle and most of the simulation domain lies in the UTC timezone, UTC times were considered for simplicity.

2.2 Model and parameter selection

In this study, the model used to perform experiments is the Weather Research and Forecasting (WRF) model version 4.6.0, (Skamarock et al., 2021). WRF is a state of the art mesoscale numerical weather prediction (NWP) system designed for both atmospheric research and operational forecasting applications. It is a mass conservative finite difference model which uses non-hydrostatic compressible Euler equations and terrain-following hydrostatic pressure vertical coordinate. It is a versatile, open-source system used globally, and it is developed by a partnership of U.S. federal agencies and universities including National Center for Atmospheric Research (NCAR) and the National Oceanic and Atmospheric Administration (NOAA).
WRF has many different parameterizations for each of its seven physics schemes (first column in Table 2.1), which need to be optimized for the selected climatology and variables of interest. The parameterizations that were tested follow research on the performance of different WRF land surface schemes in West Africa (Achugbu et al., 2020; Achugbu et al., 2024), as well as microphysics and atmospheric schemes that best reproduce extreme rainfall events in the region (Oyegbile et al., 2024; Vizy & Cook, 2023). An overview of the possible parameterizations can be seen in Table 2.1.

Schemes\Papers	Achugbu et al., 2020, Achugbu et al., 2024	Oyegbile et al., 2024	Vizy and Cook, 2023	Selected parameterization
WRF version	WRF3.9.1.1	WRF4.1	WRF4.1.3	WRF4.6.0
Resolution	D1: 12 km	D1: 27 km D2: 9 km D3: 3 km	D1: 3 km	D1: 1km
Microphysics	WSM6	Thompson	Thompson	Thompson
Longwave	RRTMG	RRTM	RRTM	RRTM
Shortwave	RRTMG	RRTM	Dudhia	RRTMG
Surface layer	MYNN	-	Monin-Obukhov	Monin-Obukhov
Land surface	Noah-MP	Noah	Noah	Noah-MP
PBL	MYNN2.5	Yonsei University	Yonsei University	MYNN2.5
Cumulus	Tiedtke	KFC	-	-

Table 2.1: Overview of possible WRF parametrization options from literature and selected model parameterization. The referred papers focus on the WRF parameterization evaluation in West Africa and assess (from left to right) the land surface scheme, the microphysics scheme and the parametrization that best reproduces extreme precipitation respectively.

Different combinations of physics parameterizations were tested to reproduce the case studies as in Section 2.1 (see tested parameterizations in Appendix A.1). The cloud formation produced by WRF, assessed through the Outgoing Longwave Radiation (OLR) vari-

able, was visually compared both to the ERA5 net top longwave radiation and the SEVIRI cloud top temperature (section 2.1.1). It was immediately evident that the ERA5 dataset significantly differed from the SEVIRI observations and as such the WRF simulation could not reproduce the selected event exactly. Nonetheless, by calibrating the physics parameters key aspects of the event were able to be reproduced, such as timing and location of convective initiation.

The parameterizations that resulted optimal were the Thompson parametrization for microphysics (Thompson et al., 2008), RRTM for longwave radiation (Mlawer et al., 1997), RRTMG for shortwave radiation (Iacono et al., 2008), Monin-Obukov for surface layer dynamics (Monin and Obukhov, 1954Janjić, 1994, Janjic, 1996, Janjic, 2002), Noah-MP for land surface (Niu et al., 2011, He et al., n.d., Yang et al., 2011) and MYNN 2.5 for physical boundary layer (Nakanishi and Niino, 2006, Nakanishi and Niino, 2006, Olson et al., 2021). The dynamic vegetation option of Noah-MP was set to 3 (dynamic vegetation = off; Leaf Area Index (LAI) from table; vegetation fraction (FVEG) calculated). The cumulous parameterization was turned off, as the model can resolve convection explicitly at the selected model grid resolution.

The model was run at 1 km spatial resolution and 15 min temporal resolution, with 70 vertical atmospheric levels and 6 s time integration.

2.2.1 Datasets

ERA5 and ERA-Land

The datasets used to initialize the nature runs are the ECMWF ERA5 and ERA5-Land datasets (Hersbach et al., 2020) for atmospheric and land variables respectively.
ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate and weather covering the period from January 1940 to present. ERA5 is hosted by the Copernicus Climate Change Service (C3S) and produced by ECMWF, and provides hourly estimates of a large number of atmospheric, land and oceanic climate variables.
Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics, based on the Integrated Forecasting System (IFS). This principle, called data assimilation, is based on the method used by numerical weather prediction centers, where every so many hours (12 hours at ECMWF) a previous forecast is combined with newly available observations in an optimal way to produce a new best estimate of the state of the atmosphere, called analysis, from which an updated, improved forecast is issued. Reanalysis works in the same way, but at reduced resolution to allow for the provision of a dataset spanning back several decades. Reanalysis does not have the constraint of issuing timely forecasts, so there is more time to collect observations, and when going further back in time, to allow for the ingestion of improved versions of the original observations, which all benefit the quality of the reanalysis product.

The ERA5 dataset has a $0.25° \times 0.25°$ grid resolution, equivalent to approximately 27 km$\times$27 km at the equator, while ERA-Land has a 9 km grid resolution. These datasets were selected for their easy accessibility and prior equilibrium between land and atmo-

sphere.

SoilGrids soil texture

In the Sahel region, the default soil type categorisation provided by WRF (FAO/UNESCO, 1971–1981) has a very coarse resolution which defines the soil moisture patterns and dominates over the small-scale variability created by precipitation. Fine-scale information on soil texture is therefore necessary to resolve fine-scale variations in soil moisture.

In order to obtain a finer SSM resolution, the default soil texture was replaced by an adapted version of the SoilGrids soil type dataset (https://soilgrids.org/) following Pedruzzi et al., 2022.

SoilGrids is a global three-dimensional soil information system developed by the International Soil Reference and Information Centre (ISRIC) – World Soil Information. The soil classification is based on the automated geostatistic data processing method and machine learning, which can determine soil properties (e.g., clay, sand, and silt content) and soil classes, with minimum human intervention. SoilGrids has the same soil percentage information as the U.S. Department of Agriculture (USDA) classification system split into six soil layers (e.g., 0, 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) with a resolution of 250 m (Hengl et al., 2014). For compatibility reasons with the WRF model resolution, the 1 km aggregated SoilGrids product was used (https://files.isric.org/soilgrids/latest/data_aggregated/1000m/) (Hengl et al., 2014).

WRF utilizes soil texture information for two layers: the top soil layer, which ranges from 0 to 30 cm, and the bottom soil layer, ranging from 30 to 100 cm. The soil texture data is split into 16 classes based on the percentage of silt, sand, and clay within the soil, following the methods of the United States Department of Agriculture (USDA) (Soil Survey Staff, 1999).

To apply SoilGrids in WRF, the 0–5, 5–15, and 15–30 cm layers were merged using a weighted average by soil depth to generate Noah MP's top soil layer, while the 30–60 and 60–100 cm layers were combined similarly to build the soil bottom layer. The SoilGrids layer at zero (0 cm) was not used due to the lack of depth, and the 100–200 cm layer was not used either by WRF. SoilGrids data were then mapped to USDA soil texture classes by obtaining the percentage value of silt, sand, and clay from the SoilGrids raster files. Consequently, the soil type was adjusted so that the water bodies indicated by the soil type matched the land mask provided by WRF.

2.3 Definition of Convective Cores from model output

To assess the effect of SSM in triggering convection within the model, it is necessary to define the location where convection is triggered, hereafter referred to as a convective core.

In order to define a convective core in the Sahel, Klein et al., 2018 used the TRMM product 2A25 (Z. Liu et al., 2012; Tropical Rainfall Measuring Mission (TRMM), 2011). This rainfall radar product provided surface rainfall rates at a resolution of about 5 km. Klein et al., 2018 defined extreme precipitation as the 99th percentile of the observed hourly

precipitation between 2004 and 2014, which amounted to a threshold precipitation rate of 30 mm/h. In the context of their research, it was necessary to correlate extreme precipitation with cloud top temperature, to define a threshold of cover area and cloud top temperature value more likely to result in extreme precipitation.

In the context of this research however, it is only necessary to understand where convection that results in sufficiently high precipitation occurs. The precipitation threshold used by Klein et al., 2018 is also not applicable, as the precipitation resolution they considered was lower (5 km); it is expected that the 30 mm/hr threshold of precipitation is too low to classify extreme precipitation as the model is likely to produce much higher precipitation rates over a smaller region. We also have precipitation data availability at a much finer temporal resolution (15 min), which allows us to consider an alternative definition.

We define a convective core as the grid point where accumulated precipitation over 15 min exceeds 10 mm for the first time in a 30 km radius and in the same day between the hours 12 and 21 UTC.

This definition follows several considerations. Firstly, previous research has shown that the afternoon convection is favored by soil moisture anomalies and is most closely related to the formation of MCSs (Taylor et al., 2011). Secondly, we assume that the location where precipitation first occurs is close to where the convection that resulted in the cloud formation was initiated. We also want to disregard any convection by wet soil anomaly, so we disregard the grid points where precipitation occurred earlier than 12:00 in the same day.

Lastly, a distinction between primary and secondary convection is necessary. We consider secondary convection the convection which occurs due to the uplift created at the edge of cold pools, namely the cooled dense air that forms beneath a cloud when rain evaporates during intense precipitation. In contrast, primary convection is more likely to occur due to instabilities linked to anomalies in evaporative fraction, which is what we are interested in investigating. Therefore, we disregard any precipitation event which occurs in a 30 km radius from the first precipitation event. 30 km is considered a large enough distance over which convection events are unlikely to be linked to uplift by cold pools as a frequent diameter of an MSC is on the order of around 50 km and follows previous research that links sub-cloud features with convective cores (Klein et al., 2018).

2.4 Experiment set-up

Preliminary experiments were carried out to evaluate the sensitivity of convective activity generated by the model to SSM variation. The soil moisture was set to wilting capacity and porosity, but in both case studies used to calibrate the physics parameterizations, convection was initiated regardless of the soil conditions, similarly to the research carried out by Gantner and Kalthoff, 2010. It was assumed that the events were primarily boundary-driven rather than due to SSM anomalies. Specifically in the case on the 2022-06-11, the boundary-prescribed wind was deemed too strong for the model to be significantly sensitive to soil moisture variations, as shown by the development of very similar cloud structures

regardless of the soil conditions. Therefore only the case study over northern Nigeria was retained, as the boundary-induced winds were less prominent and the case over Burkina Faso was only used for calibration of WRF parameterizations.

Additionally, to increase the model sensitivity to soil conditions and hopefully limit the boundary interaction effects, the decision was made to only initialize the model with the ERA5 and ERA-Land initial conditions and boundary conditions and not update the boundary conditions for the whole duration of the run. This essentially creates a thermally-driven diurnal system, where the solar heating cycles are the main driver of mesoscale circulations within the domain and thermally driven circulations become prominent. In this context, the soil moisture control on atmospheric processes is isolated as it is the main variable influencing the surface energy partitioning. To simulate the effect of assimilating observations at different overpass times over multiple days, a 4-day spinup period was considered previous to the day of the event of interest (2023-06-01). The total simulation time is 5 days: 2023-05-28 00 UTC to 2023-06-02 00 UTC, of which we will focus only on the last simulation day.

An overview of the experiment set-up can be seen in Fig. 2.3 and is further explained in the following sections.

2.4.1 ASCAT - overpass times and resolution

ASCAT is a real-aperture radar sensor measuring radar backscattering coefficient at C-band (5.255 GHz) in VV polarization with a radiometric accuracy better than about 0.3 dB. The scatterometers are mounted on the series of Metop satellites, of which the first was launched in 2006 (Metop-A), the second one in 2012 (Metop-B) and the third in 2018 (Metop-C). All three Metop satellites are identical and operated by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). The scatterometers scan the globe in a push-broom mode by six side-looking antennae (three left-hand and three right-hand beams). Each antenna observes a specific location on the earth's surface under a different azimuth and viewing angle. The two swath widths are approximately 550 km separated by a gap of 360 km, and there are 14 orbit revolutions per day resulting in a global coverage achieved in $\sim$ 1.5 days (Brocca et al., 2011; Srivastava et al., 2016).

Soil moisture is retrieved from the ASCAT backscatter measurements using a time series-based change detection approach previously developed for the ERS-1/2 scatterometer by Wagner, Lemoine, and Rott, 1999, Wagner, Lemoine, Borgeaud, and Rott, 1999 and Wagner, Noll, et al., 1999. In this approach soil moisture is considered to have a linear relationship to backscatter in the decibel space, while the noise sources include the instrument noise, speckle and azimuthal anisotropies. The surface roughness is assumed to have a constant contribution in time, and therefore is not accounted for in the change detection algorithm. By knowing the typical yearly vegetation cycle and how it influences the backscatter-incidence angle relationship for each location on the Earth, the vegetation effects can be removed (Wagner, Lemoine, and Rott, 1999), revealing the soil moisture variations. As a last step, the historically lowest and highest values of observed backscatter are assigned to the 0% (dry) and 100% (wet) references respectively, thereby yielding time

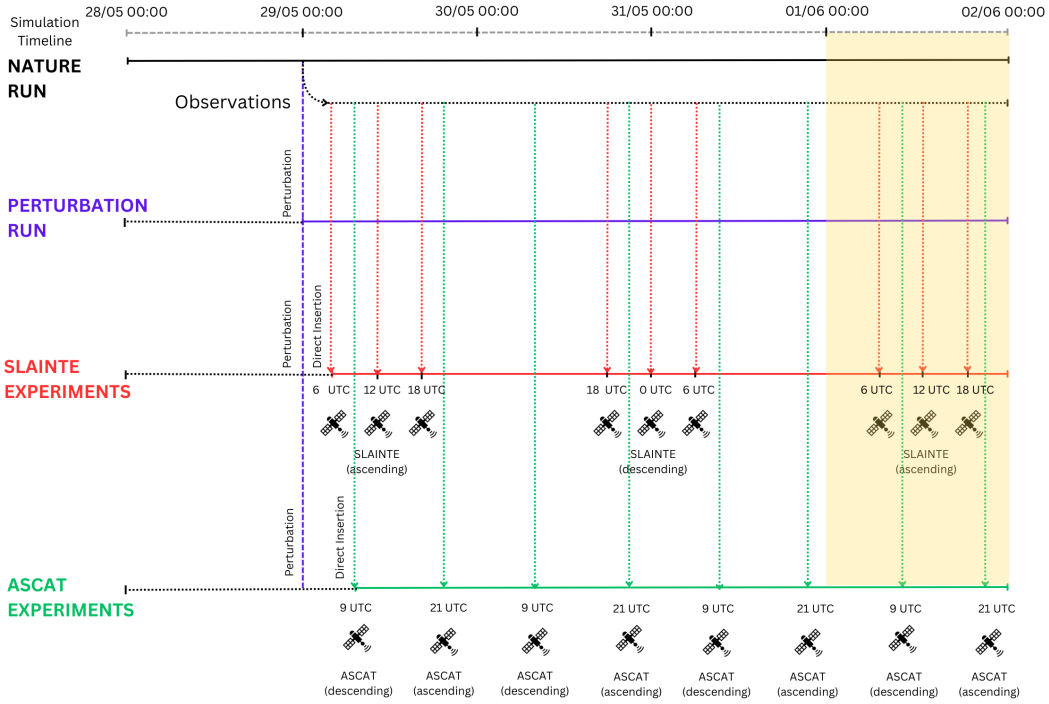


Figure 2.3: Overview of the experiment set up. The nature run runs freely from the 2023-05-28 00 UTC to 2023-06-02T 00 UTC, and from the top layer soil moisture output the SSM synthetic observations are derived. The top layer soil moisture is perturbed at 2023-05-29 00 UTC to create the perturbation run and then the simulation is left to run until 2023-06-02 00 UTC. During the assimilation experiment runs, the soil moisture observations are directly inserted into the perturbation run. The yellow box indicates the last simulation day, of which we will study the results (focus period).

series of relative soil moisture percentage values for the first few centimeters of the soil (Brocca et al., 2011).

ASCAT is the most natural candidate to use as a reference for synthetic SSM observations. It is used operationally by ECMWF, MetéoFrance and in the UK center for Hydrology and Meteorology nowcasting portal and is based on data from operational satellites, with a uniquely long and consistent record guaranteed until 2040+. Compared to other current satellite products, it has an optimal combination of temporal and spatial resolution. For reference, its passive microwave counterparts such as the SMAP satellite and SMOS have a 3-day revisit time but provide a daily product where only the values in the observed swath are included; SMOS, launched in 2009, has a much coarser spatial resolution (36 km) (ESA, 2021, SMOS L2 SM V700, Version 700), while SMAP was launched in 2015 and has an enhanced resolution 9 km SSM product, and a 9 km data assimilation product (smap.jpl.nasa.gov/data). Products with finer temporal resolution such as TMI are no longer operational.

The Sahel region is observed by ASCAT in wide swaths, of which the overpass time is slightly different each day. There is a morning (approx. 9 UTC) overpass during the satellite's descending orbit, and the evening (approx. 21 UTC) overpass during the satellite's ascending orbit, with one observation approximately every 24 h in the region. Most of the region is only observed once per day with the exception of few locations were the observations overlap. The central region of the case study area happens to be in a location where morning and evening observations overlap. Therefore, in order to evaluate the best-case scenario and the effect of assimilating at both overpass times, the simulated overpass time for ASCAT in the context of this experiment was set to 9 UTC and 21 UTC every 24 h.

The resolution of the ASCAT SSM NRT 12.5 km (H29) product is 12.5 kilometers (https://navigator.eumetsat.int/product/EO:EUM:DAT:METOP:SOMO12). Since the WRF model is run at 1 km grid, the resolution of ASCAT synthetic observations was set to 12 km for simplicity.

The ASCAT electromagnetic waves penetrate the soil of about 5 cm (Rötzer et al., 2014), depending on the soil water content. In contrast, the Noah-MP land surface scheme used in the WRF simulations provides soil moisture output in four layers, the uppermost of which is 10 cm thick. The difference in soil depth is a difference that is significant but that cannot be overcome and we thus need to neglect by assuming the two coincide.

Previous studies have shown that the root mean squared error of ASCAT-derived SSM can be considered of the order of 0.05 m^3/m^3 in regions such as Spain (Wanders et al., 2012). The assumption is made that this error is representative in the region of interest of this study. Therefore, the standard deviation of the simulated noise was set to σ =0.05 m^3/m^3. This normal error was applied to the perturbation and observations alike.

2.4.2 SLAINTE - overpass times and resolution

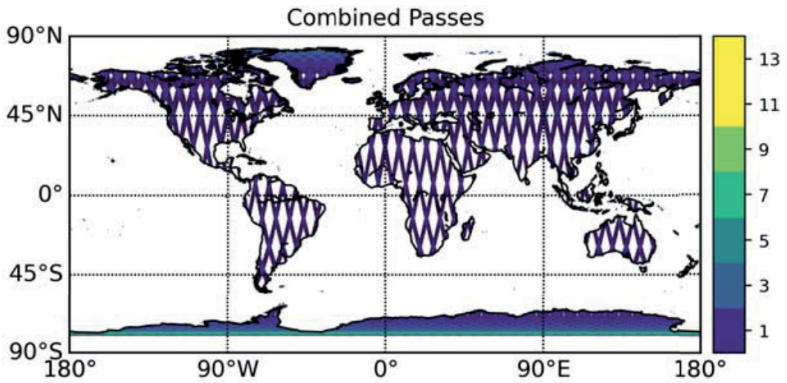


Figure 2.4: Access of the SLAINTE instruments over a 3-day period, covering about 25% of the globe (Steele-Dunne et al., 2024).

Based on its first technical concept, SLAINTE would consist of a low-earth orbit (LEO) constellation of decametric resolution monostatic L-band (1-2 GHz) SARs with high radio-

metric and interferometric accuracy (Steele-Dunne et al., 2024). The satellites would be displaced in orbit to provide observations separated by six hours. Each spacecraft would fly in a polar 3-day repeat orbit allowing for a revisit shorter than 1.5 days where ascending and descending passes overlap, with a static coverage of about 25% of the globe (see Fig. 2.4) (Steele-Dunne et al., 2024). With that in mind, an example of the SLAINTE simulated observation times over a 4-day period would be:
Day 1: 06, 12, 18 UTC
Day 2: 18 UTC
Day 3: 00, 06 UTC
Day 4: 06, 12, 18 UTC
The resolution that will be tested for SLAINTE in this research is 1 km and 5 km, to evaluate whether a coarser resolution could still be valuable for MCS forecasting.
The same standard error as for ASCAT was applied to SLAINTE observations as this is the accuracy required by application users ("WMO OSCAR Details for Variable: Soil moisture at surface", 2025). Therefore, for SLAINTE observations and perturbation alike a standard error equal to $\sigma = 0.05\ m^3/m^3$ was applied.

2.4.3 Nature run and synthetic observations

The nature run was initialized with initial and boundary conditions from ERA5 and ERA-Land on the 2023-05-28 0 UTC and left to run freely with no update of the boundary conditions until 2023-06-02 0 UTC. This nature run represents a synthetic reality, based on which soil moisture observations are generated by adding observation noise. The lack of boundary updates imply that the solar cycle is the main driver of atmospheric events, which results in a daily repetition of convective activity that varies slightly each day primarily due to the changing surface soil moisture conditions.
The top layer soil moisture output was left as is at 1 km resolution and coarsened to a 5 km and 12 km resolution, and consequently perturbed with standard error $\sigma = 0.05\ m^3/m^3$ to create synthetic observations at 1, 5 and 12 km resolution respectively. To coarsen the resolution of the dataset, the top layer soil moisture data was aggregated by dividing it into 5×5 and 12×12 grid squares. Each individual grid point within a square was then replaced with the average soil moisture value calculated for that entire square, taking care to mask the water bodies.

2.4.4 Perturbation Runs

To create perturbed simulations, the model was initially allowed to run for 24 model hours, so as to generate (through precipitation) a SSM pattern of finer resolution than the 9 km ERA-Land initial conditions. Consequently, the SSM of the restart file generated by the nature run at 2023-05-29 00 UTC was coarsened to the resolution of the observations that would be assimilated into the respective perturbation run (1 km, 5 km or 12 km) and perturbed with a normal error equal to $\sigma = N(0, 0.05)\ m^3/m^3$ as shown in Fig. 2.5. The model

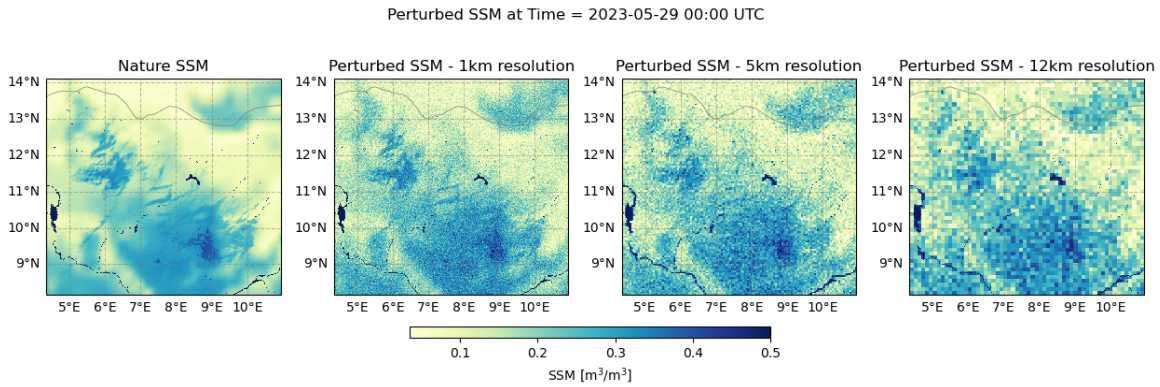


Figure 2.5: Depiction of the perturbed SSM applied to generate different perturbation runs, with perturbation applied to the (coarsened) nature top-layer soil moisture at 1 km, 5 km and 12 km resolution on 2023-05-29 0 UTC. Shown from left to right: nature SSM, 1 km-resolution perturbed SSM, 5 km-resolution coarsened and perturbed SSM and 12 km-resolution coarsened and perturbed SSM.

grid resolution was always kept at 1 km, with only the resolution of top-layer soil moisture being manually degraded.
Different perturbations were applied for different assimilation experiments because artifacts of the observation resolution can be observed in the soil moisture pattern days after their assimilation; the effect of coarser-resolution artifacts would defy the purpose of assimilating finer-resolution observations into the model.
The simulations were continued and ran with no further updates or modifications until 2023-06-02 00 UTC.

The name with which these runs are consequently referred to reflect the resolution of the coarsened SSM applied to perturb them, namely:

- 1 km perturbation run
- 5 km perturbation run
- 12 km perturbation run

2.4.5 Assimilation Experiments

The assimilation experiment runs were obtained by directly inserting the soil moisture observations in the perturbation runs at the time of the simulated overpass and at the resolution of the simulated observations. The model grid resolution remained 1 km for all simulations, but the direct insertion updated each 1 km grid cell with the value of the (coarser) resolution of the observation.
Four different experiments were conducted, with the following characteristics:

Experiment name	SSM observation resolution	Overpass Time
12 km ASCAT	12 km	ASCAT
5 km SLAINTE	5 km	SLAINTE
1 km SLAINTE	1 km	SLAINTE
1 km ASCAT	1 km	ASCAT

Table 2.2: Overview of the SSM observation resolution and overpass time characteristics of the four assimilation experiments.

- 12 km soil moisture resolution observations with ASCAT overpass times
- 5 km soil moisture resolution observation with SLAINTE overpass times
- 1 km soil moisture resolution observation with SLAINTE overpass times
- 1 km soil moisture resolution observation with ASCAT overpass times

The name with which the experiment runs are hereafter referred to reflects the SSM observation resolution and satellite overpass time, as shown in Table 2.2.

These four experiments allow us to compare different aspects of the satellite characteristics, specifically the observation timing, frequency and resolution. The 1 km ASCAT experiment, while not corresponding to any current or future satellite mission, is considered in order to facilitate the comparison between runs that differ both in observation resolution and timing.

We can evaluate the effect of spatial resolution for the SLAINTE mission by comparing the 1 km SLAINTE and 5 km SLAINTE experiments. In addition, the comparison of 1 km ASCAT to 1 km SLAINTE will provide information on the importance of observation time, while comparing 12 km ASCAT with 1 km ASCAT will allow comparison or the observations' spatial resolution when considering the ASCAT observation times. Finally, the comparison of 12 km ASCAT and 1 km SLAINTE will allow to compare the additional value of both the higher spatial resolution and the different observation time of SLAINTE.

2.5 Evaluation metrics

2.5.1 Root Mean Square Error

The root mean square error (RMSE) metric was used to compare model outputs from different runs, with

$$RMSE = \sqrt{\frac{\sum_{i,j\in D}^{N}[x(i,j) - x'(i,j)]^2}{N}} \tag{2.1}$$

where i, j coordinates in a specified (sub)domain, N the number of gridpoints in the (sub)domain D and and x, x' variables from two different model run outputs.

2.5.2 Evaluation metrics of precipitation forecasting: SAL

In order to evaluate the impact of assimilation on forecasts different methods were applied. To compare the precipitation forecast, the SAL metrics (Wernli et al., 2009) are used. These are feature-based quality verification metrics which provide information about the structure (S), amplitude (A), and location (L) of a quantitative precipitation forecast (QPF) in a pre-specified domain area D. The amplitude component A measures the relative deviation of the domain-averaged QPF from observations. Positive values of A indicate an overestimation of total precipitation; negative values indicate an underestimation. For the components S and L, coherent precipitation objects are separately identified in the forecast and observations; however, no matching is performed of the objects in the two datasets. The location component L combines information about the displacement of the predicted (compared to the observed) precipitation field's center of mass and about the error in the weighted-average distance of the precipitation objects from the total field's center of mass. The structure component S is constructed in such a way that positive values occur if precipitation objects are too large and/or too flat, and negative values if the objects are too small and/or too peaked. Perfect QPFs are characterized by zero values for all components of SAL (Wernli et al., 2009). Importantly, these metrics do not take into account the orientation of the precipitation objects.

More specifically, the **amplitude** component A corresponds to the normalized difference of the domain-averaged precipitation values of the model R_{mod} and the observations R_{obs}.

$$A = \frac{D(R_{mod}) - D(R_{obs})}{0.5[D(R_{mod}) + D(R_{obs})]} \tag{2.2}$$

where $D(R)$ denotes the domain average of the precipitation field R. The values of A are within [-2,+2] and 0 denotes a perfect forecast in terms of amplitude.

The **location** component L consists of two parts, i.e. $L = L_1 + L_2$. L_1 represents the normalized distance between the centers of mass between the two precipitation fields

$$L_1 = \frac{x(R_{mod}) - x(R_{obs})}{d} \tag{2.3}$$

where d denotes the largest distance between two boundary points of the considered domain, and $x(R)$ is the center of mass of the precipitation field R. However, even if the precipitation fields are completely different, the gravity centers may be similar. For this reason, another parameter L_2 was introduced to account for the average distance between the center of mass of the total precipitation and the individual precipitation objects, as follows:

$$L_2 = \frac{|r(R_{mod}) - r(R_{obs}))|}{d} \tag{2.4}$$

where $r = \frac{\sum_{n=1}^{m} R_n \cdot |x - x_n|}{\sum_{n=1}^{m} R_n} = \Sigma_{i,j \in R_n} R_{i,j}$, with m the number of precipitation objects in the domain and R_n the total precipitation of the n^{th} precipitation object. Both L_1 and L_2 have values in the range of [0,1], so $L \in$[0,2].

Finally, the **structure** component S compares the volume of the normalized precipitation objects. First, the "scaled volume" V_n is calculated for every precipitation object

$$V_n = \frac{\Sigma_{i,j \in R_n} R_{i,j}}{R_{max}} = \frac{R_n}{R_{max}} \tag{2.5}$$

where R_{max} is the the maximum grid-point precipitation within the object n. Then, the weighted mean V of all objects' scaled precipitation object is calculated as

$$V(R) = \frac{\sum_{n=1}^{m} R_n V_n}{R} \tag{2.6}$$

which allows us to calculate the structure component as

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0.5[V(R_{mod}) + V(R_{obs})]} \tag{2.7}$$

The possible range of values is $S \in$[-2,2]. Positive values of S indicate that the predicted precipitation objects are too large and/or too flat; in contrast, negative values occur for too small and/or too peaked objects (Wernli et al., 2009).

These metrics were considered to assess the performance of the assimilation runs compared to their respective perturbation runs by comparing the accumulated precipitation fields of each to the nature run. The precipitation forecast is said to improve with the assimilation if the SAL metric variation compared to the perturbation run is at least 0.05:

$$|S/A/L|_{perturbation} - |S/A/L|_{assimilation} > 0.05 \tag{2.8}$$

Similarly, the assimilation is said to degrade the forecast if

$$|S/A/L|_{assimilation} - |S/A/L|_{perturbation} > 0.05 \tag{2.9}$$

If the perturbation is too weak ($|S/A/L|_{perturbation} < 0.05$), the assimilation is said to degrade the forecast only if $|S/A/L|_{assimilation} > 0.10$.

2.5.3 Location metric for convective cores

In order to quantify whether the location of the convective cores improves with the assimilations, the use of the length metric (L) as above was modified. Specifically, the precipitation in the core grid points was set equal to 1 mm and NaN elsewhere, thus creating a binary precipitation field. This was considered in order to isolate the location characteristics of the cores without including information on the intensity of precipitation. This metric, however, does not take into account the orientation of the objects in space. Similar consideration as in section 2.5.2 were considered to quantify forecast improvement or degradation.

2.6 Selection of evaluation area

Due to the presence of water bodies, orographic lifting and boundary effects, it is necessary to define sub-regions within the domain where SSM is more likely to be the sole control

on atmospheric processes. To estimate the location where the precipitation field is mostly affected by the soil moisture, the total accumulated precipitation difference between the nature and perturbation runs throughout the five-day simulated period is evaluated in Fig. 2.6. From the figure emerges an area in the north-center region of the domain with no lakes nor topography where there is maximum variation of accumulated precipitation. Based on this, three focus regions are identified:

- **Region A**: Most sensitive area, which is furthest from both boundaries and water bodies.
- **Region B**: Intermediate region just excluding water bodies but including major precipitation events.
- **Region C**: Larger area encompassing most precipitation events but closer to northern and western boundaries and just excluding water bodies.

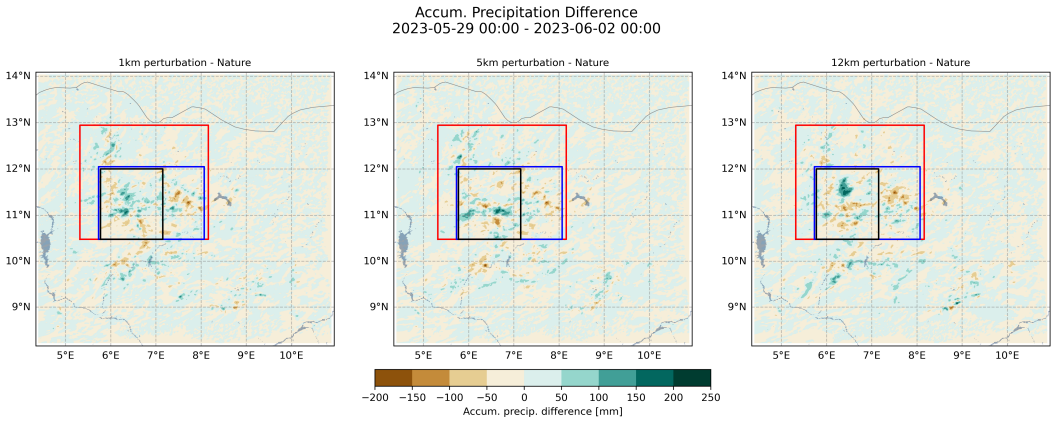


Figure 2.6: Total accumulated precipitation variation of the perturbation runs compared to the nature run over the four-day simulation period. The black (**region A**), blue (**region B**) and red (**region C**) boxes indicate the focus areas where it is assumed that the sensitivity to soil moisture variations is strongest. They include areas where no topography or water bodies are present and which are far enough from the boundaries that their influence can be disregarded.

Chapter 3

Results & Discussion

In this section the results of this research are presented and discussed. First the nature run and its main features are described, followed by a description of the perturbation runs and of the assimilation experiments. Then the effect of observation resolution and observation time and frequency is discussed by comparing different assimilation experiments. Finally, the overall additional value of SLAINTE compared to ASCAT is discussed, with focus on the implications for the SLAINTE mission requirements.

The simulations are evaluated using a combination of figures and metrics. Visual assessment is carried out through images of cloud top temperature, accumulated precipitation fields and scatter plots of convective core location and timing. Qualitative assessment is given by the SAL accumulated precipitation forecasting metrics, the modified core location metric and the time series of SSM and accumulated precipitation.

3.1 Nature run

3.1.1 Event description

The nature run reproduces key aspects of the mesoscale convective system development in the Sahel. During the five-day simulation period (2023-05-28 to 2023-06-01), convective activity and cloud formation repeat daily due to the solar forcing and constant boundary conditions, with changes that can be attributed primarily to changing soil conditions. Orographic convection due to lifting originates daily on the southeastern part of the domain around 13 UTC, while pressure gradients due to boundary interactions trigger cloud formation in the northern part of the domain which persist throughout most simulation days. Both cloud systems move towards the center of the domain, where between 17 and 21 UTC new clouds are also triggered near the center of the domain. The cloud systems dissipate around 22 UTC. The main features of the cloud system evolution can be visualized through cloud top temperature in Fig. 3.1.

Besides cloud shape and location, the cloud top temperature value can provide information on the nature of the clouds themselves. Deep convective clouds have cloud-top temperatures typically well below -40 °C. By this definition, we can qualitatively see that

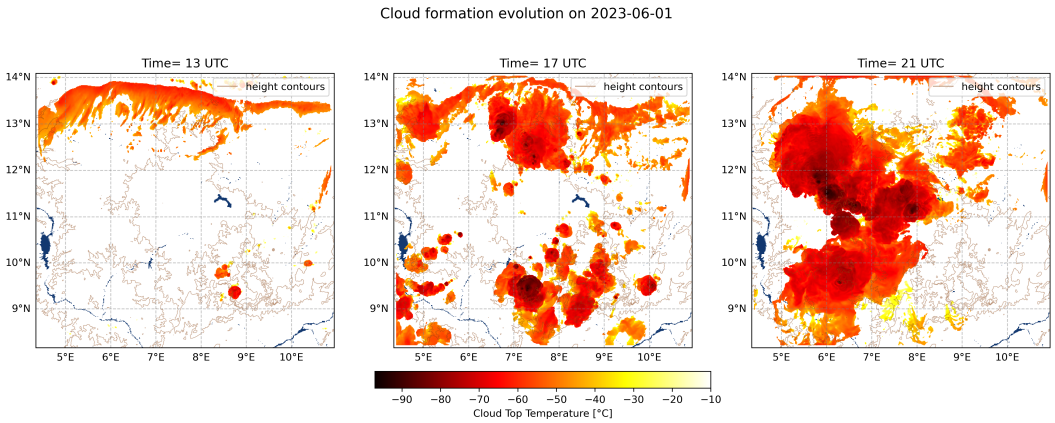


Figure 3.1: Nature run cloud top temperature, showing cloud formation at different times on 2023-06-01 over 250 m height contour plots. (Left) 13 UTC, orographic convection in the south-eastern region of the domain is initiated, as well as cloud formation due to interaction with the boundary in the north. (Center) 17 UTC, cloud formation in the central part of the domain, development of clouds caused by orographic lifting and boundary interactions. (Right) 21 UTC, clouds move to the center-west of the domain and more convection develops at the center.

most of the clouds formed during the nature run are deep convective clouds. The overall cloud area with $CTT < -40$ °C reaches an extent of over 100,000 km^2 around 20 UTC, effectively reaching the mesoscale.

Precipitation during the hours of most intense convective activity (12 - 21 UTC) result in the accumulated precipitation field in Figure 3.2. Most of the precipitation of the nature run occurs in the focus areas additionally to the precipitation in the location of mountainous terrain in the south-east. The precipitation during the nature run is locally severe, with 120 mm of accumulated precipitation reached locally.

It should be noted that no precipitation occurs in the focus areas between the hours 1 and 12 UTC on each simulation day, including the focus period. During these hours the only process effecting the soil condition is evaporation, and the soil dries on average. This implies that the SSM observations at 6, 9 and 12 UTC retain a similar spatial pattern.

Boundary effects are unfortunately prominent and influence significant aspect of the system. Due to the lack of boundary updates, artificial pressure gradients generate momentarily elevated wind speeds at the boundaries, which increase the boundary layer height and favor convection initiation at the northeastern side of the domain, as can be clearly visualized in the 10-meter wind field in Fig. 3.3a and planetary boundary layer height in Fig. 3.3b. This phenomena occurs shortly before cloud formation in the northern region of the domain at 13 UTC.

The choice to turn off the updates of the boundaries implies that the system does not evolve in a fully realistic way and needs to be treated as a highly idealized scenario. Important regional features with a daily cycle such as the diurnal cycle of the West African monsoon circulation, which induces a diurnal cycle in wind flows and can influence surface-

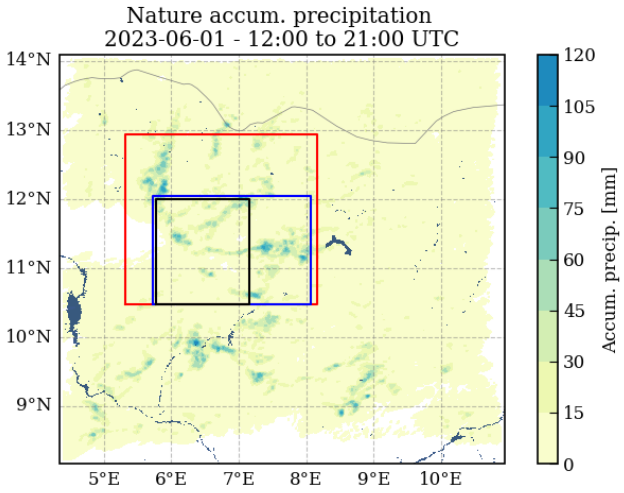
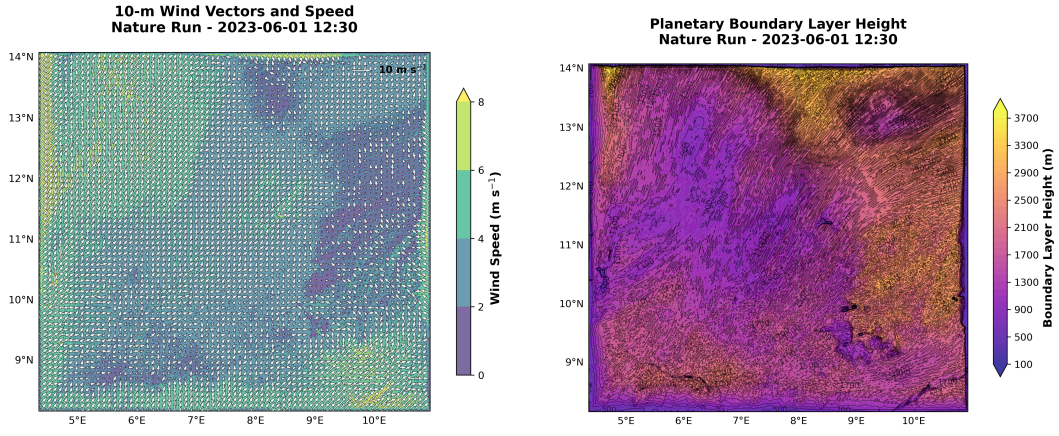


Figure 3.2: Accumulated precipitation of the nature run between 12 and 21 UTC on 2023-06-01. The boxes indicate the focus areas: **region A**, **region B** and **region C**. The regions in white indicate regions where no precipitation occurred.

induced circulations (Parker et al., 2005), are unaccounted for. It also implies that the output parameters of the simulation cannot be validated by comparing them to other datasets such as ERA5, or CTT and LST from geo-stationary satellites.



(a) (b)

Figure 3.3: Boundary effects on 2023-06-01 at 12:30 UTC visualized through (a) 10-meter wind vector and speed and (b) planetary boundary layer height, showcasing the winds due to pressure gradients at the boundaries and the influence of the boundaries in convection initiation.

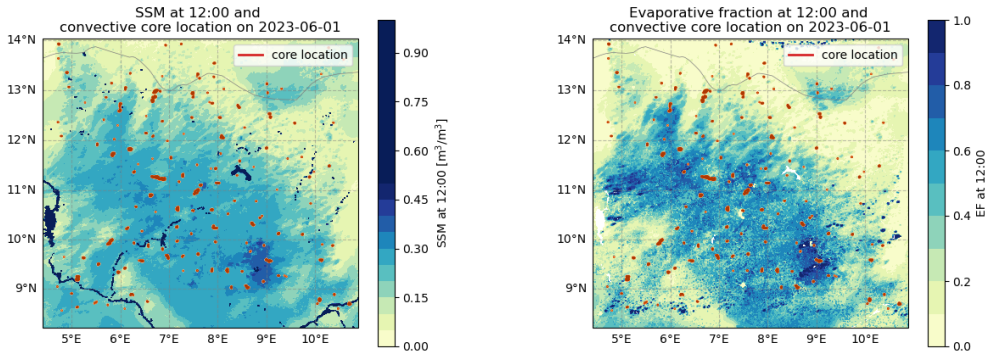
3.1.2 Convective core location and soil moisture

Following the definition of convective core established in section 2.3, the convective cores formed on the last simulation day of the nature run are shown in Fig. 3.4, overlaid on surface soil moisture (Fig. 3.4a) and evaporative fraction (Fig. 3.4b) at 12 UTC, before any precipitation occurs. The soil moisture pattern can be assumed to remain spatially consistent throughout the day until precipitation begins in each location, as evaporation may decrease absolute SSM values but is unlikely to alter its overall spatial distribution.

The evaporative fraction (EF), defined as $EF = \frac{LH}{LH+H}$ where LH is latent heat flux and H is sensible heat flux is a dimensionless ratio representing the proportion of available energy at a land surface that is used for evaporation. It is strongly correlated with soil moisture but also depends on additional variables such as relative humidity and vegetation cover (X. Liu et al., 2020). While the EF is more likely to vary throughout the day than soil moisture, it provides valuable complementary information as the convective activity we are interested in occurs where evaporation is water-limited, hence where the EF is lower.

Visual assessment reveals that convective cores predominantly form in areas characterized by strong soil moisture gradients, and where soils are drier and EF is lower than the surrounding regions. A clear example occurs at approximately (11.2°N, 7°E), where an area of significantly drier SSM corresponds to the formation of a large convective core. This spatial relationship aligns with the theoretical understanding that mesoscale convective systems can be triggered by negative soil moisture anomalies (see Fig. 1.1). The corresponding evaporative fraction map (Fig. 3.4b) confirms that this same area exhibits markedly lower EF values compared to its surroundings.

It is however possible that the requirement of a 30 km radius threshold surrounding each core is not sufficient to isolate primary convection, and that some of the cores are generated



(a) (b)

Figure 3.4: Convective core location of the nature run formed between the hours of 12 and 21 UTC on 2023-06-01, overlayed to (a) the SSM and (b) the evaporative fraction (EF) at 12 UTC on the same day (before any precipitation events occur), showing the correlation between SSM, EF and and convective core location.

by secondary convection. The MCSs in the simulation are displaced beyond 30 km and could be the source of secondary convection through cold pools.

Nonetheless, these results support the convective core definition used in this study and demonstrate that cloud formation and subsequent precipitation within the model exhibit clear sensitivity to SSM patterns. It also reproduces results correlating negative soil moisture anomalies and soil moisture gradients to convective activity observed in previous research with different models (COSMO (Gantner & Kalthoff, 2010) and REMS (Garcia-Carreras et al., 2011))

3.2 Baseline perturbation characteristics

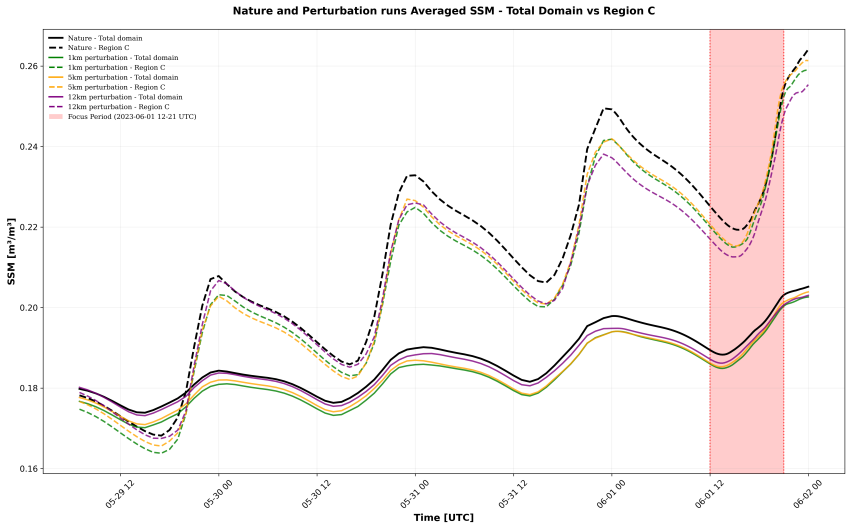


Figure 3.5: Time series of average SSM in the total domain and **region C** for the nature, 1 km perturbation, 5 km perturbation and 12 km perturbation runs during the entire simulation period (2023-05-29 00 UTC - 2023-06-02 00 UTC), showing the effect of the different resolution of SSM perturbation.

This section examines how different spatial resolutions of SSM perturbations affect the model's convective behavior. We analyze the three SSM perturbation scales (1 km, 5 km, 12 km) to establish baseline conditions before the assimilation experiments. The effect of the perturbations vary significantly with the imposed SSM perturbation resolution, which directly impacts the potential for improvement through data assimilation.

Looking at the entire simulation period, from the time series of SSM across different perturbation runs in Figure 3.5 it is evident that on average the soil conditions in the perturbation runs are drier both in the overall domain and the focus area compared to the nature run (for figure clarity, only **region C** is shown as it encapsulates both **region A** and **region B**). It also has a diurnal cycle due to evaporation and precipitation cycles.

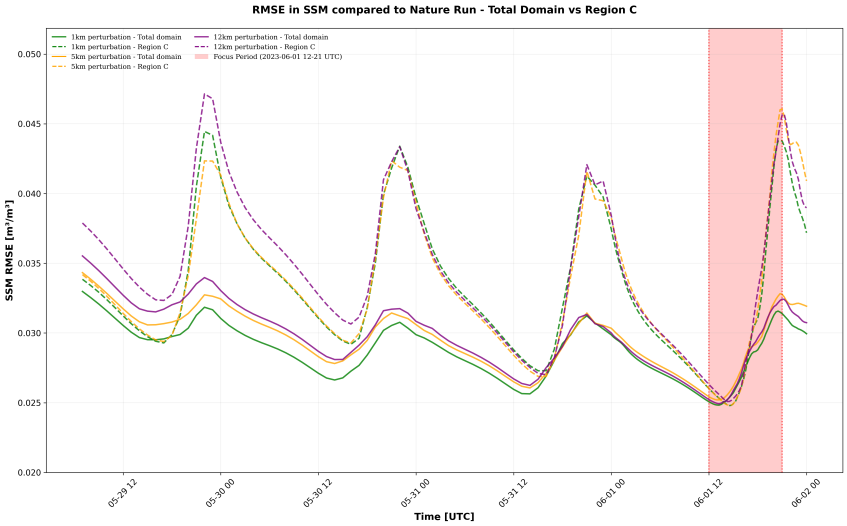


Figure 3.6: Time series of SSM RMSE compared to the nature run in the total domain and **region C** for the 1 km perturbation, 5 km perturbation and 12 km perturbation runs during the entire simulation period (2023-05-29 00 UTC - 2023-06-02 00 UTC) showing the effect of the different resolution of SSM perturbation.

From the time series of SSM RMSE compared to the nature run in Figure 3.6 it is also clear that the perturbation is initially strongest for the 12 km coarsened SSM perturbation, followed by the 1 km and 5 km perturbation runs respectively. However this initial perturbation difference across different perturbation runs quickly dissipates after the first simulation day. It should be noted that the RMSE for all runs remains below 0.05 m^3/m^3, which is the observation error that we have assumed.

Focusing now on the afternoon hours of the last simulation day (focus period), the 12 km perturbation run appears to generate the strongest variation in the focus regions. The cloud systems in Figure 3.7 have a different depth (colder cloud top temperature, see 18 UTC and 19 UTC) and the SAL metrics of accumulated precipitation in table 3.1 have the highest structure and location variation values, which are up to 0.88 and 0.37 respectively in **region B**. The core location in Table 3.2 is moderately displaced, with values between 0.2 and 0.4 in the focus regions. The amplitude and timing of precipitation in Figure 3.8 shows a moderate variation compared to the nature run, particularly within **region A**. However, the structure of accumulated precipitation in the overall domain varies the least (S=0.003), suggesting a redistribution of precipitation that is not dissimilar in intensity and shape from the nature run.

The 5 km perturbation also introduces significant perturbation effects. Clouds are triggered in the center of the domain (evident at 19 UTC) that are not present in the same location in the nature run, and which evolve into a larger central cloud (20 UTC, Figure 3.7). The precipitation field sees local intense precipitation variations (Figure 3.9), with precipitation structure SAL metrics between 0.1 and 0.4 (Table 3.1). The amplitude and location SAL metrics are not significantly affected and stay below 0.1, while the core location

is affected the most, with values around 0.3 in all three sub-regions (Table 3.2). The amplitude and timing of precipitation differs most from the nature run in all three sub-regions and most prominently in **region A** (Figure 3.8)

Finally, the 1 km perturbation has limited impact, triggering cloud development earlier than in the nature run (see panel at 17 UTC in Fig. 3.7) as well as additional convective activity in the center-north area of the domain (see panel at 19 UTC). Quantitative assessment using SAL metrics in Table 3.1 reveals relatively weak perturbations of accumulated precipitation and core location across all focus regions, with structure, amplitude, and location metrics remaining below 0.25, 0.2, and 0.12 respectively and core location metric below 0.14 (Table 3.2).

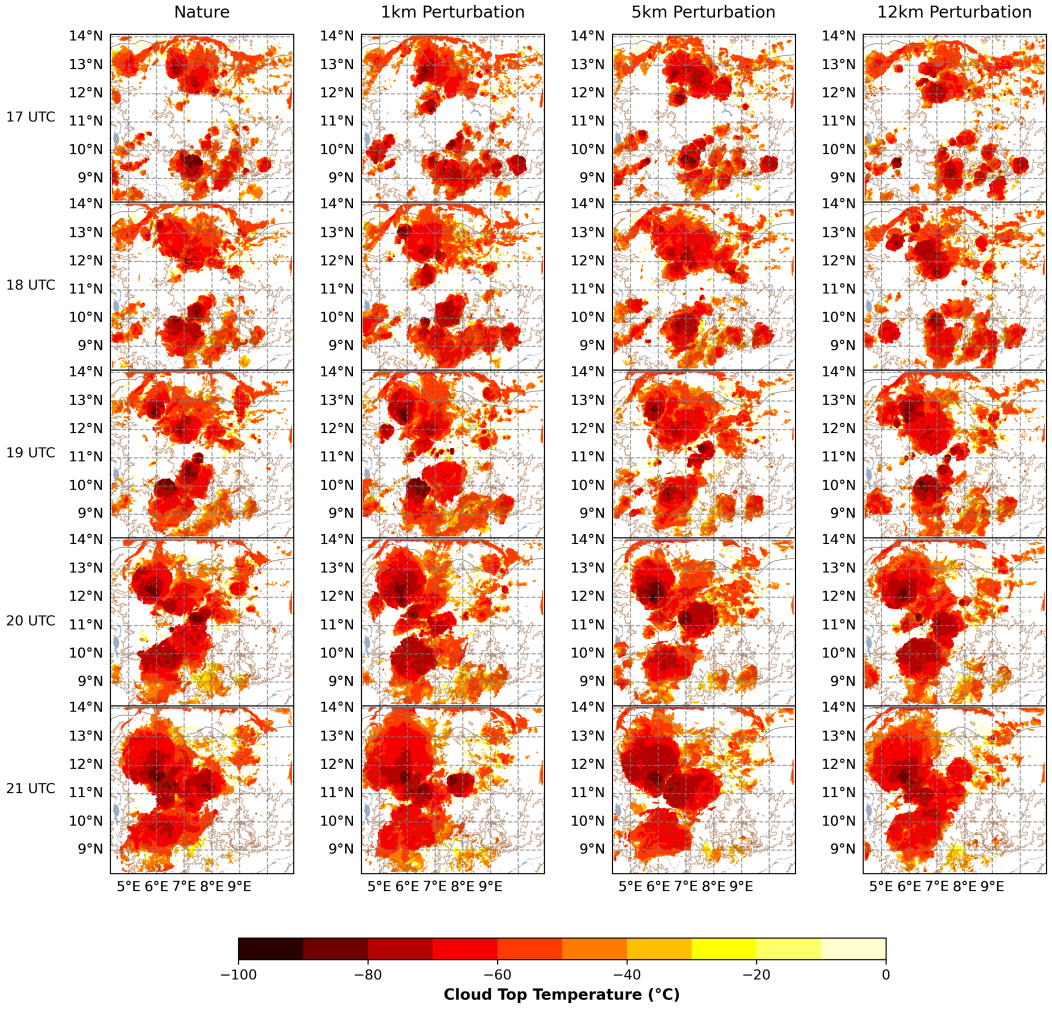


Figure 3.7: Cloud top temperature comparison for the nature, 1 km perturbation, 5 km and 12 km perturbation runs at key timesteps on 2023-06-01, showing the effect of the different resolution of SSM perturbations on cloud formation and cloud depth.

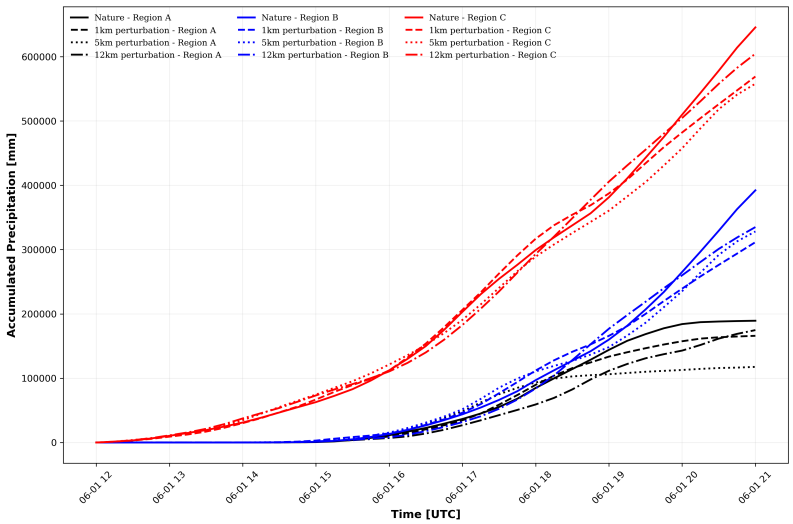


Figure 3.8: Time series of accumulated precipitation on 2023-06-01 between 12 and 21 UTC in **region A**, **region B** and **region C** for the nature (solid line), 1 km perturbation (dashed line), 5 km perturbation (dotted line) and 12 km perturbation runs (dotted and dashed line), showcasing the perturbation in amplitude and timing of precipitation.

Overall it is evident that the perturbations have a limited effect, as they change the cloud formation by only slightly displacing the cloud location and modifying their depth but not the overall cloud structure. The convective system due to orographic lifting remains always present, and clouds near the edge of the domain persist due to the lingering interaction with the boundaries.

The coarser the SSM resolution to perturb the system, the stronger the perturbation, as seen primarily by the SAL metrics in the focus areas. However, this perturbation is not reflected in the metrics of the precipitation field when considering the total domain; for example, the structure metric for the 12 km perturbation run in the total domain is much smaller than the corresponding value for the 1 km and 5 km perturbation runs.

These differences in perturbation strength must be considered when interpreting assimilation performance, as larger perturbations naturally allow for greater potential improvement.

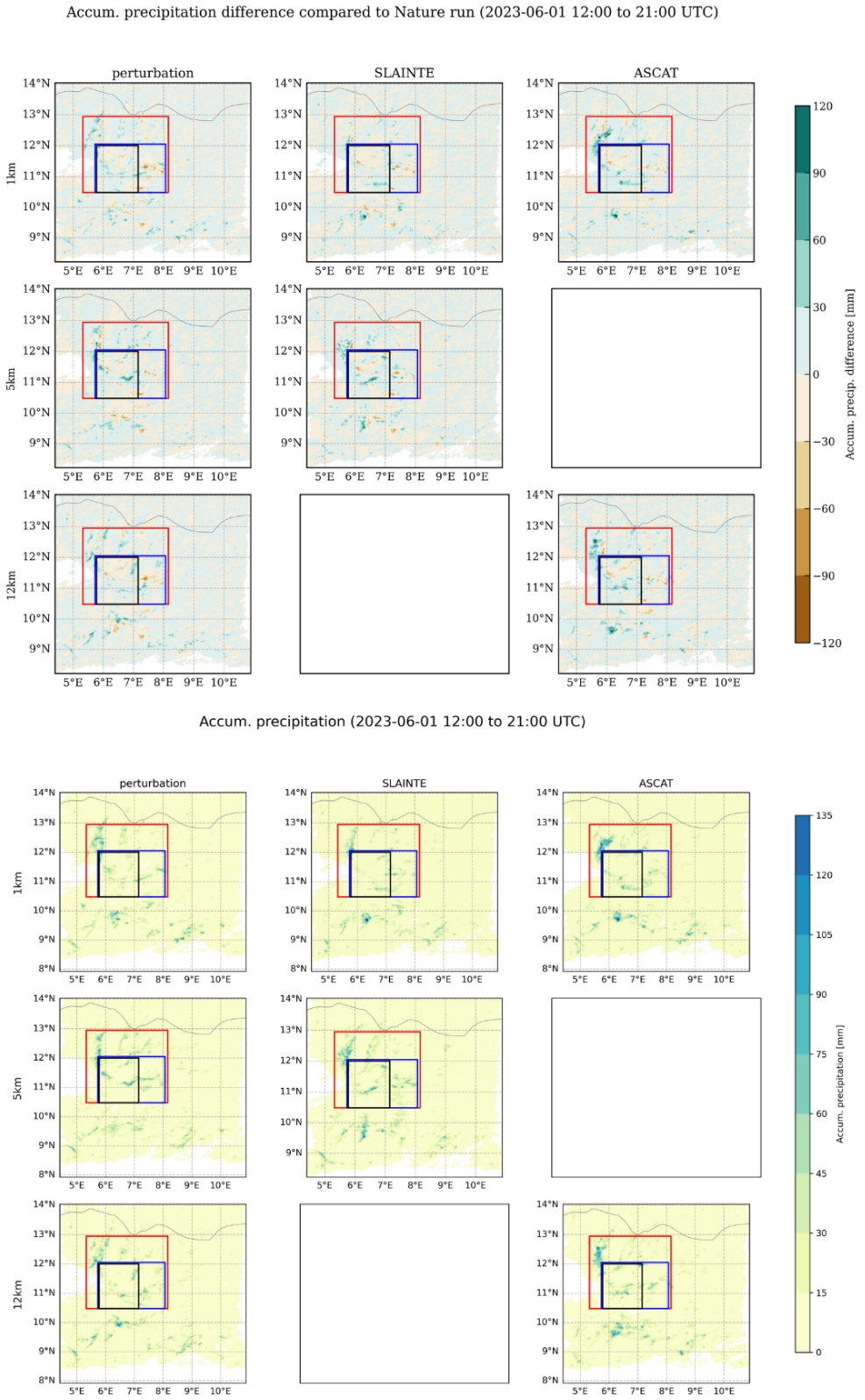


Figure 3.9: Accumulated precipitation variation compared to the nature run (top) and accumulated precipitation (bottom) for the perturbation and assimilation experiment runs during the hours of peak convective activity between 12 and 21 UTC on 2023-06-01. The boxes indicate the focus areas: **region A**, **region B** and **region C**. The regions in white indicate regions where no precipitation occurred.

		region A		region B		region C		Total Domain	
		S_{pert}	S_{assim}	S_{pert}	S_{assim}	S_{pert}	S_{assim}	S_{pert}	S_{assim}
Structure	1km SLAINTE	-0.25	0.14	-0.17	0.01	-0.04	0.06	0.14	-0.45
	1km ASCAT	-0.25	-0.008	-0.17	-0.01	-0.04	-0.16	0.14	-0.24
	5km SLAINTE	-0.19	-0.08	-0.31	0.07	-0.28	0.04	0.17	0.04
	12km ASCAT	-0.54	-0.52	-0.88	-0.54	-0.75	-0.52	0.003	0.09
		A_{pert}	A_{assim}	A_{pert}	A_{assim}	A_{pert}	A_{assim}	A_{pert}	A_{assim}
Amplitude	1km SLAINTE	-0.05	0.09	-0.16	-0.06	-0.11	-0.03	0.01	-0.03
	1km ASCAT	-0.05	0.07	-0.16	-0.11	-0.11	-0.07	0.01	0.07
	5km SLAINTE	-0.16	0.10	-0.002	0.03	-0.04	0.02	0.02	0.06
	12km ASCAT	0.06	-0.06	-0.02	-0.11	-0.004	-0.07	0.03	0.11
		L_{pert}	L_{assim}	L_{pert}	L_{assim}	L_{pert}	L_{assim}	L_{pert}	L_{assim}
Location	1km SLAINTE	0.09	0.08	0.08	0.08	0.11	0.09	0.01	0.01
	1km ASCAT	0.09	0.05	0.08	0.10	0.11	0.10	0.01	0.03
	5km SLAINTE	0.10	0.05	0.06	0.08	0.02	0.09	0.03	0.03
	12km ASCAT	0.33	0.32	0.37	0.45	0.37	0.37	0.02	0.03

Table 3.1: Structure (S), Amplitude (A) and Location (L) forecasting metrics for accumulated precipitation between 12 and 21 UTC in sub-regions and total domain, which compare the perturbation runs to the nature run (*pert*) and the assimilation runs to the nature run (*assim*). The values colored in green indicate improvement compared to the perturbation baseline, while values in red indicate degradation. No color indicates negligible variation compared to the perturbation baseline.

Region	**Run Type**	**1 km SLAINTE**	**1 km ASCAT**	**5 km SLAINTE**	**12 km ASCAT**
region A	Pert	0.14	0.14	0.27	0.20
	Assim	0.18	0.44	0.20	0.34
region B	Pert	0.08	0.08	0.36	0.39
	Assim	0.14	0.19	0.24	0.29
region C	Pert	0.09	0.09	0.35	0.28
	Assim	0.24	0.11	0.23	0.19
Total domain	Pert	0.05	0.05	0.02	0.04
	Assim	0.06	0.03	0.05	0.06

Table 3.2: Location metric (L) for convective cores. Quantifies the difference between either the perturbation or assimilation experiment runs from the nature run in the center of mass of convective cores and the distance of individual cores from their respective center of mass. The values colored in green indicate improvement compared to the perturbation baseline, while values in red indicate degradation. No color indicates negligible variation compared to perturbation baseline.

3.3 Data assimilation experiments

Having established that coarser SSM resolution perturbations produce stronger perturbations of cloud formation and precipitation field, we now examine whether data assimilation can improve forecasts compared to the perturbation baseline.

3.3.1 SSM observation assimilation

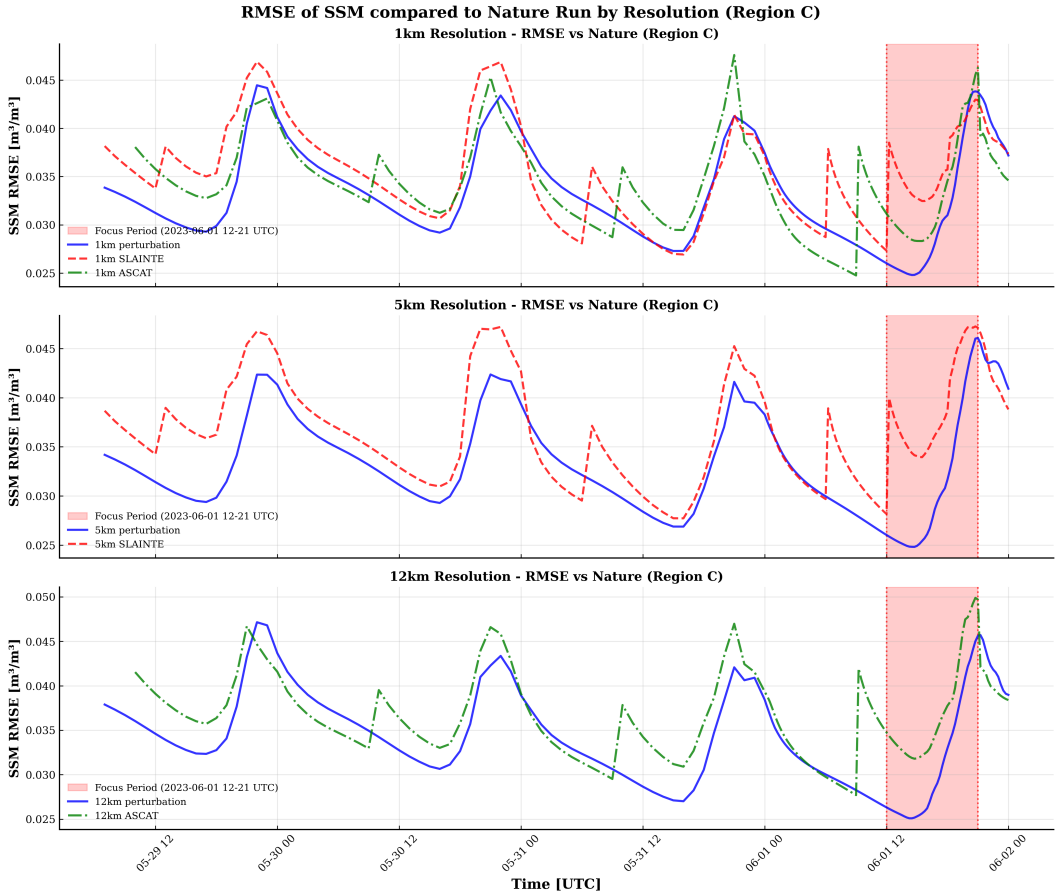


Figure 3.10: Time series of SSM RMSE compared to the nature run for the perturbation and assimilation experiment runs in **Region C** during the entire simulation period (2023-05-29 00 UTC to 2023-06-02 00 UTC)

.

The time series of RMSE in SSM compared to the nature run for the different perturbation and assimilation runs in Figure 3.10 provides preliminary information on the effect of assimilating observations at different resolutions and times. As **region A** and **region B** are enclosed in **region C**, only the latter is shown for figure clarity.

It is evident that generally the RMSE associated with the assimilation runs is higher than for the perturbation runs. Specifically, the synthetic observations assimilated during

times of the day when the soil are in the process of drying (0 to 13 UTC), the observations tend to have a much higher RMSE and the added noise could potentially perturb the system further. However, the point-wise RMSE does not take into account spatial correlation, so a higher noise level does not necessarily imply that the spatial pattern is not retained in the observations, and equivalently a lower RMSE in the perturbation run does not necessarily provide information on the spatial pattern. It is equally important to note that the noise applied to the SMM to generate synthetic observations does not take into account spatial correlations, which instead occurs in a real observation scenario.

In addition, SLAINTE would be operating at L-band, unlike ASCAT's C-band, which generally implies higher sensitivity to SSM and penetration depth than C-band (Brunelli & Mancini, 2024; Zhou et al., 2025) and therefore potentially has a lower error than the one that we have applied.

3.3.2 Low SSM resolution (12 km) runs

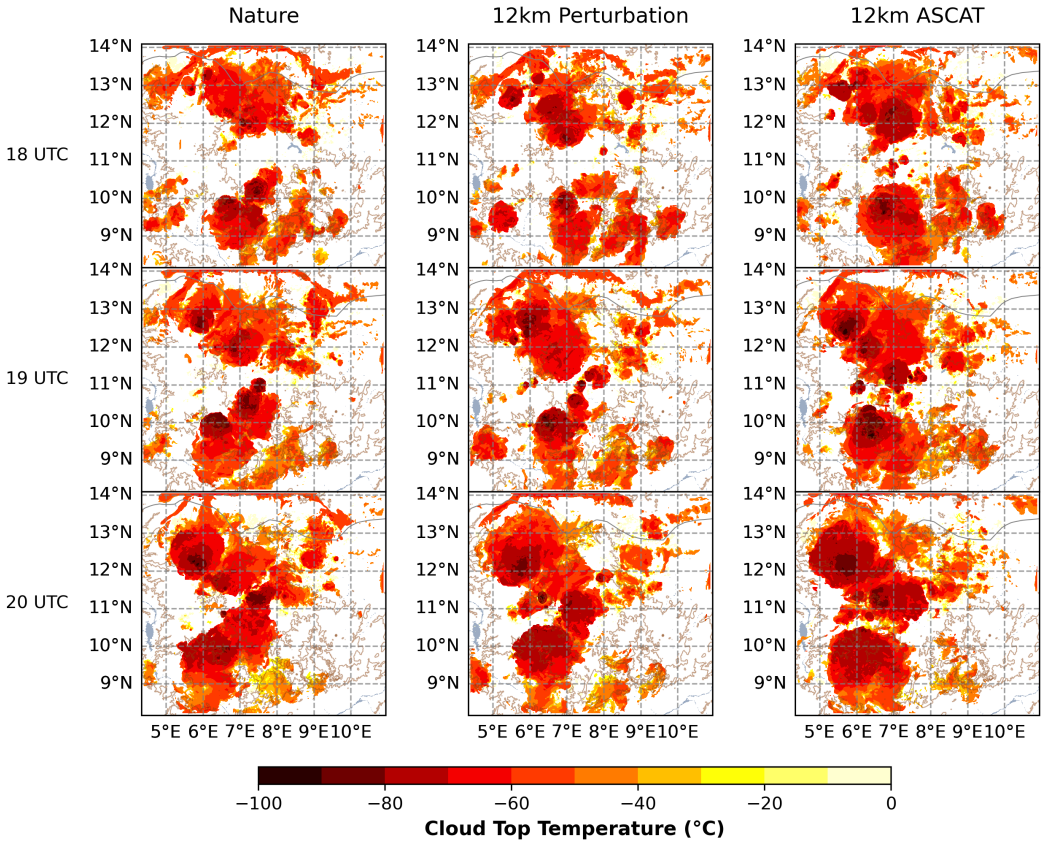


Figure 3.11: Cloud top temperature comparison for the nature, 12 km perturbation and 12 km ASCAT runs at key timesteps on 2023-06-01, showing the evolution of cloud formation and cloud depth across the different runs. The full cloud evolution at more timesteps can be found in Appendix A.1.

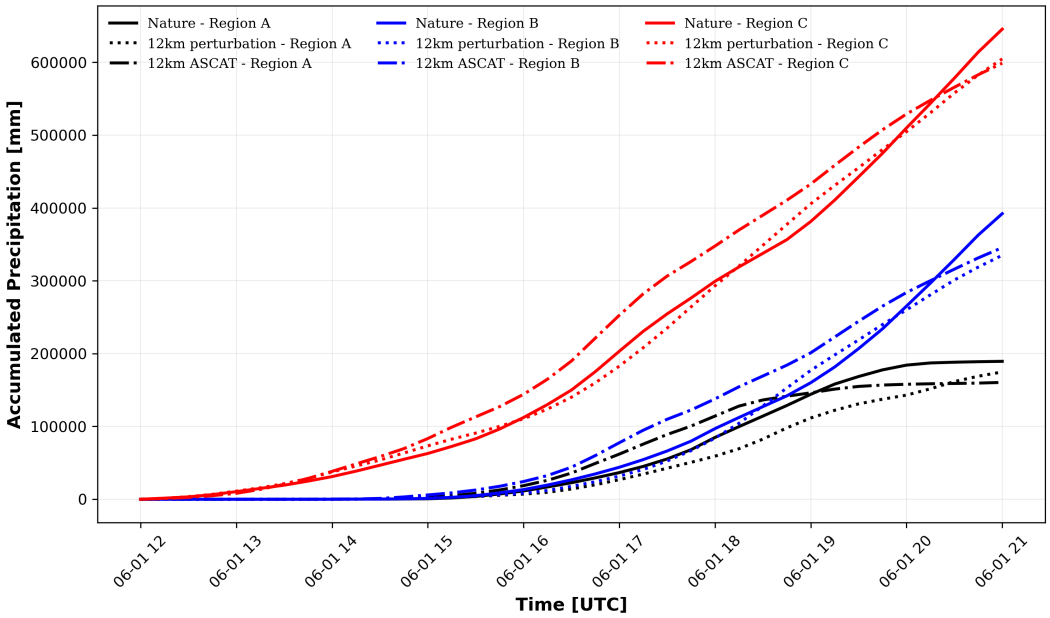


Figure 3.12: Time series of accumulated precipitation for the nature, 12 km perturbation and 12 km ASCAT runs in **region A**, **region B** and **region C** of the domain during the hours between 12 and 21 UTC on 2023-06-01.

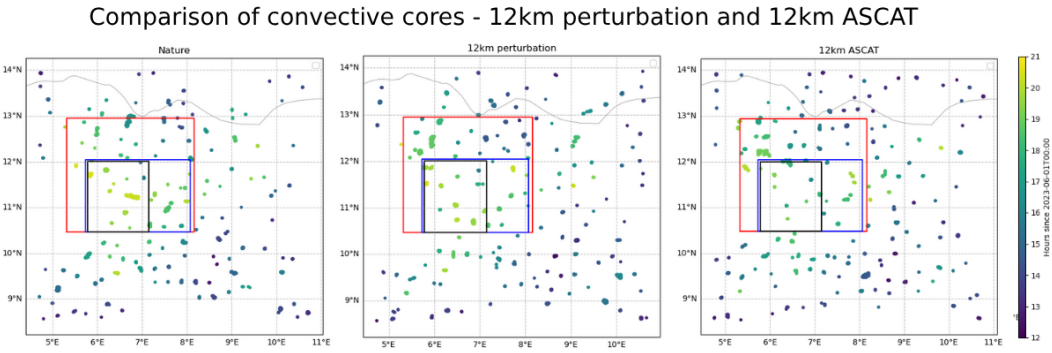


Figure 3.13: Comparison of convective core location and timing between the nature, 12 km perturbation and 12 km ASCAT runs. The colorbar indicates the time of the core formation, while the boxes indicate the focus regions: **region A**, **region B** and **region C**.

The 12 km ASCAT experiment showcases the limitations of coarse-resolution SSM observation assimilation, even under idealized observation conditions. Visual comparison of cloud evolution (Fig. 3.11) shows that the assimilation experiment produces cloud structures that are more dissimilar from the nature run than the perturbation baseline both in cloud depth and location. For example, at 18 UTC the cloud tops of the cloud system in the north of the domain are much deeper, while the same cloud system at 20 UTC extends further then its counterpart in the nature and experiment run. Additionally, at 19 UTC, the

clouds move towards the center of the domain faster than in the nature run. This is reflected in the temporal evolution of accumulated precipitation (Fig. 3.12) indicating degraded timing compared to the perturbation run, with premature onset around 17 UTC followed by reduced total accumulation relative to both the nature run and perturbation baseline.

Quantitative assessment using SAL metrics in Table 3.1 reveals mixed results, with some improvements in individual regions offset by degradation in others. For example, in **region B** the structure metric improves from -0.88 to -0.54 but the location metric degrades from 0.37 to 0.45. While convective core location metrics show improvements in **regions B** and **C** (Table 3.2), the initial perturbation is also significant, and these gains may reflect coincidental spatial correlation rather than physically meaningful improvements. Finally, in **region A** the timing of core formation occurs much earlier than in the nature and perturbation run (Figure 3.13).

3.3.3 Medium SSM resolution (5 km) runs

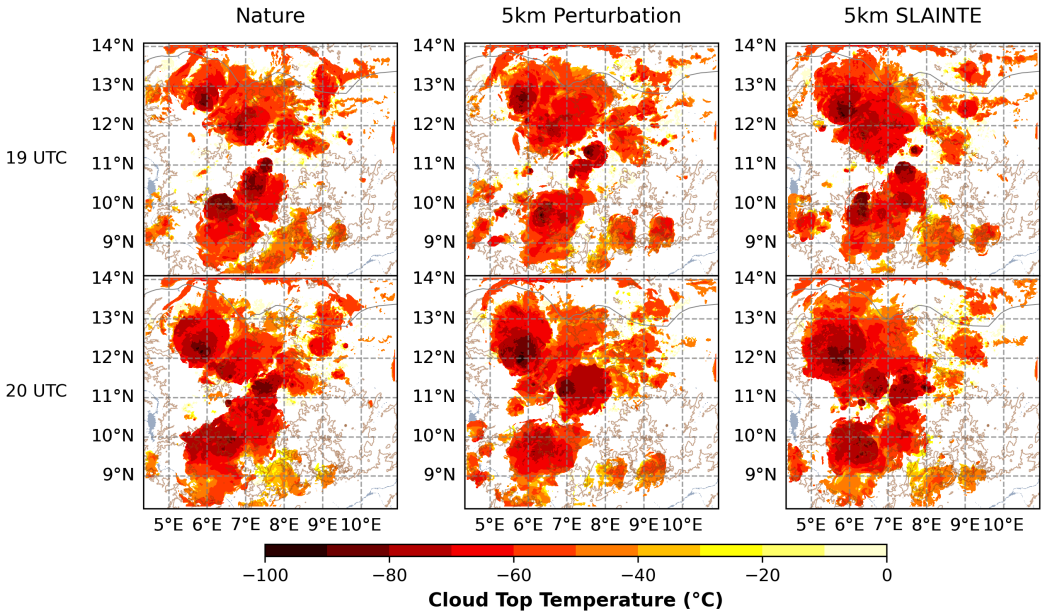


Figure 3.14: Cloud top temperature comparison for the nature, 5 km perturbation and 5 km SLAINTE runs at key timesteps on 2023-06-01, showing the variation of cloud formation and cloud depth across the different runs. The full cloud evolution at more timesteps can be found in Appendix A.2.

The 5 km SLAINTE experiment showcases the most comprehensive and consistent improvements across all evaluation metrics, suggesting an optimal balance of SSM spatial heterogeneity preservation and spatial resolution. The SAL metrics in Table 3.1 show that the accumulated precipitation field improves consistently across all focus regions, and it is the only simulation for which the core location metric improves in all focus areas (Table 3.2).

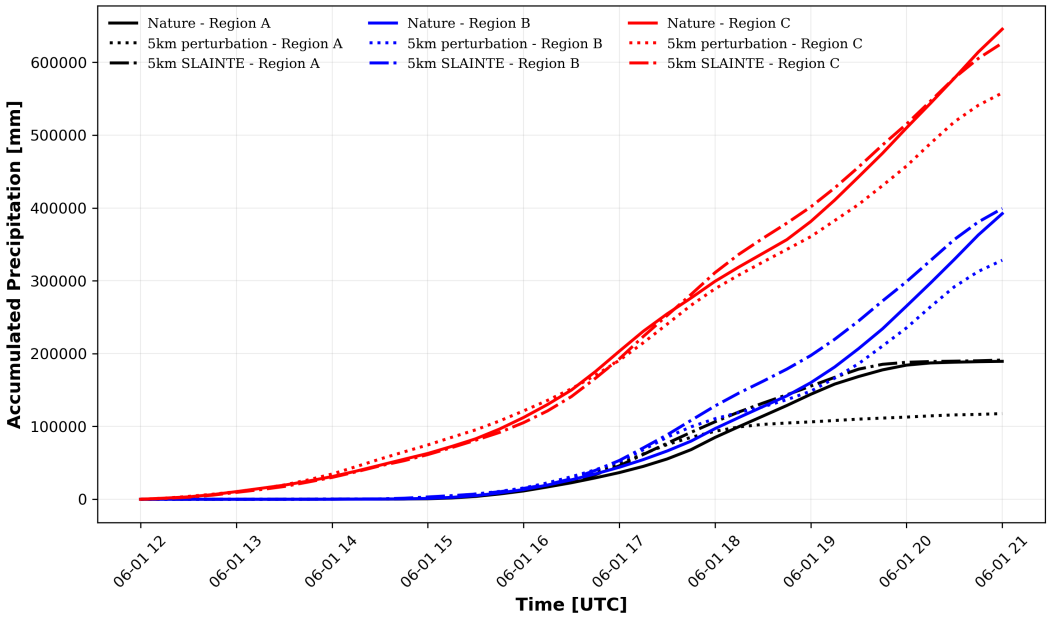


Figure 3.15: Time series of accumulated precipitation for the nature, 5 km perturbation and 5 km SLAINTE runs in **region A**, **region B** and **region C** of the domain during the hours between 12 and 21 UTC on 2023-06-01.

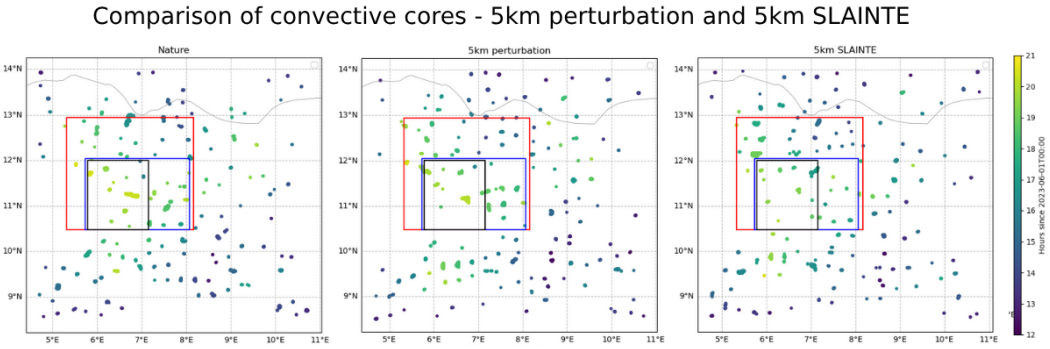


Figure 3.16: Comparison of convective core location and timing between the nature, 5 km perturbation and 5 km SLAINTE runs. The colorbar indicates the time of the core formation, while the boxes indicate the focus regions: **region A**, **region B** and **region C**.

The time series of accumulated precipitation (Fig. 3.15) shows that the timing of precipitation events are reproduced well in regions **A** and **C** and that the precipitation magnitude is well captured in all regions. Visual comparison of cloud top temperature (Fig. 3.14) reveals that it maintains cloud structure and timing closer to the nature run than the perturbation baseline (see 19 UTC in Fig 3.14).

Some visual differences not captured by the evaluation metrics include a larger extent in cloud area at 20 UTC (Figure 3.14) and locally intense precipitation variations (Figure 3.9),

as well as an anticipated timing of core formation, specifically in **region A** (Figure 3.16).

3.3.4 High resolution (1 km) runs

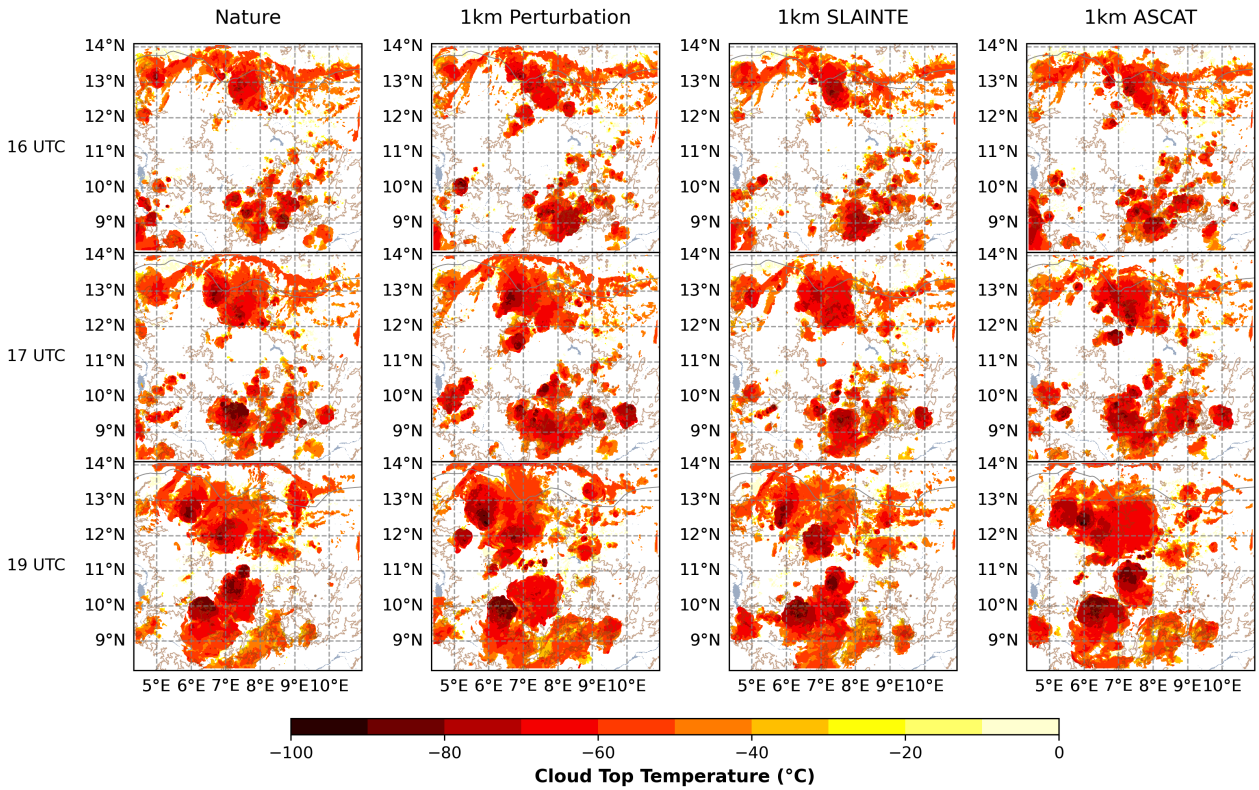


Figure 3.17: Cloud top temperature comparison for the nature, 1 km perturbation, 1 km SLAINTE and 1 km ASCAT runs at key timesteps on 2023-06-01, showing the variation of cloud formation and cloud depth across the different runs. The full cloud evolution at more timesteps can be found in the Appendix A.3.

1 km SLAINTE run

The 1 km SLAINTE experiment shows excellent precipitation field recovery in the focus regions with little visual variation from the nature run pattern (Fig. 3.9). The most significant quantitative improvements are found in **Region B**, where precipitation distribution and intensity are nearly perfectly reproduced, with both structure and amplitude SAL metrics approaching zero (Table 3.1). The temporal evolution of cloud formation shows that SLAINTE assimilation successfully restores the timing of central domain convection, particularly evident in the 16 to 17 UTC panels of Fig. 3.17 where the perturbation run shows premature cloud development that is corrected in the assimilation run. The precipitation timing and amplitude also significantly improves and follows the nature run closely in all three sub-regions (Figure 3.18).

However, despite the precipitation improvement, convective core location does not improve and in some cases deteriorates compared to the perturbation baseline (Table 3.2). This contradiction suggests that while the overall precipitation distribution and timing is well-captured, the precise location of convective initiation may be sensitive to observation noise at the 1 km scale.

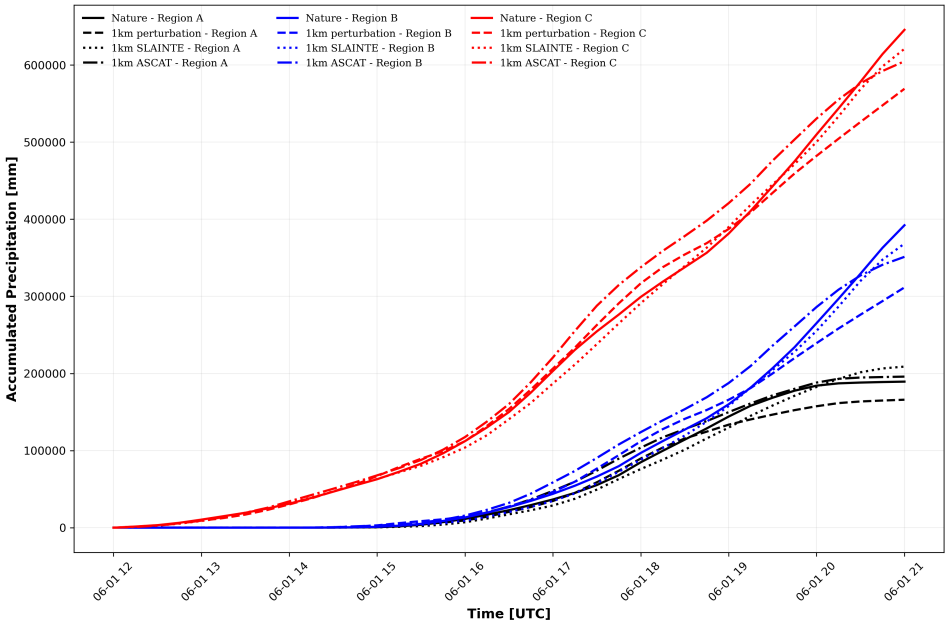


Figure 3.18: Time series of accumulated precipitation on 2023-06-01 between 12 and 21 UTC in **region A**, **region B** and **region C** for the nature (solid line), 1 km perturbation (dashed line), 1 km SLAINTE (dotted line) and 1 km ASCAT runs (dotted and dashed line), showcasing the amplitude and timing of precipitation.

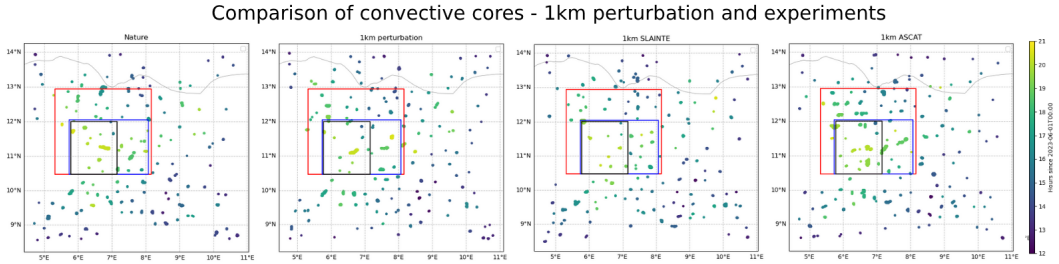


Figure 3.19: Comparison of convective core location and timing between the nature, 1 km perturbation, 1 km SLAINTE and 1 km ASCAT runs. The colorbar indicates the time of the core formation, while the boxes indicate the focus regions: **region A**, **region B** and **region C**.

1 km ASCAT run
The 1 km ASCAT experiment exhibits different behavior. There is an intense visual precipitation variation in **region C** (Fig. 3.9) which nonetheless does not significantly degrade the structure metric beyond $S_{assim} = -0.16$ in **region C** (Table 3.1), suggesting that the center of mass of the event is not strongly displaced. The timing of the cloud formation is also restored but not to the extent of the 1 km SLAINTE experiment (Figure 3.17). This discrepancy in timing is also revealed in Fig. 3.18, showing rainfall onset approximately 1 to 2 hours before the nature run timing across all focus regions. This also results in larger and more numerous core development (see Fig. 3.19) of which the location metric nonetheless degrades particularly in **region A** where it increases from 0.14 to 0.44 (Table 3.2).

3.4 Impact of observation spatial resolution

The spatial resolution of soil moisture observations significantly influences data assimilation effectiveness, with different resolutions showing distinct advantages depending on the strength of initial perturbations and the preservation of soil moisture gradients necessary for triggering convection. Comparing the pairs of ASCAT and SLAINTE experiments isolates the effect of spatial resolution while maintaining the timing of the SLAINTE and ASCAT synthetic observations.

3.4.1 ASCAT spatial resolution: 1 km ASCAT – 12 km ASCAT comparison

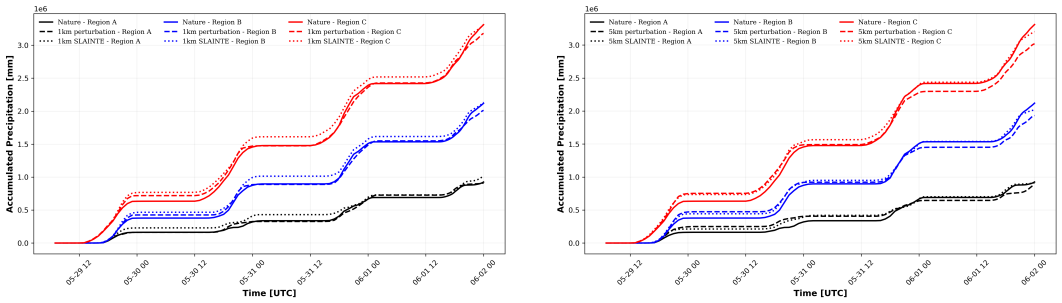
Comparison of these two runs provides information on the effect of spatial resolution at ASCAT observation times. Because of the vastly different perturbation effects, comparing them is not straightforward. There is indication that the 1 km ASCAT simulation performs better: cloud formation and precipitation forecasting metrics are generally better than for the 12 km assimilation, which instead perturbed the simulation further. For the 1 km case when the perturbation is strong enough ($|S/A/L| > 0.1$) the SAL metrics improve, while only the precipitation structure improves in the 12 km case. However, the location of the cores does not improve at all with the 1 km ASCAT experiment, while they improve in **region B** and **region C** for the 12 km ASCAT experiments. Improvement in core location alone does not arguably counter the poor performance of all the other comparative assessments, therefore we can infer that assimilation at 1 km resolution at ASCAT times performs better than assimilation at 12 km resolution.

Both simulations showcase an onset of precipitation of about 1-2 h compared to both the nature and perturbation runs (see Figures 3.12, 3.18) and the most locally intense precipitation variations compared to the nature run (Figure 3.9), suggesting that the observation time might be a more important requirement than resolution, as even the 1 km observation resolution, albeit better than the 12 km resolution, is still not sufficient to significantly improve the simulation.

3.4.2 SLAINTE spatial resolution: 1 km SLAINTE – 5 km SLAINTE comparison

The comparison of the 1 km SLAINTE and 5 km SLAINTE assimilation runs allows us to consider the effect of spatial resolution at SLAINTE observation times.

It is evident that the 5 km SLAINTE run performs significantly better than 1 km SLAINTE. As previously analyzed, the 5 km SLAINTE experiment shows unequivocal improvement of all metrics, both in precipitation field and core location. In contrast, the 1 km SLAINTE run shows improvement in precipitation field only when the perturbation exceeds sufficient strength thresholds (Table 3.1), and the core location never improves (Table 3.2).



(a) (b)

Figure 3.20: Time series of accumulated precipitation during the entire simulation period (2023-05-29 00 to 2023-06-02 00) in **region A**, **region B** and **region C** for (a) the nature (solid line), 1 km perturbation (dashed line) and 1 km SLAINTE (dotted line) runs and (b) the nature (solid line), 5 km perturbation (dashed line) and 5 km SLAINTE (dotted line) runs.

The limited performance at 1 km resolution can be further explained by the time series of accumulated precipitation over the entire simulation period (Fig. 3.20). Figure 3.20a indicates that the perturbation induced by 1 km SSM resolution perturbation has a limited effect on precipitation, and that the assimilation might instead perturb the system further. Specifically, the accumulated precipitation initially detaches from that of the nature run on the 2023-05-29 in all three regions, but quickly returns to following the nature run after 2023-05-30. In contrast, the observations on 2023-05-30 at 18 UTC seem to perturb the assimilation run further and distance it from the nature run.

The equivalent time series for the 5 km perturbation and 5 km SLAINTE experiment in Fig. 3.20b shows that the perturbation persists but that the assimilation of observations tethers the 5 km SLAINTE precipitation progressively closer to the nature run until the two overlap completely on the last simulation day.

3.4.3 Impact of observation spatial resolution - summary

These results suggests a fine balance between observation noise and resolution. Assimilation at 5 km resolution clearly improves both the convective core location and the precipitation field, which can be attributed to the combination of a sufficiently strong perturbation

and a resolution fine enough to capture meaningful soil moisture gradients. As such, we can infer that 5 km is a sufficient resolution to improve forecasts.

In contrast, although the 1 km observations improve the precipitation fields in both the ASCAT and SLAINTE experiments, they do not lead to improvements in core location. One likely explanation is that the perturbation in these runs was not strong enough to differentiate the system sufficiently from the nature run, and the assimilation instead perturbs the system further. Additionally, the added noise at 1 km resolution may have disrupted the soil moisture gradients and suppressed the secondary circulations required to trigger convection (further discussed in the following section). This aligns with findings in previous literature that suggest that soil moisture anomalies must extend over horizontal scales of approximately 10 km to effectively initiate convection (Pielke Sr., 2001; Taylor et al., 2011).

Finally, the 12 km ASCAT experiment hardly improves the simulation. Neither the cloud structures nor the precipitation fields show notable improvement compared to the perturbation run. The improvement of core location metrics must be interpreted cautiously as the metric does not take into account the orientation of the objects: the 12 km resolution could create artificial gradients due to observation noise different than those in the nature run, meaning the apparent improvement in core location may be coincidental rather than physically meaningful. In addition, the perturbation of the location metric is very strong (0.39 and 0.29 in **region B** and **region C**) respectively, allowing for considerable room for improvement.

These results suggest that while moderate resolutions such as 5 km can already provide meaningful improvements, finer resolutions may only be beneficial if observation noise is minimized enough to retain larger scale variability necessary to trigger convection.

3.5 Impact of observation time and observation frequency

Additionally to the observation spatial resolution, the timing and frequency of the observations is equally important as convection follows a daily cycle and is most sensitive to soil conditions in the afternoon-evening transition.

3.5.1 Impact of sub-daily observations: 1 km SLAINTE – 1 km ASCAT comparison

To isolate the impact of observation timing from spatial resolution effects, we compare the 1 km SLAINTE and 1 km ASCAT experiments, which differ only in their observation time while maintaining identical spatial resolution and perturbation characteristics. As discussed previously, the perturbation of the 1 km perturbation run is not very strong so there is limited possibility for improvement of the forecasts.

To assess the significance of morning observation timing, it should first be noted that no precipitation occurs between 6 and 12 UTC on each simulation day. Taking into account the time series of average SSM in figure 3.5, we can also infer that SSM later in those hours

is lower on average but retains the same spatial pattern as we assume the evaporation rate to be largely uniform.

Since evaporation and the drying of the soil occurs naturally within the model, the assimilation of the 12 UTC observation in addition to the 6 UTC observation in the 1 km SLAINTE experiment provides little additional information. In fact, it may perturb the system further by introducing additional random noise, as suggested by the time series of SSM RMSE in Figure 3.10.

Similarly, the 9 UTC synthetic observation assimilated in the 1 km ASCAT experiment closely resembles the 6 and 12 UTC observations, differing only in added noise and average SSM due to evaporation. Therefore, differences in the two experiments due to these morning assimilations should be attributed to the model response time to the SSM assimilation rather than the observation time itself.

The evening overpass may create more distinct differences. The 18 UTC SLAINTE observation occurs during the time when convection is most intense while the 21 UTC ASCAT overpass misses this window of strong convective activity. These considerations are essential for interpreting the performance variation between these two runs.

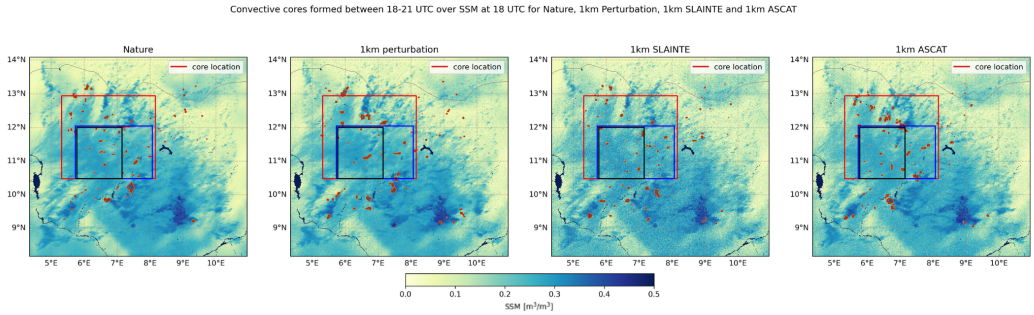


Figure 3.21: Convective cores generated between 18 and 21 UTC in the nature, 1 km perturbation, 1 km SLAINTE and 1 km ASCAT runs, overlayed to the SSM at time 18 UTC, showcasing the effect of the 18 UTC observation assimilation in the SLAINTE assimilation run. The boxes indicate the focus areas: **region A**, **region B** and **region C**.

	Region A	**Region B**	**Region C**
1 km pert	0.10	0.18	0.15
1 km SLAINTE	0.08	0.31	0.29
1 km ASCAT	0.40	0.26	0.17

Table 3.3: Location metric for convective cores formed between 18 and 21 UTC in the 1 km perturbation, 1 km SLAINTE and 1 km ASCAT runs. The red text color indicates degraded performance compared to the perturbation baseline, while no color indicates negligible performance variation compared to the perturbation run.

The most significant difference between the two runs is in fact the timing of precipitation

onset (Fig. 3.18). The ASCAT experiment consistently produces premature precipitation around 17 UTC in all focus areas. This early development is accompanied by reduced total precipitation accumulation, suggesting that the premature triggering may exhaust available atmospheric moisture prematurely. In contrast, the SLAINTE experiment maintains precipitation timing and amplitude much closer to the nature run.

Notably, the additional precipitation in the 1 km ASCAT run occurs before 18 UTC, indicating that the better performance of the 1 km SLAINTE run cannot be attributed solely to the assimilation of the 18 UTC observation. However, since such assimilations occurred over the course of the three-day spinup period as well, we could assume that consistent assimilation of observations at the SLAINTE timing is more suited to improve the timing of precipitation events.

Analysis of convective cores developing between 18 and 21 UTC (Fig. 3.21, Table 3.3) provides additional insight into the effect of timing and the value of the 18 UTC observation in the SLAINTE experiment. The SSM in the nature run presents significant gradients in **region A** and **region B**, which coincide with the location of the cores. The assimilation of the SSM observation in the 1 km SLAINTE experiment reproduces the shape of the larger scale variability in the northern part of **region C**. However, the gradients in **region A** are less pronounced because of the added noise, which could disrupt secondary circulations in the area. This could perhaps explain the absence of cores in the region and why the core location does not improve compared to the 1 km ASCAT run during the same time window (see Table 3.3).

3.5.2 Impact of observation time and frequency – Summary

This analysis reveals that the most significant difference between the ASCAT and SLAINTE assimilation runs lies in precipitation onset timing. Time series of accumulated precipitation show that both the 1 km ASCAT and 12 km ASCAT experiments produce precipitation earlier than the nature run -a trend that does not appear in either of the SLAINTE experiments. This suggests that assimilation at SLAINTE overpass times better captures the timing of convection and subsequent precipitation.

This improvement stems from the effect of cumulative assimilations over multiple days rather than a single assimilation. This is demonstrated by the lack of improvement in core location immediately after the 18 UTC assimilation in the 1 km SLAINTE run.

Nonetheless, the assimilation of the 18 UTC observation over multiple days appears to have a crucial role in constraining the simulations. This overpass is in fact the main difference between the 1 km SLAINTE and the 1 km ASCAT assimilations and is the most likely reason why the former performs better. Since most precipitation in the simulations occurs between 13 and 22 UTC, it is reasonable to infer that the earlier 18 UTC SLAINTE assimilation provides valuable information at the critical stage of convective activity, whereas the 21 UTC ASCAT overpass partly misses this window.

Evaluating the morning overpass timing is challenging due to the absence of precipitation and minimal soil condition changes during these hours. Under these circumstances, a single observation in the morning hours may be sufficient to constrain the system, but its optimal timing cannot be studied further due to minimal changes in the simulated soil sys-

tem. The second SLAINTE morning assimilation may in fact introduce unnecessary noise and disrupt the system further, making a single earlier observation preferable under these conditions.

Finally, it is logical to assume that a higher number of assimilated observations generally leads to better results, which further explains why the cumulative impact of SLAINTE updates generally outperforms the ASCAT assimilation.

3.6 SLAINTE – ASCAT comparative assessment

The preceding analyses of spatial resolution effects (Section 3.4.1) and observation timing impacts (Section 3.5) allow a comprehensive comparison between the SLAINTE mission concept and ASCAT mission capabilities for MCS forecasting. This section synthesizes these findings to evaluate the attributes of the SLAINTE observing system based on its first technical concept and identify optimal mission configurations.

3.6.1 Key performance differences

This comparative analysis demonstrates that SLAINTE configurations consistently outperform ASCAT experiments across multiple forecast metrics, with the improvement primarily attributable to observation timing rather than spatial resolution alone. As established in Section 3.5, SLAINTE's 18 UTC overpass occurs during peak convective initiation, while ASCAT's 21 UTC observation partially misses this critical atmospheric window. The cumulative effect of SLAINTE assimilations leads to better performance in precipitation timing.

Regarding spatial resolution, the 5 km configuration provides the optimal balance between spatial heterogeneity preservation and noise levels and would unequivocally improve the forecasts. If finer observation resolutions were to be applied, observation noise should be carefully considered: the 1 km SSM resolution assimilation caused greater perturbation throughout the entire domain than any other assimilation configuration —far exceeding even the 1 km perturbation itself. This impact is evident in the structure metric degradation across the overall domain (Table 3.1), which varied from 0.14 in the 1 km perturbation to -0.45 in the 1 km SLAINTE run.

Finally, the poor performance of the 12 km ASCAT run across most qualitative and quantitative metrics confirms that there is limited value to assimilating ASCAT SSM observations for MCS forecasting even under such idealized observational conditions.

3.6.2 Operational considerations and limitations

Several factors must be considered when interpreting these results. In operational data assimilation setting, sophisticated observation operators can preserve spatial heterogeneity at the model grid even when assimilating coarser resolution data. As such, the 12 km resolution measurements would not degrade the overall SSM resolution as drastically in the model, resulting in a shift in average values rather than direct value substitution. This

suggests that ASCAT observations could complement SLAINTE coverage in regions where the latter provides insufficient spatial sampling.

Additionally, ASCAT operationally provides only daily observations with non-overlapping overpass areas, and an overlap area as large as the one considered in this study is unrealistic. The choice to include both was considered to evaluate both observation times and show that even in the best scenario of overlapping overpasses, ASCAT observations have limited applicability for assimilation, as the results of this study indicated.

Similarly, depending on the resolution of SLAINTE, it is unlikely that an area as large as the studied domain will be observed in its entirety. In the current mission concept, the swath width would be 40 km along a transect of a few 1000 km, making the assimilation of observations only possible for a much smaller area. However, if a resolution coarser than 1 km is considered a wider swath would be applicable.

It is equally important to note that the perturbation runs are but a single realization of possible random model evolutions. This constrains the validity of the study results to these model realizations and should be instead verified by perturbation ensembles.

3.6.3 Implications for SLAINTE mission concept requirements

The design of the mission requirements should consider the combined value of assimilating products from current and future missions. ASCAT could be used in combination with SLAINTE to provide additional spatial coverage in regions SLAINTE would not observe. A new Sentinel-1 C-band SAR product was also lunched in early 2025, with a spatial resolution of 1 km which provides observations every 6-12 days over the Sahel. (Fan et al., 2025), and the ROSE-L (Radar Observing System for Europe in L-band) mission is currently being designed by the European Space Agency (ESA). The latter aims to achieve a 50 m^2 spatial resolution with a 3-6 day revisit time (Pierdicca et al., 2019).

In this context, the clear additional value of SLAINTE would be its temporal resolution. One observation at 18 local time in addition to at least one morning observation before 12 should be sufficient to constrain the forecasts sufficiently, depending on the model response time to SSM variations.

As for the spatial resolution, because of the scale ($>$10 km) of the inhomogeneities involved in triggering convection and the fact that 5 km appears to be a fine enough resolution to improve forecasts, it may be preferable to opt for a coarser observation resolution in favor of a larger swath area.

Due to the very fast evolution of MCSs, the SLAINTE data should also be made available shortly after the overpass time. Specifically, the 18 UTC overpass is only useful because it captures the soil conditions at a time of peak convective activity. If that window is missed even by a few hours, the effect would be more similar to assimilating at the 21 UTC overpass time which we concluded achieves less optimal results.

Chapter 4

Conclusions and Recommendations

This research investigated the potential value of the Sub-daily Land-Atmosphere INTEractions (SLAINTE) satellite mission concept for improving numerical weather prediction (NWP) forecasting of Mesoscale Convective Systems (MCSs) in the Sahel through synthetic data assimilation experiments using the WRF model.

Soil moisture assimilation demonstrates measurable but limited impacts on cloud formation and precipitation characteristics under the idealized conditions tested. The coarser the resolution of the SSM perturbation and assimilation, the stronger the observed variations. Aspects of cloud formation are affected, specifically cloud depth and cloud location, as well as precipitation characteristics, mainly intensity and timing. However, convection was triggered regardless of soil conditions in all experiments, indicating that the convection initiation sensitivity to soil moisture perturbations within the model remains limited.

An optimal spatial resolution of 5 km emerged from this analysis. The coarser 12 km resolution representing current ASCAT capabilities showed minimal forecast improvement and even degraded performance. The 1 km resolution, while improving precipitation field characteristics, did not improve convective core location and introduced excessive noise that degraded domain-wide precipitation structure metrics. The 5 km resolution proved optimal, seemingly preserving necessary soil moisture gradients for convection initiation while limiting observation noise.

Sub-daily observations at SLAINTE overpass times consistently outperformed ASCAT timing observations across multiple metrics. The key improvements result from the 18 UTC overpass occurring during peak convective activity, compared to ASCAT's 21 UTC timing that partially misses this critical window. Time series analysis revealed that ASCAT experiments produced premature precipitation onset by 1-2 hours across all focus regions, while SLAINTE experiments maintained timing closer to the nature run.

The 18 UTC observation proved critical for constraining MCS forecasts, capturing soil conditions during peak convective activity. Morning observations (6-12 UTC) provided equally important constraints, though the optimal morning timing could not be precisely determined due to minimal soil condition changes during these hours in the selected case study. Under these conditions, the assimilation of a single early morning observation instead of two may be preferable to avoid unnecessary noise introduction in the system. The cumulative effect of consistent sub-daily observations over the 3-day spin-up period proved

more important than individual observation impacts.

Overall, the 5 km SLAINTE experiment showed the most comprehensive improvements, with precipitation Structure, Amplitude and Location (SAL) metrics improving across all focus regions and convective core location metrics improving in all three sub-regions (Tables 3.1, 3.2). Cloud formation timing was better preserved, with precipitation onset matching the nature run compared to 1-2 hour anticipated precipitation observed in ASCAT experiments.

Several factors limit the generalizability of these results. Primarily, the decision to disable boundary condition updates created a highly idealized thermally-driven simulation that does not represent operational forecasting conditions. This approach, while isolating soil moisture effects, introduced unrealistic pressure gradients that significantly influenced convection initiation. In addition, results are based on a single MCS event over northern Nigeria on 2023-06-01. The specific characteristics of this case (eg. no morning precipitation, orographic lifting in the south-east) may not represent the full spectrum of MCS-triggering conditions across the Sahel.

It is therefore recommended to carry out additional simulations with more frequent boundary updates that better represent reality, perhaps by including observational datasets from satellites such as LST to the initial conditions and boundary conditions to generate more realistic simulations. As the interaction with the boundaries significantly influence the model output, if computational resources allow it it would be optimal to consider an even larger domain, possibly with a second nested domain where the event of interest takes place. It would also be valuable to carry out a wide range of experiments in different regions of the Sahel reproducing different case studies to account for the diverse range of conditions in which MCSs occur.

It is equally necessary to implement perturbation ensembles to quantify model sensitivity to soil moisture variations and provide robust uncertainty estimate for assimilation impacts. In such an assimilation setup, more sophisticated assimilation schemes could be tested which would preserve spatial heterogeneity to the model grid when incorporating coarser-resolution observation, rather than direct insertion methods used in this study.

To improve and constrain the definition of convective cores, more variables such as vertical wind velocity, CIN and CAPE could be integrated to differentiate those events that are not related to SSM but happen to occur in conditions that fit the definition. The 30 km radius in which precipitation events are considered could also be studied further, as it is a qualitative length that is not always befitting; for example, it does not consider the case in which a $\sim$60 km wide cloud system is displaced further than those 30 km and triggers secondary convection through uplift at the edge of cold pools.

Previous studies (Santanello Jr. et al., 2019, Kerr et al., 2010) showed that WRF is sensitive to variations of SSM initial conditions, but it seems that the model is less sensitive to modifications of SSM mid-run. A possible alternative could be to spin up the model for at least a day and use only the SSM output as initial conditions. This new SSM could then be coarsened and used to initialize the perturbed runs, and the simulation would be somewhat more realistic as there would be less elapsed time since the initialization of the boundaries. Another option could be to perturb not only the SSM but also other variables, such as soil

temperature.

Finally, it is important to note that the improvement of weather forecasts is futile if it is not accompanied by close collaboration with governmental and research centers of the region, which can combine the gained information with facilitation of timely and clear alerts to the local population.

Bibliography

Achugbu, I. C., Dudhia, J., Olufayo, A. A., Balogun, I. A., Adefisan, E. A., & Gbode, I. E. (2020). Assessment of WRF Land Surface Model Performance over West Africa. *Advances in Meteorology*, *2020*(1), 6205308. https://doi.org/10.1155/2020/6205308

Achugbu, I. C., Laux, P., Chen, L., Dudhia, J., Balogun, I. A., Arnault, J., Adeyewa, Z. D., Akintola, O. A., & Kunstmann, H. (2024). Performance evaluation of a high-resolution regional climate model in West Africa: Sensitivity to land surface schemes. *Theoretical and Applied Climatology*, *155*(4), 3099–3118. https://doi.org/10.1007/s00704-023-04800-x

Alessi, M. J., Herrera, D. A., Evans, C. P., DeGaetano, A. T., & Ault, T. R. (2022). Soil Moisture Conditions Determine Land-Atmosphere Coupling and Drought Risk in the Northeastern United States. *Journal of Geophysical Research: Atmospheres*, *127*(6), e2021JD034740. https://doi.org/10.1029/2021JD034740

Anderson, S. R., Cole, S. J., Klein, C., Taylor, C. M., Diop, C. A., & Kamara, M. (2024). Nowcasting convective activity for the Sahel: A simple probabilistic approach using real-time and historical satellite data on cloud-top temperature. *Quarterly Journal of the Royal Meteorological Society*, *150*(759), 597–617. https://doi.org/10.1002/qj.4607

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G. C., & Riemer, M. (2019). Quantitative View on the Processes Governing the Upscale Error Growth up to the Planetary Scale Using a Stochastic Convection Scheme. *Monthly Weather Review*, *147*(5), 1713–1731. https://doi.org/10.1175/MWR-D-18-0292.1

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., & Bittelli, M. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sensing of Environment*, *115*(12), 3390–3408. https://doi.org/10.1016/j.rse.2011.08.003

Brunelli, B., & Mancini, F. (2024). Comparative analysis of SAOCOM and Sentinel-1 data for surface soil moisture retrieval using a change detection method in a semiarid region (Douro River's basin, Spain). *International Journal of Applied Earth Observation and Geoinformation*, *129*. https://doi.org/10.1016/j.jag.2024.103874

Chen, F., & Avissar, R. (1994). Impact of Land-Surface Moisture Variability on Local Shallow Convective Cumulus and Precipitation in Large-Scale Models. *Journal of Applied Meteorology and Climatology*, *33*(12), 1382–1401. https://doi.org/10.1175/1520-0450(1994)033⟨1382:IOLSMV⟩2.0.CO;2

Di Baldassarre, G., Montanari, A., Lins, H., Koutsoyiannis, D., Brandimarte, L., & Blöschl, G. (2010). Flood fatalities in Africa: From diagnosis to mitigation. *Geophysical Research Letters*, *37*(22). https://doi.org/10.1029/2010GL045467

Drusch, M. (2007). Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. *Journal of Geophysical Research: Atmospheres*, *112*(D3). https://doi.org/10.1029/2006JD007478

Fan, D., Zhao, T., Jiang, X., García-García, A., Schmidt, T., Samaniego, L., Attinger, S., Wu, H., Jiang, Y., Shi, J., Fan, L., Tang, B.-H., Wagner, W., Dorigo, W., Gruber, A., Mattia, F., Balenzano, A., Brocca, L., Jagdhuber, T., … Peng, J. (2025). A Sentinel-1 SAR-based global 1-km resolution soil moisture data product: Algorithm and preliminary assessment. *Remote Sensing of Environment*, *318*, 114579. https://doi.org/10.1016/j.rse.2024.114579

FAO/UNESCO. (1971–1981). *Soil map of the world* [Legend and 9 volumes]. UNESCO.

Gantner, L., & Kalthoff, N. (2010). Sensitivity of a modelled life cycle of a mesoscale convective system to soil conditions over West Africa. *Quarterly Journal of the Royal Meteorological Society*, *136*(S1), 471–482. https://doi.org/10.1002/qj.425

Gao, H., Wood, E. F., Jackson, T. J., Drusch, M., & Bindlish, R. (2006). Using TRMM/TMI to Retrieve Surface Soil Moisture over the Southern United States from 1998 to 2002. *Journal of Hydrometeorology*, *7*(1), 23–38. https://doi.org/10.1175/JHM473.1

Garcia-Carreras, L., Parker, D. J., & Marsham, J. H. (2011). What is the Mechanism for the Modification of Convective Cloud Distributions by Land Surface–Induced Flows? *Journal of the Atmospheric Sciences*, *68*(3), 619–634. https://doi.org/10.1175/2010JAS3604.1

Graf, M., Arnault, J., Fersch, B., & Kunstmann, H. (2021). Is the soil moisture precipitation feedback enhanced by heterogeneity and dry soils? A comparative study [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.14332]. *Hydrological Processes*, *35*(9), e14332. https://doi.org/10.1002/hyp.14332

H SAF. (2021). ASCAT surface soil moisture climate data record v7 12.5 km sampling - Metop [Dataset]. https://doi.org/10.15770/EUM_SAF_H_0009

He, A. C., Valayamkunnath, A. P., Barlage, A. M., Chen, A. F., Gochis, A. D., Cabell, A. R. S., Schneider, A. T., Rasmussen, A. R. M., Niu, A. G., Yang, A. Z., Niyogi, A. D., & Ek, A. M. B. (n.d.). The Community Noah-MP Land Surface Modeling System Technical Description Version 5.0. Retrieved September 6, 2025, from https://opensky.ucar.edu/islandora/object/%3A3912

Hengl, T., Jesus, J. M. d., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLOS ONE*, *9*(8), e105992. https://doi.org/10.1371/journal.pone.0105992

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., & Stocker, E. F. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales [Section: Journal of Hydrometeorology]. https://doi.org/10.1175/JHM560.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13). https://doi.org/10.1029/2008JD009944

Janjic, Z. I. (1996). The surface layer in the NCEP Eta model. *Eleventh Conference on Numerical Weather Prediction*, 354–355.

Janjic, Z. I. (2002). *Nonsingular implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model* (NCEP Office Note No. 437) (Available from NCEP). National Centers for Environmental Prediction. Camp Springs, MD.

Janjić, Z. I. (1994). The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Monthly Weather Review*, *122*(5), 927–945. https://doi.org/10.1175/1520-0493(1994)122⟨0927:TSMECM⟩2.0.CO;2

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S. E., Drinkwater, M. R., Hahne, A., Martín-Neira, M., & Mecklenburg, S. (2010). The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle. *Proceedings of the IEEE*, *98*(5), 666–687. https://doi.org/10.1109/JPROC.2010.2043032

Klein, C., Belušić, D., & Taylor, C. M. (2018). Wavelet Scale Analysis of Mesoscale Convective Systems for Detecting Deep Convection From Infrared Imagery. *Journal of Geophysical Research: Atmospheres*, *123*(6), 3035–3050. https://doi.org/10.1002/2017JD027432

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., … GLACE Team. (2004). Regions of strong coupling between soil moisture and precipitation. *Science (New York, N.Y.)*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Lin, L.-F., Ebtehaj, A. M., Flores, A. N., Bastola, S., & Bras, R. L. (2017). Combined Assimilation of Satellite Precipitation and Soil Moisture: A Case Study Using TRMM and SMOS Data [Section: Monthly Weather Review]. https://doi.org/10.1175/MWR-D-17-0125.1

Liu, W., Zhang, Q., Li, C., Xu, L., & Xiao, W. (2022). The influence of soil moisture on convective activity: A review. *Theoretical and Applied Climatology*, *149*(1), 221–232. https://doi.org/10.1007/s00704-022-04046-z

Liu, X., Xu, J., Zhou, X., Wang, W., & Yang, S. (2020). Evaporative fraction and its application in estimating daily evapotranspiration of water-saving irrigated rice field. *Journal of Hydrology*, *584*, 124317. https://doi.org/10.1016/j.jhydrol.2019.124317

Liu, Z., Ostrenga, D., Teng, W., & Kempler, S. (2012). Tropical Rainfall Measuring Mission (TRMM) Precipitation Data and Services for Research and Applications. *Bulletin of the American Meteorological Society*, *93*(9), 1317–1325. https://doi.org/10.1175/BAMS-D-11-00152.1

Matar, J., Sanjuan-Ferrer, M. J., Rodriguez-Cassola, M., Steele-Dunne, S., & De Zan, F. (2024). A Concept for an Interferometric SAR Mission with Sub-daily Revisit. *EUSAR 2024; 15th European Conference on Synthetic Aperture Radar*, 18–22.

Mathon, V., Laurent, H., & Lebel, T. (2002). Mesoscale Convective System Rainfall in the Sahel. *Journal of Applied Meteorology*, *41*(11), 1081–1092. https://doi.org/10.1175/1520-0450(2002)041⟨1081:MCSRIT⟩2.0.CO;2

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Monin, A. S., & Obukhov, A. M. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere [Originally published in Russian]. *Contributions of the Geophysical Institute, Academy of Sciences of the USSR*, *151*, 163–187.

Nakanishi, M., & Niino, H. (2006). An Improved Mellor–Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog. *Boundary-Layer Meteorology*, *119*(2), 397–407. https://doi.org/10.1007/s10546-005-9030-8

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*, D12109. https://doi.org/10.1029/2010JD015139

Olson, J. B., Smirnova, T., Kenyon, J. S., Turner, D. D., Brown, J. M., Zheng, W., & Green, B. W. (2021). A Description of the MYNN Surface-Layer Scheme. https://repository.library.noaa.gov/view/noaa/30605

Oyegbile, O., Chan, A., Ooi, M., Anwar, P., Mohamed, A. A., & Li, L. (2024). Evaluation of WRF model performance with different microphysics schemes for extreme rainfall prediction in Lagos, Nigeria: Implications for urban flood risk management. *Bulletin of Atmospheric Science and Technology*, *5*(1), 19. https://doi.org/10.1007/s42865-024-00081-y

Parker, D. J., Burton, R. R., Diongue-Niang, A., Ellis, R. J., Felton, M., Taylor, C. M., Thorncroft, C. D., Bessemoulin, P., & Tompkins, A. M. (2005). The diurnal cycle of the West African monsoon circulation. *Quarterly Journal of the Royal Meteorological Society*, *131*(611), 2839–2860. https://doi.org/10.1256/qj.04.52

Pedruzzi, R., Andreão, W. L., Baek, B. H., Hudke, A. P., Glotfelty, T. W., Dias de Freitas, E., Martins, J. A., Bowden, J. H., Pinto, J. A., Alonso, M. F., & Toledo de Almeida Abuquerque, T. (2022). Update of land use/land cover and soil texture for Brazil:

Impact on WRF modeling results over São Paulo. *Atmospheric Environment*, *268*, 118760. https://doi.org/10.1016/j.atmosenv.2021.118760

Pielke Sr., R. A. (2001). Influence of the spatial distribution of vegetation and soils on the prediction of cumulus Convective rainfall. *Reviews of Geophysics*, *39*(2), 151–177. https://doi.org/10.1029/1999RG000072

Pierdicca, N., Davidson, M., Chini, M., Dierking, W., Djavidnia, S., Haarpaintner, J., Hajduch, G., Laurin, G. V., Lavalle, M., López-Martínez, C., Nagler, T., & Su, B. (2019). The Copernicus L-band SAR mission ROSE-L (Radar Observing System for Europe) (Conference Presentation). In F. Bovenga, C. Notarnicola, N. Pierdicca, & E. Santi (Eds.), *Active and passive microwave remote sensing for environmental monitoring iii* (111540E, Vol. 11154). SPIE. https://doi.org/10.1117/12.2534743

Rötzer, K., Montzka, C., Bogena, H., Wagner, W., Kerr, Y. H., Kidd, R., & Vereecken, H. (2014). Catchment scale validation of SMOS and ASCAT soil moisture products using hydrological modeling and temporal stability analysis. *Journal of Hydrology*, *519*, 934–946. https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.07.065

Santanello Jr., J., Lawston, P., Kumar, S., & Dennis, E. (2019). Understanding the Impacts of Soil Moisture Initial Conditions on NWP in the Context of Land–Atmosphere Coupling. *Journal of Hydrometeorology*, *20*(5), 793–819. https://doi.org/10.1175/JHM-D-18-0186.1

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., & Huang, X.-Y. (2021). A Description of the Advanced Research WRF Model Version 4. *National Center for Atmospheric Research (NCAR), Boulder, USA*. https://doi.org/10.5065/1dfh-6p97

Soil Survey Staff. (1999). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys* (2nd). Natural Resources Conservation Service, U.S. Department of Agriculture. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/class/taxonomy/

Srivastava, P. K., Pandey, V., Suman, S., Gupta, M., & Islam, T. (2016). Available Data Sets and Satellites for Terrestrial Soil Moisture Estimation. In *Satellite soil moisture retrieval* (pp. 29–44). Elsevier. https://doi.org/10.1016/B978-0-12-803388-3.00002-4

Steele-Dunne, S., Basto, A., de Zan, F., Dorigo, W., Lhermitte, S., Massari, C., Matar, J., Milodowski, D., Miralle, D., Monteith, A., Rodriguez Cassola, M., Taylor, C., Tebaldin, S., & Ulander, L. (2024). SLAINTE: A SAR mission concept for sub-daily microwave remote sensing of vegetation. *EUSAR 2024; 15th European Conference on Synthetic Aperture Radar*, 870–872.

Taylor, C. M., Klein, C., & Harris, B. L. (2024). Multiday Soil Moisture Persistence and Atmospheric Predictability Resulting From Sahelian Mesoscale Convective Systems. *Geophysical Research Letters*, *51*(20), e2024GL109709. https://doi.org/10.1029/2024GL109709

Taylor, C. M., Belušić, D., Guichard, F., Parker, D. J., Vischel, T., Bock, O., Harris, P. P., Janicot, S., Klein, C., & Panthou, G. (2017). Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. *Nature*, *544*(7651), 475–478. https://doi.org/10.1038/nature22069

Taylor, C. M., de Jeu, R. A. M., Guichard, F., Harris, P. P., & Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, *489*(7416), 423–426. https://doi.org/10.1038/nature11377

Taylor, C. M., Gounou, A., Guichard, F., Harris, P. P., Ellis, R. J., Couvreux, F., & De Kauwe, M. (2011). Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns. *Nature Geoscience*, *4*(7), 430–433. https://doi.org/10.1038/ngeo1173

Taylor, C. M., Klein, C., Dione, C., Parker, D. J., Marsham, J., Diop, C. A., Fletcher, J., Chaibou, A. A. S., Nafissa, D. B., Semeena, V. S., Cole, S. J., & Anderson, S. R. (2022). Nowcasting tracks of severe convective storms in West Africa from observations of land surface state. *Environmental Research Letters*, *17*(3), 34016. https://doi.org/10.1088/1748-9326/ac536d

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization [Section: Monthly Weather Review]. https://doi.org/10.1175/2008MWR2387.1

Tropical Rainfall Measuring Mission (TRMM). (2011, July). *Trmm precipitation radar rainfall rate and profile l2 1.5 hours v7* [Digital Science Data]. Greenbelt, MD, Goddard Earth Sciences Data; Information Services Center (GES DISC). https://disc.gsfc.nasa.gov/datacollection/TRMM_2A25_7.html

U.S. Geological Survey. (2010). Global multi-resolution terrain elevation data 2010 (GMTED2010) [Digital elevation dataset]. https://doi.org/10.5066/F7J38R2N

Vizy, E. K., & Cook, K. H. (2023). Extreme rainfall events in the West African Sahel: Understanding storm development over the Damergou gap using convection-permitting simulations in the Weather Research and Forecasting model. *Quarterly Journal of the Royal Meteorological Society*, *149*(752), 959–983. https://doi.org/10.1002/qj.4443

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., & Gneiting, T. (2018). Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall over Northern Tropical Africa. *Weather and Forecasting*, *33*(2), 369–388. https://doi.org/10.1175/WAF-D-17-0127.1

Wagner, W., Lemoine, G., Borgeaud, M., & Rott, H. (1999). A study of vegetation cover effects on ers scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2), 938–948. https://doi.org/10.1109/36.752212

Wagner, W., Noll, J., et al. (1999). Monitoring soil moisture over the Canadian Prairies with the ERS scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(1), 206–216. https://doi.org/10.1109/36.739154

Wagner, W., Lemoine, G., & Rott, H. (1999). A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data. *Remote Sensing of Environment*, *70*(2), 191–207. https://doi.org/https://doi.org/10.1016/S0034-4257(99)00036-X

Wanders, N., Karssenberg, D., Bierkens, M., Parinussa, R., de Jeu, R., van Dam, J., & de Jong, S. (2012). Observation uncertainty of satellite soil moisture products determined with physically-based modeling. *Remote Sensing of Environment*, *127*, 341–356. https://doi.org/10.1016/j.rse.2012.09.004

Wernli, H., Hofmann, C., & Zimmer, M. (2009). Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique [Section: Weather and Forecasting]. https://doi.org/10.1175/2009WAF2222271.1

WMO OSCAR Details for Variable: Soil moisture at surface. (2025). https://space.oscar.wmo.int/variables/view/soil_moisture_at_surface

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research: Atmospheres*, *116*(D12). https://doi.org/10.1029/2010JD015140

Zhou, X., Wang, J., Shan, B., He, Y., & Xing, M. (2025). Sensitivity of multi-frequency and multi-polarization SAR to soil moisture at different depths in agricultural regions [Type: Article]. *Journal of Hydrology*, *660*. https://doi.org/10.1016/j.jhydrol.2025.133513

Zipser, E. J., Cecil, D. J., Liu, C., Nesbitt, S. W., & Yorty, D. P. (2006). WHERE ARE THE MOST INTENSE THUNDERSTORMS ON EARTH? *Bulletin of the American Meteorological Society*, *87*(8), 1057–1072. https://doi.org/10.1175/BAMS-87-8-1057

Appendix A

Additional Tables and Figures

A.1 Tested WRF parameterizations

Schemes/Trials	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7
Microphysics	Thompson	Thompson	WSM6	Thompson	Thompson	Thompson	Thompson
Longwave	RRTM	RRTM	RRTM	RRTM	RRTM	RRTM	RRTMG
Shortwave	Dudhia	Dudhia	Dudhia	Dudhia	RRTMG	RRTMG	RRTMG
Surface layer	Monin-Obukhov	Monin-Obukhov	Monin-Obukhov	Monin-Obukhov	Monin-Obukhov	Monin-Obukhov	Monin-Obukhov
Land surface	Noah-MP	Noah-MP	Noah-MP	Noah-MP	Noah-MP	Noah-MP	Noah-MP
PBL	MYNN2.5	MYNN2.5	MYNN2.5	MYNN2.5	MYNN2.5	YSU	MYNN2.5
Cumulus	none	none	none	none	none	none	none
dveg option	4	2	4	4	2	3	2

Table A.1: Tested WRF parameterizations of which the performance was considered before selecting the parameterization combination in table 2.1.

A.2 Full time evolution of cloud top temperature

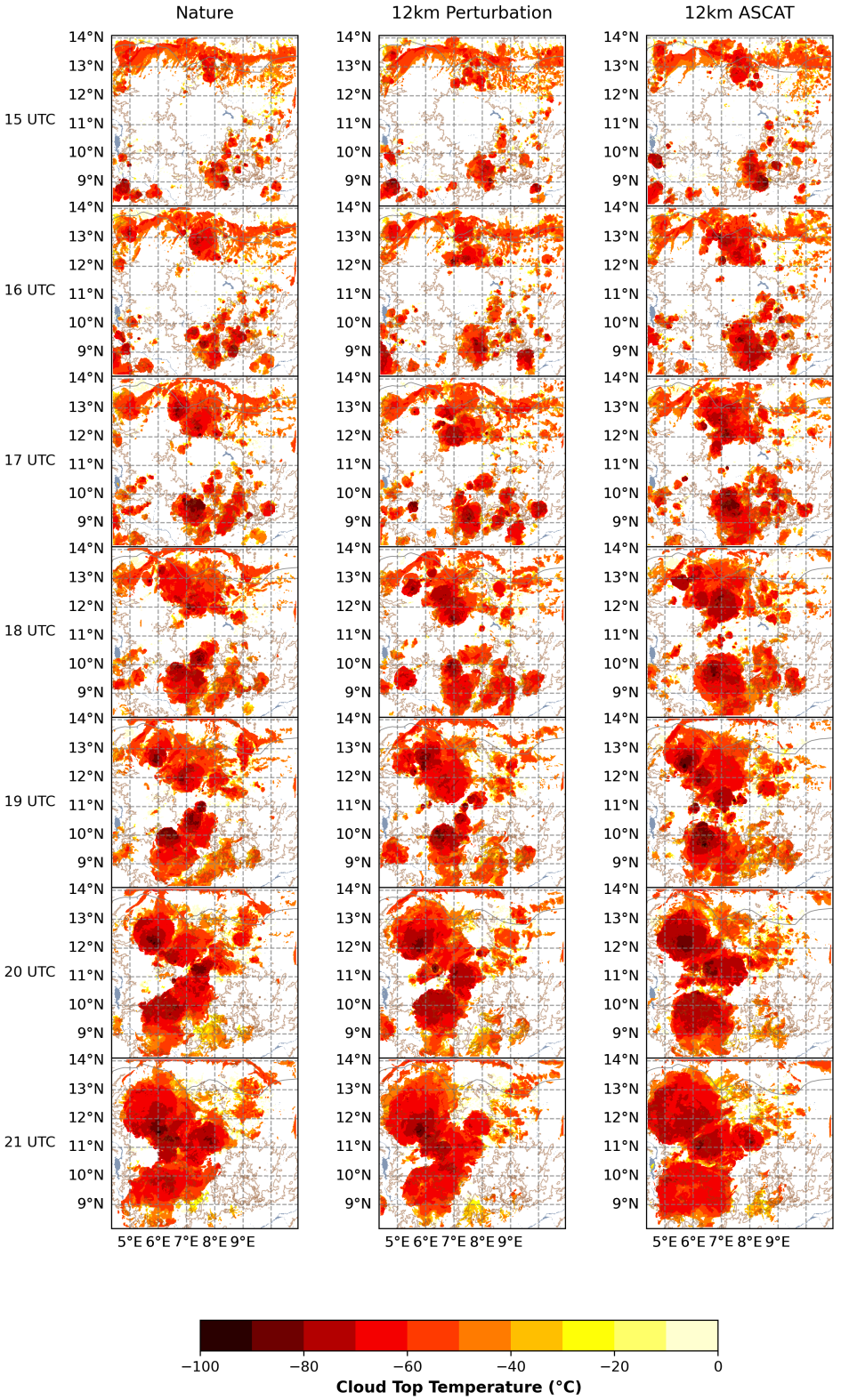


Figure A.1: Cloud top temperature comparison for the Nature, 12 km perturbation and 12 km ASCAT on 2023-06-01, showing the cloud evolution and the variation of cloud formation and cloud depth across the different runs.

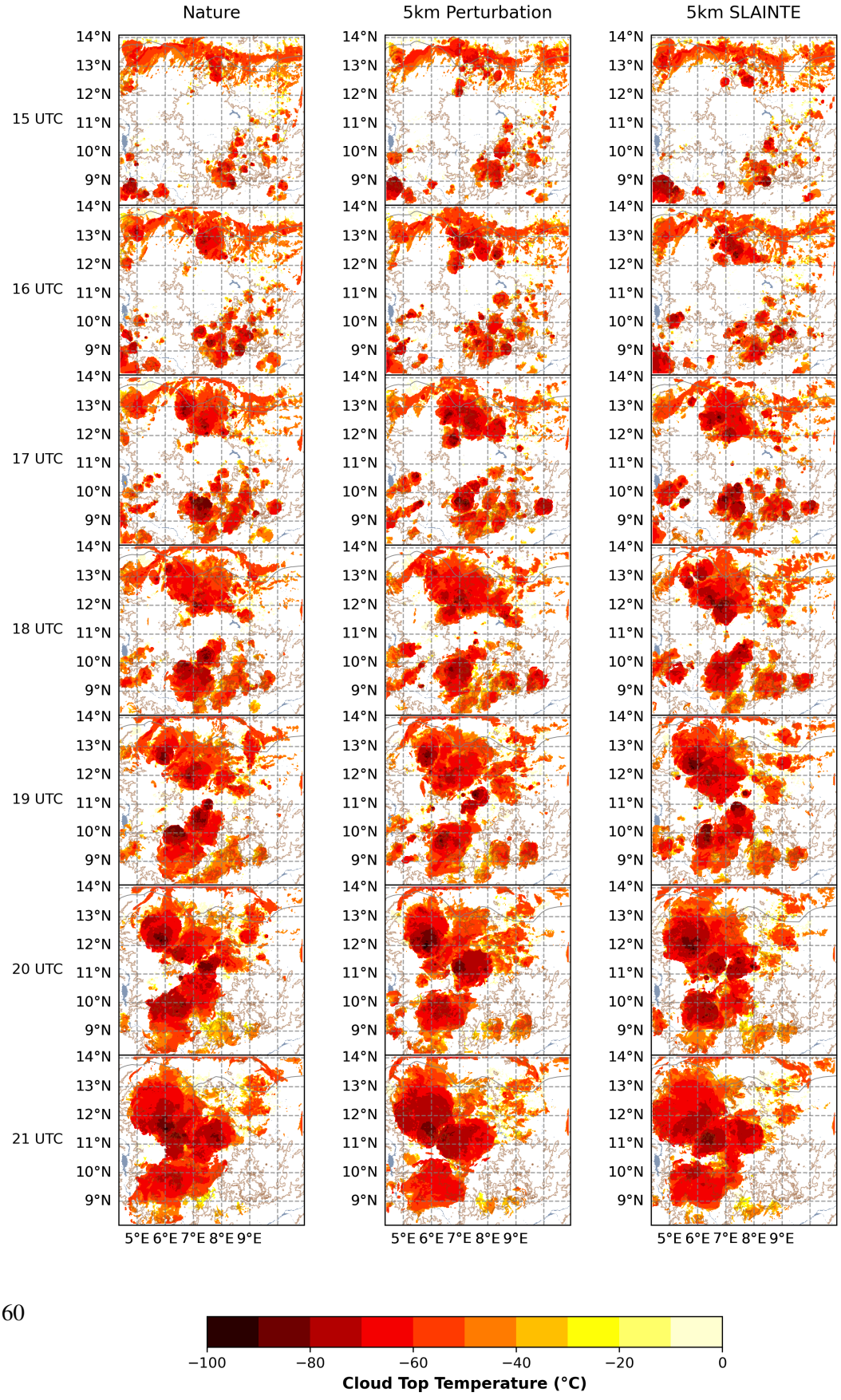


Figure A.2: Cloud top temperature comparison for the Nature, 5 km perturbation and 5 km SLAINTE on 2023-06-01, showing the cloud evolution and the variation of cloud formation and cloud depth across the different runs.

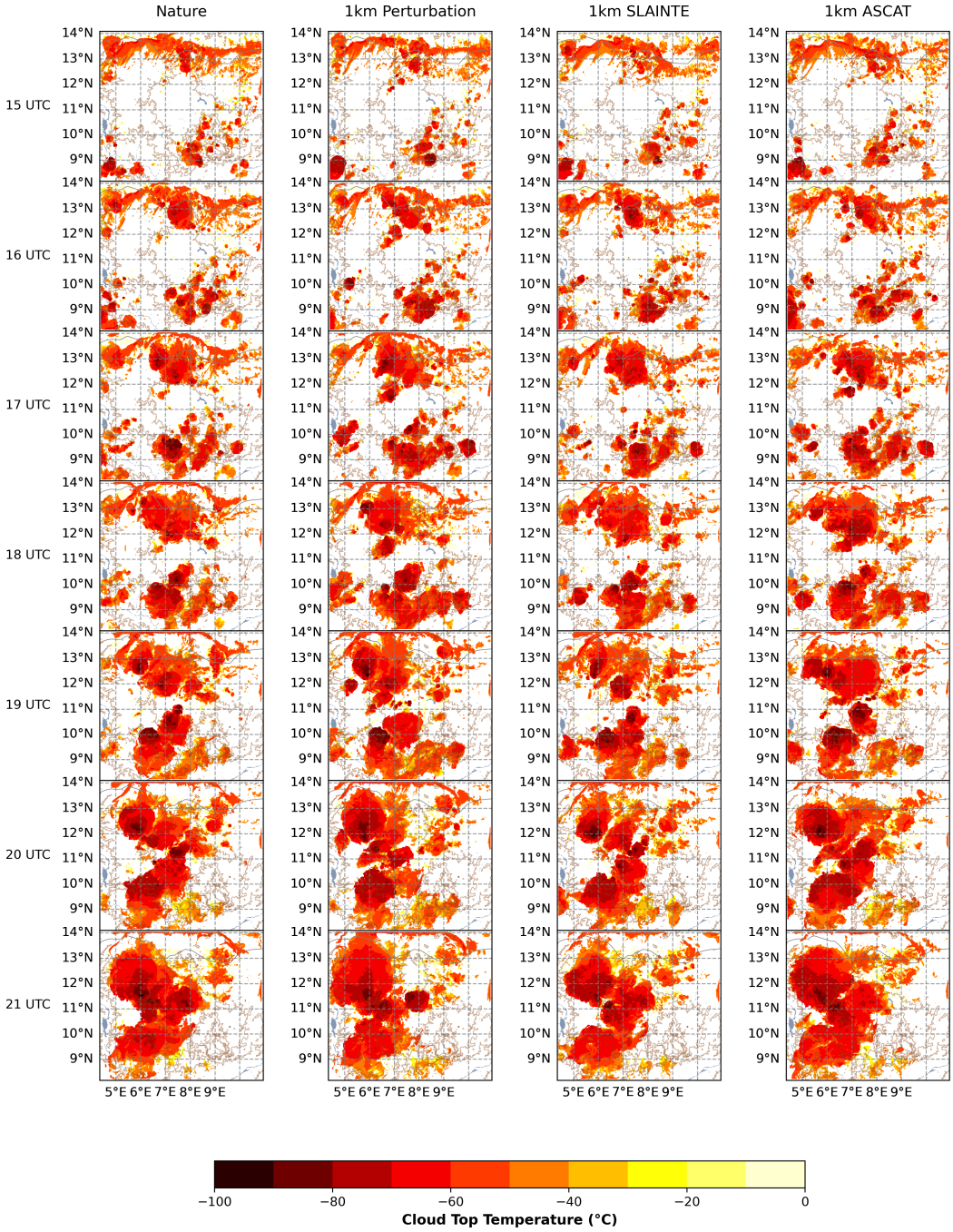


Figure A.3: Cloud top temperature comparison for the Nature, 1 km perturbation, 1 km SLAINTE and 1 km ASCAT runs on 2023-06-01, showing the cloud evolution and the variation of cloud formation and cloud depth across the different runs.

A.3 Time series of accumulated precipitation

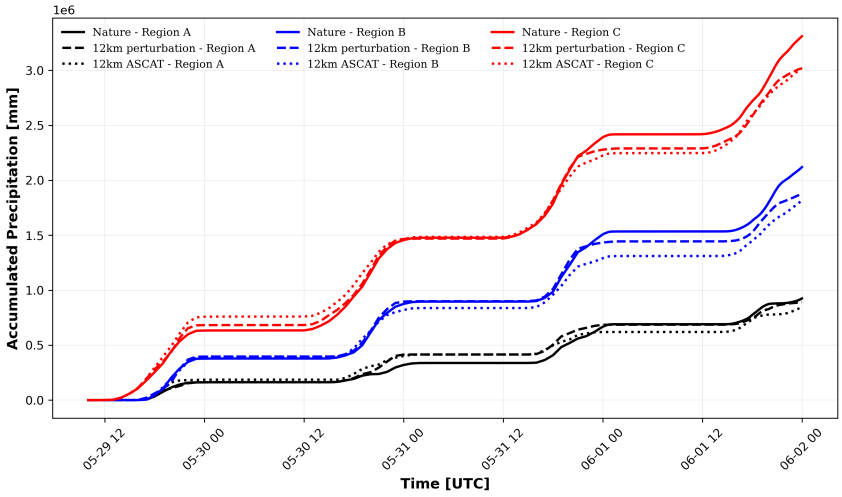


Figure A.4: Time series of total accumulated precipitation during the entire simulation period (2023-05-29 00 to 2023-06-02 00) in **region A**, **region B** and **region C** for (a) the nature (solid line), 12 km perturbation (dashed line) and 12 km ASCAT (dotted line) runs

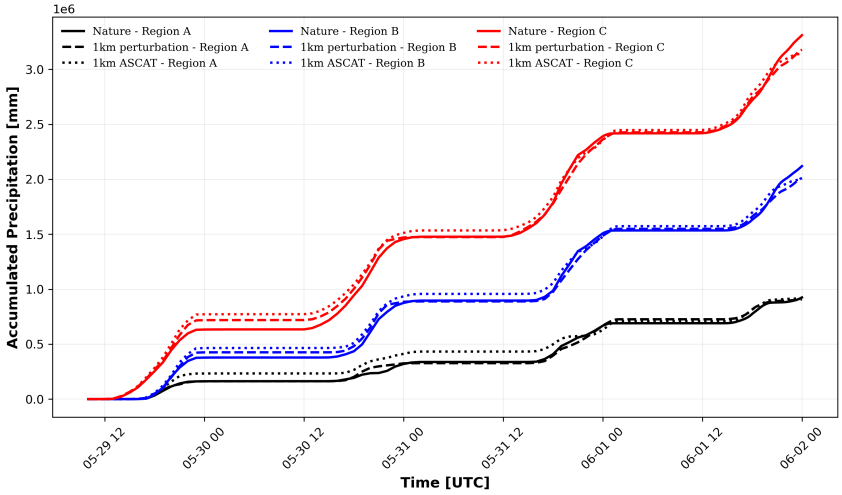


Figure A.5: Time series of total accumulated precipitation during the entire simulation period (2023-05-29 00 to 2023-06-02 00) in **region A**, **region B** and **region C** for (a) the nature (solid line), 1 km perturbation (dashed line) and 1 km ASCAT (dotted line) runs