# Exploring the potential of using remote sensing data to model agricultural systems in data-limited areas

**Luleka Dlamini**

(DLMLUL001)

A dissertation submitted towards the degree of

Master of Science in Environmental and Geographical Science

Faculty of Science

University of Cape Town

Supervisor: Dr Olivier Crespo

December 2019

## *Table of contents*

## *Plagiarism declaration*

1. I recognize that plagiarism is immoral and against the University of Cape Town's (UCT) rules and regulations.
2. I am familiar with UCT's document to avoid plagiarism and I have utilized it to guarantee that I refrain from all kinds of plagiarism methods.
3. For both citation and referencing, I have used the UCT Harvard style. I have used quotation marks where the actual words of others have been borrowed. I have also ensured that all quotations and contributions from other work(s) have been properly cited and referenced.
4. I have not and won't enable anybody to duplicate my work.
5. I therefore, confirm that this dissertation is my own work

Name: Luleka Dlamini

Student No: DLMLUL001

**Signature** Signed by candidate

**Date** ____19/12/2019________________________

## *Abstract*

Crop models (CMs) can be a key component in addressing issues of global food security as they can be used to monitor and improve crop production. Regardless of their wide utilization, the employment of these models, particularly in isolated and rural areas, is often limited by the lack of reliable input data. This data scarcity increases uncertainties in model outputs. Nevertheless, some of these uncertainties can be mitigated by integrating remotely sensed data into the CMs. As such, increasing efforts are being made globally to integrate remotely sensed data into CMs to improve their overall performance and use. However, very few such studies have been done in South Africa. Therefore, this research assesses how well a crop model assimilated with remotely sensed data compares with a model calibrated with actual ground data (Maize_control). Ultimately leading to improved local cropping systems knowledge and the capacity to use CMs.

As such, the study calibrated the DSSAT-CERES-Maize model using two generic soils (i.e. heavy clay soil and medium sandy soil) which were selected based on literature, to measure soil moisture from 1985 to 2015 in Bloemfontein. Using the data assimilation approach, the model's soil parameters were then adjusted based on remotely sensed soil moisture (SM) observations. The observed improvement was mainly assessed through the lens of SM simulations from the original generic set up to the final remotely sensed informed soil profile set up. The study also gave some measure of comparison with Maize_control and finally explored the impacts of this specific SM improvement on evapotranspiration (ET) and maize yield.

The result shows that when compared to the observed data, assimilating remotely sensed data with the model significantly improved the mean simulation of SM while maintaining the representation of its variability. The improved SM, as a result of assimilation of remotely sensed data, closely compares with the Maize_control in terms of mean but there was no improvement in terms of variability. Data assimilation also improved the mean and variability of ET simulation when compared that of Maize_control, but only with heavy clay soil. However, maize yield was not improved in comparison. This confirms that these outputs were influenced by other factors aside from SM or the soil profile parameters. It was concluded that remote sensing data can be used to bias correct model inputs, thus improve certain model outputs.

## *Acknowledgments*

I wish to offer my genuine thanks and appreciation to the following people and organizations:

- Firstly, to my supervisor Dr. Oliver Crespo for his assistance, guidance, and expertise throughout this journey. I am utterly grateful for always availing himself to answer my million questions. I would have not made it this far without his constant motivation and encouragement. Thank you for having faith in my abilities.
- To the University of Cape Town, the Climate System Analysis Group (CSAG), and the South African National Space Agency (SANSA) thank you for the financial support.
- To Wiltrud Durand for sharing her DSSAT model calibration for Bloemfontein
- To the South African Weather Services (SAWS) thank you for providing climate data and the Department of Agriculture, Forestry, and Fisheries (DAFF) for providing observed maize yield data for the Bloemfontein district.
- To Philip Mukwena, and my fellow students Farirai, Temitope, Siyabusa, and Stefaan thank you for selflessly providing both technical and editorial assistance.
- Lastly, to my family and close friends thank you for your constant prayers, love, support, and encouragement you have shared throughout this journey.

## *List of acronyms and abbreviations*

| | |
|---|---|
| AMSR-E | Advanced Microwave Scanning Radiometer-EOS |
| APSIM | Agricultural Production Systems sIMulator |
| ASCAT | Advanced scatterometer |
| BD | Bulk density |
| BV | Bainsvlei |
| CGE | Crop growth engine |
| CL | Clay content |
| CMs | Crop models |
| CSM-CERES | Cropping System Model-Crop Environment Resource Synthesis |
| DAFF | Department of Agriculture, Forestry, and Fisheries |
| DBL | Depth of the base layer |
| DLL | Drainage lower limit |
| DSSAT | Decision Support System for Agrotechnology Transfer |
| DSSAT-CERES | Decision Support System for Agrotechnology Transfer-Crop Environment Resource Synthesis |
| DUL | Drained upper limit |
| ESA CCI | European Space Agency's Climate Change Initiative |
| ET | Evapotranspiration |
| GLAI | Green leaf area index |
| HClay | Heavy clay soil |
| Hu | Hutton |
| Ks.test | Kolmogorov-Smirnov Tests |
| LAI | Leaf area index |
| MBE | Mean bias error |
| MCRM | Markov Chain canopy Reflectance Model |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| OC | Organic carbon |
| PAW | Profile (or Plant) available water |
| PDF | Probability distribution function |
| POWER | Prediction of Worldwide Energy Resources |
| R | Pearson correlation |
| $R^2$ | Coefficient of determination |

| | |
|---|---|
| RMSE | Root Mean Square Error |
| ROC | Runoff curve |
| RS | Remote sensing |
| RS–P–YEC | Remote-Sensing–Photosynthesis–Yield Estimation for Crops |
| SA | South Africa |
| SADC | Southern African Development Community |
| Sandy | Medium sandy soil |
| SAT | Saturated upper limit |
| SAVI | Synthesized soil adjusted vegetation index |
| SAWS | South African Weather Services |
| SI | Silt content |
| SM | Soil moisture |
| SMAP | Soil Moisture Active Passive |
| SMOS | Soil Moisture and Ocean Salinity |
| SSE | Sum of squared differences |
| SSR | Regression sum of squares |
| SST | Total sum of squared deviations |
| SW | Swartland |
| UCT | University of Cape Town |
| USA | United State of America |
| WOFOST | World Food Studies |

## *List of figures*

## *List of tables*

# *Chapter one: Introduction*

The global populace is estimated to reach 9.15 billion by 2050 and concurrently food demand is expected to increase (Alexandratos and Bruinsma, 2012; Xu et al., 2016). This implies that food generation should increase by at least 60% above 2005/2007 production levels to ensure that human nutritional needs are met (Alexandratos and Bruinsma, 2012; Xu et al., 2016; Chivasa et al., 2017). Moreover, food production is further challenged by other contemporary issues including climate change, land use management, arable land degradation, or environmental responsibility (Mishra et al., 2013; Ray et al., 2013; Battude et al., 2016; Leroux et al., 2017). Therefore, a sustainable increase in food production is particularly urgent in countries whose economies are predominantly agriculture-based and where the population is expected to grow at an even higher rate than globally like it is the case for many African countries and rural South Africa (SA). Traditional methods of assessing agricultural productions (i.e. estimating crop yield) can be expensive, time-consuming and laborious (Fang et al., 2008; Leroux et al., 2017; He et al., 2018). Alternatively, numerical crop models can be used for the same purpose as they are relatively quicker and cheaper. As such, modelling and monitoring agricultural systems will be essential for agricultural management, economic growth and for tracking global food security in these countries

## 1.1.Crop modelling

Crop models (CMs) are one of only a few tools used to explore sustainable ways to increase production at various locations. These numerical tools, particularly process-based CMs, attempt to capture the relationship between the surrounding environmental conditions and the biophysical processes within a crop (Yuping et al., 2008; Zinyengere et al., 2015; Nagamani and Mariappan, 2017; He et al., 2018). By capturing these biophysical processes, commonly in daily time steps, the model then estimates crop growth, development, and crop yield. There are numerous CMs, with different complexity, that have been developed over the years (Rivington and Koo, 2010; Figure 1.1). Some CMs have the capacity to model only a single crop or represent in detail only a part of a crop production process, whereas others have been developed to simulate various crops in complex rotations under various weather conditions, crop management and soil characteristics (Rivington and Koo, 2010; Nagamani and Mariappan, 2017). The model's ability to capture crop biophysical processes and then simulate crop growth and yield depend heavily on the crop growth engine (CGE) used (Steduto, 2003; Bauböck, 2014). The CGEs of all CMs are based either on solar radiation (radiation use efficiency-RUE), carbon, or water productivity (Figure 1.2). The advantages and disadvantages of these engines, together with the models that use them are outlined in Table 1.1.


**Figure 1.1:** A word cloud of known crop models (Source: Rivington and Koo, 2010).

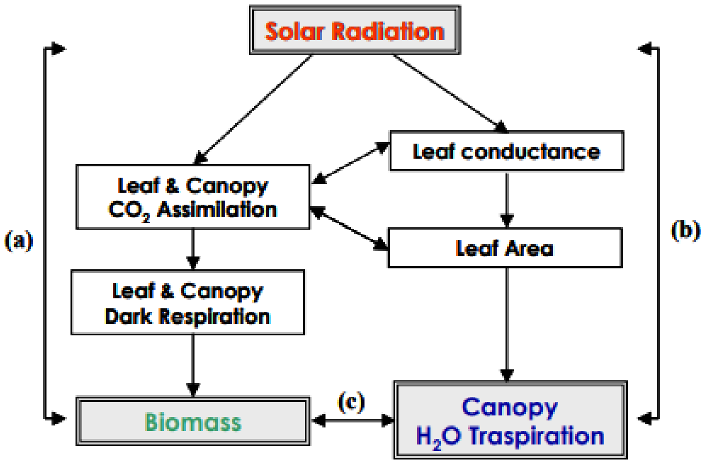


**Figure 1.2:** The underlying processes involved in biomass production (a) and canopy transpiration (b). Carbon-based and solar-based engines use pathway (a), whereas water productivity-based engine uses pathway (c) (Source: Steduto, 2003).

**Table 1.1:** The advantages and weaknesses of the three common CGEs (Adapted from Steduto, 2003; Bauböck, 2014).

| | Carbon-based | Solar-based | Water-based |
|---|---|---|---|
| **Advantages** | Excellent subdivision in hierarchical levels of the system organisation<br>Processes have sound physical and physiological basis | Robust RUE relationship<br>Constant under non-stressed conditions<br>Less complex, easy-to-drive values | Very robust (esp. under water-limited conditions)<br>Stable under water and salinity stress conditions |
| **Disadvantages** | Complex<br>Cultivar specific<br>Maintenance and growth respiration processes introduce uncertainties | Inconsistent variability amongst crops, location and years<br>Non-linear under stress conditions | Has difficulty in deriving actual transpiration<br>Has been less implemented in crop models due to difficulties in measuring actual canopy transpiration |
| **Crop models** | Wageningen models (BACROS, SUCROS, ARID CROP, WOFOST, MACROS, PAPRAN, SWAP, etc)<br>American CROPGRO (DSSAT) series<br>BioSTAR | CERES (DSSAT)<br>LINTUL<br>EPIC<br>STIC<br>CropSyst<br>APSIM | AquaCrop<br>CropSyst<br>CropWat<br>SWAMP<br>ACRU<br>SAPWAT<br>PUTU |

CMs can provide valuable scientific knowledge of certain interactions between crops and biophysical processes in order to improve our understanding of local cropping systems (Zinyengere et al., 2015). Therefore, they can be a key tool in addressing issues of global food security, including monitoring and prediction of agricultural droughts and its impacts; crop production (i.e. yield); precision agriculture; impacts of crop management; and agriculture water resources (Li et al., 2010; Nagamani and Mariappan, 2017; Kasampalis et al., 2018). However, CMs typically depend on accurate estimation of input parameters, which for several areas worldwide, including most South African rural areas, are commonly not available (Yuping et al., 2008; Mishra et al., 2013; Wang et al., 2014). For instance, in many African countries, data limitation (e.g. sparse meteorological data and diverse crop management strategies) is the main challenge that hinders accurate calibration and evaluation of these models (Zinyengere et al., 2015; Battude et al., 2016; Leroux et al., 2017; Kasampalis et al., 2018). If the data is available, it is often of insufficient quality and/or quantity to drive and produce acceptable simulations. This ultimately increases the uncertainties within the CMs results and thus impacts the confidence of the interpretation of the model outputs. Moreover, this lack of appropriate data makes the use of CMs rare in locations that would highly benefit from a deeper understanding and a wider exploration capacity, of local cropping and farming systems. Hence, there is a need to combine CMs with other tools to improve their performance and use especially in data-limited areas.

While those models are very efficient at simulating mainstream crops such as maize, rice, or wheat, they do not deal very well with underutilized crops such as groundnuts, black nightshade, cleome and grain sorghum (Mabhaudhi et al., 2016; Mabhaudhi et al., 2018). Concurrently while those models translate accurately large-scale crop management, for instance with irrigation systems or a motorized fertilisation application, they poorly represent typical smallholder farming systems management alternatives such as micro fertilisation, water harvesting or mixed cropping (Nagamani and Mariappan, 2017). This is often the result of developing the models while relying on existing large sets of data (as it is often not the case in many rural poorer locations) and the potential gain of mainstream crops under commercial farming system conditions. SA's unique profile of productive commercial agriculture and extended smallholder farming communities offers two opposite conditions, one which can be used as a known benchmark on the one hand, and the other which would highly benefit from the improved modelling capacity in areas with little field data.

Given the known limitations of those models, they still have a unique capacity to simulate conditions that are either too costly or risky to implement in the field or conditions that cannot be experimented physically. As such, they have been successfully used worldwide at various scales, for over four decades (Kasampalis et al., 2018). They have evolved from being used as a collection of knowledge into expert tools. For example, they are used to better understand cropping systems and their interactions with historical and future climate (Knox et al., 2010; Jones et al., 2015; Schulze and Durand, 2016), the impact of crop management (Li et al., 2010; García-Vila and Fereres, 2012; Basche et al., 2016), especially looking for sustainable ways to intensify production, as well as their interaction with other farm activities such as mixed crop-livestock systems (Gbetibouo et al., 2010; Descheemaeker et al., 2016; Descheemaeker et al., 2018) or accounting for household vulnerability (Ncube et al., 2016; Zougmoré et al., 2016; Hammond et al., 2017). The improved access to these models together with improved computing capacities has allowed for the success of fine-tuned large-scale studies such as AgMIP (Rosenzweig and Hillel, 2015), thus, has also democratized the use of CMs across various areas.

### 1.2. Remote sensing and crop models

Remote sensing (RS) can produce timely, promptly, relatively cheap (or free) and precise approximations of the earth's surface at various spatial and temporal resolutions (Fang et al., 2007; Battude et al., 2016; Leroux et al., 2017; Kasampalis et al., 2018). Due to the advancement in technology, RS data (including leaf area index, soil moisture, and many other optical vegetation indices) has become relatively easy to access and use. Moreover, it has been recorded for over three decades which allows for both short-term and long-term analysis. This has resulted in many innovative agricultural applications, including using biophysical processes to estimate and monitor crop growth, development, and yield (Kasampalis et al., 2018).

Numerous studies have recognized that RS data can provide even more valuable information when combined with CMs (Delecolle et al., 1992; Fang et al., 2007; Yaping et al., 2008; Inas et al., 2013; Leroux et al., 2017; Jin et al., 2018). Thus, improve crop growth and yields monitoring and estimation, even in the data-limited area. The scarce nature of data in remote rural areas limits the accuracy achievable by crop simulations and consequently produces low confidence outputs with limited value for agricultural advisors, with extensive local expertise which is often poorly represented in CMs. Therefore, RS data can directly or indirectly be used to fill the data-scarce gap (Mishra et al, 2013; Leroux et al., 2017; Jin et al., 2018). As such, improving the integration of CMs with RS can produce new scientific knowledge gained as well as potentially improve local capacity to better use CMs. For example, apart from improving simulation capacity, democratising CMs in areas where their use is not optimal could expose local experts to new relevant agricultural and technical concepts. Moreover, it can also reveal CM aspects that could be strengthened by using local expert knowledge, thus potentially improving the capacity of global CMs to better represent and be more relevant to local cropping systems. Hence, combining RS data with CMs can directly offer better calibration and validation potential of these modelling tools, which will lead to higher confidence in the model outputs, including the better representation of regional and local cropping system characteristics. Ultimately, increasing both the value of the model outputs for decision making and the confidence of the decision-makers, particularly in data-limited and remote areas, as it is the case for many South African farming areas.

Increasing efforts are being made globally to benefit from combining RM data with the simulation capacity of CM tools. The two most common methods of integrating RM into CMs are *forcing* or *recalibration* (also known as data assimilation) approach (Yaping et al, 2008; Mishra et al., 2013; Jin et al., 2018). The *forcing* approach involves substituting the CMs parameters with the RS observations, while these parameters are adjusted based on RS observations when the data assimilation approach is used. Yaping et al. (2008), Thorp et al. (2010), Mishra et al. (3013), Ines et al. (2013) or Yao et al. (2015) for instance have used various remotely sensed data such as leaf area index, soil moisture, and other vegetation indices, to infer agricultural information relating to productivity potential (e.g. water stress, evapotranspiration and mostly yield). Some efforts in that direction also exist in Africa, for example, with maize in Burkina Faso (Leroux et al., 2017), and maize water requirements in the Free State, SA (Moeletsi and Walker, 2012). There are also some interesting works that are associating remotely sensed data with either crop yield or water use in SA (Singels et al., 2014; Durand and Ferreira, 2017). Nevertheless, these studies are very few to fully understand the potential and constraints of integrating these tools, particularly in data-limited areas.

### 1.3. Aim and Objectives

Most studies combine remotely sensed data with CMs to evaluate and advance the models' capacity to accurately estimate crop yield. However, there is also a need for studies that can use remotely sensed data to advance poor model calibrations or improve the confidence in model outputs even when there is no data to necessarily evaluate the model performance. Therefore, the aim of this study is to assess how well the model assimilated with remotely sensed data exclusively compares with a model calibrated with sufficient ground data. Hence, improve the capacity and use of CMs in data-limited areas. For the purpose of this research, remotely sensed soil moisture (SM) will be integrated with the DSSAT-CERES-Maize model. The recalibration method will be utilized; therefore, the model soil parameters are adjusted based on remotely sensed SM observations. The key objectives of this study are to:

1. measure how well the DSSAT-CERES-Maize model calibrated with generic soils captures the temporal variation of SM between 1985 and 2015 in Bloemfontein.
2. develop a framework to and ultimately assimilate remotely sensed SM into the model
3. assess the improvement and ultimate SM simulation, in comparison to the control
4. evaluate how SM driven improvement impacted other key model outputs (i.e. evapotranspiration and maize yield)

Based on the aim and objectives of this study, and the assumption that remotely sensed data will offer a relatively good surface volumetric soil moisture measurements, the following research questions are formulated:

- Does RS data allow one to improve model representation (i.e. SM) and build confidence to an extent comparable to a model calibrated with traditional field data?
- Can remotely sensed data be integrated with CMs to potentially model agricultural systems in data-limited areas?

## 1.4. The structure of the thesis

This thesis is divided into five (5) chapters. After the introduction, *Chapter two* reviews the literature by initially highlighting the significance of maize production in SA. This is followed by the current theoretical knowledge and pragmatic evidence around CMs and RS, including the different approaches that can be used to combine these tools. The methods and data employed to accomplish the aim and objectives of this dissertation are described and discussed in *Chapter three*. *Chapter four* presents and discusses in-depth the results of this study, by particularly assessing the improvement of SM simulations from the original generic set up to the final RS informed soil profile set up, how this improved SM compares with Maize_control as well as the impacts on evapotranspiration and maize yield. This dissertation concludes with *Chapter five*, which summarises the key findings of this study, and discusses whether assimilating remotely sensed data with CMs improves model representation and build confidence to an extent comparable to a model calibrated with traditional field data. This chapter also acknowledges the limitations of the current study as well as briefly discuss future research avenues.

# *Chapter two: Literature review*

## 2.1. Introduction

This chapter firstly highlights the importance of maize production in South Africa (SA). The study then reviews key characteristics that affect maize production in the country. The advancement in technology combined with greater knowledge on modelled and observed agricultural systems has resulted in substantial development of crop models and remote sensing outputs. Therefore, this literature review then explores the current theoretical knowledge and pragmatic evidence around these tools. This includes the review of different types of crop models and the use of remotely sensed products to monitor agricultural systems. This chapter concludes by reviewing various methods that can be used to integrate remotely sensed data with crop models, including the advantages and limitations of such an integration.

## 2.2. Importance of maize production in South Africa

Maize is one of extremely significant grain grown and produced in southern Africa as it is both the main staple food and major animal feed. The major producer of maize (about 50%) within the Southern African Development Community (SADC) area is South Africa (SA) (Davis, 2006; Akpalu et al., 2008). Maize is the most widely produced field crop in SA and it remains the main source of carbohydrates for local people (FAO, 2018; DAFF, 2019). Moreover, it accounts for approximately 70% of grain production and covers approximately 60% of the cultivated land (Akpalu et al., 2008; DAFF, 2017). Over the past five seasons, maize remained the largest contributor to the South African field crop gross value (DAFF, 2019). The latest report on national agricultural production states that during the 2017/18 growing season, the maize sector produced a gross value of about R25 491 million, which was 14,4% lower than the R29 781 million obtained during the 2016/2017 season (DAFF, 2019). Furthermore, the maize sector provides direct employment to over 128 000 South Africans (DAFF, 2017). Therefore, maize production is a major driver of food inflation in the country and contributes significantly to the economy.

Maize is cultivated across all nine South African provinces. The major provinces that produce maize are Mpumalanga, North West, and Free State, as they together account for more than 81% of total maize production (Figure 2.1; Karuaihe, 2006; DAFF, 2019). The individual provincial contributions to maize production during the 2017/2018 season are illustrated in Figure 2.2. The maize sector consists of both commercial and non-commercial farmers. In SA, maize is mainly produced under rain-fed conditions (~90%) and less than 10% under irrigation (DAFF, 2019). The major commercial farmers (estimated area planted 2,319 million ha for

2017/2018 season) are found within the three major producing provinces, while the major non-commercial farmers are in the northern KwaZulu-Natal, Eastern Cape and Limpopo (DAFF, 2017; DAFF, 2019). SA produces mostly white maize (~53%) while the remaining percentage is yellow maize (~47%, DAFF, 2019). White maize is mainly produced for human consumption and yellow maize is primarily produced for animal feed. The current ratio of areas cultivated with maize is 55% white maize to 45% yellow maize (DAFF, 2019). Approximately 5.9% of white maize production in cultivated in irrigated areas and 94.1% in rain-fed drylands along the western part of the maize belt, whereas about 13.5% of yellow maize is irrigated and 86.5% is rain-fed along the eastern part of the maize belt (DAFF, 2017; DAFF, 2019).

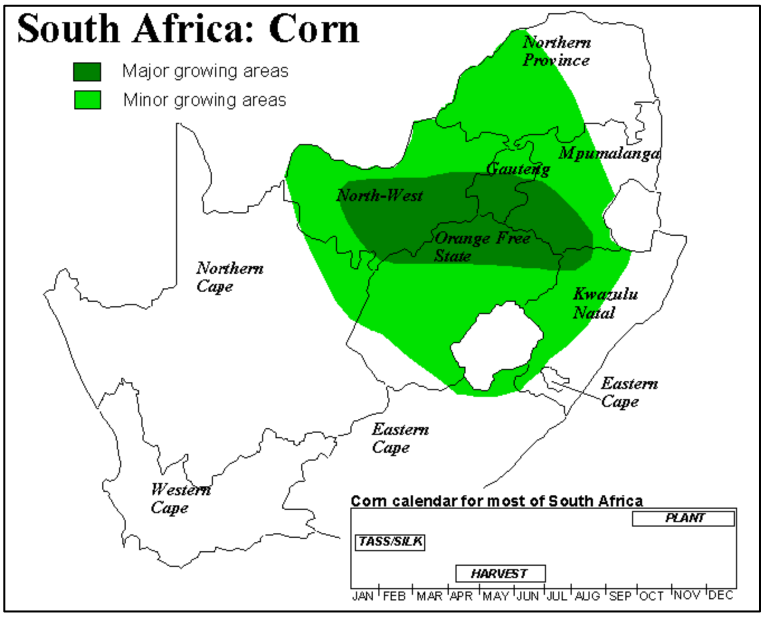


**Figure 2.1:** Major and minor maize growing areas in South Africa (Karuaihe et al., 2006).

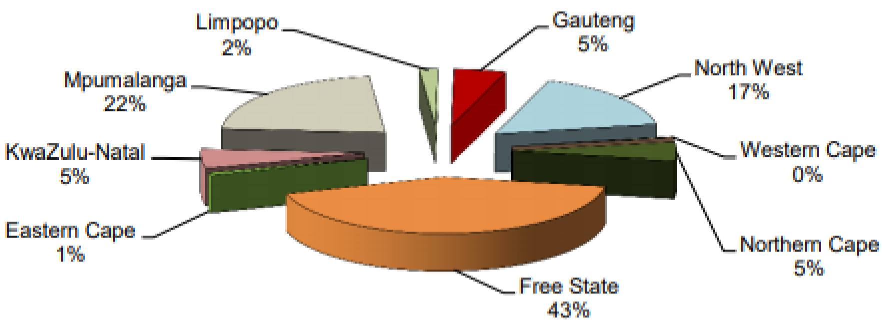


**Figure 2.2:** Provincial maize production during the 2017/2018 season (DAFF, 2019).

## 2.3. Key characteristics that affect maize production

Maize is cultivated under various climatic and soil characteristics in SA (du Plessis, 2003; Davis, 2006). Effective production of maize relies heavily on appropriate crop management strategies (e.g. correct production inputs and timing of activities) that will ensure and promote both agricultural production and environmental sustainability. The crop management factors that may affect maize yields range from planting date or period, seed quality, plant population density, weed management, plant and soil nutrition, and suitable cultivar (Sangio, 2001; Subedi and Ma, 2009; Beiragi et al., 2011). In the semi-arid area, mostly rain-fed areas, maize production is generally limited by multiple integrated conditions including infertile soils, erratic climate as well as poor environmental management (Mati, 2000; Du Plessis, 2003; Ramadoss et al., 2004). This study mainly discussed four characteristics that affect maize production, particularly in SA, namely climate, soil, planting period, and plant population density.

### *2.3.1. Climate*

There is a consensus amongst literature that changes in-, and unpredictability of- climate already have had and will further have- a negative effect on agricultural (i.e. maize) production (Mati, 2000; du Plessis, 2003; Molua and Lambi, 2006; DAFF, 2017). Consequently, an amplified increase in food insecurity within Southern African Development Community (SADC) region may be expected as a result of decreased availability and decreased access to agricultural production. In SA, maize is one of the main drivers of food inflation (Akpalu et al., 2008). Therefore, a decrease in maize production may potentially result in a rise in total revenue due to its inelastic demand. However, this would ultimately increase food insecurity within SA and the whole SADC region (Akpalu et al., 2008).

There are many climatic factors that affect maize production. However, in semi-arid areas, the two most important climatic factors that significantly affect maize production are daily temperature and seasonal rainfall conditions (Ramadoss et al., 2004). For example, most plant processes that have a substantial effect on crop development and productivity, highly depend on temperature (Molua and Lambi, 2006; Subedi and Ma, 2009). This is because optimum temperatures are required for photosynthesis, which enables growth to take place. As such, maize cannot be planted in regions with an average daily temperature that is lower than 19ºC or average monthly temperature that is lower than 23ºC during summer (du Plessis, 2003). On the other hand, relatively high temperatures (above 45ºC) may lead to shorter life cycle maize crop as well as lesser and lighter grains, hence leading to relatively low grain quality and yield

(Ramadoss et al., 2004; Molua and Lambi, 2006). Furthermore, low temperatures (below 4ºC) could also result in frost conditions that could be detrimental to all maize development phases. As such, 120-140 days without frost are needed to avoid any probable damages (du Plessis, 2003).

Maize production is also affected by the intensity, frequency, reliability, and distribution of rainfall that is received throughout the growing season. In non-optimal conditions, for instance, a minimum of 350-450 mm of annual rainfall is needed to harvest about 3 ton/ha of maize yield (du Plessis, 2003). A study in SA found that an observed 10% decrease in average seasonal rainfall can result in about a 4% decrease in average yield (Akpalu et al., 2008). Moreover, du Plessis (2003) argues that an average of 450-600 mm of water, particularly from soil moisture (SM) reserves, is needed throughout the maize growing season. The total length of the maize roots can exceed 2 meters, therefore, allowing the crop to tap into the SM reserves. However, even though the SM reserve can be full towards the beginning of the developing stages of maize, water stress coupled with high temperatures (drought) can significantly affect rain-fed maize production (Ramadoss et al., 2004).

#### *2.3.2. Soil*

Soils that are suitable for healthy maize cultivation need to have good drainage, appropriate chemical properties, optimal moisture regime, efficient depth as well as adequate amounts of crop nutrients (du Plessis, 2003). In SA, large-scale maize production is mainly cultivated in soils with clay content that is less than 10% (i.e. sandy soil) or more than 30% clay content (i.e. clay/clay-loam soils, du Plessis, 2003). However, maize can also be cultivated in soils with texture classes that range from 10 to 30% clay content as they have air and moisture regime that is optimum for healthy maize growth. For example, in some areas of Free State and North West rainfed maize is cultivated in soils with 5-20% clay content (Haarhoff et al., 2020). Moreover, maize yield can also be affected by the amount of soil water available throughout the growing season. Molua and Lambi (2006) expressed that during the growing season about 50% of the water should be available to the rooting zone, otherwise this may have a detrimental effect on the maize development stages and ultimately reduce yields. As such, soil properties at the root zone are critical because they influence the accessibility of nutrients and water (Molua and Lambi, 2006). Maize production can also be directly and indirectly affected by soil compaction and soil pH (which should be between 6.5 and 7.5), as they reduce the ability of the root's penetration, percolation and access water and nutrients (Motavalli et al., 2003; Lui et al., 2010).

### *2.3.3. Planting period*

Appropriate planting period is very critical for optimum maize development since it affects the timing and length of maize vegetative and reproductive phases (Beiragi et al., 2011). Generally, the planting of maize begins when groundwater and soil temperature are appropriate for optimum germination (DAFF, 2017). In SA, particularly, determining an appropriate planting window is of utmost importance to have successful rain-fed maize production (du Plessis, 2003). However, Amekudzi *et al.* (2015) expressed that maize farmers sometimes have difficulties in determining an appropriate planting period "that is neither too early nor too late". This is because, in a specific season, the planting window depends mostly on rainfall patterns (i.e. rainfall onset, cessation and duration of the rainy season). Generally, in SA, maize is planted around late spring to early summer, with ideal planting period being between November and December (Figure 2.1; du Plessis, 2003; DAFF, 2017; DAFF, 2019). However, sometimes the planting period can start as early as October and extend to January. Maize is then harvested from as early as April and up to the end of August (Figure 2.1).

Planting periods are mainly associated with prolonged climatic conditions of a respective area. However, regardless of the planting period, unforeseen weather conditions may occur just before/after planting or during the growing season of which may negatively affect maize production (Pswarayi and Vivek, 2007; Beiragi et al., 2011). As such, it is recommended that farmers utilize numerous planting periods with lengthy timeframes (i.e. October-January) to evade crop failure or unprofitable yields (Beiragi et al., 2011). Moreover, the planting schedule must be prepared in such a way that the heat and water sensitive phases of maize (i.e. flowering phase) do not fall well within the middle of the summer drought periods (du Plessis, 2003). Pswarayi and Vivek (2007) recommend that farmers use early maturing maize varieties as they can be planted across multiple dates. For example, such varieties can be planted relatively late during a growing season that is characterized by late rainfall onset or early cessation in order to reduce the impact of water stress (or drought).

### *2.3.4. Plant population density*

Maize production can also be influenced by plant population density (PPD). Generally, maize PPD depends on soil fertility, length of the growing season, water availability, time of planting, crop variety, and row spacing (Sangio, 2001; Azam et al., 2007). For example, an increase in PPD may result in low maize yield, particularly when there is high competition for limited natural resources (Sangio, 2001). High PPD, can particularly increase crop canopy, thus, reduce the amount of solar radiation required for optimum net photosynthesis (Azam et al., 2007). It also increases competition for limited nutrients and water resources. On the other hand, low PPD can promote the growth of weeds and can slow the tillering process of the crop, hence further reduce maize yield (Sangio, 2001; Azam et al., 2007).

In SA drylands where water is most limiting factor, PPD for maize crop ranges between 10 000 to 43 000 plants ha$^{-1}$ and this depends on seasonal rainfall and whether the region has a cooler, temperate, or warmer climate (du Plessis, 2003). For example, to potentially produce 4 ton/ha of maize in dryland areas with cooler, temperate or warmer climate, a PPD of 25 000, 16 000 and 14 000 per hectare are recommended, respectively (du Plessis, 2003). While areas of high rainfall or with irrigation tend to have high PPD per area. For example, to potentially produce 8-10 ton/ha maize yield in an irrigated area that has cooler, temperate or warmer climate, PPD of 55 000, 50 000 and 45 000 per hectare is recommended, respectively (du Plessis, 2003). Sangio (2001) suggested that the combination of both management and environmental factors together with reducing row width can result in optimum plant population density. Therefore, improving the potential to obtain maximum maize yield.

### 2.4. Crop models

Consistent and appropriate crop yield and growth predictions across various scales are critical for supporting decision-makers as they sustainably design and assign required agricultural resources (Fang et al., 2008; He et al., 2018). Agricultural survey (i.e. counting field quantities of standing crops) is a traditional and reliable approach to estimate regional crop yields (Lobell et al., 2003; Leroux et al., 2017; He et al., 2018). However, literature shows that such methods are laborious, long and expensive especially over large regional scales, and their results can only be obtained after the harvest period (Lobell et al., 2003; Moriondo et al., 2007; Leroux et al., 2017; He et al., 2018). Alternatively, numerical crop models (CMs), which are relatively quicker and cheaper, can be used to explore numerous probable experiments during a specific growing season.

CMs are also known as “crop yield models, crop growth models, or agricultural system models” (Kasampalis et al., 2018). Regardless of how researchers call CMs, there is a consensus that CMs partially represent a real-world agricultural system of interest (Nagamani and Mariappan, 2017; Kasampalis et al., 2018). In general, crop models computationally use mathematical equations to represent the relationship between crops and environments (Murthy, 2004; Kasampalis et al., 2018). Therefore, they can simulate “crop growth, development, and yield as a function of weather, soil conditions and crop management practices” (Nagamani and

Mariappan, 2017). Such information is important at various scales because it helps outline many agronomic practices and socio-economic choices that may potentially influence people's livelihoods. Therefore, crop models are very important as they can provide useful and decisive information to policymakers and land managers.

### *2.4.1. Types of crop models*

Numerous types of CMs have been created throughout the years and they can be organised into different groups (Murthy, 2011; Rauff and Bello, 2015). These models vary in structure, mathematical formulation, and degree of detail:

- *Empirical vs mechanistic models*

Empirical (statistical) crop models use a regression equation to express the relationship between the dependent and independent variables. For example, a specific variable (e.g. rainfall) can be altered to investigate its influence on crop yield. Statistical models were essentially intended to function at the interseasonal and regional scale, therefore, they are mostly suitable for investigating interannual variability of crop production at a regional scale (Hertel and Rosch, 2010; Zinyengere et al., 2014). Although these models are simple and require fewer input data to run, they can also provide useful insights about historical yield and how they were impacted by various factors (Nagamani and Mariappan, 2017). They can also be used to update other types of CMs. Nevertheless, empirical models are usually limited by their simplistic nature as they do not account for the interactions (e.g. soil-plant-atmosphere interactions) between the variables (Hertel and Rosch, 2010; Murthy, 2011; Nagamani and Mariappan, 2017). Another problem with the empirical models is that they cannot simulate yield response in areas whereby the historical data is not accessible and/or the environmental conditions are different from the area for which the model was developed (Nagamani and Mariappan, 2017; Kasampalis et al., 2018). Moreover, they may not be used to investigate future climate impacts on yield since future drivers of change may be different from those observed in the past and currently (Kasampalis et al., 2018).

On the contrary, the mechanistic models (also known as process-based models) do not only explain the relationship between the independent variable (e.g. temperature) and agricultural productivity (i.e. dependent variable) but are based on relevant physical, chemical and biological processes. Mechanistic models try to mimic these processes and numerically explain in what manner and why a certain response takes place (Murthy, 2011; Rauff and Bello, 2015). Such models typically explain the instantaneous rate of crop processes that quickly change over a short period of time (Nagamani and Mariappan, 2017). For instance, they can be used to explore crop transpiration processes because the variation of such processes is quicker at daytime as a result of the changes in temperature as well as radiation. Unlike, empirical models, mechanistic models are very complex and can describe the soil-plant-atmosphere system relationship in greater detail. They also need a relatively large amount of input data compared to empirical models, some of which are usually not obtainable (Rauff and Bello, 2015; Nagamani and Mariappan, 2017).

- *Deterministic vs stochastic models*

Deterministic models are that one that makes use of a defined coefficients/predictions to estimate a precise value of the dependent parameter, without associating them with a probability distribution, variance, or random variable (Murthy, 2011; Nagamani and Mariappan, 2017). Therefore, even though there may be errors in the observed data, these models will simulate the exact value for yield or any other dependent variable. On the other hand, stochastic models either use or attach a probability component to the individual outputs. Thus, for every input dataset, various outputs are specified together with their respective probabilities that specifies an expected average and variance (Rauff and Bello, 2015). As such, stochastic models estimate crop yield or any other dependent parameter at a specified rate. However, these models can easily become complex, thus very difficult to handle technically (Nagamani and Mariappan, 2017).

- *Dynamic vs static models*

The main difference between dynamic and static models is their capability to consider time as a factor. For example, dynamic models incorporate time as one of the variables in the equations. Therefore, they use differential equations to predict how crops respond to various ecological conditions over time (Nagamani and Mariappan, 2017; Kasampalis et al., 2018). In contrast to dynamic models, static models do not have time as the variable in the model. So over a given period, the value of both the dependent and independent parameters remains constant, regardless of whether the final output of the respective system has grown over a certain time period (Nagamani and Mariappan, 2017). Nevertheless, both these types of CMs can be used to explore historical and future conditions (i.e. changes in yield as a result of climate or management practice) (Zinyengere et al., 2014).

- *Simulation models*

Simulation models (also known as crop simulation models, CSMs) are defined as computer models that are based on differential mathematical equations that represent real cropping system processes that are linked through physical, physiological, and biochemical laws. Therefore, they provide an empirical systematic understanding of both internal and external processes (and their relations) that influence crop production (Rauff and Bello, 2015). Generally, such models predict crop development, growth, and yield as a result of weather, soil and management conditions (Nagamani and Mariappan, 2017). They generally use differential equations to predict the rate and state of different variables over time (e.g. as short as daily intervals), which is usually from the planting to maturity period.  Therefore, such models can be a combination of the above-mentioned crop model types. For example, DSSAT is a mechanistic, deterministic, dynamical field-based simulation model. CSMs enable researchers to evaluate multiple environmental constraints including the effects of droughts, crop management, and soil management (with or without climate changes or extreme events) on crop development and performance (Doraiswamy et al., 2004; Fang et al., 2008; Rauff and Bello, 2015). However, such models usually need extensive quantities of input data to sufficiently run and/or produce useful outputs. Moreover, these models simulate potential yield instead of actual yield and they also have difficulty capturing other processes such as pests, diseases, and weeds (Nagamani and Mariappan, 2017).

The advancement of computer technologies together with the increase in understanding of biophysical processes has led to a user-friendly development of various types of complex models. A summary of the development of some CMs as captured by Jin et al. (2018) is shown in figure 2.3. This timeline shows that DSSAT, WOFOST, MONICA, STICS, APSIM, AquaCrop, and DAISY CMs have improved the most to accurately simulate crop yield and crop growth. Although crop models can not completely represent the actual biophysical conditions during crop development, they still provide valuable scientific knowledge of certain crop biophysical processes, and particularly how these processes interact with each other. Thus, ultimately improving the understanding of the performance of cropping systems across many environmental constraints and crop management strategies.

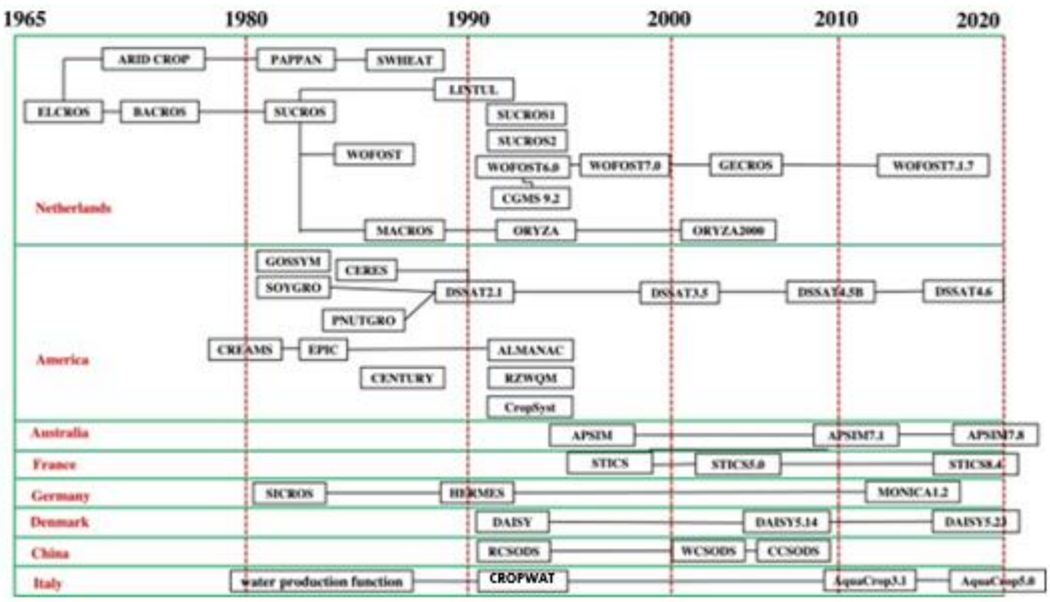


**Figure 2.3:** The advancement of the fundamental crop models over time (Jin et al., 2018).

### *2.4.2. Uses of crop models*

Researchers started using CMs in the 1960s and mainly to model the processes taking place at plant scale. According to Kasampalis et al. (2018), CMs can now be utilized as tools to inform important subjects, research as well as education, and technology transfer. For example, in crop management and decision making, CMs can be used as tools to obtain ideal planting dates and to explore the influence of weather-related risks on crop production (Murthy, 2004). Moreover, they can also be used to evaluate crop performance and suitability where crops have previously not been cultivated. Literature also shows that CMs have been utilized to create preparation maps for applying fertiliser (Werner et al., 2000). Such applications are particularly important in developing countries, like South Africa, as they are characterized by limited environmental and financial resources.

In general, researchers use CMs to study the relationship between the physiological crop processes and its surrounding environment (for instance, at daily intervals) to predict yield at maturity level (Asseng et al., 2013). For instance, Singels et al. (2010) reviewed the ability of major South African CMs, including SWAMP, PUTU, CANESIM, and SAPWAT, to model crop growth and crop water requirement in SA. While Schulze (2003) used ACRU to optimize maize planting dates across SA. Other researchers also use CMs to test plant breeding processes since these models include the interaction between crop development, environmental change, and genetic constraints. For example, In SA Singels (1992) used PUTU model to identify optimal wheat cultivar traits and planting strategies for various environments. Yin et al. (2004) updated the model coefficients to understand the role of crop physiology (i.e. quantitative trait loci in the DNA) in estimating gene-to-phenotype relationships across the various environmental condition. More recently, Messina et al. (2011) used the Agricultural Production Systems sIMulator (APSIM) crop model in the USA to create a yield-trait performance in order to improve the breeding cycle of maize varieties that are tolerant to droughts.

Climate change is amongst the contemporary issues that significantly threaten global food production. For example, an increase in this environmental change may consequently lower or increase the productivity of some crops. As such, CMs have been utilized to evaluate the potential effects of climate change on crop growth, development, and yield across various temporal and spatial scales, including SA (Lobell et al., 2006; Knox et al., 2010; Thornton et al., 2011; Beleste et al., 2015; Jones et al., 2015; Schulze and Durand, 2016). Considering the effects of environmental change on crops very early is important as it can assist farmers as well as decision-makers reduce the associated risks, thus improve food security. Nevertheless, there are relatively few studies that have utilized crop models in SA.

### *2.4.3. Limitations of crop models*

Modelling and monitoring crop production is an essential numerical step toward ensuring food security (Kasampalis et al., 2018). The fundamental advantage of CMs is that they can infer the temporal variation of crop growth, development, and yield beyond one experimental area (Nagamani and Mariappan, 2017). Nevertheless, CMs are just tools mostly used at research and policy level to understand various cross-scale crop production problems, so they have some limitations (Kasampalis et al., 2018). The fundamental constraint of CMs is linked with the quality and quantity of input data. Generally, the extent of input data needed depends on the CM used as well as the complexity of the system being modelled (Nagamani and Mariappan, 2017). For example, when working with a less complex system, an empirical model can be used because it is simple and requires less amount of input data (Kasampalis et al., 2018). As such, the accuracy of the outputs is sufficient. Whereas mechanistic models can better mimic complex systems, however, they need large amounts of data input (Nagamani and Mariappan, 2017). Unfortunately, some of these data are often either not available or very costly to acquire (Mishra et al., 2013, Kasampalis et al., 2018). This ultimately results in uncertainties in the simulated outputs. Although empirical models can be used as a backup in cases whereby data is hard to get, they are enormously limited in terms of simulating complex systems (e.g. soil-plant-atmosphere interactions) and information produced (Nagamani and Mariappan, 2017;

Kasampalis et al., 2018). Thus, ultimately producing outputs which are less useful. Moreover, incorrect model validation might result from relating the CMs outputs with the ones from more generic models (Di Paola et al., 2016). Therefore, researchers that use these models should ensure that they always adhere to the model's initial conditions that were used to test and develop them (Bregaglio et al., 2015; Kasampalis et al., 2018).

Furthermore, CMs are limited by the shortage of spatial information, particularly at the regional scale. For example, after comparing eight different models for wheat yield Palosuo et al. (2011) found that field-based CMs cannot be used in regional-scale studies unless appropriate parameterization is maintained, or important input factors of the model are ignored. Crop models are also enormously limited in terms of the information they produce. For example, CMs simulate potential yield instead of actual yield. Generally, CMs are efficient at simulating mainstream crops and large-scale crop management (Nagamani and Mariappan, 2017). However, comparatively, they poorly represent uncommon crops and crop management alternatives used by smallholder farmers. Moreover, CMs are also limited in biophysical factors that they can account for (Nagamani and Mariappan, 2017; Mabhaudhi et al., 2018). For example, many CMs do not include essential biophysical factors like tillage, insects, diseases, weeds, and certain nutrients (e.g. phosphorus). Regardless of these known limitations, CMs still have a unique capacity to simulate crop growth, development, and yield under various conditions. Thus, they can also provide valuable scientific knowledge of diverse small-holder farming systems, particularly in remote and rural areas**.**

### 2.5. Remote sensing and agricultural monitoring

Remote sensing (RS) has been recognized as a valuable tool for various agricultural applications. As such, they have been used to monitor crop development and estimate yield production across various scales. This may be attributed to RS multispectral range, relatively low cost, quick and recurring nature in coverage as well as fine spatial and temporal resolution (Doraiswamy et al., 2004; Leroux et al., 2017; Chivasa, 2017; Kasampalis et al., 2018). As of late, the RS community has seen an enormous development of sensing technologies, including expanding spectral bands, radars, sensors, and laser-induced fluorescence (Yaping et al., 2008; Jin et al., 2018; Kasampalis et al., 2018). This has increased the potential of these tools to offer precise and consistent estimates of crop systems and soil properties at various scales (Jin et al., 2018). As such, RS can now be applied in agriculture to identify and approximate farming lands and growing stock (Chakrabarti et al., 2014). It can also be used to detect crop nutrients deficiency, to map soil properties (e.g. soil moisture; Chakrabarti et al., 2014), and estimate leaf area index (LAI; Fang et al., 2008; Yao et al., 2015). While others use RS to monitor agricultural drought assessment by means of normalized difference vegetation index (NDVI) as well as predict crop yield production across different scales (Rao and Ranga, 2001; Doraiswamy et al., 2004; Maes and Steppe, 2012; Courault et al., 2017). RS can also be utilized to record climate parameters such as precipitation and temperature data, especially where weather stations are not available (Mishra et al, 2013). However, RS data generally does not account for the long-term impacts of the soil characteristics, climate, and surrounding environment on crop growth but rather represent the immediate state of a crop's canopy cover (Wang et al., 2014). Moreover, only a partial measure of RS data could be recorded throughout the entire crop development cycle because of the restrictions forced by climatic conditions and time it takes for the sensor to pass the same spot on earth (Doraiswamy et al., 2004; Wang et al., 2014).

### *2.5.1. Remote sensing and soil moisture monitoring*

Soil moisture (SM) is one of the important parameters within the earth system. It assumes a fundamental role in several processes and feedbacks that occur in the earth system (Fang et al., 2016). For example, it affects the relationship between the land surface and the surrounding environment (Peng and Loew, 2017). More specifically, SM undertakes a significant role in crop growth and development. It controls the amount of water that is available to the plant and drives photosynthesis, ecosystem dynamics, and soil respiration which then affect the optimal growth of the plant (Fu et al., 2003; Feng, 2016; Dorigo et al., 2017). The importance of SM was also recently acknowledged by the Global Climate Observing System when it was flagged as a second "Essential Climate Variable" after precipitation (Wagner et al., 2012; Dorigo et al., 2017).

Enormous improvement has been made on utilizing RS tools to retrieve near-surface SM across various scales. As such, for decades numerous microwave and optical remote sensing satellites have been successfully collecting SM data (Liu et al., 2012; Fang et al., 2016; Peng and Loew, 2017). SM observations from microwave satellites particularly have been greatly advanced, thus generating various high-quality SM global products. These include the advanced scatterometer (ASCAT, Naeimi et al., 2009), the Advanced Microwave Scanning Radiometer-EOS (AMSR-E, Owe et al., 2008), the Soil Moisture and Ocean Salinity (SMOS, Jacquette et al, 2010), the Soil Moisture Active Passive (SMAP, Entekhabi et al., 2010) or the European Space Agency's Climate Change Initiative (ESA CCI, Wagner et al., 2012). The accuracy of these SM products has been extensively quantified and validated, either by intercomparing various remotely sensed products or directly comparing them with *in situ* measures where they are available (Albergel et al., 2012; Peng et al., 2015; Dorigo et al., 2017). Moreover, such satellite SM outputs have been employed extensively in various disciplines including climate sciences, agriculture, and hydrology (Loew et al., 2013; AghaKouchak et al., 2015; Li et al., 2016).

In crop modelling, a good representation of SM estimation is required to detect the effect of moisture stress on crop growth, and at maturity to estimate yields (Mishra et al., 2013). However, obtaining accurate *in situ* SM is a challenge, especially in Africa where there is a dearth of relevant SM information (Jamali et al., 2011; Mishra et al., 2013). Consequently, CMs estimate SM based on precipitation data, but in data-limited regions as in most African countries obtaining accurate precipitation data can be difficult (Mishra et al., 2013; Zinyengere et al., 2015). Therefore, RS can bridge this gap by providing accurate and timely surface soil moisture even in data-limited areas, therefore, improve CMs outputs. Nonetheless, very few studies that have evaluated the use of remotely sensed SM data into CMs in African countries.

## 2.6. Combining remote sensing data with crop models

### *2.6.1. Integration of remote sensing data into a crop model*

Both CMs and RS tools have great potential to precisely monitor crop development and estimate crop yield (Yaping et al., 2008). Consequently, many researchers try to integrate these two numerical tools to predict crop growth, development and yield at various scales (Delecolle et al., 1992; Yaping et al, 2008; Mishra et al., 2013; Jin et al., 2018). Both tools have many similarities and together they have a large potential to better develop the capacity and accuracy of CMs simulations, especially in a data-scarce environment. Moreover, RS data can potentially be utilized to sufficiently set up the CMs where no data exits. Jin et al. (2018) argued that integrating RS data with crop models offer researchers and decision support groups an opportunity to utilize advanced and sophisticated tools that were formally developed for point-based analyses. Hence, improve the current understanding of many cropping systems.

There are mainly two approaches that can be used to integrate remotely sensed data into CMs: *forcing* and *recalibration* (Delecolle et al., 1992; Yaping et al, 2008; Jin et al., 2018). In *forcing*, the one or more model state variable is updated or replaced with the remotely sensed data estimates. This is the easiest approach to use but obtaining daily time steps of RS data that is required to drive the model is a challenge as the satellite can be blocked by clouds and only overpass repeat cycle (Yaping et al, 2008; Jin et al., 2018). Nevertheless, there are various studies that have successfully forced remotely sensed data into crop models. For example, remotely sensed green leaf area index (GLAI) data was integrated with the CSM-CROPSIM-CERES-Wheat model to simulate yield, canopy weight, and evapotranspiration in Maricopa, Arizona (Thorp et al., 2010). While in the Punjab state of India, Tripathy et al. (2013) used a leaf area index (LAI) to drive World Food Studies (WOFOST) model to predict spatial wheat yield. A study by Mishra et al. (2013) firstly used thermal infrared RS to approximate total root zone SM in South-eastern U.S. The subsequent SM profile was then forced into the Decision Support System for Agrotechnology Transfer (DSSAT) model, instead of the precipitation data, to improve maize yield simulations. Yao et al. (2015) modified Remote-Sensing–Photosynthesis–Yield Estimation for Crops (RS–P–YEC) model using Moderate Resolution Imaging Spectroradiometer (MODIS) LAI data to quantitatively simulate maize yield in the Northeast China Plain. Nevertheless, very few studies have used this method in the Africa context (Moeletsi and Walker, 2012; Leroux et al., 2017).

The *recalibration* (also referred to as data assimilation) integration approach involves adjusting the model parameters or initial conditions with RS data during the development stages of the crop. The recalibration method is normally used to reduce the difference between the model's simulated parameter and the RS data (Yaping et al, 2008; Jin et al., 2018). There are many studies that have successfully used this method to improve crop yield and biomass estimates. For example, Yaping et al. (2008) regionalized and adjusted the WOFOST model by reducing the difference between simulated and synthesized remotely sensed soil adjusted vegetation index (SAVI) data to monitor winter wheat growth in North China. In the same area and using the same model, Jin et al. (2015) recalibrated this model using remotely sensed LAI to evaluate the accumulation of heavy metals in rice crops. Neiring et al. (2012) and Ines et al. (2013) integrated observed remotely sensed LAI and SM into the Decision Support System for Agrotechnology Transfer-CropSim-Ceres model (DSSAT-CERES) to respectively improve wheat and maize yield estimates. In Indiana (USA), the Cropping System Model (CSM)–Crop Environment Resource Synthesis (CERES)–Maize was coupled with the Markov Chain canopy Reflectance Model (MCRM), and the resulting LAI and vegetation index products were used to estimate maize yield (Fang et al., 2011). This method has also been used in other studies using various simulation models and crops across the world (Ma et al., 2013; Wang et al., 2014; Battude et al., 2016). However, to the best of the study's knowledge, no studies have used this method in the Africa context.

Jin et al. (2018) proposed a third method of combining RS data and CMs (which is exclusive to monitoring), called *updating* methods. In this method, the crop model simulation input data is continuously being updated and Jin et al. (2018) assume "that a better simulation data at day *t* will improve the accuracy of the simulation data at succeeding days". Ines et al. (2013) modified the DSSAT-CERES model such that, for any given growing season, the Ensemble Kalman Filter framework updated the model state variables (i.e. rainfall) as LAI and soil moisture RS data becomes available, to therefore simulate maize yield in Story Country, USA. While El Sharif et al. (2015) integrated Soil Moisture Active Passive (SMAP) SM data into DSSAT by updating the weather inputs to simulate regional agricultural yield and irrigation needs in Ames, Iowa.

Recalibration and updating are regarded as better methods to integrate RS data into a CM due to their flexibility and minimum errors (Jin et al., 2018). Whereas the forcing approach may bring about additional errors as certain values of the model parameters are replaced by RS observational data which generally have some errors. In Africa, some efforts have been made to combine RS data with biomass in SA (Moeletsi and Walker, 2012; Singels et al., 2014) and with/for maize yield prediction in a scarce field-data environment of Burkina Faso (Leroux et al., 2017). Nonetheless, these studies are very few to fully understand and evaluate the potential of integrating RS with CMs, especially in data-scarce areas like rural SA.

Most of the literature reviewed combined RS data with CMs to improve the accuracy and capacity of the model to simulate crop yield. Such studies were conducted in the areas whereby there was at least some data available to sufficiently calibrate the model and then evaluate its performance by comparing the results (i.e. yield) with the available observed national/district data. However, a problem may arise when there is poor-to-no input data (field data) to either calibrate or assess the performance of the model, which is the case for many (especially rural and remote area) locations in Africa (Zinyengere et al., 2015). Therefore, there is a need for studies that explore whether using RS data to improve poor model calibration can sufficiently improve the confidence in CM outputs to a level that is meaningful and useful, even when there is limited data to evaluate the model performance.

### *2.6.2. Improving remote sensing observations with crop model outputs*

Although this approach is beside the aim of this study, crop model outputs can be used to improve remote sensing estimates. Accurate identification of the biophysical processes that are responsible for the soil and crop conditions are important for applying relevant management

(Jin et., 2018). However, these processes cannot be identified from remotely sensed data as it only represents the actual soil-crop status. Jin et al. (2018) suggest that CMs can be combined with multitemporal and/or multispectral remotely sensed data to improve the identification of appropriate processes and management. This means that CM outputs may be used to interpret and improve remote sensing data. Hence, in a case whereby RS data or images are not available, model-derived information on crop and soil parameters at various scales may be useful in filling the data gaps in remote sensing observations (Jin et., 2018). For example, soil conditions (e.g. SM) and crop characteristics (including LAI, biomass, evapotranspiration and canopy cover/structure) from CMs can potentially be utilized to fill the gaps in the remotely sensed data. The missing data could have been caused by unfavourable climatic conditions (i.e. clouds) and/or technical issues (Yaping et al, 2008; Jin et al., 2018). Thus, CMs can improve the accuracy of remote sensing data.

#### *2.6.3. The advantages and limitations of combining remote sensing data with crop models*

Generally, CMs sometimes need spatial input data to sufficiently run and produce useful outputs (Kasampalis et al., 2018). Therefore, a noteworthy advantage of combining RS data with CMs is that RS can provide additional spatial data, which is usually not available from non-dimensional models (Seidl et al., 2004; Launay and Guerif, 2005; Thorp et al., 2010). Spatial information is essential for many applications, such as to monitor and predict site-specific crop development and production across varying scales, in order to inform decision making (e.g. trading, logistics, and insurance; Kasampalis et al., 2018). Another advantage of combining RS data with CMs is that RS can quantify and provide missing crop information during the actual growing season (Delecolle et al., 1992; Doraiswamy, 2004; Courault et al., 2017; Kasampalis et al., 2018). This is because CMs tend to fail when the growing conditions change from the optimum, which can be a result of either biotic or abiotic stresses.

A reported limitation of using RS data is that satellites can provide ready-to-use products with a relatively high temporal resolution (hence they can easily be combined with CMs), but low spatial resolution (Rembold et al., 2013). As such, Kasampalis et al. (2018) proposed that high spatial resolution data, such as MODIS of PROBA-V, must be used. Another limitation of CMs, particularly process-based models, is that the model success (due to its complex nature), relies significantly on the capacity to set up the crop processes through the integration of connected variables (Kasampalis et al., 2018). This is mainly due to the complexity of process-based models. This means that, firstly, one must recognise that to produce useful outputs, it is very likely that more than one variable must be recalibrated (Rembold et al., 2013; Kasampalis et al., 2018). Secondly, those multiple variables recalibration should be done in some sort of coherence to indeed improve the modelling of the crop processes. This is ultimately the key in process-based models in order to improve the confidence in model outputs.

# Chapter three: Methodology

## 3.1. Introduction

Chapter three describes the data and methodology used in this study. Firstly, the characteristics and climate of the study area are presented. Then the Decision Support System for Agrotechnology Transfer (DSSAT) model is described in depth. The study then describes the methods used to achieve the aim of this thesis, including the recalibration procedure used to integrate remotely sensed soil moisture into DSSAT. The data inputs used to successfully run the DSSAT model are also highlighted. This chapter concludes with data analysis procedures which were employed to primarily assess the improvement of the key model outputs.

## 3.2. Study Area

### *3.2.1. The Free State Province*

The Free State province (28.45ºS, 26.80ºE) is the third-largest province (out of nine provinces) in South Africa (SA). It is situated in the east-central part of SA and covers about 129 464 $km^2$, which is just over 10% of the SA's entire land area (Davis et al., 2006; Free State Province, 2007; Beleste et al., 2015). In 2011, the population in this province was 2.7 million across one metropolitan municipality (Mangaung metropolitan) and four district municipalities, specifically Fezile Dabi, Lejweleleputswa, Thabo Mofutsanyana, and Xhariep municipalities (Free State Province, 2007; StatsSA, 2011). Agriculture is the main sector that dominates this province's land. Most of the agricultural production within the province is rain-fed with only about 10% irrigated (DAFF, 2017). About 32 000 $km^2$ of the land is covered by cultivated land while 87 000 $km^2$ accounts for the natural veld and grazing land (Beleste et al., 2015). Field crop production contributes about two-thirds of the province's gross agricultural revenue, mainly through maize, soybeans, sorghum, sunflowers, and wheat production. This province produces about 40% of the country's total white maize production, and 38% of the yellow maize (Beleste et al., 2015; DAFF, 2017). Hence, agricultural production in this province contributes significantly to the country's economy.

### *3.2.2. Bloemfontein*

Bloemfontein (29.07°S, 26.13°E) is the capital city of the Free State province and the sixth-largest city in SA. Bloemfontein forms part of the Mangaung metropolitan municipality and more than 50% of the population of this municipality is concentrated within this city. It has been estimated that by 2030 the population in this city will be 736 052, which is about a third above the 2011 census population count (IDP, 2018). Bloemfontein is the economic power hub of the local communities and a significant proportion of land remains dedicated to agricultural production. This area is semi-arid with marginal and erratic rainfall, and it is exposed to excessive runoff and evaporation loss, and experiences low crop productivity if is not irrigated (Baiphethi et al., 2003; Beletse et al., 2015). Nevertheless, the major crops cultivated in this area are mostly maize, wheat, and sunflower. The other crops planted in this area include sorghum, soybeans, groundnuts, and potatoes. The community members also share communal pasture for livestock rearing, mostly sheep, cattle and horses (Baiphethi et al., 2003).

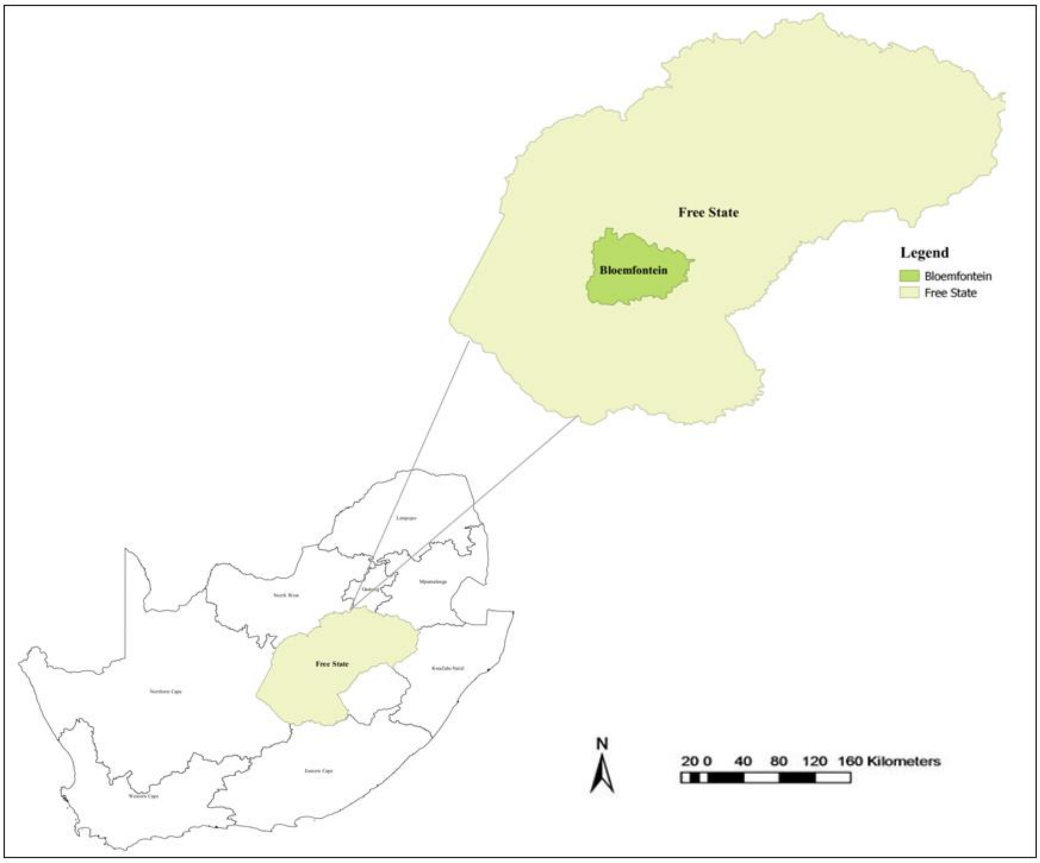


**Figure 3.1:** The map showing the Free State province and Bloemfontein district.

In terms of climate, Bloemfontein experiences hot and wet summers while winters are cold and dry (Figure 3.2). The highest average maximum temperature is recorded in January (31°C) and the lowermost mean minimum temperature is recorded in July (-2°C). In Bloemfontein, the highest monthly average rainfall is received in February averaging 96 mm, while the average annual precipitation is about 560 mm, with over 70% of precipitation falling during the summer months (i.e. October to April) as short thunderstorms in the afternoon (Moeletsi, 2010; Beletse et al., 2015).

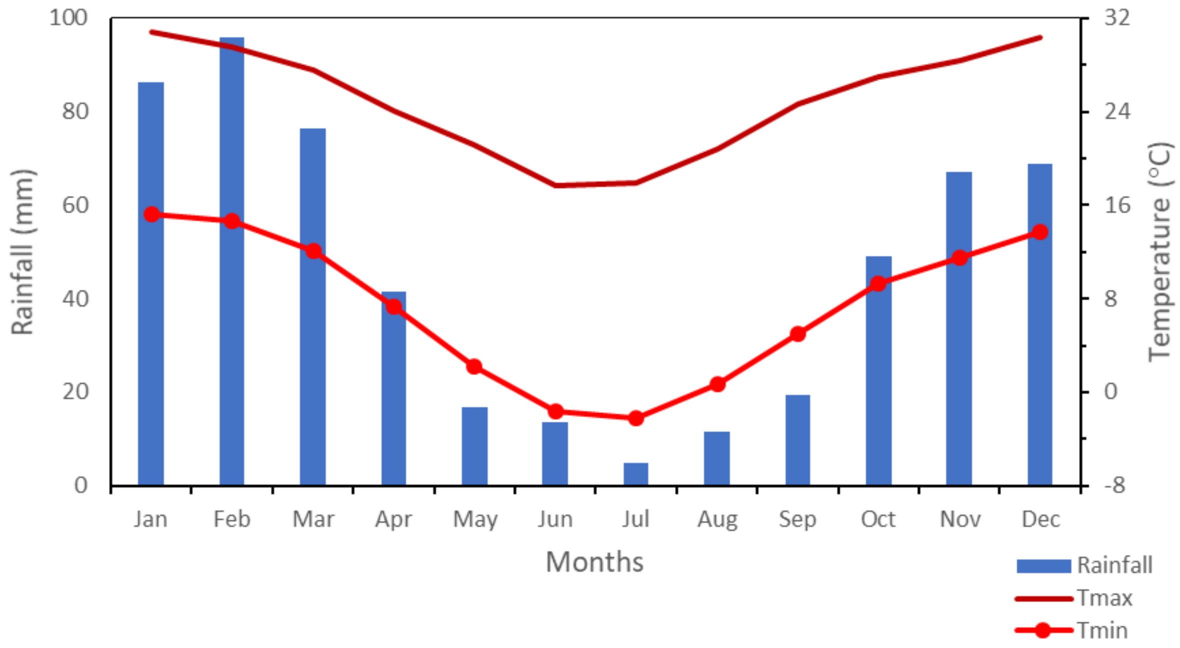


**Figure 3.2:** The monthly climatology of Bloemfontein (1985-2015). The bars represent the average monthly total rainfall (mm) and lines represent the average maximum (dark red) and minimum (red) temperature. Data source: South Africa Weather Services (SAWS).

### 3.3. Model description

The study used the Decision Support System for Agrotechnology Transfer (DSSAT) crop model to explore the potential of integrating remotely sensed data in conditions of unavailable ground data. The study used this model as it has been extensively used and evaluated throughout the world (Neiring et al., 2012; Ines et al., 2013; Mishra et al., 2013), including in southern Africa (Thornton et al., 2011; Beleste et al., 2015; Zinyengere et al., 2015; Durand and Ferreira, 2017; Leroux et al., 2017). These studies vary from field to regional scale including climate impact and adaptation studies, water management plans, drought monitoring, as well as the estimation of maize yield in data-limited areas in southern Africa.

DSSAT is a suite of process-based crop models, for which it simulates various biophysical processes including daily crop growth, crop phenological development, and crop yield as a response to the crop management, soil and weather conditions (Jones et al., 2003; Mishra et al., 2013; Zinyengere et al., 2015; Nagamani and Mariappan, 2017). This crop models include, but are not limited to, CERES (for maize, sorghum, wheat, and rice), SubSor (for potatoes), Canegro (for sugarcane) or CROPGRO (for soybeans, cotton, peanut, and dry beans), and many more (Jones et al., 2003). For the model to initially run, it requires at least daily weather, layered soil characteristics, plant growth characteristics as well as the applied crop management practices. The model simulates several biophysical states of the plant and immediate environment throughout the growing season (e.g. soil moisture, leaf area index, nutrients stress, water stress, and other crop stresses), and harvest or maturity level yield (which most studies are interested in) (Jones et al., 2003).

The DSSAT's most important variable for this thesis is the soil moisture (SM), as the model was integrated with remotely sensed SM data. The model's SM storage, as well as permeation characteristics, are founded on an easy vertical soil storage direction notion (Jones et al., 2003; Mishra et al., 2013). Moisture is stored in one layer until full saturation is reached, and then moisture infiltrates the next layer. The crop is then allowed to take up moisture till the soil wilting point is reached. Mishra et al. (2013) further state that as an essentially a point-based plant models DSSAT does not reflect sideways percolation of moisture.

### 3.4. Methods

The study used the DSSAT-CERES-Maize model (version 4.7), in sequential application to ensure that the model considers the previous climatic conditions at the beginning of each new year. The schematic flowchart of the overall method followed in this study is represented in Figure 3.3. Based on the key objective of the current study, which is to explore the potential of using remotely sensed data to model agricultural systems in areas with limited ground data, it was assumed that there is no physical access to- or existence of- reliable soil data. In-depth details of the input data used are discussed in section 3.4. Therefore, the study, firstly, calibrated the DSSAT-CERES-Maize model using the two generic soils; heavy clay soil and medium sandy soil that were selected based only on the previous literature, together with the weather information and crop management practices to estimate SM from 1985 to 2015 (i.e. Baseline experiment). The simulated SM was then compared to the corresponding European Space Agency's (ESA) remotely sensed SM (hereafter referred to as observed SM). The intra- and inter-annual variation of SM produced by these two soils were used to guide assimilation of the observed remotely sensed data into the model. Specifically, the study then calculated the mean bias error (MBE) during the growing season (i.e. October to March) between the simulated and observed SM, with an aim to minimize it by adjusting the input parameters (Figure 3.4).

The study recalibrated the model by adjusting three parameters of the soil profiles, namely, the drainage lower limit (DLL), the drained upper limit (DUL), and the saturated upper limit (SAT). This was done with an aim to minimise the mean difference between the observed and simulated SM during the growing season (Figure 3.4). The soil profile parameters were altered in such a way that the model retained less water for heavy clay soil (i.e. decreased DLL, DUL, and SAT) and retained more water for medium sandy soil (i.e. increased DLL, DUL, and SAT). Therefore, the study reduced or increased the plant available water (PAW) depending on the soil used (O'Green, 2013). Once the minimization threshold of MBE was satisfied (i.e. Final experiment), the study used the optimized soil profile parameters to simulate the improved SM. Finally, the study explored the impacts of this specific SM improvement and soil profile optimization on other model outputs, namely, evapotranspiration (ET) and maize yield.

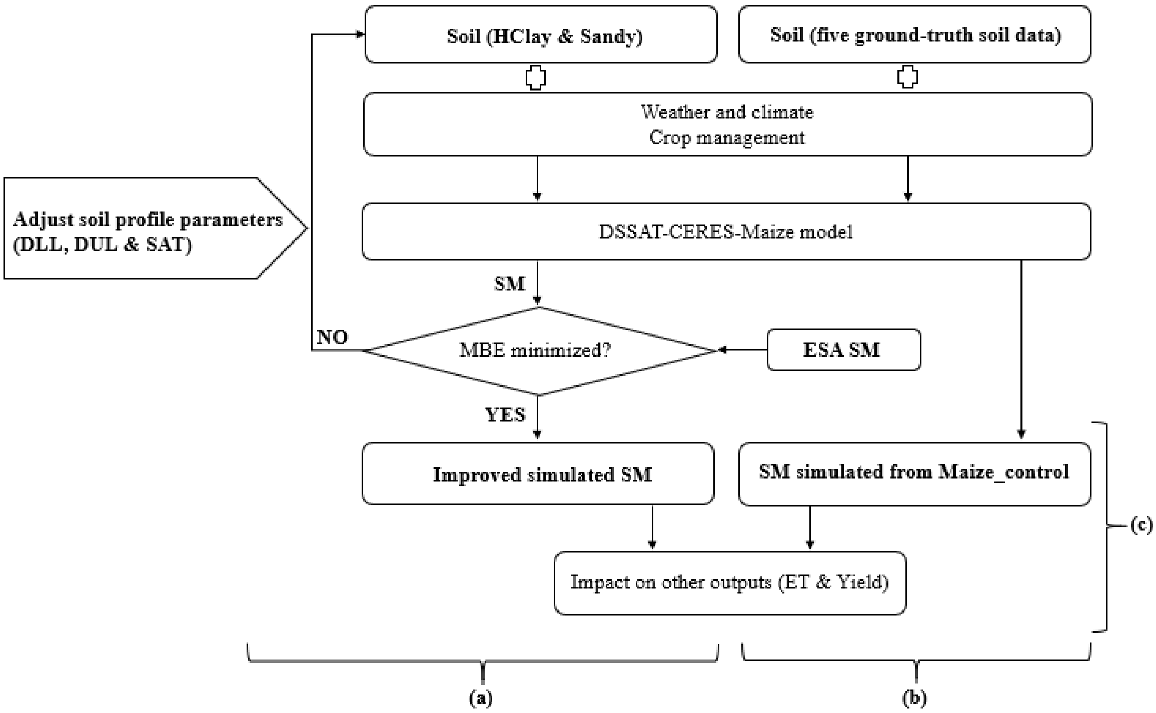


**Figure 3.3:** The schematic flowchart of the methods followed in this study. The data assimilation process is represented by (a), (b) the model calibrated with actual ground data (Maize_control), and (c) represents the improvement assessment and comparison part of the study methods.

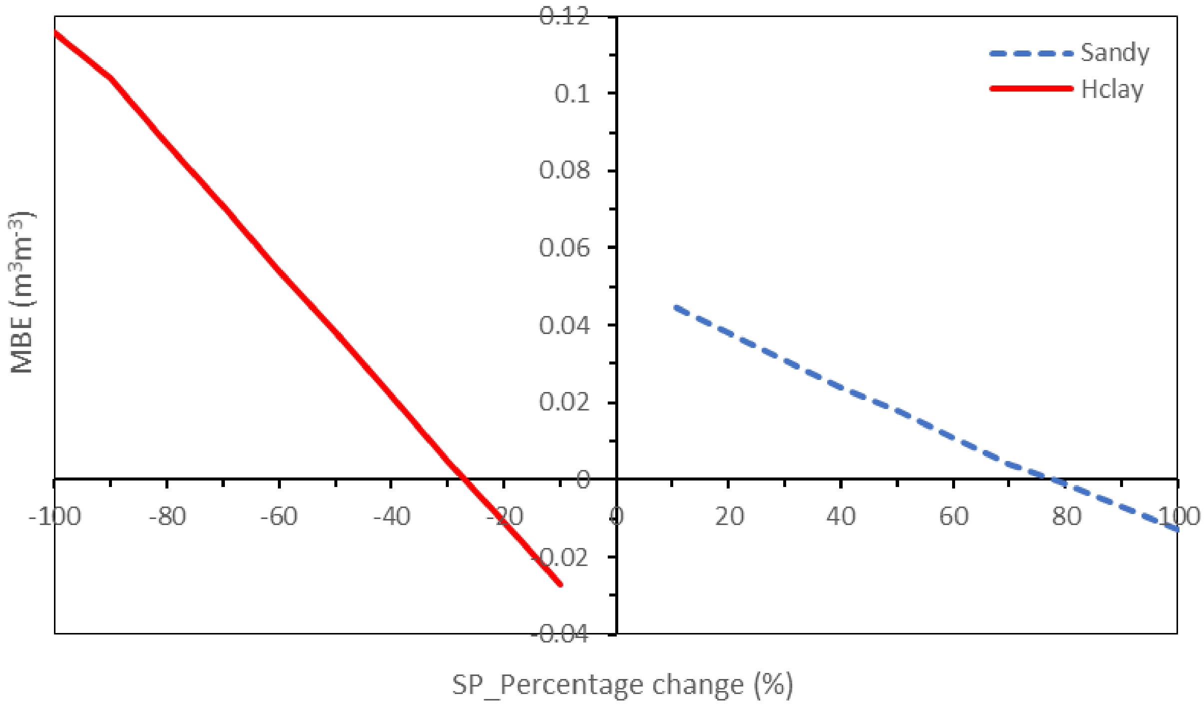


**Figure 3.4:** The mean bias error (MBE) between the observed and simulated soil moisture ($m^3m^{-3}$) for heavy clay (red) and medium sandy soil (blue).

Normally, the performance of a CM is assessed by comparing the results of the model with real data and, therefore, determine its suitability for the intended objective (Zinyengere et al., 2015). However, since the aim of this study is to assess how well the model assimilated with RM data compares with a CM calibrated with actual ground data, the study also calibrated the DSSAT-CERES-Maize model using extensive ground reference soil (i.e. five soil datasets) data (hereafter referred to as Maize_control). As such, the main difference between the data assimilated and Maize_control experiments set up was soil data, while the climate, crop variety, and crop management strategies were kept the same. The Maize_control experiment also simulated SM, ET, and maize yield. The study then mainly compared improved SM simulated by the data assimilated model (i.e. Final experiment) with that of Maize_control. Additionally, it also compared the ET and maize yield simulated by Final experiments and Maize_control experiment.

### 3.5. Data input

The minimum input data needed to run DSSAT include climate (i.e. temperature, rainfall, and solar radiation) data, soil characteristics for the study area, crop variety, and crop management strategies used. According to Zinyengere et al. (2015), such data is normally available at experimental levels, however, often not continuously available over long enough duration as required by the model. Such challenges further increase with the extent of the study. The data used to complete this study was acquired from field experiments, remote sensing, grey literature, national government departments, and international records.

### *3.5.1. Climate*

Minimum and maximum daily air temperature (ºC) as well as total daily rainfall (mm) data (1985-2015) for Bloemfontein were acquired from the South Africa Weather Services (SAWS). On the days where the data was missing, it was obtained from the nearby station as advised by SAWS or interpolated based on the expert judgment of the available information. The details of the main weather station used to obtain this climate data are represented in Table 3.1. The SAWS stations do not record daily solar radiation (MJ/m²-day), therefore, the study obtained this data from the NASA Prediction of Worldwide Energy Resources (POWER) global agroclimatology database (NASA, 2014; https://power.larc.nasa.gov/data-access-viewer/). This data is freely available, and it estimates near-real-time radiation at 0.5x0.5 degrees resolution at a single point (latitude and longitude). The accuracy of solar radiation data has been validated at a large scale globally, although not necessarily in South Africa (El Sharif et al., 2015; Hoell et al., 2015).

**Table 3.1:** The characteristics of the weather station data for Bloemfontein – mean annual minimum (min) and maximum (max) temperature (ºC) and total mean annual rainfall (mm). In the square brackets are the means over October to March growing season.

| District | Longitude (ºS) | Latitude (ºE) | Elevation (m) | Temperature (ºC) | | Rainfall (mm) |
|---|---|---|---|---|---|---|
| | | | | Min | Max | |
| Bloemfontein | -29.1 | 26.29 | 1354 | 7.3[12.8] | 25[29] | 561[450] |

### *3.5.2. Crop variety and crop management*

The study obtained the crop variety and crop management strategies used in Bloemfontein from datasets that were collected by Durand and Ferreira (2017) across five locations in the Bloemfontein area (mostly from commercial farmers). These data included planting dates, plant population density, crop variety, fertilizer application, irrigation, and timing of their application (Table 3.2). This data was supplemented with data from the independent and local experts' agronomists from the University of the Free State (W Abraha 2018, personal communications, 14 March). Within the DSSAT-CERES-Maize module, the study used the 2650-2700 GDD maize cultivar with a medium growth period (i.e. 121-145 days). This is a representation of the PAN 6479 variety, which reaches maturity around 119-140 days after planting, and it is usually planted in the Free State province (Beleste et al., 2015; Durand and Ferreira, 2017). The planting period ranged from the 1st of November to the 17th of December of each year, and a planting rule only proceeded when the temperature was between 10-36 °C and soil water content at 30 cm depth was between 85-100%. The plant population density was 15000 plants ha$^{-1}$. 50 kg ha$^{-1}$ nitrogen fertiliser was directly applied twice over the growing season, specifically, on 0 and 42 days after planting to improve nitrogen use efficiency and maize grain yield (Tadesse et al., 2013; Durand and Ferreira, 2017). Most of the maize cultivation within the district is generally produced under rain-fed condition, therefore, no irrigation was applied in the model.

**Table 3.2:** The crop variety and crop management practices used in the DSSAT-CERES-Maize model.

| Management practice | Treatment |
|---|---|
| Maize cultivar | 2650-2700 GDD (Medium) |
| Fertiliser (N, kg ha$^{-1}$) | 100 |
| Plant population density (plant ha$^{-1}$) | 15000 |
| Planting date (Julian days) | 1 Nov-17 Dec (305-351) |
| Irrigation | No |

### *3.5.3. Soil*

For the specific conditions of this study, it is assumed that there is no physical access to- or the existence of- reliable soil data or expertise of soil information. While expertise or field measurements will definitely improve this situation, the study proceeds with only a literature base definition of soil, knowing the limitation of this setup and purposefully dedicated this thesis to assess the capacity to improve this situation. Based on the literature, maize in South Africa is mainly cultivated under soils with low clay content (i.e. ≤ 10%) or heavy clay content that exceeds 30%, therefore, in sandy or clay soils respectively (du Plessis, 2003). On this basis alone, without any particular soil or crop expertise, and without soil information to calibrate further the soil module in DSSAT, the study picked two generic soils within the DSSAT soil database, based on their clay content and through their soil type and name. The two soils are medium sandy soil (i.e. Default-Medium sand soil) and heavy clay soil (i.e. Weakly self-mulching grey medium to heavy clay soil). Some of the chemical and physical characteristics of these soils before and after data assimilation are described in Table 3.3 and Table 3.4.

The study also used the soil data collected by Durand and Ferreira (2017) from five different locations within the Bloemfontein district, to represent the ground “truth” data. They used the characteristics of the dominant soil in each land type, including the soil form and series from the 1977 edition of the South African Soil Classification System (MacVicar et al, 1977), the horizons occurring per soil profile, the average depth range (mm) of each horizon and the clay content range. The dominant soil forms ranged from Hutton (Hu), Swartland (SW) to Bainsvlei (BV) types. More detailed descriptions of these soil types may be acquired in MacVicar et al. (1977) and Fanourakis (2012). Furthermore, the land type data and expert knowledge were used to estimate organic carbon, bulk density, and pH. The rooting depth was taken as an average depth of the subsoil horizon. The soil water content characteristics were estimated using parameters given by Schulze *et al.* (1985), while the profile/plant available water (PAW) per horizon was estimated using the formula (0.31 x estimated silt + clay %) - (0.03 x 10 cm depth index) + 7.93 given by Boedt & Laker (1985). The soil runoff curve numbers were based on information from the Soil Conservation Service in the USA, which is freely accessible on the web (http://age-web.age.uiuc.edu/classes/). The land type data was also used to estimate the soil albedo fraction. The authors also assumed that 0.1 represents very light and reflective soil, while 1.0 represents a very dark to black soil. The key chemical and physical characteristics of these reference soils that were used in the study are defined in Table 3.5.

**Table 3.3:** The chemical and physical characteristics of the generic soils (Baseline set up) used in the study.

| Soil type | DBL (cm) | BD (g cm$^{-3}$) | OC (%) | CL (%) | SI (%) | ROC (%) | pH (water) | DLL (cm cm$^{-3}$) | DUL (cm cm$^{-3}$) | SAT (cm cm$^{-3}$) |
|---|---|---|---|---|---|---|---|---|---|---|
| Heavy clay | 5-130 | -99 | -99 | -99 | -99 | 60 | -99 | 0.153-0.180 | 0.288-0.256 | 0.273-0.414 |
| Medium sandy | 5-120 | 1.66 | 0.11-0.29 | 5 | 5 | 70 | 6.5 | 0.018-0.039 | 0.091-0.097 | 0.345-0.347 |

**Table 3.4:** The chemical and physical characteristics of the optimized soil profile of the generic soils (Final set up) used in the study.

| Soil type | DBL (cm) | BD (g cm$^{-3}$) | OC (%) | CL (%) | SI (%) | ROC (%) | pH (water) | DLL (cm cm$^{-3}$) | DUL (cm cm$^{-3}$) | SAT (cm cm$^{-3}$) |
|---|---|---|---|---|---|---|---|---|---|---|
| HClay | 5-130 | -99 | -99 | -99 | -99 | 60 | -99 | 0.115-0.135 | 0.171-0.192 | 0.205-0.311 |
| Sandy | 5-120 | 1.66 | 0.11-0.29 | 5 | 5 | 70 | 6.5 | 0.032-0.045 | 0.164-0.175 | 0.621-0.623 |

**Table 3.5:** The chemical and physical characteristics of the reference soils used to calibrate the Maize_control experiment (Reference set up).

| Soil type | DBL (cm) | BD (g cm-3) | OC (%) | CL (%) | SI (%) | ROC (%) | pH (water) | DLL (cm cm-3) | DUL (cm cm-3) | SAT (cm cm-3) |
|---|---|---|---|---|---|---|---|---|---|---|
| S1-Hu33/34 | 30-120 | 1.55 | 0.8 | -99 | -99 | 65 | 7.2 | 0.071-0.082 | 0.175-0.190 | 0.184-0.200 |
| S2-Sw31/41/42 | 30-60 | 1.35-1.45 | 0.8 | -99 | -99 | 85 | 7.2 | 0.132-0.244 | 0.236-0.345 | 0.248-0.362 |
| S3-Hu36 | 30-80 | 1.45-1.55 | 0.8 | -99 | -99 | 70 | 7.2 | 0.104-0.142 | 0.208-0.247 | 0.218-2.59 |
| S4-Hu33/34 | 30-120 | 1.50-1.55 | 0.8 | -99 | -99 | 65 | 7.2 | 0.070-0.082 | 0.175-0.190 | 0.184-0.200 |
| S5-Hu33/36 + Bv36 | 20-120 | 1.40-1.55 | 0.8 | -99 | -99 | 75 | 7.2 | 0.082-0.148 | 0.184-0.556 | 0.193-0.584 |

*Legend: DBL = depth of the base layer, BD = bulk density, OC = organic carbon, CL = clay content, SI = silt content, ROC = runoff curve, DLL = drainage lower limit, DUL = drainage upper limit, and SAT = saturation.*

### *3.5.4. Remotely sensed soil moisture*

Soil moisture (SM) is the main parameter in the data assimilation process. For this study, surface volumetric SM ($m^3m^{-3}$) was obtained from the European Space Agency's Climate Change Initiative (ESA CCI) passive and active sensors for the duration of the study period (Liu et al., 2012; Dorigo et al., 2017; Gruber et al., 2017). The data was obtained from the http://www.esa-soilmoisture-cci.org/ (v04.2) website and only the data for Bloemfontein (26.29ºE**,** -29.1ºS) was extracted. This global data is freely available and has a length of over 30 years. The satellites have a spatial resolution of 25 km and a grid spacing of 0.25 degrees (Dorigo et al., 2017). The data meet the accuracy (0.04 $m^3m^{-3}$) standards of the satellite-based SM products. Although the satellites have a revisit time of shorter than 3 days, the data sometimes had missing values for over 20 consecutive days (particularly between 1985 and mid-1990s) which may potentially be a result of unfavourable conditions (i.e. cloud cover) or technical errors. As such, the study used average soil moisture (i.e. monthly climatology and growing season average) to handle missing values.

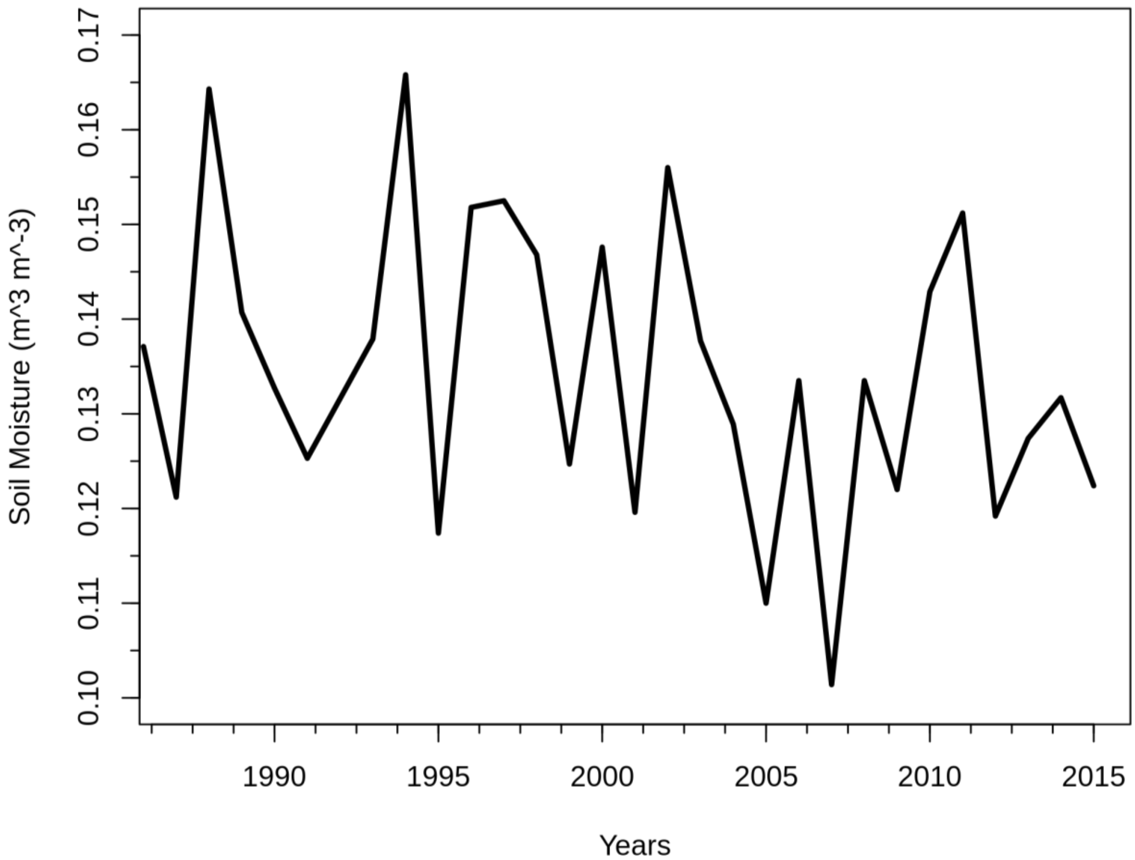


**Figure 3.5:** Average annual variation of remotely sensed soil moisture for Bloemfontein district (1985-2015).

### 3.6. Data analysis

The study mainly assessed the improvement made by assimilating remotely sensed data into a crop model primarily through the lens of SM simulation improvement from the original, coarse, knowingly generic, set up to the final RS informed soil profile set up. Therefore, the study compared both the baseline and final improved SM to the observed remotely sensed SM, using monthly climatology and growing season interannual variation, as well as the standard goodness of fit statistics that are described below. The study also gave some measure of comparison with a model calibrated with sufficient field data (Maize_control), and finally explored impacts of this specific soil moisture improvement, on other indicators; e.g. evapotranspiration (ET) and maize yield. As such, the simulated SM, ET, and yield from the Maize_control experiment were averaged, respectively, and compared the same results from the data assimilation (Baseline and Final) process. The performances of the baseline and final experiments were evaluated by using the following statistical measures: the *Mean bias error* (MBE, equation 1), the *Root Mean Square Error* (RMSE, equation 2), the *Coefficient of determination* ($R^2$, equation 3) and *Pearson correlation* (R, equation 4). The study also used the *D value* from the Kolmogorov-Smirnov Tests (equation 5) to qualify the distance and difference between the probability distribution function (PDF) of the Maize_control and data assimilation experiments. This is because this nonparametric statistical method is also sensitive to the shape and location of the empirical PDF of two samples (Marrozi, 2013). These statistical variables were respectively calculated as follows:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i) \quad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\tilde{y}_i - y_i)^2}{n}} \quad (2)$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (3)$$

$$r = \frac{n(\sum \tilde{y}_i y_i) - (\sum \tilde{y}_i)(\sum y_i)}{\sqrt{[n\sum \tilde{y}_i^2 - (\sum \tilde{y}_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} \quad (4)$$

$$D_n = \sup_{\tilde{y}_i, y_i} \left|F_{1,n}(\tilde{y}_i) - F_{2,n}(y_i)\right| \quad (5)$$

Where $n$ is the number of years (i.e. number of observations), $\tilde{y_i}$ and $y_i$ are the output simulated by Maize_control and data assimilation experiments, respectively. In the case whereby the model simulations were compared with the observed datasets (i.e. remotely sensed SM or district recorded maize yield), the $\tilde{y_i}$ and $y_i$ represented the simulated and observed outputs, respectively. According to Wilks (1995) parameters used to calculate $R^2$ respectively represent: SST = total sum of squared deviations of the simulated outputs from their average, SSR = regression sum of squares, and SSE = sum of squared differences between the residuals and their averages. $F_{1,n}$ and $F_{2,n}$ represents the empirical PDF of the Maize_control and data assimilation respectively, and $=_{\tilde{y_i},y_i}^{sup}$ is the supremum function.

## 4.1. Temporal variation of soil moisture before and after data assimilation

### *4.1.1. Monthly variation*

Figure 4.1 shows the model's ability to simulate monthly volumetric soil moisture (SM) before and after assimilating remotely sensed data over the Bloemfontein district. The crop model without data assimilation, regardless of the soil type, captured the monthly variation of SM very well (Figure 4.1a). However, when heavy clay soil was used, the model overestimated SM specifically between October and April and underestimated it between May and September. These periods, respectively, correspond with the extended growing season and non-growing seasons of maize within the district (DAFF, 2019). On the other hand, when the medium sandy soil was used, the model generally underestimated SM. Complementary to heavy clay soil results, the crop model with medium sandy soil also maintained relatively higher moisture during the maize growing season and relatively lower moisture during the non-growing season (Figure 4.1a).

Guided by minimizing the mean bias error (MBE) during the growing season, assimilating remotely sensed data with the crop models improved the simulated soil moisture between October and April (Figure 4.1b). As such, the difference between the observed and simulated monthly SM during the growing season was reduced. Nevertheless, when heavy clay soil was used, the model slightly overestimated SM during the first five months (October to February) of the growing season and slightly underestimated it over the last two months (March to April) of the season. On the other hand, the model with medium sandy soil also overestimated SM at the beginning of the growing season (October to December) and underestimated it during the last four months of the season (January to April).

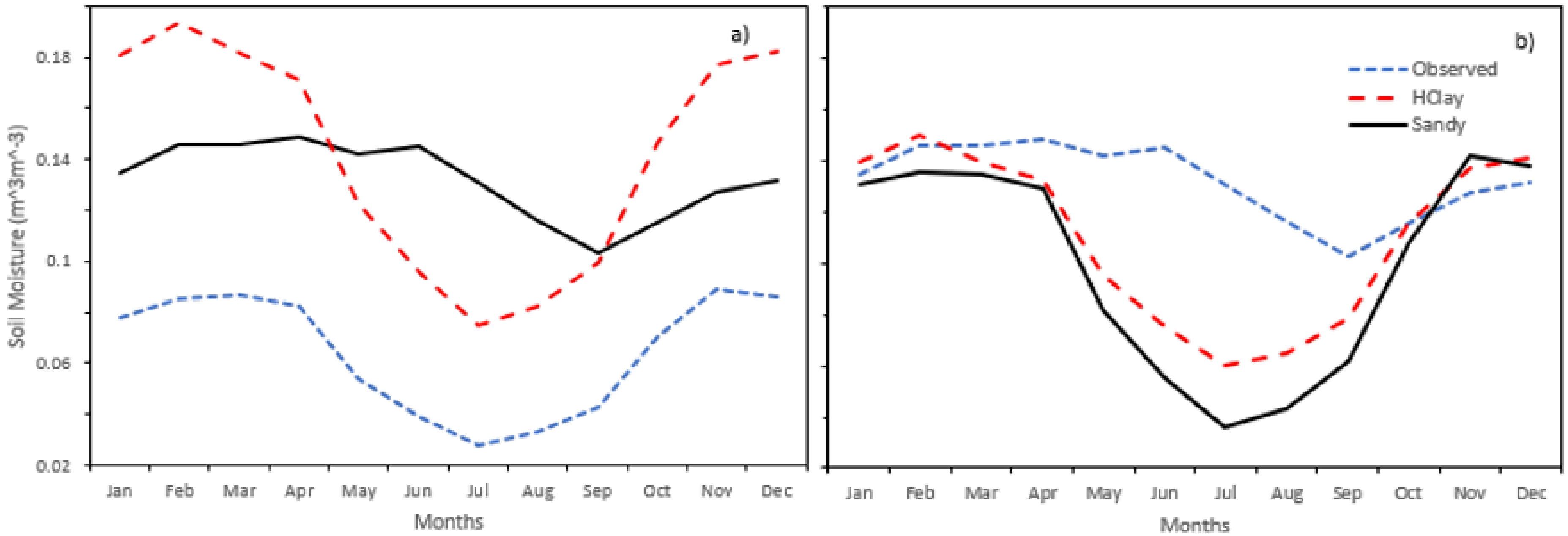


**Figure 4.1:** Performance of DSSAT in simulating monthly volumetric soil moisture ($m^3m^{-3}$) before (a) and after (b) data assimilation. The black line represents observed remotely sensed SM while red and green lines represent SM simulated by the model when heavy clay soil and medium sandy soils are respectively used.

#### *4.1.2. Interannual variation of soil moisture before and after data assimilation*

Figure 4.2 shows how well the model simulates the interannual variation of volumetric soil moisture (SM) over the years. Regardless of the soil used, the crop model without data simulation already captured the year-to-year variation of growing season SM very well (Figure 4.2a). For example, the model was able to capture the growing seasons with the lowest (i.e. 1996, 2002, 2008, 2013) and highest SM (i.e. 1995, 1997, 2001, 2003 and 2007)  However, the model with heavy clay soil generally overestimated the average seasonal SM when compared to the observed remotely sensed SM (0.178 and 0.135 $m^3m^{-3}$, respectively). While with medium sandy soil profile, the model underestimated the growing season moisture (0.082 $m^3m^{-3}$). Moreover, the model had a higher range of seasonal SM with heavy clay soil when compared to the observed SM as well as SM simulated with medium sandy soil (0.150-0.218, 0.112-0.163, and 0.051-0.115 $m^3m^{-3}$, respectively).

Data assimilation improved the model's ability to simulate seasonal SM (Figure 4.2b). This process significantly reduced the mean difference between the observed and simulated seasonal SM. Nevertheless, overall the model with heavy clay soil slightly overestimated the seasonal SM (Mean = 0.137 $m^3m^{-3}$) while the model with medium sandy soil slightly underestimated it (Mean= 0.131 $m^3m^{-3}$). The model, regardless of the soil, better captured the year-to-year variation of seasonal SM. For example, the model captured better the low SM that was recorded during the 1996 and 2013 growing season as well as the relatively high SM during the 2001 growing season, respectively. While the model with medium sandy soil captured better the relatively low SM that was recorded during the 2002 and 2006 growing season as well as the relatively high SM during the 1997 and 2003 growing season, respectively. Assimilating remotely sensed data also improved the range of simulated moisture over the years. For example, the range of SM during the growing season is 0.118-0.160 $m^3m^{-3}$ for heavy clay soil and 0.083-0.170 $m^3m^{-3}$ for medium sandy soil, respectively.

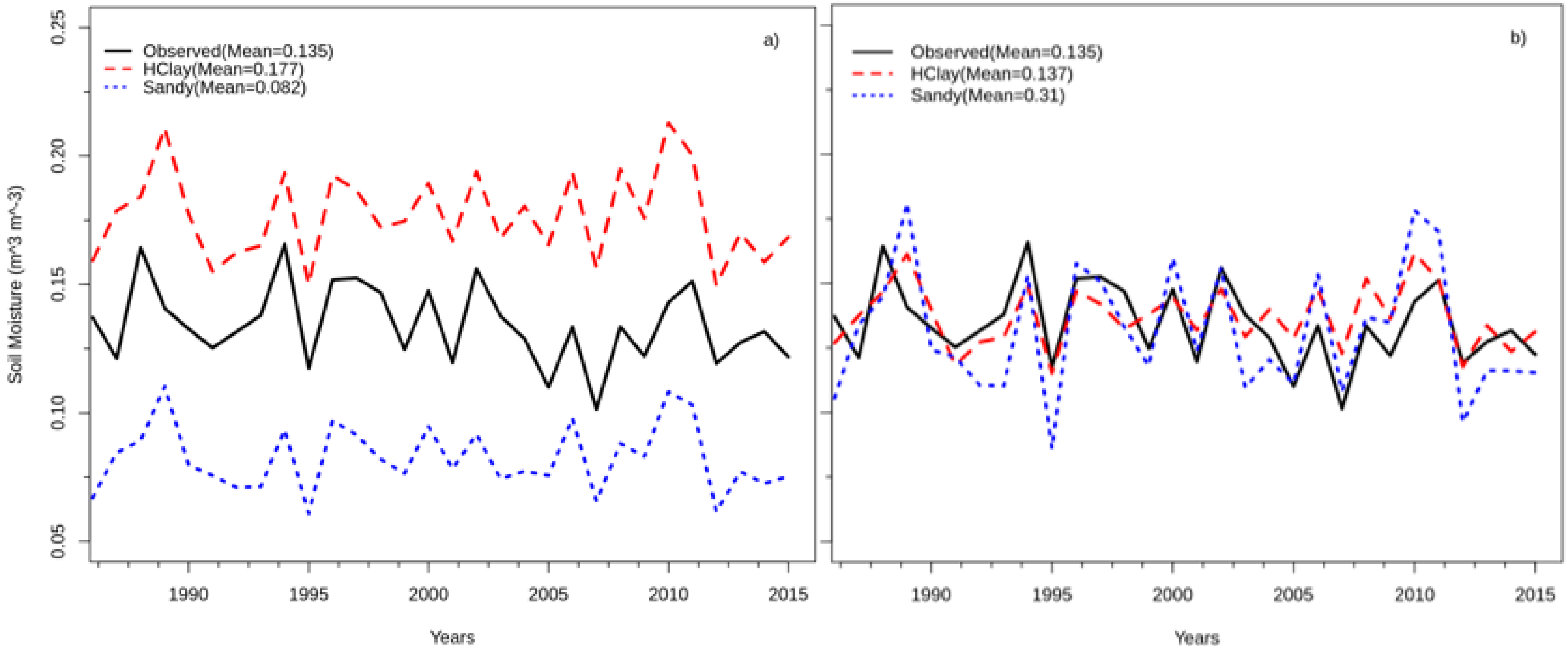


**Figure 4.2:** DSSAT simulated seasonal volumetric soil moisture ($m^3m^{-3}$) before (a) and after (b) data assimilation (1986-2015). The black line represents observed remotely sensed SM while red and green lines represent moisture simulated by the model when heavy clay soil and medium sandy soils are respectively used.

## 4.2. Assessing the improvement of model representation

Due to the data-scarce context of this study, the aim is not to improve the accuracy of a poorly calibrated model but to improve confidence in model simulations by recalibrating it using remotely sensed soil moisture (SM). Thereafter, assess how well the model assimilated with remotely sensed data compares with a model that is traditionally calibrated with ground data (Maize_control). In that context, the robustness of assimilating remotely sensed SM with the crop model was evaluated by mainly comparing the growing season volumetric soil moisture simulated from generic heavy clay and medium sandy soil experiments with a version calibrated with extensive ground data (Maize_control). The study also gave a measure of comparison beyond soil moisture, by exploring the impacts of the above specific soil moisture improvement, on other indicators; evapotranspiration (ET) and maize yield.

### *4.2.1. Soil moisture*

Data assimilation significantly improved the mean comparison between the soil moisture (SM) simulated by the heavy clay (i.e. final set-ups) and Maize_control experiments (Table 4.1, Figure 4.3a, and Figure 4.4a). As such, the MBE (-0.009 $m^3m^{-3}$) and RMSE (0.009 $m^3m^{-3}$) were significantly reduced compared to the baseline MBE and RMSE (-0.05 and 0.05 $m^3m^{-3}$, respectively). Moreover, the final experiment captured better the interannual variation of average growing season SM as simulated by Maize_control (R = 0.99, Table 4.1 and Figure 4.3a). For instance, during the 1991 and 2001 growing season, the final experiment simulated the same amount of SM as that of Maize_control, respectively (Figure 4.3a). Nevertheless, compared to Maize_control, the final experiment generally simulated a relatively higher SM. Complementary to the baseline experiment, the final experiment simulated above-normal SM in 1987 and 1999 growing season while the opposite is true for Maize_control. Moreover, the final experiment did not simulate SM during the 1990 growing season. This indicates that the model's planting rule conditions (temperature and soil water) were not met during that season.

Assimilating remotely sensed data with the model also improved the variability comparison between the SM simulated by the final heavy clay and Maize_control experiments ($R^2$ = 0.99, Table 4.1, Figure 4.4a). For example, data assimilation improved the comparison between the location and the shape of the distribution of the simulated SM from the two experiments (Table 4.1 and Figure 4.4a). As such, the difference (D) between the location of the heavy clay and Maize_control distributions were reduced from 0.98 to 0.3. While, the peak of SM simulated using heavy clay soil improved from 20 to 27.5, which is very close to that of Maize_control (27). Unlike the range of the baseline experiment (0.13-0.24 $m^3m^{-3}$), the range of the final simulated SM (0.10-0.17 $m^3m^{-3}$) is well within that of the Maize_control (0.09-0.15 $m^3m^{-3}$), although some moisture was still located at the higher end of the distribution.

**Table 4.1:** The statistical analysis between simulated volumetric soil moisture ($m^3m^{-3}$) from Maize_control and heavy clay soil or medium sandy soil.

| Statistical measure | HClay | | Sandy | |
|---|---|---|---|---|
| | Baseline | Final | Baseline | Final |
| **MBE** | -0.049 | -0.009 | 0.047 | -0.003 |
| **RMSE** | 0.05 | 0.009 | 0.05 | 0.01 |
| **R** | 0.98 | 0.99 | 0.91 | 0.78 |
| **$R^2$** | 0.96 | 0.97 | 0.84 | 0.79 |
| **KS.test (D)** | 0.93 | 0.30 | 0.89 | 0.27 |

Assimilating remotely sensed data with the crop model also improved the mean difference between the growing season SM simulated by final medium sandy and Maize_control experiment. Therefore, the final MBE (-0.003 $m^3m^{-3}$) and RMSE (0.01) were significantly reduced, compared to the baseline MBE and RMSE (0.047 and 0.05 $m^3m^-3$ respectively, Table 4.1). Data assimilation also led to a more comparable location of the moisture distribution (Table 4.1 and Figure 4.4b). As such, the difference between the medium sandy and Maize_control SM distributions was reduced from 0.89 to 0.27 (baseline and final experiment, respectively). Hence, a substantial range of the final simulated SM fall within that of Maize_control, unlike the baseline experiment which only had a small portion of the higher end of the SM distribution falling within that of Maize_control.

Assimilating remotely sensed data into a model with medium sandy soil, however, resulted in a mismatch in the variability of the simulated SM. Therefore, SM variability was compromised to increase the fit in the overall mean. As such, the shape of the distribution of the final experiment becomes flattened (peak density of 15) and widespread (range of 0.05-0.22 $m^3m^{-3}$, Figure 4.4b). This increase in variability of the SM may have led to a reduction in the correlation and coefficient of determination (R = 0.78 and $R^2$ = 0.79 respectively, Table 4.1) between the two experiments. Moreover, although this experiment also substantially captured the interannual variation of SM, the year-to-year variations of SM were relatively higher than that of the Maize_control experiment (Figure 4.3b). The discrepancy in simulated SM is particularly noticeable during the 1989, 2010, and 2011 growing season. Moreover, during the 1991, 1999, and 2004 growing season, the final experiment simulated above-normal SM when Maize_control simulated below-normal SM, respectively. This suggests that the SM from the baseline medium sandy soil model was more comparable, in terms of variability, to one produced using the Maize_control ($R = 0.91$ and $R^2 = 0.84$). For example, there is very little difference between the year-to-year variation of SM that is simulated by these experiments, the highest difference being 0.02 $m^3m^{-3}$ which was recorded during the 1991 growing season. In fact, during 1992, 1993, 2003, 2013 growing seasons, these experiments respectively simulated the same amount of SM.

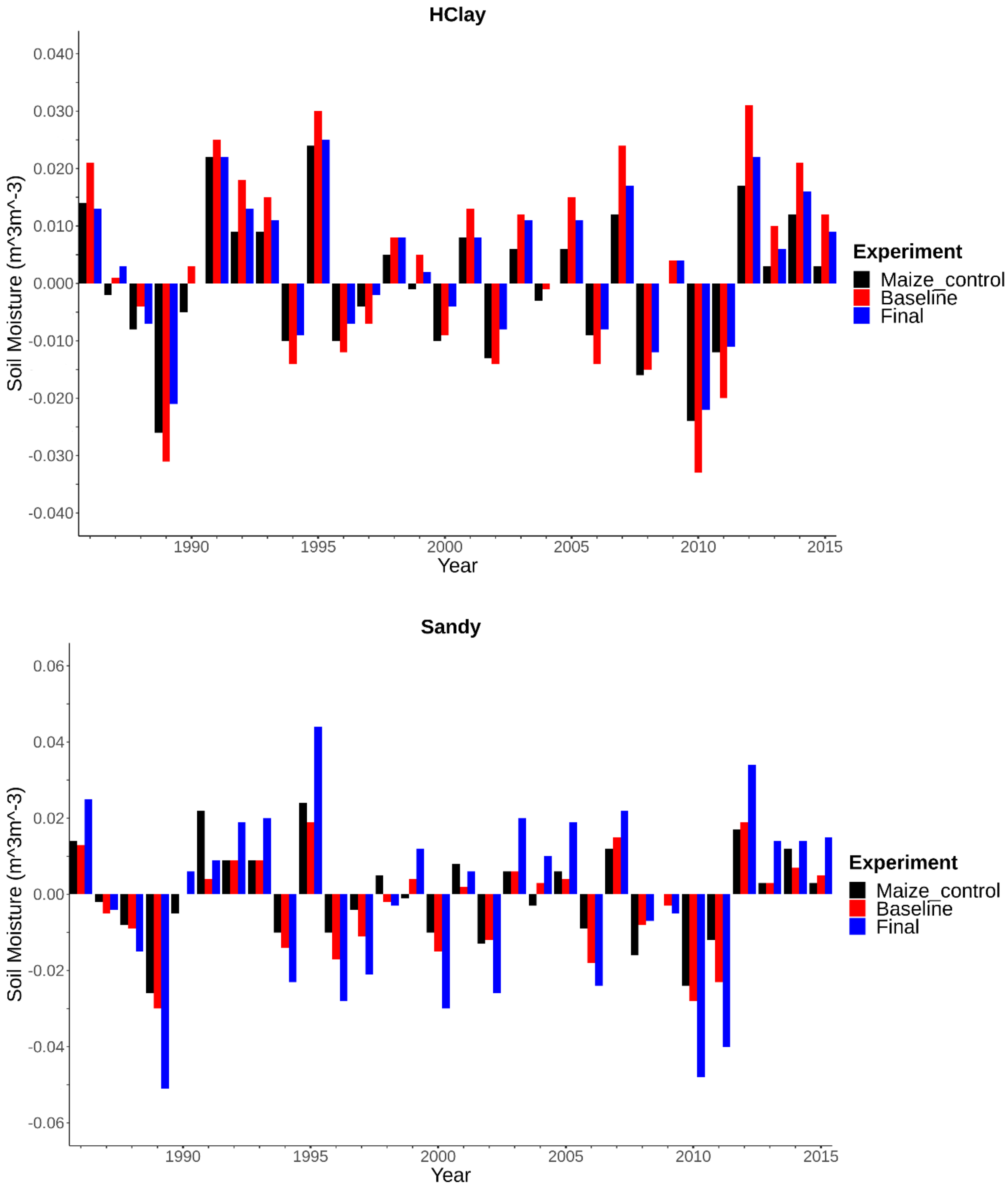


**Figure 4.3:** The interannual variability of volumetric soil moisture anomaly averaged over the growing season (1985-2015) for Maize_control (black), baseline (red), and final (blue) for heavy clay soil (top) and medium sandy soil (bottom). *In 2009 soil moisture anomaly Maize_control is zero.

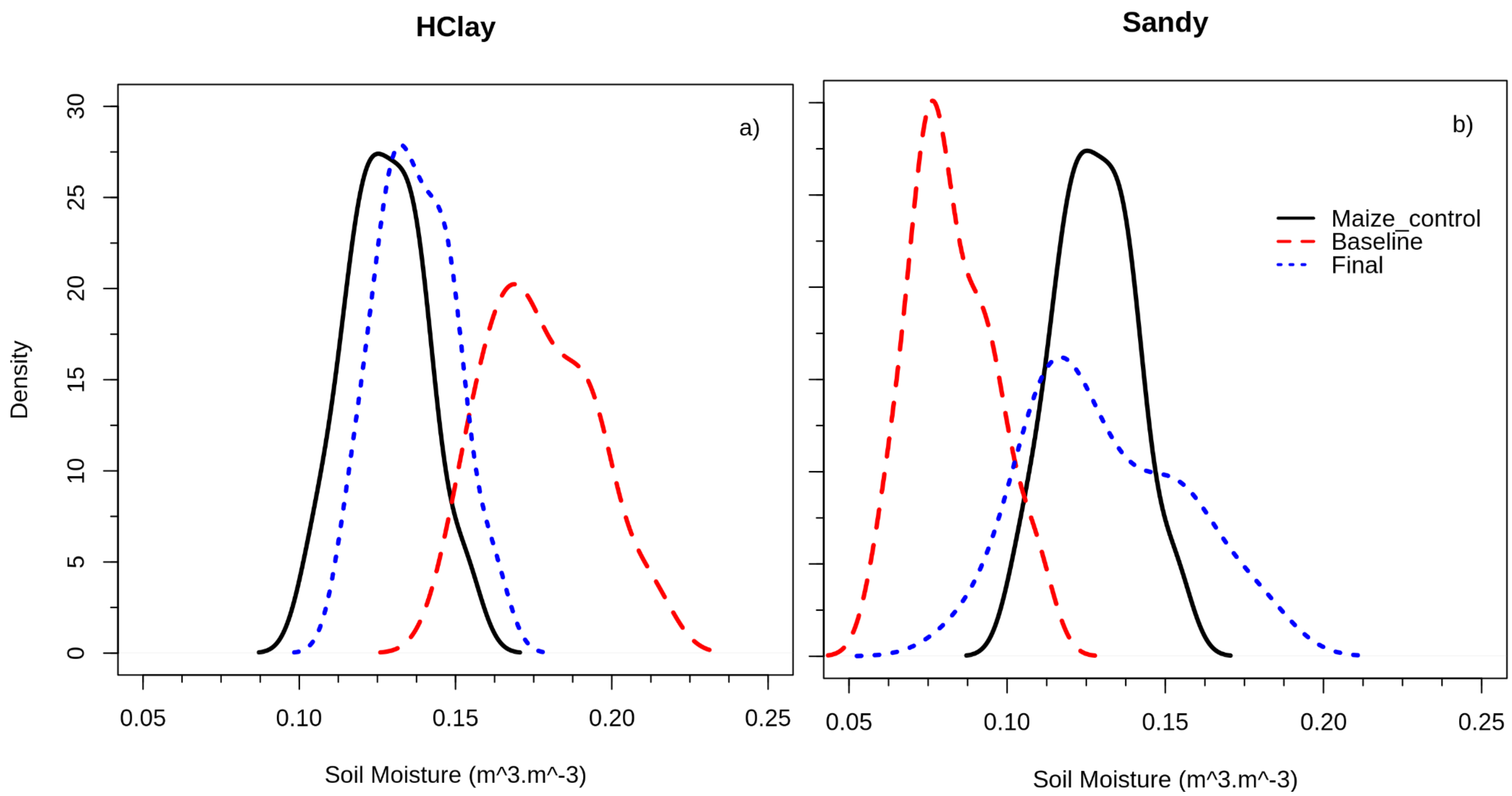


**Figure 4.4:** The probability density function of the simulated volumetric soil moisture ($m^3m^{-3}$) for Maize_control (black), heavy clay soil (a), and medium sandy soils (b) before (red) and after (blue) the data assimilation.

#### *4.2.2. Evapotranspiration*

Assimilating remotely sensed SM data into the model significantly improved both the mean and variability comparison between simulated evapotranspiration (ET) from the Maize_control and heavy clay soil experiments (Table 4.2, Figure 4.5a and Figure 4.6a). For example, it led to a relatively lower MBE (-48.83 mm) and RMSE (66.01 mm) compared to results from the baseline experiment (MBE = -62.26 mm and RMSE = 115.38 mm). Furthermore, the location and variability of the final experiments' distribution were also more comparable to that of Maize_control (D = 0.33, Table 4.2, and Figure 4.6a). As such, the variability in Maize_control ET can be explained by heavy clay soil ET which improved from 53% to 85%. Moreover, a greater proportion of ET distribution from the final experiment (Range = 80-580 mm) falls well within that of Maize_control (Range = 100-400 mm), even though a certain portion of its range fall on the higher end of the scale. Hence, the final experiment captured better the low to medium values of ET.

The main peak of the ET distribution from the heavy clay soil (final experiment) was also improved and got closer to that of Maize_control (0.0047 and 0.0068, respectively). Moreover, a greater proportion of ET distribution from the final experiment fell well within that of Maize_control, even though a certain portion of the range of final experiments' distribution falls on the higher end of the scale. The final experiment captured better the year-to-year variation of ET as simulated by Maize_control (R = 0.92). Nevertheless, overall it generally simulated relatively higher ET values compared to Maize_control (Figure 4.5a). This was particularly pronounced during the 1993 and 2002 growing seasons, respectively. Moreover, during the 1990 and 2004 growing season, this experiment maintained an above-normal ET, while the Maize_control simulated the opposite.

**Table 4.2:** The statistical analysis between the simulated evapotranspiration (mm) from Maize_control and heavy clay soil or medium sandy soil before and after data assimilation.

| **Statistical measure** | **HClay** | | **Sandy** | |
|---|---|---|---|---|
| | **Baseline** | **Final** | **Baseline** | **Final** |
| **MBE** | -62.26 | -48.83 | -75.39 | -103.71 |
| **RMSE** | 115.38 | 66.01 | 85.08 | 130.97 |
| **R** | 0.73 | 0.92 | 0.90 | 0.79 |
| **$R^2$** | 0.53 | 0.85 | 0.81 | 0.63 |
| **KS.test (D)** | 0.40 | 0.33 | 0.43 | 0.57 |

On the other hand, assimilating remotely sensed soil moisture data to the model with medium sandy soil improved neither the mean nor the variability comparison between the simulated ET by this model and the Maize_control experiment (Table 4.2, Figure 4.5b and Figure 4.6b). For example, it led to a relatively higher MBE (-103.71 mm) and RMSE (130.97 mm, Table 4.2). The ability of the model to capture the year-to-year variation of ET as simulated by Maize_control was also reduced. For instance, the correlation between the simulated ET from medium sandy soil and Maize_control decreased from 0.90 to 0.79, while the coefficient of determination decreased from 81% to 63% (for baseline and final experiment, respectively). Moreover, the final experiment simulated ET with higher variability (wider spread) and flatter distribution compared to the one from the Maize_control and baseline experiment (range = 0-680 mm and peak density of 0.0039, Figure 4.6b). As such, the difference in the location of the final experiment and Maize_control ET distributions was increased to 0.57.

The results show that the ET simulated by medium sandy soil experiment without data assimilation (baseline) was more comparable, in terms of both mean and variability, to the one simulated by Maize_control (Table 4.2, Figure 4.5b, and Figure 4.6b). For instance, even though the baseline experiment was poorly calibrated it simulated ET that is better correlated to that of Maize_control (R = 0.90). Moreover, it captured better the interannual variation of ET as simulated by Maize_control (Figure 4.5b). As such, the MBE (-75.89 mm) and RMSE (85.08 mm, Table 4.2) were relatively lower. Furthermore, about half of the range of the distribution of ET, with a peak of 0.0052, fell within that of Maize_control (D = 0.43, Table 4.2 and Figure 4.6b). Therefore, 81% of the variability in ET from Maize_control can be explained by the changes in ET from the baseline experiment.

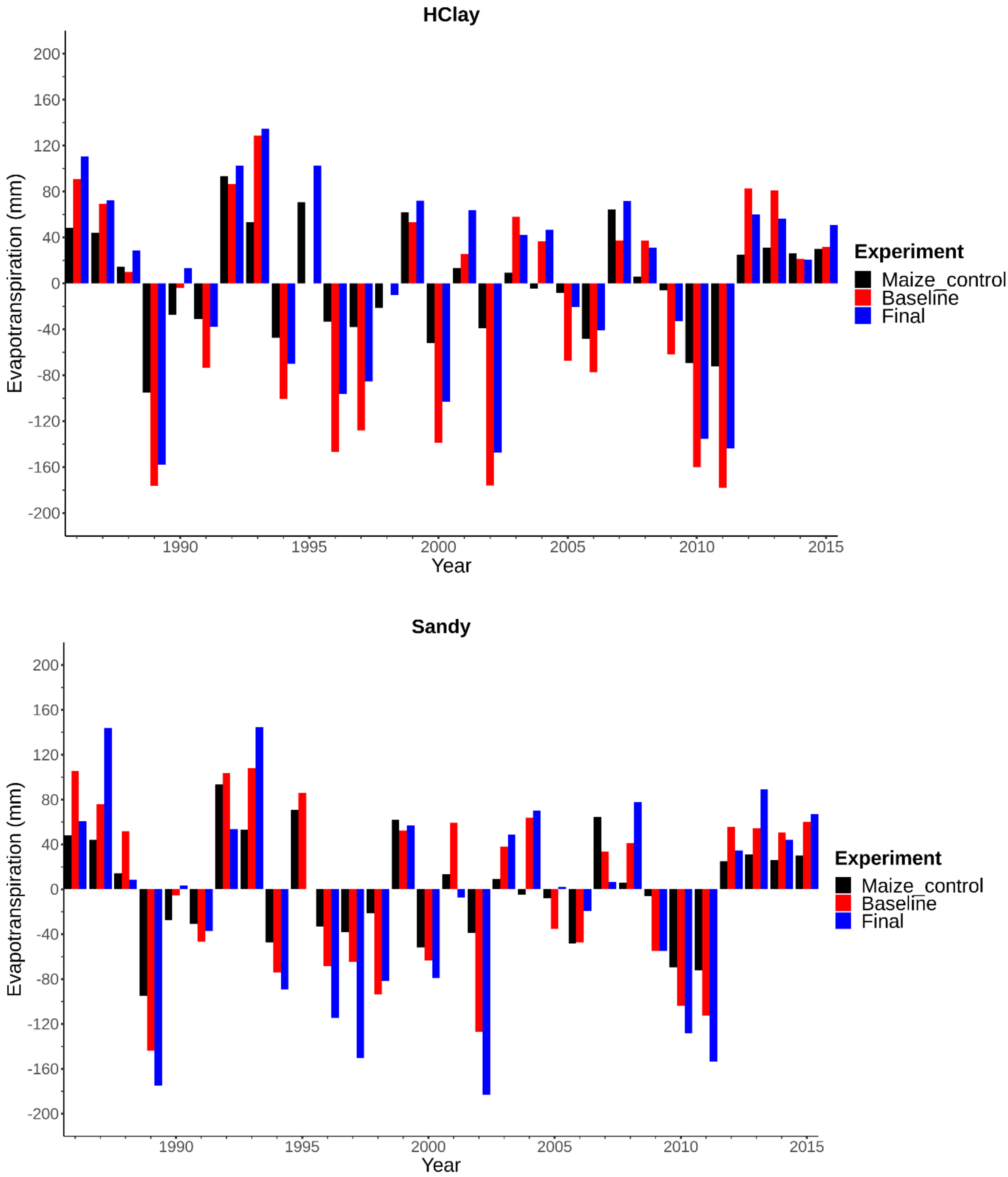


**Figure 4.5:** The interannual variability of evapotranspiration anomaly averaged over the growing season (1985-2015) for Maize_control (black), baseline (red), and final (blue) for heavy clay soil (top) and medium sandy soil (bottom).

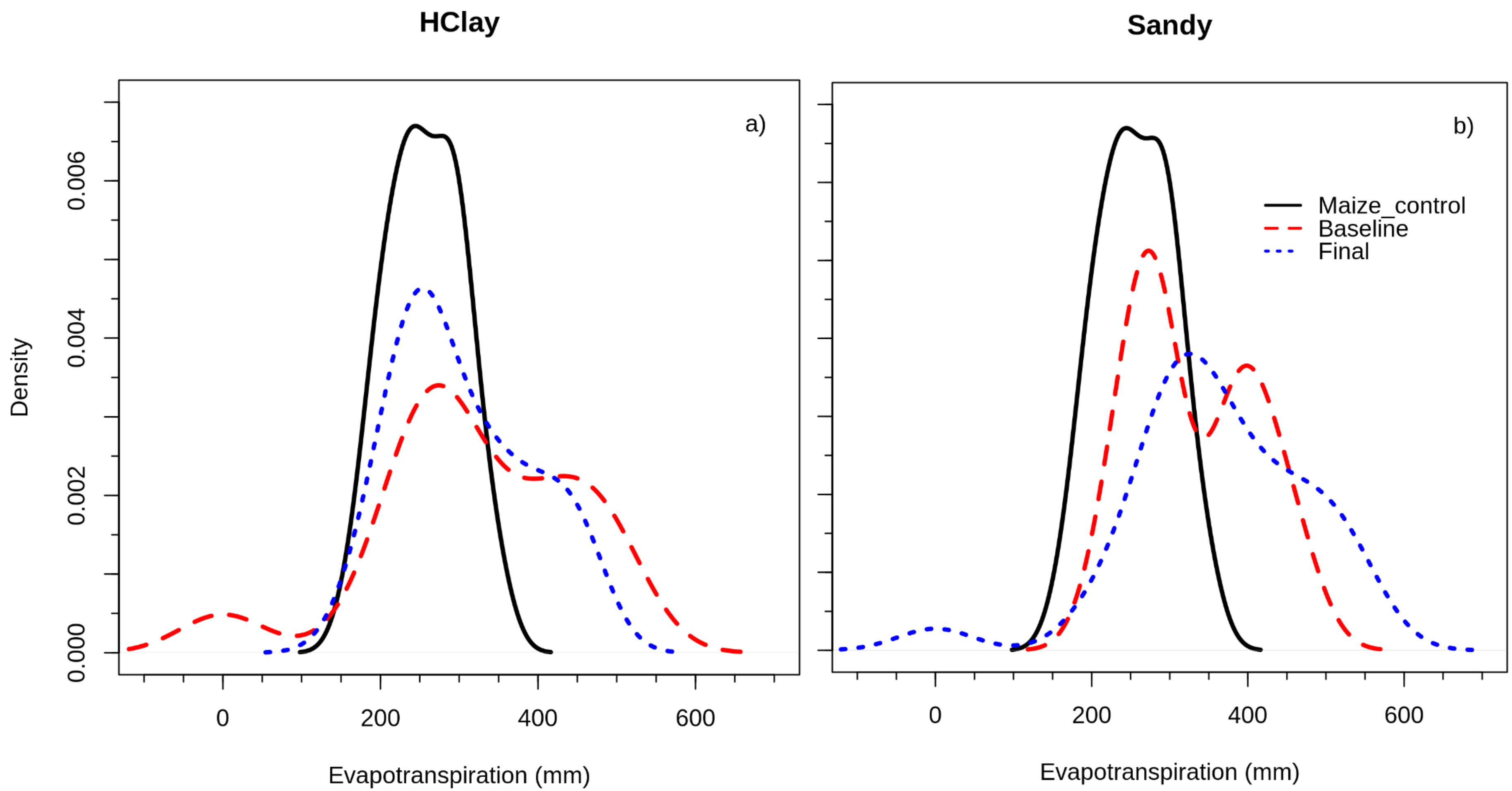


**Figure 4.6:** The probability density function of the simulated evapotranspiration (mm) averaged over the growing season (1985-2015) for Maize_control (black), heavy clay soil (a), and medium sandy soils (b) before (red) and after (blue) the data assimilation.

### *4.2.3. Maize yield*

The study also gave a measure of comparison with a model calibrated with extensive field data, by exploring the impacts of the above specific soil moisture improvement on maize yield. Regardless of the soil type used, data assimilation generally had great difficulty in improving the comparison between the simulated yield from final and Maize_control experiments (Table 4.3, Figure 4.7, and Figure 4.8). For example, when heavy clay soil was used, the model generally simulated yield that is relatively lower than the one simulated for the Maize_control and further maintained the same mean. As such, there is little difference between the MBE (1.75 and 1.73 ton/ha) and RMSE (2.28 and 2.03 ton/ha) for the baseline and final experiment, respectively (Table 4.3). Although, data assimilation slightly improved the variability of the simulated yield and captured better the interannual variation of yield (R = 0.67 and $R^2$ = 0.45), the final experiment of the model with heavy clay soil still maintained a broader spread and relatively low yield, with small range of yield falling within that of Maize_control (D = 0.63; Figure 4.8a).

**Table 4.3:** The statistical analysis between the simulated maize yield (ton/ha) from Maize_control and heavy clay soil or medium sandy soil before and after data assimilation.

| Statistical measure | HClay | | Sandy | |
|---|---|---|---|---|
| | Baseline | Final | Baseline | Final |
| **MBE** | 1.75 | 1.73 | -0.02 | 0.09 |
| **RMSE** | 2.28 | 2.03 | 0.71 | 1.16 |
| **R** | 0.58 | 0.67 | 0.76 | 0.61 |
| **$R^2$** | 0.34 | 0.45 | 0.57 | 0.37 |
| **KS.test (D)** | 0.60 | 0.63 | 0.23 | 0.23 |

On the other hand, assimilating remotely sensed soil moisture data into the model with medium sandy soil improved neither mean nor variability comparison between the maize yield simulated by final and Maize_control experiment (Table 4.3, Figure 4.7b and Figure 4.8b). Rather, it led to a relatively higher MBE (0.09 ton/ha) and RMSE (1.16 ton/ha) compared to the baseline MBE (-0.02 ton/ha) and RMSE (0.71 ton/ha). The difference between the location of the final and Maize_control experiments' yield distributions did not change (D = 0.23). Nevertheless, the final experiment still captured the interannual variability of yield as simulated by Maize_control (R= 0.61, Figure 4.7b) and a significant portion of its distribution falls well within the range of the Maize_control (Figure 4.8b). However, the distribution of the yield from the final experiment has a broader spread, which possibly resulted in a decrease of the correlation between the maize yield from the medium sandy soil and Maize_control (from 0.76 to 0.61 for baseline and final experiment, respectively) and the coefficient of determination (from 0.57 to 0.37 for baseline and final experiment, respectively). These results further suggest that the maize yield estimates from the generically calibrated medium sandy soil model (without data assimilation) were more comparable to the one produced by Maize_control.

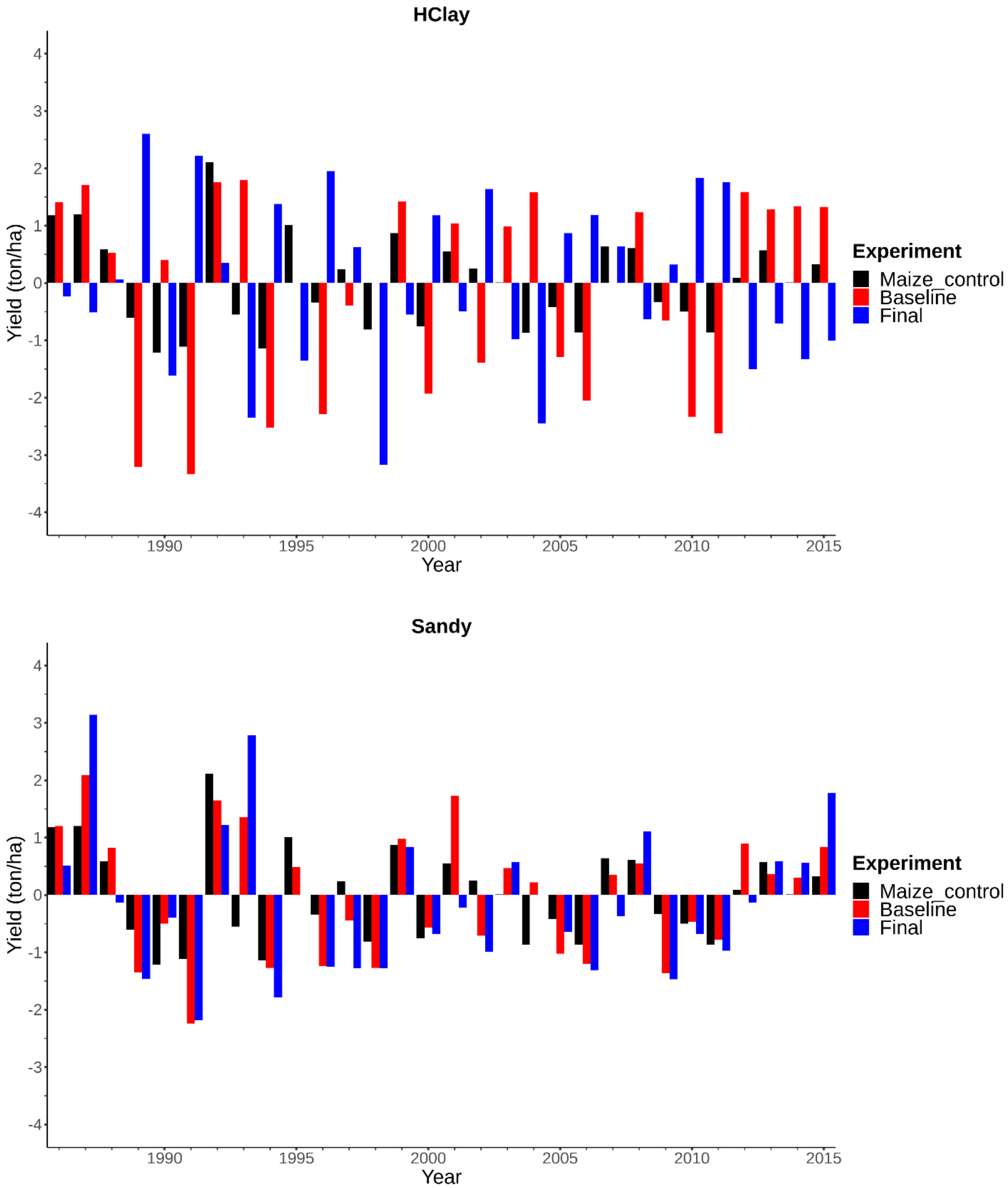


**Figure 4.7.** The interannual variability of maize yield (ton/ha) anomaly averaged over the growing season (1985-2015) for Maize_control (black), baseline (red), and final (blue) for heavy clay soil (top) and medium sandy soil (bottom).

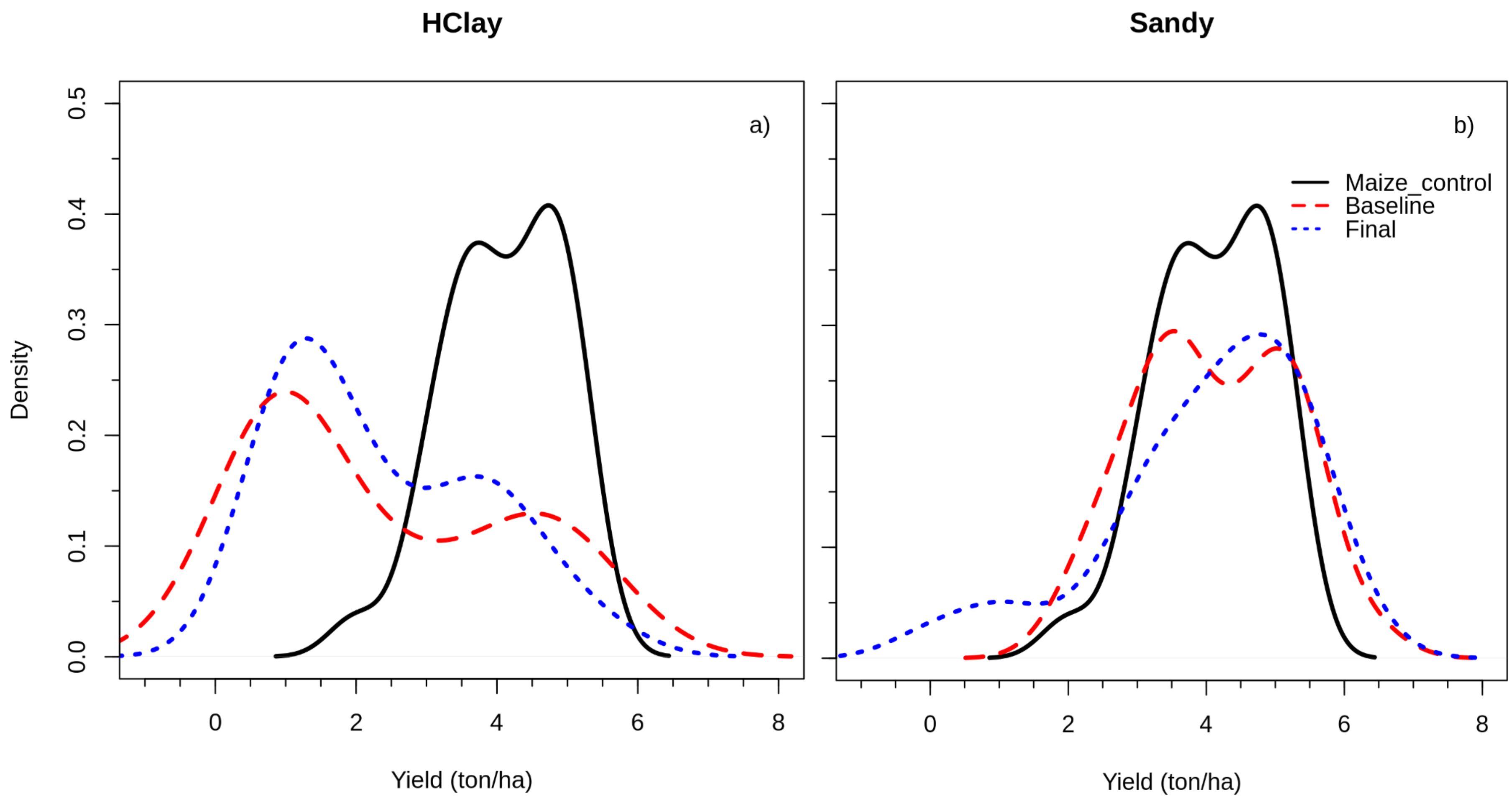


**Figure 4.8.** The probability density function of the simulated maize yield (ton/ha) averaged over the growing season (1985-2015) for Maize_control (black), heavy clay soil (a), and medium sandy soils (b) before (red) and after (blue) the data assimilation.

## 4.3. Discussion

The results show that the model experiments that were calibrated with generic soils (i.e. without the data assimilation) captured both the monthly and interannual variability of the observed remotely sensed volumetric soil moisture (SM) very well. Specifically, the crop model simulated relatively higher SM, regardless of the soil type, during the maize growing season and relatively low SM during the non-growing season (Figure 4.1). According to the O'Green (2013) and USDA (1999) global soil moisture regime classification, when the "soil is moist during most of the growing season and then followed by the prolonged dry season", it falls under the *ustic* SM regime. Therefore, the model was able to capture that the Bloemfontein district has an *ustic SM* regime as illustrated by global SM regime classification (Appendix A). Moreover, the results show that the model simulated relatively high soil moisture content between October and April, which corresponds to the months whereby the district receives most of its rainfall (Figure 3.2). This is because the SM dynamics (i.e. amount, availability, and flow) are influenced by the amount as well as the timing of rainfall. As such during the rainy season, a surplus in SM may occur when soils become saturated (O'Green, 2013).

Overall, the baseline model (without data assimilation) with heavy clay soil generally overestimated SM, thus leading to a relatively high MBE (-0.042 $m^3m^{-3}$) and RMSE (0.045 $m^3m^{-3}$, Table 4.4, Figure 4.9a). Whereas the medium sandy soil, overall, underestimated SM (MBE = 0.049 and RMSE = 0.053 $m^3m^{-3}$, Table 4.4, Figure 4.9b). These results are expected because according to O'Green (2013) fine-textured soils, like clay soils, have a more noteworthy capacity to hold water as they have a high distribution of small pore sizes (i.e. micropores). While sandy soils have a coarse texture and high distribution of large pore sizes, which ultimately limit the retention of water. Therefore, the model captured the differences in the properties of the two soils very well. Nevertheless, the overall correlation between the observed and simulated seasonal SM, from both soils, was 0.63 (Table 4.4).

Even though the model calibrated with two generic soils had produced opposite initial conditions, the study successfully improved the model calibration by assimilating remotely sensed SM. As such, the model captured better both the monthly and interannual variability of the observed SM. Moreover, regardless of soil type data assimilation significantly improved the model's ability to capture the overall observed growing season mean of SM (Table 4.4, Figure 4.9). Furthermore, the difference between the baseline and final SM for both soils was statistically significant (p-value= $2.85 \times 10^{-14}$, respectively). The final experiment for both heavy clay soil and medium sandy soil also had a relatively low MBE (0.002, 0.003 $m^3m^{-3}$) and RSME

(0.012, 0.018 $m^3m^{-3}$), respectively. Moreover, data assimilation improved the ability of the model with medium sandy soil to represent low to medium values of SM, while it was biased (underestimated) at capturing higher values.

The results from this study further showed that assimilating remotely sensed data into a crop model can possibly improve certain key model simulations. For example, when compared with an experiment that was calibrated with extensive ground-data (Maize_control), data assimilation in this study significantly improved SM. Moreover, this assimilation of SM had a larger effect on improving ET assimilation of only heavy clay soil was used. The poor improvement of ET for sandy soil may be attributed to the limited capacity of such a coarse-textured soil to retain water (i.e. moisture; O'Green, 2013). Regardless of the soil used, this study also struggled to improve maize yield comparisons. A study by Thorp et al. (2010) also found that assimilating remotely sensed data into a crop model can potentially improve certain model outputs, in their case, it was canopy weight and ET while the model had great difficulty at improving maize yield. This confirms that other simulated outputs, particularly yield, the importance of factoring more variables besides soil moisture which was used to adjust certain soil profile parameters. Furthermore, the study only changed three parameters of the soil profile (i.e. drainage lower limit (DLL), the drained upper limit (DUL), and the saturated upper limit (SAT)), hence adjusting other soil parameters such as the depth of the base layer, drainage rate, and runoff curve number and many other none soil parameters could have produced refined impacts. The other model outputs could have also been affected by the errors in remotely sensed SM data used in the study. For example, the spatial and temporal resolution of remotely sensed SM used to recalibrate the model (Son et al., 2016). As such, the greatest improvement was observed with SM comparison, and marginally with ET simulation for only heavy clay soil.

When compared to outputs simulated by Maize_control, assimilating remotely sensed data with the model generally improved the overall growing season mean of the outputs. However, it did not necessarily improve the model's ability to represent the variability of the outputs. This is particularly noticeable in the model's ability to represent SM. For example, the model with the heavy clay soil overall maintained relatively the same correlation (R = 0.62) and coefficient of determination ($R^2$ = 0.39) and reduced the MBE to 0.002 $m^3m^{-3}$ (Table 4.4; Figure 4.9). While assimilating remote sensing data into the model with medium sandy soil slightly improved the ability of this experiment to represent variability (R= 0.66 and $R^2$= 0.44) while it reduced MBE to 0.003 $m^3m^{-3}$.

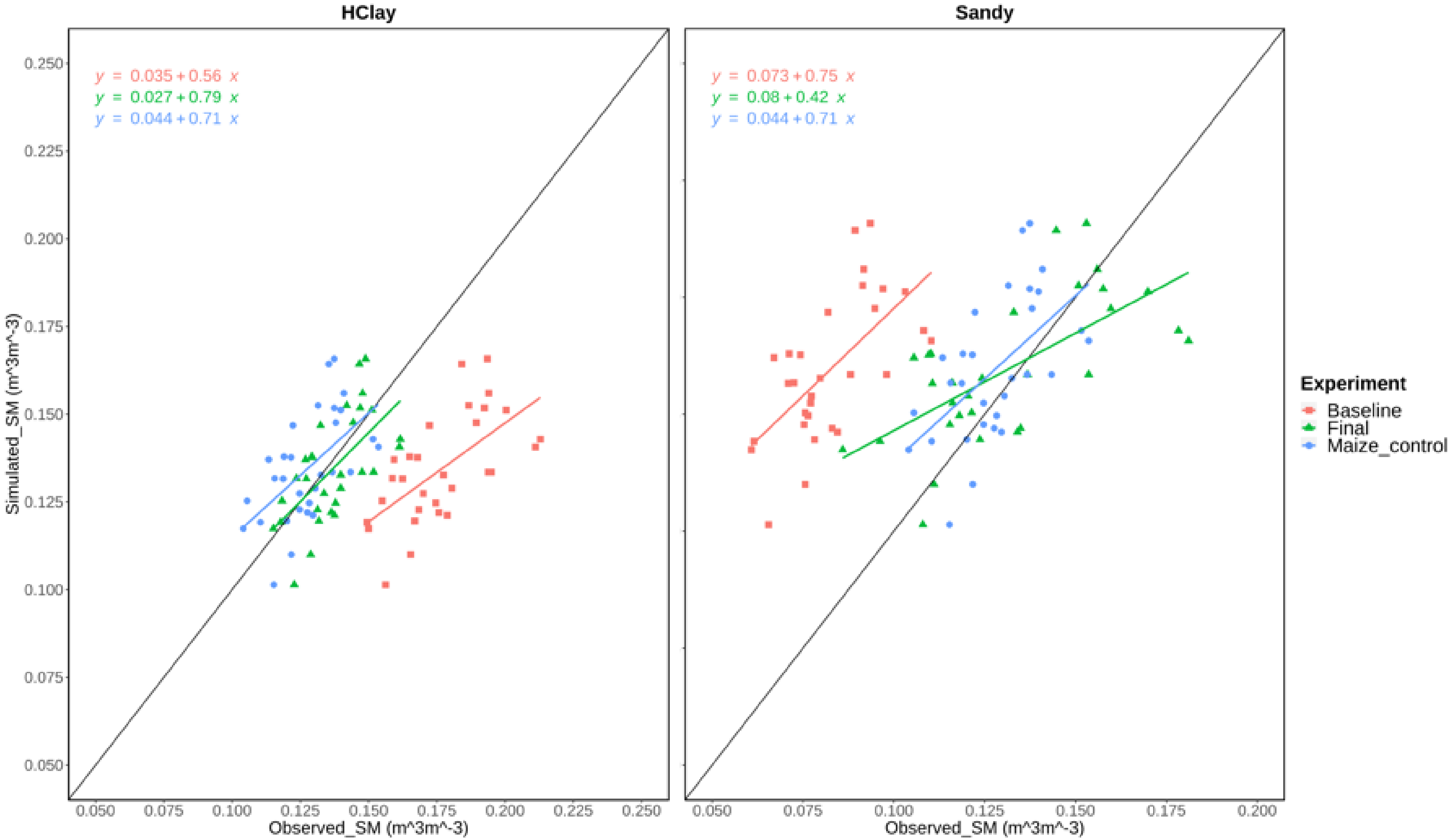


**Figure 4.9:** The relationship between the observed and simulated seasonal volumetric soil moisture ($m^3m^{-3}$) from the model with heavy clay soil (a) and medium sandy soil (b). The red and green dots and lines represent the baseline and final experiment for each soil type. While blue dots and line represent the Maize_control experiment.

**Table 4.4:** Statistical analysis of the relationship between observed and simulated volumetric soil moisture ($m^3m^{-3}$).

| Statistical measure | Maize_control | HClay | | Sandy | |
|---|---|---|---|---|---|
| | | Baseline | Final | Baseline | Final |
| **MBE** | 0.007 | 0.042 | 0.002 | 0.049 | 0.003 |
| **RMSE** | 0.015 | 0.045 | 0.012 | 0.053 | 0.018 |
| **R** | 0.57 | 0.63 | 0.62 | 0.63 | 0.66 |
| **$R^2$** | 0.33 | 0.40 | 0.39 | 0.40 | 0.44 |

These results were expected as the data assimilation approach was intended to minimise the mean bias error between the simulated and observed SM during the growing season. Similarly, one would have observed an improvement in variability if the study minimised a variability index. Therefore, the approach used in this study can be utilized to help improve the model's ability to capture long-term variations (i.e. mean) of certain outputs, including understanding long-term historical trends and/or the effect of predicted climate change on certain crop development and biophysical processes (i.e. ET). Furthermore, although a crop model integrated with remotely sensed data may not necessarily improve all the outputs (e.g. yield), it can still be used to disaggregate valuable knowledge embedded in the model. For example, there could be a high value in using the SM modelling capacity that was observed in this study, even though one would not have full confidence in the model's ability to simulate yield. Thus, allowing one to explore the influence of environmental change or crop management strategies on SM. Moreover, the improved SM simulations from this study can be used to inform other research as it can be used as input data in other models.

It is also important to note that even though Maize_control was calibrated with extensive field data; it still has its own biases. For example, when the simulated yield of the Maize_control experiment was compared to the observed district yield, it was found that the model generally overestimated maize yield (MBE = -1.85 and RMSE =2.10 ton/ha, Appendix B). However, this was likely anticipated because the crop model does not incorporate the yield-limiting factors, including weeds, pests, and diseases (Durand and Ferreira, 2017; Nagamani and Mariappan, 2017). Therefore, the simulated yield from the model is rather a reflection of potential yield other than actual yield. Furthermore, the soil moisture (SM) simulated from the Maize_control experiment also had some biases as well (Table 4.4 and Figure 4.9). Overall, this model slightly underestimated the observed SM (MBE = 0.007, RMSE = 0.015 $m^3m^{-3}$). More specifically, it relatively overestimated low-to-medium and underestimated the high values of SM. Contrary to the relationship between the observed and the simulated SM from the two soils (regardless of the experiment type), the correlation between Maize_control and observed SM is slightly lower (R= 0.57, Table 4.1). Moreover, only 33% of the total variance of the observed SM can be described by the SM simulated from the Maize_control experiment.

# *Chapter five: Conclusion*

Crop models can be a key component in addressing issues of global food security as they can be utilized to monitor and improve crop production. However, their use, particularly in isolated and rural areas, is commonly limited by the lack of sufficient and reliable field data. This data scarcity increases uncertainties in crop simulations, therefore, the outputs produced can be of limited value to agricultural advisors. However, some of this uncertainty can be mitigated by using remotely sensed data. As part of the continuing efforts to explore the potential of using remote sensing data to calibrate and validate crop models, as well as to integrate the two to improve model simulations, especially in data-limited areas, the study assessed how well a crop model assimilated with remotely sensed (RS) data performed compared with one calibrated with actual ground data. This was achieved by calibrating the DSSAT-CERES-Maize model using two generic soils (i.e. heavy clay soil and medium sandy soils), to firstly, measure soil moisture from 1985 to 2015 in Bloemfontein, before improving soil profile with RS data. The two generic soils were selected only on a literature basis and assuming a no field data condition, recognising that in South Africa, maize is mainly cultivated under sandy soil or clay/clay-loam soils. Intra- and inter-annual soil moisture served to guide the assimilation of the observed RS data into the model. The improvement was assessed primarily through the lens of SM simulation improvement from the knowingly generic set up to the final RS informed soil profile set up. The study also gave some measure of comparison with a model calibrated with extensive field data (Maize_control), and finally explored impacts of this specific soil moisture improvement, on other indicators; evapotranspiration (ET) and maize yield.

## 5.1. Summary of the key findings

Even though the model was calibrated with generic soil (baseline), it still captured both the monthly and interannual variation of soil moisture (SM) very well. However, overall the model with heavy clay soil overestimated SM, while the medium sandy soil underestimated SM during the growing season (October to March). This is because heavy clay soils have a higher capacity to hold water than sandy soils since they are fine-textured and have a higher distribution of micropores.

The study successfully improved the model calibration using remotely sensed SM, even though the two soils had produced opposite initial conditions. When compared to the observed data, assimilating remotely sensed data with the model significantly improved the intra- and inter-annual as well as the overall growing season mean of SM while maintaining the model's capacity to represent the variability of SM. Furthermore, when the simulated SM from the two soils was compared with a model calibrated with sufficient field data (Maize_control), it was found that data assimilation also significantly improved the model's ability to represent the growing season mean of SM. However, it did not improve the model's capacity to represent the variability of SM. This is an expected result as the approach was intended to minimise the mean base error between the simulated and observed SM during the growing season.

When the study evaluated how SM driven improvement impacted other key model outputs, it was found that the ability of data assimilation procedure to improve ET simulations depended on the soil type used. For example, data assimilation improved the comparison between the model with heavy clay soil and Maize_control, in terms of both mean and variability. However, when medium sandy soil was used, the comparison with Maize_control was reduced both in mean and variability. Furthermore, regardless of the soil type used, data assimilation did not improve the maize yield comparison in terms of both mean and variability (although a slight improvement in variability was observed when heavy clay soil was used). This confirms that yield and ET were influenced by other factors aside from soil moisture or the soil profile parameters.

Overall, the simple assimilation of remotely sensed data into a crop model method used in this study has the potential to improve and capture long -term variations (i.e. mean) of certain key model outputs (i.e. SM), while it is relatively poor at representing variability. Overall, the SM from the crop model which was assimilated with remotely sensed data compares very well with one calibrated with actual ground data, particularly in terms of the overall mean. As such, this approach can be adopted to understand historical and/or further effects of environmental change on crop production in data-limited areas. Hence, it can be concluded that RS data can be used together with a crop model to bias correct model inputs or improve the calibration of the model in the absence of observed field data. However, the biases of the RS data used must be carefully considered.

## 5.2. Recommendations for future studies

The study acknowledges the complexity of the model used, the cropping system being modelled, as well as the possible errors in the RS data used. Thus, to provide more vigorous and useful information to the agricultural policymakers, the study acknowledges that the current results can be improved in multiple ways. For example, one of the limitations of this study is that it only used one remote sensing data, which may have impacted the results.

Therefore, it recommends that similar future studies should use multiple remotely sensed data sets to complement these results (e.g. LAI, NDVI, etc.) in order to potentially consider the operationalization possible with higher resolution. Another limitation of this study is that it only adjusted the three soil parameters to reduce the mean bias error between the observations and simulations. As such, the study improved SM and the confidence in the SM produced, however, it improved marginally ET and did not improve yield. Hence, if one is interested in more than SM, one would need to recalibrate multiple variables. As a result, in the future, this study recommends that a sensitivity analysis test of the input parameters is conducted to identify the multiple variables that are very critical in determining the final output of interest (e.g. crop yield, evapotranspiration, water stress, nutrient stress, etc.). For example, according to a study by Fang et al. (2008), users can input multiple combinations of ecological conditions and crop management strategies into the DSSAT model, thus allowing them to test the sensitivity of intended output to certain parameters. The other limitation of this study is that it used a manual approach to recalibrate the model which is generally time-consuming and can be erroneous. Hence, the study recommends that the use of a more robust optimization process or algorithm (e.g. Powell or the particle swarm optimization) to adjust these initial parameters.

### 5.3. The importance of this study

This study showed that assimilating remotely sensed data into a generically calibrated CM can produce some model outputs that are comparable to those obtained using the traditional ground data calibration (particularly SM). As such, this piece of work can potentially open the opportunity to democratise the use of CMs and improve the capacity of local experts to use these models. Therefore, this approach can improve the meaning, relevance, and confidence of CMs outputs, even in data-limited areas. Even though field data will always be better at a small-scale level (e.g. the plant, its processes and the field management), remote sensing could bring so much in terms of spatial and temporal variability as well as administrative monitoring. Hence, regardless of the limitations of each of these methods, playing at each strength there is so much that can be achieved. Nevertheless, this framework can offer rural agricultural advisors decision support that is currently pending on high data for input, making it unavailable in many places. Thus, igniting an interest to access and use crop models, particularly in local small and rural agricultural communities where field data is inconsistently available.

The use of crop models is very important to better understand the local cropping systems and how they are affected by environmental conditions as well as management strategies. Currently, crop models are immensely useful and do very well in cases whereby they deal with western or commercial studies but could also benefit so much from being to handle small-holder farming systems and process. Thus, integrating these models with remote sensing data can attract and improve the use of CMs by small-holder farmers, extension offers, and researchers in data-limited area such as rural and remote areas.

# *References*


Abera, K., Crespo, O., Seid, J. and Mequanent, F., 2018. Simulating the impact of climate change on maize production in Ethiopia, East Africa. *Environmental Systems Research*, *7*(1), p.4.

AghaKouchak, A., Farahmand, A., Melton, F.S., Teixeira, J., Anderson, M.C., Wardlow, B.D. and Hain, C.R., 2015. Remote sensing of drought: Progress, challenges, and opportunities. *Reviews of Geophysics*, *53*(2), pp.452-480.

Akpalu, W., Hassan, R.M. and Ringler, C., 2008. Climate variability and maize yield in South Africa. *Environment and Production Technology Division. IFPRI Paper*, *843*. https://core.ac.uk/download/pdf/6337698.pdf

Albergel, C., De Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y. and Wagner, W., 2012. Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, *118*, pp.215-226.

Alexandratos, N. and Bruinsma, J., 2012. *World agriculture towards 2030/2050: the 2012 revision* (Vol. 12, No. 3). FAO, Rome: ESA Working paper.

Amekudzi, L., Yamba, E., Preko, K., Asare, E., Aryee, J., Baidu, M. and Codjoe, S., 2015. Variabilities in rainfall onset, cessation and length of rainy season for the various agro-ecological zones of Ghana. *Climate*, *3*(2), pp.416-434.

Araujo, J.A., Abiodun, B.J. and Crespo, O., 2016. Impacts of drought on grape yields in Western Cape, South Africa. *Theoretical and applied climatology*, *123*(1-2), pp.117-130.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D. and Brisson, N., 2013. Uncertainty in simulating wheat yields under climate change. *Nature climate change*, *3*(9), p.827.

Azam, S., Ali, M., Amin, M., Bidi, S. and Arif, M., 2007. Effect of plant population on maize hybrids. J. Agric. Biol. Sci., 2(1): 13-20

Baiphethi, M.N., Viljoen, M.F. and Kundhlande, G., 2003. *Creating agribusiness Opportunities for small scale Farmers in Thaba Nchu by introducing water harvesting Techniques: a profitability and risk Analysis* (No. 344-2016-15087).

Basche, A.D., Archontoulis, S.V., Kaspar, T.C., Jaynes, D.B., Parkin, T.B. and Miguez, F.E., 2016. Simulating long-term impacts of cover crops and climate change on crop production and environmental outcomes in the Midwestern United States. *Agriculture, Ecosystems & Environment*, *218,* pp.95-106.

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V. and Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sensing of Environment*, *184*, pp.668-681.

Bauböck, R., 2014. Simulating the yields of bioenergy and food crops with the crop modeling software BioSTAR: the carbon-based growth engine and the BioSTAR ET 0 method. *Environmental Sciences Europe*, *26*(1), p.1.

Beiragi, M.A., Khorasani, S.K., Shojaei, S.H., Dadresan, M., Khodadad Mostafavi, K. and Golbashy, M., 2011. A study on effects of planting dates on growth and Yield of 18 Corn Hybrids (Zea mays L.). Am. J. Experimen. Agric., 1(3): 110-120

Beletse, Y.G., Durand, W., Nhemachena, C., Crespo, O., Tesfuhuney, W.A., Jones, M.R., Teweldemedhin, M.Y., Gamedze, S.M., Bonolo, P.M., Jonas, S. and Walker, S., 2015. Projected impacts of climate change scenarios on the production of maize in southern Africa: an integrated assessment case study of the Bethlehem district, central Free State, South Africa.

Boedt, L.J.J. and Laker, M.C., 1985. The development of profile available water capacity models. WRC Report No. 98/1/85, Water Research Commission, Pretoria.

Bregaglio, S., Frasso, N., Pagani, V., Stella, T., Francone, C., Cappelli, G., Acutis, M., Balaghi, R., Ouabbou, H., Paleari, L. and Confalonieri, R., 2015. New multi-model approach gives good estimations of wheat yield under semi-arid climate in Morocco. *Agronomy for sustainable development*, *35*(1), pp.157-167.

Chakrabarti, S., Bongiovanni, T., Judge, J., Zotarelli, L. and Bayer, C., 2014. Assimilation of SMOS soil moisture for quantifying drought impacts on crop yield in agricultural regions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(9), pp.3867-3879.

Chipanshi, A.C., Chanda, R. and Totolo, O., 2003. Vulnerability assessment of the maize and sorghum crops to climate change in Botswana. *Climatic change*, *61*(3), pp.339-360.

Chivasa, W., Mutanga, O. and Biradar, C., 2017. Application of remote sensing in estimating maize grain yield in heterogeneous African agricultural landscapes: a review. *International Journal of Remote Sensing*, *38*(23), pp.6816-6845.

Courault, D., Demarez, V., Guérif, M., Le Page, M., Simonneaux, V., Ferrant, S. and Veloso, A., 2017. Contribution of remote sensing for crop and water monitoring. In *Land Surface Remote Sensing in Agriculture and Forest* (pp. 113-177).

Davis, J., Tavasci, D. and Marais, L., 2006. Fostering rural and local economic development in the Free State of South Africa. *Natural Resources Institute, University of Greenwich, UK*, pp.2-10.

Delecolle, R., Maas, S.J., Guerif, M., Baret, F., 1992. Remote sensing and crop production models: present trends. ISPRS J. Photogrammetry Remote Sens. 47, 145–161.

Department of Agriculture Fisheries and Forestry (DAFF), 2017. *Trends in the Agricultural Sector 2016*. Department of Agriculture Fisheries and Forestry, Directorate Communication Services, Pretoria, RSA.

Department of Agriculture Fisheries and Forestry (DAFF), 2019. *Trends in the Agricultural Sector 2018*. Department of Agriculture Fisheries and Forestry, Directorate Communication Services, Pretoria, RSA.

Descheemaeker, K., Oosting, S.J., Tui, S.H.K., Masikati, P., Falconnier, G.N. and Giller, K.E., 2016. Climate change adaptation and mitigation in smallholder crop-livestock systems in sub-Saharan Africa: a call for integrated impact assessments. *Regional Environmental Change*, *16*(8), pp.2331-2343.

Descheemaeker, K., Zijlstra, M., Masikati, FP., Crespo, O. and Tui, S.H.K., 2018. Effects of climate change and adaptation on the livestock component of mixed farming systems: A modelling study from semi-arid Zimbabwe. *Agricultural Systems*, *159*, pp.282-295.

Di Paola, A., Valentini, R. and Santini, M., 2016. An overview of available crop growth and yield models for studies and assessments in agriculture. *Journal of the Science of Food and Agriculture*, *96*(3), pp.709-714.

Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J. and Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote sensing of environment*, *92*(4), pp.548-559.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A. and Haas, E., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the-art and future directions. *Remote Sensing of Environment*, *203*, pp.185-215.

du Plessis, J., 2003. Maize Production, Directorate Agricultural Information Services Department of Agriculture in cooperation with ARC-Grain Crops Institute Publication. Available at http://www.arc.agric.za/arc-gci/Fact%20Sheets%20Library/Maize%20Production.pdf, last accessed (17 Oct 2019)

Durand, W. and Ferreira, F., 2017. Better crop estimates with satellites and remote sensing. Oilseeds Focus, 3(4), pp. 28-29.

El Sharif, H., Wang, J. and Georgakakos, A.P., 2015. Modeling regional crop yield and irrigation demand using SMAP type of soil moisture data. *Journal of Hydrometeorology*, *16*(2), pp.904-916.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J. and Kimball, J., 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), pp.704-716.

Fang, H., Liang, S. and Hoogenboom, G., 2011. Integration of MODIS LAI and vegetation index products with the CSM–CERES–Maize model for corn yield estimation. *International Journal of Remote Sensing*, *32*(4), pp.1039-1065.

Fang, H., Liang, S., Hoogenboom, G., Teasdale, J. and Cavigelli, M., 2008. Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *International Journal of Remote Sensing*, *29*(10), pp.3011-3032.

Fang, L., Hain, C.R., Zhan, X. and Anderson, M.C., 2016. An inter-comparison of soil moisture data products from satellite remote sensing and a land surface model. *International journal of applied earth observation and geoinformation*, *48*, pp.37-50.

Fanourakis, G.C., 2012. Estimating soil plasticity properties from pedological data. *Journal of the South African Institution of Civil Engineering*, *54*(2), pp.117-125.

FAO., 2018. *FAOSTAT database*. Rome: FAO. Retrieved from http://www.fao.org/faostat/en/#data

Feng, H., 2016. Individual contributions of climate and vegetation change to soil moisture trends across multiple spatial scales. *Scientific Reports*, *6*, p.32782.

Free State Province., 2007. Reviewed Free State Growth and Development Strategy, 2007-2014. Approved 20 March 2007. South Africa.

Fu, B., Wang, J., Chen, L. and Qiu, Y., 2003. The effects of land use on soil moisture variation in the Danangou catchment of the Loess Plateau, China. *Catena*, *54*(1-2), pp.197-213.

García-Vila, M. and Fereres, E., 2012. Combining the simulation crop model AquaCrop with an economic model for the optimization of irrigation management at farm level. *European Journal of Agronomy*, *36*(1), pp.21-31.

Gbetibouo, G.A., Hassan, R.M. and Ringler, C., 2010. Modelling farmers' adaptation strategies for climate change and variability: The case of the Limpopo Basin, South Africa. *Agrekon*, *49*(2), pp.217-234.

Gruber, A., Dorigo, W.A., Crow, W. and Wagner, W., 2017. Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), pp.6780-6792.

Haarhoff, S.J., Kotzé, T.N. and Swanepoel, P.A., 2020. A prospectus for sustainability of rainfed maize production systems in South Africa. *Crop Science*, *60*(1), pp.14-28.

Hammond, J., Fraval, S., Van Etten, J., Suchini, J.G., Mercado, L., Pagella, T., Frelat, R., Lannerstad, M., Douxchamps, S., Teufel, N. and Valbuena, D., 2017. The Rural Household Multi-Indicator Survey (RHoMIS) for rapid characterisation of households to inform climate smart agriculture interventions: Description and applications in East Africa and Central America. *Agricultural Systems*, *151*, pp.225-233.

He, M., Kimball, J.S., Maneta, M.P., Maxwell, B.D., Moreno, A., Beguería, S. and Wu, X., 2018. Regional Crop Gross Primary Productivity and Yield Estimation Using Fused Landsat-MODIS Data. Remote Sensing, 10(3), p.372.

Hertel, T.W. and Rosch, S.D., 2010. *Climate change, agriculture, and poverty*. The World Bank.

Hoell, James, P. W. Stackhouse, Jr., D.J. Westberg, W. S. Chandler, A. J. Barnett, T. Zhang, 2015: A Global Agroclimatology Data Base: Thirty Plus Years of Satellite-Derived Solar Insolation and Assimilation Model Meteorological Parameters. ASA, CSSA, and SSSA International Annual Meeting, November 15-18, Minneapolis, Minnesota (Poster Number 310).

Ines, A.V., Das, N.N., Hansen, J.W. and Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, *138*, pp.149-164.

Integrated Development Plan (IDP) for 2018/19, 2018. Mangaung Metropolitan Municipality. Free States Province, RSA

Jacquette, E., Al Bitar, A., Mialon, A., Kerr, Y., Quesney, A., Cabot, F. and Richaume, P., 2010, October. SMOS CATDS level 3 global products over land. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XII* (Vol. 7824, p. 78240K). International Society for Optics and Photonics.

Jamali, S., Seaquist, J., Ardö, J. and Eklundh, L., 2011. Investigating temporal relationships between rainfall, soil moisture and MODIS-derived NDVI and EVI for six sites in Africa. *Savanna*, *21*(547550), p.38.

Jin, M., Liu, X., Wu, L. and Liu, M., 2015. An improved assimilation method with stress factors incorporated in the WOFOST model for the efficient assessment of heavy metal stress levels in rice. *International Journal of Applied Earth Observation and Geoinformation*, *41*, pp.118-129.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G. and Wang, J., 2018. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, *92*, pp.141-152.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J. and Ritchie, J.T., 2003. The DSSAT cropping system model. *European Journal of Agronomy*, *18*(3-4), pp.235-265.

Jones, M.R., Singels, A. and Ruane, A.C., 2015. Simulated impacts of climate change on water use and yield of irrigated sugarcane in South Africa. *Agricultural Systems*, *139*, pp.260-270.

Karuaihe, R.N., Wang, H.H. and Young, D.L., 2006. *Weather-based crop insurance contracts for African Countries* (No. 1004-2016-78902).

Kasampalis, D., Alexandridis, T., Deva, C., Challinor, A., Moshou, D. and Zalidis, G., 2018. Contribution of remote sensing on crop models: a review. *Journal of Imaging*, *4*(4), p.52.

Knox J.W., Rodriguez Diaz J.A., Nixon D.J., Mkhwananzi M., 2010. A preliminary assessment of climate change impacts on sugarcane in Swaziland. Agricultural Systems 103: 63-72.

Launay, M. and Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture, ecosystems & environment*, *111*(1-4), pp.321-339.

Leroux, L., Baron, C., Castets, M., Escorihuela, M.J., Diouf, A.A., Bégué, A. and Lo Seen, D., 2017. Estimating maize grain yield in scarce field-data environment: an approach combining remote sensing and crop modelling in Burkina Faso.

Li, H., Qiu, J., Wang, L., Tang, H., Li, C. and Van Ranst, E., 2010. Modelling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat–maize rotation system in China. *Agriculture, ecosystems & environment*, *135*(1-2), pp.24-33.

Li, Y., Grimaldi, S., Walker, J.P. and Pauwels, V., 2016. Application of remote sensing data to constrain operational rainfall-driven flood forecasting: a review. *Remote Sensing*, *8*(6), p.456.

Liu, X.B., Zhang, X.Y., Wang, Y.X., Sui, Y.Y., Zhang, S.L., Herbert, S.J. and Ding, G., 2010. Soil degradation: a problem threatening the sustainable development of agriculture in Northeast China. *Plant, Soil and Environment*, *56*(2), pp.87-97.

Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A., Wagner, W., McCabe, M.F., Evans, J.P. and Van Dijk, A.I.J.M., 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment*, *123*, pp.280-297.

Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I. and Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agriculture, Ecosystems & Environment*, *94*(2), pp.205-220.

Lobell, D.B., Field, C.B., Cahill, K.N. and Bonfils, C., 2006. Impacts of future climate change on California perennial crop yields: Model projections with climate and crop uncertainties. *Agricultural and Forest Meteorology*, *141*(2-4), pp.208-218.

Loew, A., Stacke, T., Dorigo, W., De Jeu, R. and Hagemann, S., 2013. Potential and limitations of multidecadal satellite moisture observations for climate model evaluation studies. *Hydrology and Earth System Sciences*, *17*, pp.3523-3542.

Ma, H., Huang, J., Zhu, D., Liu, J., Su, W., Zhang, C. and Fan, J., 2013. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST–ACRM model with Ensemble Kalman Filter. *Mathematical and Computer Modelling*, *58*(3-4), pp.759-770.

Mabhaudhi, T., Chibarabada, T. and Modi, A., 2016. Water-food-nutrition-health nexus: Linking water to improving food, nutrition and health in Sub-Saharan Africa. *International journal of environmental research and public health*, *13*(1), p.107.

Mabhaudhi, T., Chibarabada, T.P., Chimonyo, V.G.P. and Modi, A.T., 2018. Modelling climate change impact: a case of bambara groundnut (Vigna subterranea). *Physics and Chemistry of the Earth, Parts A/B/C*, *105*, pp.25-31.

MacVicar, C.N., 1977. *Soil classification. A binomial system for South Africa*. Department of Agricultural Technical Services.

Maes, W.H. and Steppe, K., 2012. Estimating evapotranspiration and drought stress with ground-based thermal remote sensing in agriculture: a review. *Journal of Experimental Botany*, *63*(13), pp.4671-4712.

Mashaba, Z., Chirima, G., Botai, J.O., Combrinck, L., Munghemezulu, C. and Dube, E., 2017. Forecasting winter wheat yields using MODIS NDVI data for the Central Free State region. *South African Journal of Science*, *113*(11-12), pp.1-6.

Mati, B.M., 2000. The influence of climate change on maize production in the semi-humid–semi-arid areas of Kenya. *Journal of Arid Environments*, *46*(4), pp.333-344.

Messina, C.D., Podlich, D., Dong, Z., Samples, M. and Cooper, M., 2010. Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. *Journal of experimental botany*, *62*(3), pp.855-868.

Mishra, V., Cruise, J.F., Mecikalski, J.R., Hain, C.R. and Anderson, M.C., 2013. A remote-sensing driven tool for estimating crop stress and yields. *Remote Sensing*, *5*(7), pp.3331-3356.

Moeletsi, M.E. and Walker, S., 2012. Evaluation of NASA satellite and modelled temperature data for simulating maize water requirement satisfaction index in the Free State Province of South Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, *50*, pp.157-164.

Moeletsi, M.E., 2010. *Agroclimatological risk assessment of rainfed maize production for the Free State Province of South Africa* (Doctoral dissertation, University of the Free State).

Molua, E.L. and Lambi, C.M., 2006. Assessing the impact of climate on crop water use and crop water productivity: The CROPWAT analysis of three districts in Cameroon. *University of Pretoria: Pretoria, South Africa*, p.44.

Moriondo, M., Maselli, F. and Bindi, M., 2007. A simple model of regional wheat yield based on NDVI data. *European Journal of Agronomy*, *26*(3), pp.266-274.

Motavalli, P.P., Stevens, W.E. and Hartwig, G., 2003. Remediation of subsoil compaction effects on corn N availability by deep tillage and application of poultry manure in a sandy-textured soil. Soil Till. Res., 71: 121-31.

Murthy, V.R.K., 2004. Crop growth modeling and its applications in agricultural meteorology. *Satellite remote sensing and GIS applications in agricultural meteorology*, *235*.

Murthy, V.R.K., 2011. Crop growth modeling and its applications in agricultural meteorology. Satell. Remote Sens. *GIS Appl. Agric. Meteorol*, *1*, pp.235-261.

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S. and Wagner, W., 2009. An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), pp.1999-2013.

Nagamani, K. and Mariappan, V.N., 2017. Biosciences and Plant Biology. *Int. J. Curr. Res. Biosci. Plant Biol*, *4*(8), pp.80-92.

NASA, 2014. Prediction of Worldwide Energy Resource (POWER): Agroclimatology. [Available online at https://power.larc.nasa.gov/data-access-viewer/].

Ncube, M., Madubula, N., Ngwenya, H., Zinyengere, N., Zhou, L., Francis, J., Mthunzi, T., Olivier, C. and Madzivhandila, T., 2016. Climate change, household vulnerability and smart agriculture: The case of two South African provinces. *Jàmbá: Journal of Disaster Risk Studies*, *8*(2).

Nearing, G.S., Crow, W.T., Thorp, K.R., Moran, M.S., Reichle, R.H. and Gupta, H.V., 2012. Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment. *Water Resources Research*, *48*(5).

O'Geen, A. T., 2013. Soil Water Dynamics. *Nature Education Knowledge* 4(5):9

Owe, M., de Jeu, R. and Holmes, T., 2008. Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research: Earth Surface*, *113*(F1).

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J. and Trnka, M., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *European Journal of Agronomy*, *35*(3), pp.103-114.

Peng, J. and Loew, A., 2017. Recent advances in soil moisture estimation from remote sensing.

Peng, J., Niesel, J., Loew, A., Zhang, S. and Wang, J., 2015. Evaluation of satellite and reanalysis soil moisture products over Southwest China using ground-based measurements. *Remote sensing*, *7*(11), pp.15729-15747.

Pswarayi, A. and Vivek, B.S., 2007. Combining ability amongst CIMMYT's early maturing maize (Zea mays L.) germplasm under stress and non-stress conditions and identification of testers. Eupytica, 162: 353-362.

Ramadoss, M., Birch, C.J., Carberry, P.S. and Robertson, M., 2004, September. Water and high temperature stress effects on maize production. In *Proceedings of the 4th International Crop Science Congress, Brisbane, Australia.[* http://www.regional.org.au/au/asa/2004/poster/1/3/3/1037_madhiyazhaganr.htm] 17/10/2019

Rao, J.K. and Ranga, A.N.G., 2001. Remote Sensing and Agriculture. *Kisan World*, *28*, pp.49-50.

Rauff, K.O. and Bello, R., 2015. A review of crop growth simulation models as tools for agricultural meteorology. *Agricultural Sciences*, *6*(09), p.1098.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. *PloS one*, *8*(6), p.e66428.

Rembold, F., Atzberger, C., Savin, I. and Rojas, O., 2013. Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, *5*(4), pp.1704-1733.

Rivington, M. and Koo, J., 2010. Report on the meta-analysis of crop modelling for climate change and food security survey.

Rosenzweig, C. and Hillel, D., 2015. Handbook of Climate Change and Agroecosystems The Agricultural Model Intercomparison and Improvement Project (AgMIP) Integrated Crop and Economic Assessments — Joint Publication with American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America.

Sangoi, L., 2001. Understanding plant density effects on maize growth and development: an important issue to maximize grain yield. *Ciência rural*, *31*(1), pp.159-168.

Schulze, R.E. and Durand, W., 2016. Maize Production in South Africa and Climate Change. In: Schulze, R.E. (ed.) Handbook for Farmers, Officials and Other Stakeholders on Adaptation to Climate Change in the Agriculture Sector within South Africa. Section C: Crops in South Africa and Climate Change, Chapter C1.

Schulze, R.E., 2003. Implications of "Incorrect" seasonal climate forecast scenarios on maize yields in South Africa: A sensitivity study. School of Bioresources Engineering and Environmental Hydrology, University of Natal, Pietermaritzburg, RSA. ACRU-cons Report, 43. 30p.

Schulze, R.E., Hutson, J.L. and Cass, A., 1985. Hydrological characteristics and properties of soils in Southern Africa 2: Soil water retention models. *Water SA*, *11*(3), pp.129-136.

Singels, A., 1992. Evaluating wheat planting strategies using a growth model. *Agricultural systems*, *38*(2), pp.175-184.

Singels, A., Annandale, J.G., Jager, J.D., Schulze, R.E., Inman-Bamber, N.G., Durand, W., Rensburg, L.V., Heerden, P.V., Crosby, C.T., Green, G.C. and Steyn, J.M., 2010.

Modelling crop growth and crop water relations in South Africa: Past achievements and lessons for the future. *South African Journal of Plant and Soil*, *27*(1), pp.49-65.

Son, N.T., Chen, C.F., Chen, C.R., Chang, L.Y. and Chiang, S.H., 2016. Rice Yield Estimation Through Assimilating Satellite Data Into a Crop Simulation Model. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *8*.

Stats, S.A., 2011. Statistics South Africa. *Formal census*.

Steduto, P., 2003. Biomass water-productivity. *Comparing the growth-Engines of Crop Models. FAO expert consultation on crop water productivity under deficient water supply*, pp.26-28.

Subedi, K.D. and Ma, B.L., 2009. Assessment of some major yield-limiting factors on maize production in a humid temperate environment. *Field crops research*, *110*(1), pp.21-26.

Tadesse, T., Assefa, A., Liben, M. and Tadesse, Z., 2013. Effects of nitrogen split-application on productivity, nitrogen use efficiency and economic benefits of maize production in Ethiopia. *International Journal of Agricultural Policy and Research*, *1*(4), pp.109-115.

Thornton, P.K., Jones, P.G., Ericksen, P.J. and Challinor, A.J., 2011. Agriculture and food systems in sub-Saharan Africa in a 4 C+ world. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1934), pp.117-136.

Thorp, K.R., Hunsaker, D.J. and French, A.N., 2010. Assimilating leaf area index estimates from remote sensing into the simulations of a cropping systems model. *Transactions of the ASABE*, *53*(1), pp.251-262.

Tripathy, R., Chaudhari, K.N., Mukherjee, J., Ray, S.S., Patel, N.K., Panigrahy, S. and Parihar, J.S., 2013. Forecasting wheat yield in Punjab state of India by combining crop simulation model WOFOST and remotely sensed inputs. *Remote sensing letters*, *4*(1), pp.19-28.

USDA. 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. *Agriculture Handbook Number*. *436*. U.S. Government Printing Office Washington. DC.

Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E. and Ertl, M., 2012. Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals)*, *7*, pp.315-321.

Walker, N.J. and Schulze, R.E., 2006. An assessment of sustainable maize production under different management and climate scenarios for smallholder agro-ecosystems in KwaZulu-Natal, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(15-16), pp.995-1002.

Wang, H., Zhu, Y., Li, W., Cao, W. and Tian, Y., 2014. Integrating remotely sensed leaf area index and leaf nitrogen accumulation with RiceGrow model based on particle swarm optimization algorithm for rice grain yield assessment. *Journal of Applied Remote Sensing*, *8*(1), p.083674.

Werner, A., Dölling, S., Jarfe, A., Kühn, J., Pauly, J. and Roth, R., 2000. Deriving maps of yield-potentials with crop models, site information and remote sensing. In *Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, Minnesota, USA, 16-19 July, 2000* (pp. 1-20). American Society of Agronomy.

Wessels, K.J., Prince, S.D., Zambatis, N., MacFadyen, S., Frost, P.E. and Van Zyl, D., 2006. Relationship between herbaceous biomass and 1-km2 Advanced Very High Resolution Radiometer (AVHRR) NDVI in Kruger National Park, South Africa. *International Journal of Remote Sensing*, *27*(05), pp.951-973.

Wilks, D.S., 1995. Statistical methods in the atmospheric sciences, 1995. *Library of Cataloging-in-publication, Academic Press: San Diego, CA*, p.465.

Xu, H., Twine, T.E. and Girvetz, E., 2016. Climate change and maize yield in Iowa. *PloS one*, *11*(5), p.e0156083.

Yao, F., Tang, Y., Wang, P. and Zhang, J., 2015. Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China Plain. *Physics and Chemistry of the Earth, Parts A/B/C*, *87*, pp.142-152.

Yin, X., Struik, P.C. and Kropff, M.J., 2004. Role of crop physiology in predicting gene-to-phenotype relationships. *Trends in plant science*, *9*(9), pp.426-432.

Yuan, W., Cai, W., Nguy-Robertson, A.L., Fang, H., Suyker, A.E., Chen, Y., Dong, W., Liu, S. and Zhang, H., 2015. Uncertainty in simulating gross primary production of cropland ecosystem from satellite-based models. *Agricultural and Forest Meteorology*, *207*, pp.48-57.

Yuping, M., Shili, W., Li, Z., Yingyu, H., Liwei, Z., Yanbo, H. and Futang, W., 2008. Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, *10*(4), pp.426-437.

Zinyengere, N., Crespo, O., Hachigonta, S. and Tadross, M., 2014. Local impacts of climate change and agronomic practices on dry land crops in Southern Africa. *Agriculture, ecosystems & environment*, *197*, pp.1-10.

Zinyengere, N., Crespo, O., Hachigonta, S. and Tadross, M., 2015. Crop model usefulness in drylands of southern Africa: an application of DSSAT. *South African Journal of Plant and Soil*, *32*(2), pp.95-104.

Zougmoré, R., Partey, S., Ouédraogo, M., Omitoyin, B., Thomas, T., Ayantunde, A., Ericksen, P., Said, M. and Jalloh, A., 2016. Toward climate-smart agriculture in West Africa: a review of climate change impacts, adaptation strategies and policy developments for the livestock, fishery and crop production sectors. *Agriculture & Food Security*, *5*(1), p.26.

# *Appendix A*

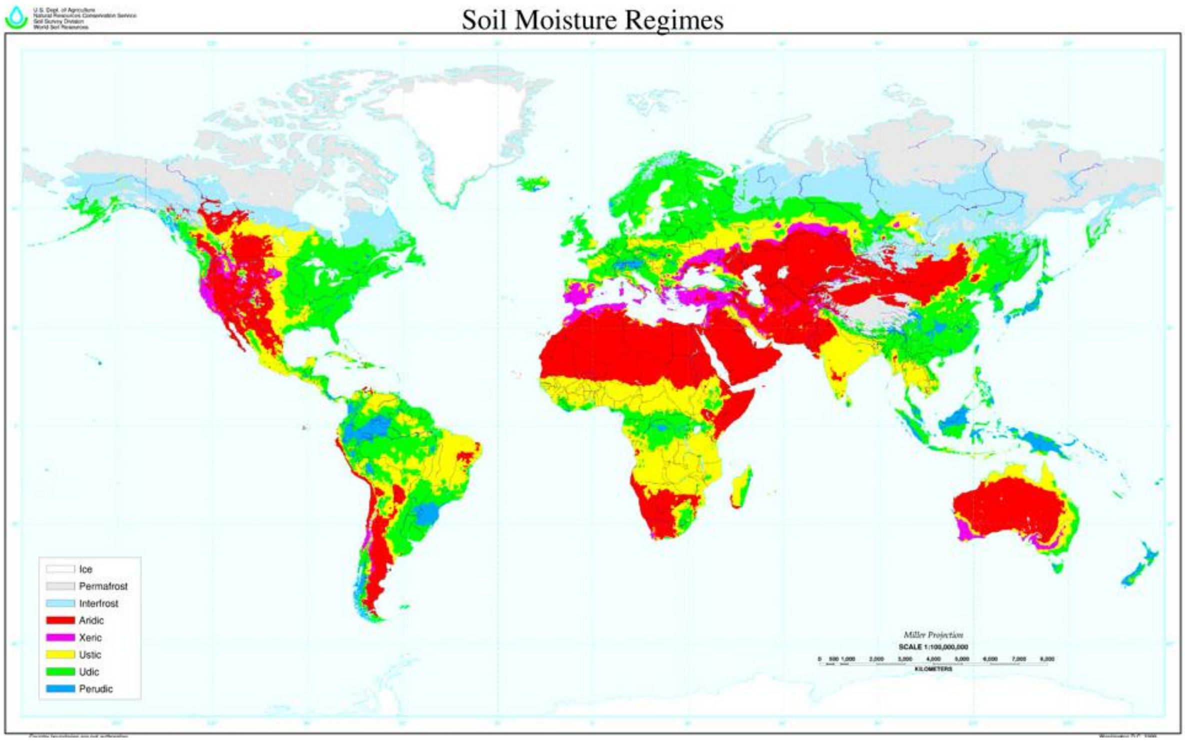


**Figure A:** Global soil moisture regimes. Data source USDA (1999).

## *Appendix B*

**Table A:** The statistical relationship between the Bloemfontein district observations and Maize_control model simulation maize yield.

| Statistical measure | Value |
|---|---|
| **MBE** | -1,85 |
| **RMSE** | 2,1 |
| **R** | 0,33 |
| **$R^2$** | 0,11 |

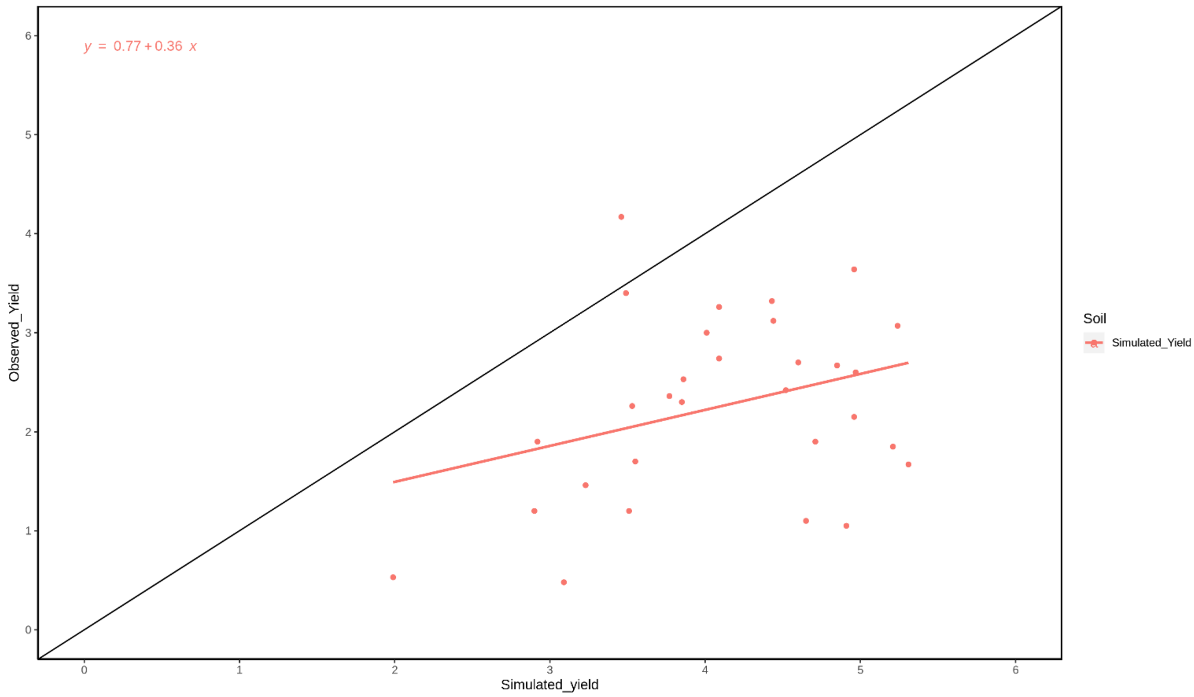


**Figure B:** The regression analysis of maize yield from the Bloemfontein district observations and Maize_control model simulation (1985-2015). Data source: South African Department of Forestry and Fisheries (DAFF).