DEVELOPMENT OF A SOUTH AFRICAN NATIONAL INPUT DATABASE TO RUN THE SWAT MODEL IN A GIS

Report to the
WATER RESEARCH COMMISSION

by

Jay le Roux[1]

and

Ndifelani Mararakanye[1], Leushantha Mudaly[2], Harold Weepener[3], Michael van der Laan[3]

[1]University of the Free State, Department of Geography
[2]University of Pretoria, Department of Plant and Soil Science
[3]Agricultural Research Council, Water Science

WRC Report No. 3053/1/22
ISBN 978-0-6392-0454-3

February 2023

Obtainable from

Water Research Commission
Lynnwood Bridge Office Park
Bloukrans Building
4 Daventry Road
Lynnwood Manor
PRETORIA

orders@wrc.org.za or download from www.wrc.org.za

This is the final report of WRC project no. C2019/20-00089.

EXECUTIVE SUMMARY

Introduction

An important requirement for hydrological modelling is spatial input datasets, including topographic data, land use-cover interactions, soil properties, and climate conditions. The combination of models and remote sensing techniques within a GIS framework is commonly utilized to assess hydrological processes such as streamflow, water erosion, sediment yield dynamics and nutrient inputs/outputs. A major limitation to model application in South Africa (SA), however, is the lack of standardized geo-spatial and "open-source" datasets developed for South African vegetation and soil types. This study collates multiple geo-spatial datasets at a national scale and interpret/format the data for use as baseline input to run the Soil and Water Assessment Tool (SWAT) in any catchment in SA. ArcSWAT is a graphical user interface for SWAT and ArcGIS® software extension, streamlining access to key databases and facilitating the preparation of input datasets. ArcSWAT was selected mainly because it is a spatially semi-distributed model that has gained international acceptance and has been applied to support various large catchment (10-10 000 km^2) modelling studies across the world. The aim of the project is threefold:

1. Collation of multiple geo-spatial datasets at a national scale.
2. Interpreting and formatting the data for use as baseline to run the SWAT model in SA.
3. Application of baseline input data in four research catchments in SA.

Materials and Methods

Interpreting and formatting national baseline data

The first step was to acquire national datasets followed by formatting the datasets for use as baseline to run the SWAT model in SA. Catchment outlines were obtained and prepared from the hydrologically corrected 90 m SRTM DEM and derived products of Weepener *et al*. (2012). National Land Cover maps (SANLC, 2014; 2018; 2020) with 72 to 73 land cover classes were linked to the land cover types in the ArcSWAT database. Soil texture and hydraulic parameter values were assigned to the Land Types of SA (Land Type Survey Staff, 1972-2006). Pedotransfer functions based on the studies of Van Tol *et al.* (2013); Van Zijl *et al.* (2016) and Van Tol *et al.* (2020) were used to generate the required hydraulic parameters, including available water capacity and saturated hydraulic conductivity. Climate data were acquired and interpreted to create 12 weather generator (WGN) files from ARC (2021) stations in different climate zones in SA.

Application of baseline input data in four research catchments

Using the national baseline datasets, the next step was to apply ArcSWAT in four (previously simulated) research catchments including Middle Olifants River Catchment (MORC), Lower Vaal River Catchment (LVRC), Mkabela Catchment (MC) and Tsitsa River Catchment (TRC). These catchments were simulated using the same weather data over the same timeframes as before. Various topographical, soil and land cover input datasets were used in the original case studies; in many cases more detailed data (spatially and temporally) than the baseline input dataset utilized here. The reason for duplicating the application of ArcSWAT in these previously modelled catchments is to compare the results (flow and sediment yield) of the two different input datasets used in each catchment, as well as the hydrological accuracy against measured streamflow data. Model performance was determined by means of the coefficient of efficiency (NSE) of Nash and Sutcliffe (1970), as well as the coefficient of determination (r^2). A per cent deviation method (*Dv*) of Martinec & Rango (1989) was used as a measure of goodness-of-fit between simulated and measured streamflow data.

Results and Discussion

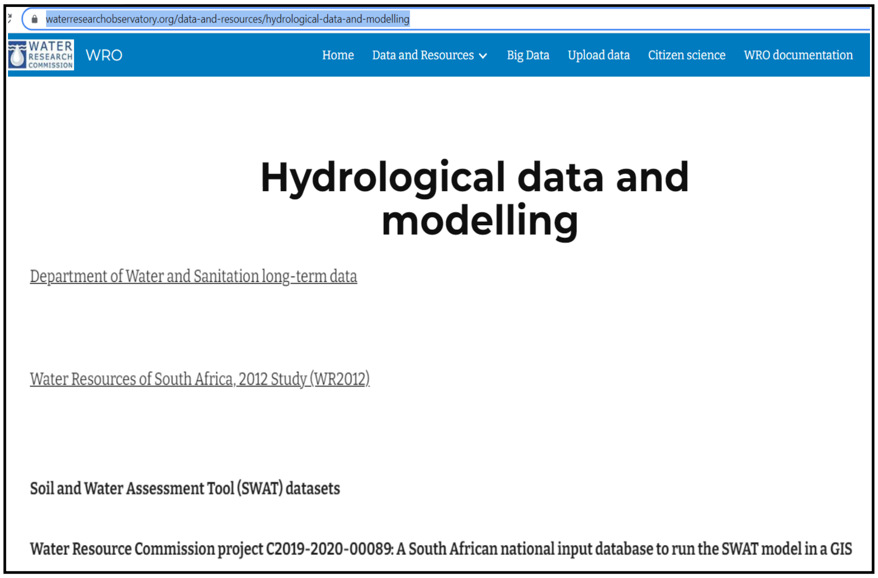
The national input database to run the SWAT model in SA is stored in the Water Research Observatory (WRO) data portal of WRC project: C2020/2021-00440 titled "Development and application of a big data platform to improve agricultural water resources management in South Africa". The URL is: https://www.waterresearchobservatory.org/data-and-resources/hydrological-data-and-modelling. The portal provides geo-spatial input datasets including:

- SWAT catchment outline data (tertiary and quaternary) including the hydrologically corrected SRTM DEM of SA at 90 m resolution (Weepener *et al.*, 2012);
- South African National Land Cover (SANLC, 2014; 2018; 2020) linked to SWAT land cover codes;
- Soil map with SWAT attribute data for each Land Type of SA (Land Type Survey Staff, 1972-2006);
- Weather statistics (WGN) files required as input by the model.

Appendix 1 lists the Metadata of the datasets, based on ISO 19115 Geospatial metadata standards.

Appendix 2 lists the post-graduate students whose studies are contributed to the project.

Appendix 3 indicates different forms of knowledge dissemination.



Title page of the national input database to run the ArcSWAT model (in a GIS) is stored in the Water Research Observatory data portal of WRC project: C2020/2021-00440.

Comparison between catchment and national data model results

Performance of the national data models was determined by comparing streamflow and sediment outputs with previous modelled catchment data models, as well as comparison of the hydrological accuracy against measured streamflow data. The catchment data models were slightly superior compared to the national data models, as shown by more accurate NSE, r^2 and *Dv* values. Between the four catchments, the LVRC performed the best, followed by the MC and then TRC, whereas the MORC performed the poorest. It is postulated that differences in the performance between catchments was largely influenced by the quality of rainfall data, since the other input datasets (DEM, soil and land use-cover) were similar in all four catchments. Graphically, in each of the four catchments, the streamflow output of both data models appear similar. Likewise, both data models show similar trends in sediment load estimations, with occasional steep peaks that can be associated with wetter months. Spatial similarities and/or differences of sediment source areas are illustrated by means of sediment yield maps of the respective catchments. Although the average sediment yield of catchment and national data models are similar, some spatial differences are noted. Spatial differences in sediment yield between catchment and national data models are mainly attributed to land use-cover variances since the other input datasets (DEMs and soil input data) are in essence similar between data models. Despite these differences, the national baseline data appears to be an efficient input dataset, capable of modelling streamflow and sediment dynamics at a catchment scale.

Conclusion and Recommendations

One of the biggest challenges to set-up and run the SWAT model in SA is to obtain appropriate input data, especially soil data. The input datasets consist of more detailed and higher resolution soil data than the global datasets of Abbaspour *et al.* (2019). Although SWAT users could use the input data 'as is', it is recommended to supplement, improve and/or replace the input data with recent/sophisticated data. Reliable rainfall data is necessary to consider the spatial distribution of rainfall throughout a catchment. Stream channel processes and hydrological structures need to be characterised, as well as ancillary information regarding management practices. Calibration of model simulations with measured data is also essential by adjusting the most sensitive model parameters. It is recommended to expand the database to include data input for cross-bordering catchments. In conclusion, the national input database is an important step forward in the application of hydrological modelling by assisting modellers to set-up and run the SWAT model anywhere in SA. The database will save time with model set-up, as well as assist in the standardization of SWAT modelling efforts in SA.

Acknowledgements

The Water Research Commission (WRC) for funding the research project, with special thanks to Mr Wandile Nomquphu and Ms Penny Jaca for MS managerial and administrative support. The project team appreciates valuable inputs from the Reference Group which benefited the research greatly:

Name	Organisation
Prof Mohamed Albesit	SPU
Dr David Clark	UKZN
Dr Julia Glenday	SAEON
Dr Alen Manyevere	UFH
Prof Dominic Mazvimavi	UWC
Dr Beason Mwaka	DWS
Ms Celiwe Ntuli	DWS
Dr Tendai Sawunyama	IUCMA
Prof Johan van Tol	UFS

Dr Stefanie Schutte at UKZN for spatial Carbon data.

Mr Adriaan Schutte at UP for the collation and uploading of data onto the Water Research Observatory data portal.

Ms Aimee Thompson assisted with obtaining soil parameters for her BSc Honours at UP.

Ms Marike Stander applied the baseline dataset for comparison with fingerprinting techniques for her PhD at the UFS.

The UFS for provision of office space, administrative support and computing equipment.

Members of the project team and collaborating organizations for sharing source datasets and their knowledge, experience and advice.

This page was intentionally left blank

TABLE OF CONTENTS

EXECUTIVE SUMMARY .... *iii*

Acknowledgements .... **vii**

1. INTRODUCTION *1*

1.1 Aims and objectives **3**

2. MATERIALS AND METHODS *5*

2.1 Model selection and description **5**

2.2 Interpreting and formatting national baseline data **7**

2.2.1 Topographic and catchment outline data 7

2.2.2 National land cover data 8

2.2.3 Soil input data 10

2.2.4 Weather statistics 15

2.3 Application of baseline input data in four research catchments **18**

2.3.1 SWAT application in the Middle Olifants River Catchment 19

2.3.2 SWAT application in the Lower Vaal River Catchment 27

2.3.3 SWAT application in the Mkabela Catchment 35

2.3.4 SWAT application in the Tsitsa River Catchment 41

3. RESULTS AND DISCUSSION *47*

3.1 Online Data Portal System for South Africa **47**

3.1.1 SWAT catchment outline data 48

3.1.2 SWAT land cover data 51

3.1.3 SWAT user soil data 54

3.1.4 SWAT weather generator files 57

3.2 Comparison between catchment and national data model results **60**

3.2.1 Comparison of data models in the Middle Olifants River Catchment 60

3.2.2 Comparison of data models in the Lower Vaal River Catchment 65

3.2.3 Comparison of data models in the Mkabela Catchment 70

3.2.4 Comparison of data models in Tsitsa River Catchment 76

3.2.5 Overall comparison of catchment and national data models 82

4. CONCLUSION AND RECOMMENDATIONS *85*

5. REFERENCES *87*

APPENDIX 1: METADATA FOR ARCSWAT BASELINE DATA *95*

APPENDIX 2: CAPACITY DEVELOPMENT *99*

APPENDIX 3: KNOWLEDGE DISSEMINATION *100*

List of Figures

Figure 1: Hydrologically corrected SRTM DEM at 90 m resolution and major flow paths (Weepener *et al.*, 2012). 8
Figure 2: National Land Cover map (SANLC, 2018) with 72 land cover classes at 20 m resolution. 9
Figure 3: Land Types of SA usable at a scale of 1:250,000 (Land Type Survey Staff, 1972-2006). 11
Figure 4: Location of the 12 ARC weather stations used to create the WGN files, superimposed over the rainfall erosivity factor (R) map of Le Roux *et al.*, 2008). 17
Figure 5: Location map of four research catchments where ArcSWAT is applied including Middle Olifants, Lower Vaal, Mkabela and Tsitsa. 18
Figure 6: The location of the MORC illustrating the 60 delineated sub-catchments, streams, outlets, reservoirs, hydrometric and weather station locations. 20
Figure 7: (a) Land cover map and (b) soil map of MORC. 23
Figure 8: The location of the LVRC illustrating the 27 delineated sub-catchments, streams, outlets, reservoir and weather stations location. 28
Figure 9: (a) Land cover map and (b) soil map of LVRC. 31
Figure 10: Sub-catchment boundaries, streams, outlets, hydrometric and weather station locations. 36
Figure 11: (a) Land cover map and (b) soil map of MC. 39
Figure 12: The TRC illustrating the 13 delineated sub-catchments, streams, outlets, hydrometric and weather station locations. 42
Figure 13: (a) Land cover map and (b) soil map of TRC. 45
Figure 14: Title page of the national input database to run the ArcSWAT model (in a GIS) is stored in the Water Research Observatory data portal of WRC project: C2020/2021-00440. 48
Figure 15: Tertiary catchment boundaries of SA (Weepener *et al.*, 2012). 49
Figure 16: Quaternary catchment boundaries of SA (Weepener *et al.*, 2012). 49
Figure 17: Tertiary and Quaternary catchment inlet and outlet locations of SA (Weepener *et al.*, 2012). 50
Figure 18: DWS monitoring stations locations of SA (DWS, 2022b). 50
Figure 19: Graphical comparison of monthly streamflow (in m^3/s) at the hydrometric station with (a) catchment and (b) national data during the observation period (1989 to 2015) in the MORC. 61
Figure 20: Graphical comparison of total annual sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1987 to 2015) in the MORC. 62
Figure 21: Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1987 to 2008) in the MORC. 63
Figure 22: Spatial comparison of average annual sediment yield (in Kg/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the MORC. 64
Figure 23: Graphical comparison of monthly streamflow (in m^3/s) for SWAT simulations with (a) catchment and (b) national data during the observation period (1980 to 2018) in the LVRC. 66
Figure 24: Graphical comparison of total annual sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1987 to 2018) in the LVRC. 67
Figure 25: Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1980-2018) in the LVRC. 68
Figure 26: Spatial comparison of average annual sediment yield (in Kg/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the LVRC. 69
Figure 27: Graphical comparison of monthly streamflow (in m^3/s) at the main catchment outlet with catchment and national data during the observation period (January 2006 to June 2008) in the MC. 70

Figure 28: Graphical comparison of monthly streamflow (in m^3/s) at the H-flume outlet of sub-catchment 2 with (a) catchment and (b) national data during the observation period (August 2006 to March 2008) in the MC. 72
Figure 29: Graphical comparison of total annual sediment load (in metric) for SWAT simulations with catchment and national data during the observation period (January 2006 to June 2008) in the MC. 73
Figure 30: Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (January 2006 to June 2008) in the MC. 74
Figure 31: Spatial comparison of average annual sediment yield (in Kg/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the MC. 75
Figure 32: Graphical comparison of monthly streamflow (in m^3/s) for SWAT simulations with (a) catchment and (b) national data during the observation period (20018-2012) in the TRC. 78
Figure 33: Graphical comparison of total annual sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (20018-2012) in the TRC. 79
Figure 34: Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (20018-2012) in the TRC. 80
Figure 35: Spatial comparison of average annual sediment yield (in t/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the TRC. 81
Figure 36: Comparison of the performance (NSE, r^2 and *Dv*) of the four catchments (MORC, LVRC, MC and TRC) with (a) catchment and (b) national data. 83

List of Tables

Table 1: Definition/description and methodology/reasoning used to assign soil parameter values to Land Types at a national scale. 12
Table 2: Description of soil albedo values given to Land type broad soil pattern codes. 14
Table 3: Clay content in the B-horizon, assigned to each soil series in the Land Type Database (according to clay factor values of the A-horizon). 15
Table 4: SWAT parameters of a weather generator file. 16
Table 5: 12 ARC weather stations used to create WGN files. 17
Table 6: Method used to assign value to required SWAT soil characteristics. 21
Table 7: Wastewater treatment works location in the catchment, flow and nutrient concentrations. 25
Table 8: Agricultural practices for the MORC. 25
Table 9: Selected SWAT parameters and values used for the calibration of SWAT model hydrology in the MORC. 27
Table 10: Definition/description and methodology/reasoning used to assign soil parameter values to Land Types in LVRC. 29
Table 11: Reservoir management operations parameters. 32
Table 12: Crop rotation and management schedule for potato-onion-maize-grass under irrigation and maize-sunflower-maize-soybean under dryland used in SWAT (Mararakanye *et al.*, 2021). 33
Table 13: Selected SWAT parameters and values used for the calibration of SWAT model hydrology in the LVRC. 34
Table 14: Definition/description and methodology/reasoning used to assign soil parameter values in MC. 37
Table 15: Parameter information used to model each of the farm dams and wetlands in the MC (Le Roux *et al.*, 2009). 40
Table 16: Definition/description and methodology/reasoning used to assign soil parameter values to Land Types in TRC. 43
Table 17: List of land cover types in the ArcSWAT database (excluding parameters) that was linked to the National Land Cover (2013/14, 2018 and 2020) maps of SA. 51
Table 18: Topsoil parameter values assigned to the Land Types of SA (only a selection of Land Types are shown). 55
Table 19: Subsoil parameter values assigned to the Land Types of SA (only a selection of Land Types are shown). 56
Table 20: Weather statistics (WGN) from 1981-2000 of 12 stations for the month of January. 58
Table 21: Weather statistics (WGN) from 2001-2020 of 11 stations for the month of January. 59
Table 22: MORC performance metrics obtained from monthly streamflow calibration and validation for each model from 1989-2015. 60
Table 23: LVRC performance metrics obtained from monthly streamflow calibration and validation for each model from 1980-2018. 66
Table 24: MC performance metrics obtained from monthly streamflow calibration and validation for each model from January 2006 to June 2008. 71
Table 25: TRC performance metrics obtained from monthly streamflow calibration and validation for each model from January 2008 to December 2012. 77

1. INTRODUCTION

A major limitation to model application in South Africa (SA) is the lack of standardized geo-spatial and "open-source" datasets. Although the Agricultural Catchment Research Model (ACRU) (Schulze, 1995) and the Atlas database (Schulze, 2007) have been developed and applied successfully in South Africa (SA), ACRU is not yet interfaced into a GIS similar to catchment scale models such as the Soil and Water Assessment Tool (SWAT). SWAT is a catchment-scale and continuous time model operating on a daily time-step to simulate water, sediment and chemical fluxes in large catchments with varying climatic conditions, soil properties, stream channel characteristics, land use and management practices (Srinivasan *et al.*, 1998; Arnold *et al.*, 2012). The model considers most hydrological and sedimentological aspects into one simulation package, including factors controlling runoff on hillslopes and streamflow in river channels, as well as sediment generation, channel transport and deposition into sinks (Gassman *et al.*, 2007). Although the SWAT model and its baseline input datasets were developed for use in the USA (not SA), SWAT is routinely coupled within a GIS which offer unprecedented flexibility in the representation and organization of spatial data (Chen and Mackay, 2004). ArcSWAT-2012 is a graphical user interface for SWAT and ArcGIS® software extension, streamlining access to key databases and facilitating the preparation of input datasets (including topography, drainage network, land cover, soil, climate and land management) (Srinivasan *et al.*, 1998). ArcSWAT-2012 has gained international acceptance and has been applied to support various large catchment (10-10 000 km^2) modelling studies across the world with minimal or no calibration effort (e.g. Srinivasan *et al.*, 2010; Gassman *et al.*, 2014; Guzha *et al.*, 2018). SWAT has also been applied in SA to support various large catchment modelling studies (Glenday *et al.*, 2021).

The WRC funded a number of projects where SWAT has been applied. From 2005 to 2013, WRC project K5/1516 was the first large catchment scale application of SWAT in a GIS, modelling the Mkabela Catchment near Wartburg to provide a comparative catchment scale alternative to ACRU-NP (Lorentz *et al.*, 2012). SWAT effectively identified NPS pollution source areas, as well as storages where connectivity is reduced at the catchment scale (Görgens *et al.*, 2012). The second WRC project K5/2402/4 where SWAT was applied followed in 2014. SWAT was applied in Two Streams and Fountainhill Estate in KwaZulu-Natal to assess the impact of sediment yield and nutrient outputs on water resources of different farming and forestry systems (Hill *et al.*, 2019; Scot-Shaw *et al.*, 2020). SWAT provided detailed spatial and temporal data and a platform was created to successfully test various land use and management scenarios at a catchment scale. The third WRC project (K5/2501) worth mentioning, applied SWAT (amongst others) to model flow, sediment and nutrient outputs in the Lower Vaal and Middle Olifants Catchments (Van der Laan and Franke, 2019). Results illustrate that irrigated cropping systems is a significant source of NPS pollution. Another WRC project (K5/22431) worth mentioning was conducted in the Mzimvubu River Catchment, the only large river network in SA without a dam. Modelling the flow and sediment yield made it possible to estimate life expectancies for 2 proposed dams on the Tsitsa River (between 43 and 55 years). More recently, WRC project K5/2927 by Glenday *et al*. (2021) reviewed the structural differences across several commonly used modelling tools in SA (ACRU, WRSM-Pitman, SPATSIM-Pitman, SWAT and MIKE-SHE) and explored the implications of these differences in various settings (Mistley Catchment, Upper Berg River Catchment, Upper Kromme River Catchment, and Middle Letaba Catchment). The biggest challenge for most (perhaps all) of the studies mentioned above was to obtain appropriate input data (especially soil data) for use in SWAT. Hydrological models interfaced in a GIS with their own geo-spatial input datasets are absent in most developing countries such as SA (Akoko *et al.* 2021).

An important requirement for hydrological modelling is spatial input datasets including topographic-drainage-network variables, land use-cover interactions, soil properties, and climate conditions. The combination of models and remote sensing techniques within a GIS framework is commonly utilized to assess hydrological processes such as streamflow, water erosion, sediment yield dynamics and nutrient inputs/outputs (e.g. Guzha *et al.*, 2018). In Europe, for example, PESERA (Pan-European Soil Erosion Risk Assessment Project) is capable of national assessments of runoff and water erosion using baseline input datasets (Kirkby *et al.*, 2004). In the USA, BASINS (Better Assessment Science Integrating Point and Nonpoint Sources) developed by the Environmental Protection Agency allows the user to import and assemble geo-spatial data required by models including digital elevation data, soil and land use input parameters (Arnold *et al.*, 1998). At a global scale, Abbaspour *et al.* (2019) prepared databases of soil, land use, actual evapotranspiration, as well as weather databases that could serve as standard inputs in SWAT models. In SA, as mentioned above, ACRU is a hydrological model (Schulze, 1995) with associated input datasets referred to as the South African Atlas of Climatology and Agrohydrology (Schulze, 2007). Although the ACRU model and Atlas database have been applied successfully in SA (e.g. Dickinson and Collins, 1998; Van Zyl and Lorentz, 2003; Görgens *et al.*, 2012), ACRU is not yet interfaced into a GIS similar to catchment scale models such as ArcSWAT. This study collates multiple geo-spatial datasets at a national scale and interpret/format the data for use as baseline input to run the ArcSWAT model in any catchment in SA. National input datasets include digital elevation data, catchment outlines, land cover map and codes, soil map and attribute table, and weather statistics for specific coordinates are required as input by the ArcSWAT model.

1.1 Aims and objectives

The aim of the project is threefold:

1. Collation of multiple geo-spatial datasets at a national scale.
2. Interpreting and formatting the data for use as baseline to run the SWAT model in SA.
3. Application of baseline input data in four research catchments in SA.

The national input datasets are made available in cloud storage as "open-source" to standardize ArcSWAT modelling efforts in SA. Input database to run the ArcSWAT model is stored in the Water Research Observatory (WRO) data portal of WRC project: C2020/2021-00440 titled "Development and application of a big data platform to improve agricultural water resources management in South Africa". The URL is: https://www.waterresearchobservatory.org/data-and-resources/hydrological-data-and-modelling. The portal provides geo-spatial input datasets including:

- SRTM DEM of SA 90 m resolution including major flow paths and catchment outlines (tertiary and quaternary) (Weepener *et al.*, 2012);
- South African National Land Cover (SANLC, 2014; 2018; 2020) linked to SWAT land cover codes;
- Soil map with SWAT attribute data for each Land Type of SA (Land Type Survey Staff, 1972-2006);
- Weather statistics (WGN) files required as input by the model.

One of the biggest challenges to set-up and run the SWAT model in SA is to obtain appropriate input data, especially soil data. A large part of modelling effort goes into the construction of input datasets (Jetten *et al.*, 2003; Glenday *et al.*, 2021). This study addressed this challenge by providing appropriate data for use in SWAT. The input datasets consist of more detailed and higher resolution soil and land cover data than the global datasets of Abbaspour *et al.* (2019). The national input database will assist researchers and students to set up and run the

SWAT model anywhere in SA. Therefore, this database will save time with model set-up, as well as assist in the standardization of SWAT modelling efforts in SA. Modellers will be able to use the input data ‘as is’, or alternatively supplement, improve and/or replace the input data with recent/sophisticated data. Such an input dataset is an important step forward in the application of hydrological models to assist agricultural water management.

Using the national baseline datasets, ArcSWAT is applied in four (previously simulated) research catchments including Middle Olifants, Lower Vaal, Mkabela and Tsitsa. These catchments are simulated using the same weather data, over the same timeframes, as before. The reason for duplicating the application of ArcSWAT in these previously modelled catchments is to compare the results (flow and sediment yield) of the two different input datasets used in each catchment, as well as hydrological accuracy against measured streamflow data. Comparing the results and accuracies of the two input datasets (original input versus baseline input), allowed appraisal of the performance of the baseline input data.

Following the Introduction Section 1 above, Section 2 provides a description of the Material and Methods, including collation and formatting of national baseline data, and application in four research catchments. Section 3 presents the Results and Discussion including a description of the Online Data Portal System for SA where the baseline data are stored, followed by a comparison between catchment and national data model results.

2. MATERIALS AND METHODS

First, national datasets were acquired, followed by formatting the datasets for use as baseline to run the SWAT model in SA. Second, the national baseline datasets are used in the modelling of four (previously simulated) research catchments including Middle Olifants, Lower Vaal, Mkabela and Tsitsa. As mentioned above, the reason for duplicating the application of ArcSWAT in these previously modelled catchments is to compare the results (flow and sediment yield) of the two different input datasets used in each catchment, as well as hydrological accuracy against measured streamflow data. Comparing the results and accuracies of the two input datasets (original input versus baseline input), allowed appraisal of the performance of the baseline input data. Section 2.1 below provides a brief motivation for selecting the SWAT model, as well as a description of the model.

2.1 Model selection and description

Hydrological modelling at a catchment scale is challenging due to spatial variability of the controlling factors, the lack of input and validation data, as well as measurement variability (Vanmaercke *et al.*, 2011). Models that represent all hydrological processes at a catchment scale are usually complex; they require extensive input data and calibration efforts; and in addition their results are usually not sufficiently accurate (De Vente *et al.*, 2013). Semi-distributed or semi-lumped models are often preferred above fully-distributed or physically-based models, since the application of the latter in large catchments lead to additional errors and uncertainty resulting from more parameters and input data requirements. The foundational strength of semi-distributed models is that they partition the catchment of interest into homogeneous morphological units thus, allowing to certain extents, the spatial variation of topography and land use to be accounted for (Lenhart *et al.*, 2005; Gassman *et al.*, 2007). Assessments are usually carried out by means semi-distributed models such as SWAT that accounts for connectivity aspects by integrating 2D-routing of sediment fluxes.

The SWAT model was selected mainly because it is a spatially semi-distributed model that has gained international acceptance and has been applied to support various large catchment (10-10 000 km^2) modelling studies across the world (e.g. Mishra *et al.*, 2007; Wang *et al.*, 2009; Srinivasan *et al.*, 2010; Gassman *et al.*, 2014). SWAT is a catchment-scale, continuous time model operating on a daily time-step developed by the US Department of Agriculture (USDA) Agricultural Research Service to simulate water, sediment and chemical fluxes in large catchments with varying climatic conditions, soil properties, stream channel characteristics, land use and management practices (Srinivasan *et al.*, 1998; Arnold *et al.*, 2012). SWAT considers most hydrological and sedimentological aspects into one simulation package, including factors controlling runoff on hillslopes and streamflow in river channels, as well as sediment generation, channel transport and deposition into sinks (Gassman *et al.*, 2007). Furthermore, SWAT is routinely coupled within a GIS which, according to Chen and Mackay (2004), offer unprecedented flexibility in the representation and organization of spatial data.

SWAT divides a catchment into multiple sub-catchments, which can be further divided into hydrological response units (HRUs) consisting of homogeneous soil and land use characteristics (Gassman *et al.*, 2007). The hydrologic component is based on the water balance equation in the soil profile integrating several processes, including surface runoff volume using the Green and Ampt (1911) infiltration method or the USDA SCS (1972) curve number method. Here, the SCS curve number method was chosen which is empirically based and relates runoff potential to land use and soil characteristics. Peak runoff rate is estimated with a modification of the rational method, where runoff rate is a function of daily surface runoff volume and a proportion of rainfall occurring until all of the catchment is contributing to flow at the outlet, known as the time of concentration (Neitsch *et al.*, 2011). The time of concentration is estimated using Manning's Formula considering both overland and channel flow. Sediment yield caused by rainfall and runoff is computed with the Modified Universal Soil Loss Equation (MUSLE) (Williams, 1975), using surface runoff and peak flow rate together with the widely used USLE (Wischmeier and Smith, 1978) factors including soil erodibility, slope length and steepness, crop cover management and erosion control practice. Certain nutrients and pesticides are also simulated by SWAT, but are outside the scope of this research and are not described here.

Once the loadings of water and sediment have been determined, they are summed to the sub-catchment level and routed through the stream network of the catchment including ponds, wetlands, depressional areas, and/or reservoirs (Neitsch *et al.*, 2011). SWAT incorporates a simple mass balance model to simulate the transport of sediment into and out of water bodies, where settling is calculated as a function of concentration and transportation out of a farm dam is a function of the final concentration (Neitsch *et al.*, 2011). Flow is routed through the channel using either the variable-rate storage method (Williams, 1969) or the Muskingum method (Overton, 1966), which are both variations of the kinematic wave model. Here the default variable storage method was chosen. Sediment is routed by means of a simplified stream power theory where the maximum amount of sediment that can be transported, deposited or re-entrained from a channel segment is a function of the peak channel velocity (Arnold *et al.*, 1995). The equations mentioned above and additional theoretical documentation for SWAT is given by Neitsch *et al.* (2011). ArcSWAT-2012 which is a graphical user interface for SWAT and ArcView® software extension (Srinivasan *et al.*, 1998) was used for this study. A description follows of the interpretation and formatting of the national baseline data.

2.2 Interpreting and formatting national baseline data

The first step was to acquire national datasets followed by formatting the datasets for use as baseline to run the SWAT model in SA, including topographic and catchment outline data, land cover data, soil input data and weather statistics. Metadata of the datasets was drafted based on the ISO 19115 metadata style for Geospatial metadata standards (see Appendix 1: Metadata).

2.2.1 Topographic and catchment outline data

Topographic and catchment outline data were obtained and prepared from the hydrologically corrected 90 m SRTM DEM and derived products (see Figure 1) of Weepener *et al.* (2012) including minor corrections/updates in 2018 (DWS, 2022a). The methodology followed to for the hydrological correction of the 90 m SRTM DEM and derived products is described in Weepener *et al.* (2012).

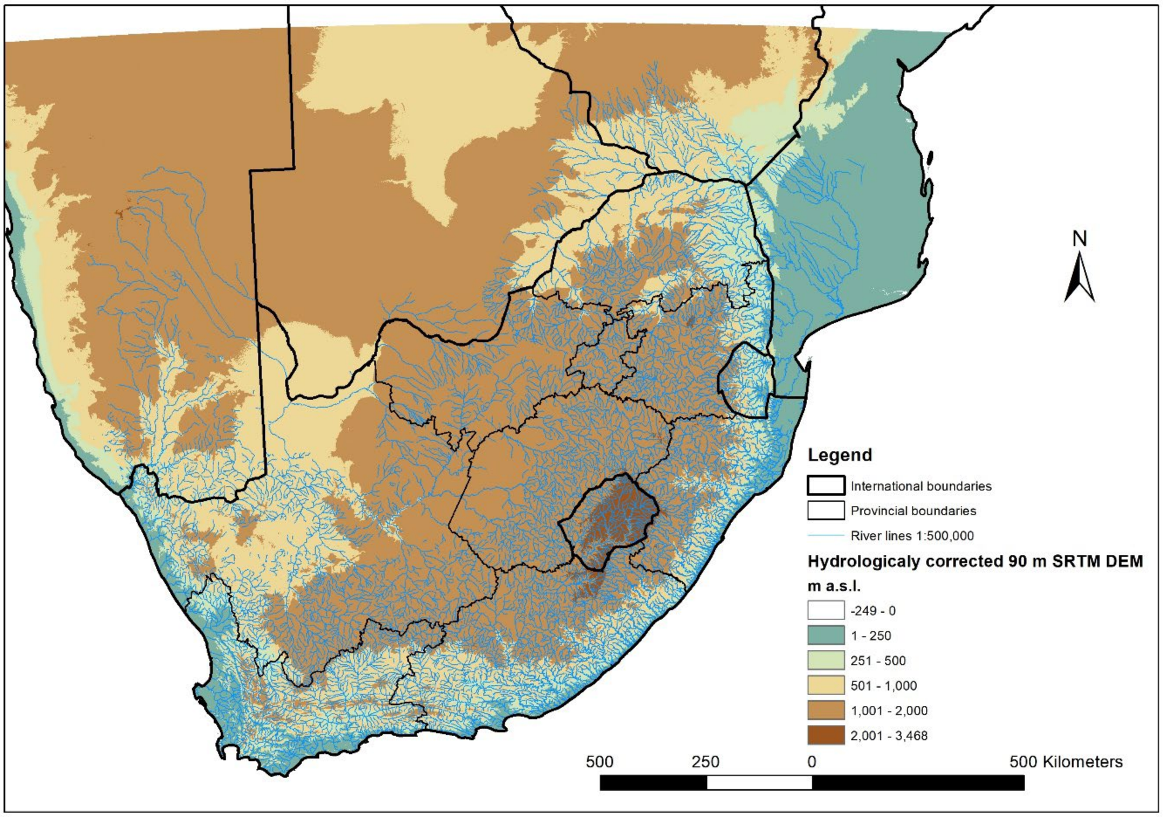


Figure 1: Hydrologically corrected SRTM DEM at 90 m resolution and major flow paths (Weepener *et al.*, 2012).

2.2.2 National land cover data

National Land Cover maps (SANLC, 2014; 2018; 2020) with 72 to 73 land cover classes were linked to the land cover types in the ArcSWAT database. Figure 2 illustrates the National Land Cover map (SANLC, 2018) with 72 land cover classes at 20 m resolution.

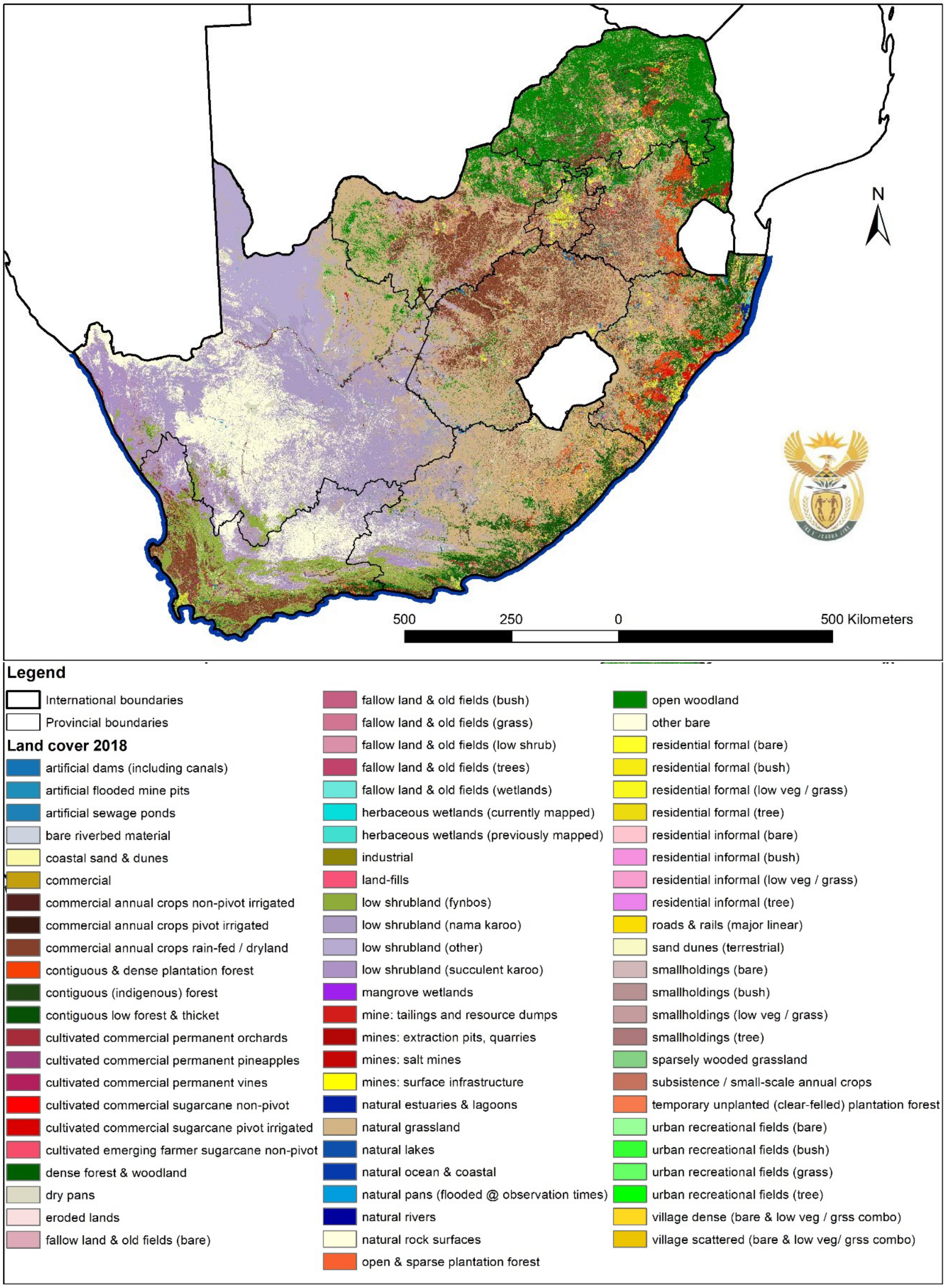


Figure 2: National Land Cover map (SANLC, 2018) with 72 land cover classes at 20 m resolution.

2.2.3 Soil input data

Soil texture and hydraulic parameter values were assigned to the Land Types of SA (Land Type Survey Staff, 1972-2006) (see Figure 3). Pedotransfer functions based on the studies of Van Tol *et al.* (2013); Van Zijl *et al.* (2016) and van Tol *et al.* (2020) were used to generate the required hydraulic parameters, including available water capacity and saturated hydraulic conductivity. Since the Land Type Database was generated at a national scale (1:250,000) and contains its own (unknown) level of uncertainty, soil properties can range significantly between and within Land Types. It was therefore decided not to disaggregate and parametrize Land Type polygons into morphological terrain units at a national scale (Du Plessis *et al.*, 2020). Tables 1, 2, and 3 indicate the definition/description and methodology/reasoning used to assign soil parameter values to Land Types at a national scale.

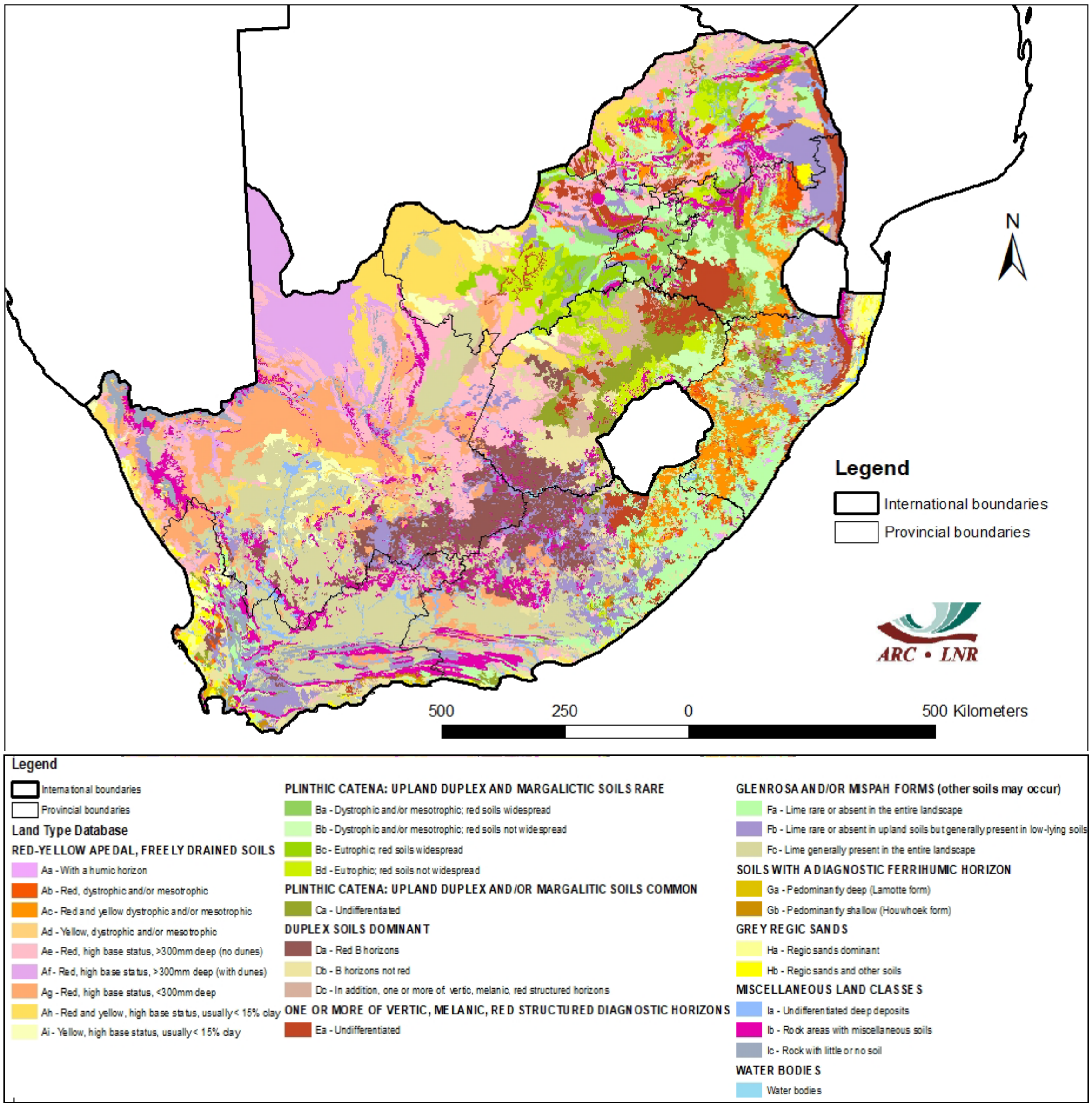


Figure 3: **Land Types of SA usable at a scale of 1:250,000 (Land Type Survey Staff, 1972-2006).**

Table 1: Definition/description and methodology/reasoning used to assign soil parameter values to Land Types at a national scale.

<table>
<tr><th>Parameter name</th><th>Definition/description</th><th>Methodology/reasoning</th></tr>
<tr><td>Number of layers in the soil</td><td>---</td><td>Two soil layers/horizons were incorporated into each soil component (Land Type).</td></tr>
<tr><td>Depth from soil surface to bottom of layer (mm)</td><td>Depth of each individual soil layer.</td><td>Depth descriptions/classes in the Land Type Database (Land Type Survey Staff, 1972-2006) and Schulze (2007) were used to assign depth to each Land Type layer.</td></tr>
<tr><td>Maximum rooting depth of soil profile (mm)</td><td>If no depth is specified, the model assumes the roots can develop throughout the entire depth of the soil profile.</td><td>As above.</td></tr>
<tr><td>Soil Hydrologic Group (A,B,C,D)</td><td>The U.S. Natural Resource Conservation Service (NRCS) classifies soils into four hydrologic groups based on infiltration characteristics of the soils. In term of runoff potential, Soil Group A = low, B = moderately low, C = moderately high, D = high.</td><td>Used the hydrological classes given in Schulze (2007) for each Land Type.</td></tr>
<tr><td>Available water capacity of the soil layer (mm H_2O/mm soil)</td><td>The plant available water, also referred to as the available water capacity, is calculated by subtracting the fraction of water present at permanent wilting point from that present at field capacity, AWC = FC - WP where AWC is the plant available water content, FC is the water content at field capacity, and WP is the water content at permanent wilting point.</td><td>AWC = FC - WP given in Schulze (2007) for each Land Type.</td></tr>
<tr><td>Saturated hydraulic conductivity (mm/hr)</td><td>The saturated hydraulic conductivity, K_{sat}, relates soil water flow rate (flux density) to the hydraulic gradient and is a measure of the ease of water movement through the soil. K_{sat} is the reciprocal of the resistance of the soil matrix to water flow.</td><td>Values were derived from the Rosetta Model (Schaap, 2001) based on the soil texture classes of each soil series in the Land Type Database of SA.
<table>
<tr><th>Clay (%)</th><th>Ks (mm/hr)</th></tr>
<tr><td>0-1</td><td>800</td></tr>
<tr><td>1-8</td><td>210</td></tr>
<tr><td>8-10</td><td>61</td></tr>
<tr><td>10-12</td><td>26</td></tr>
<tr><td>12-19</td><td>13</td></tr>
<tr><td>19-30</td><td>4,3</td></tr>
<tr><td>30-34</td><td>2,3</td></tr>
<tr><td>34-46</td><td>1,2</td></tr>
<tr><td>46-50</td><td>0,6</td></tr>
</table></td></tr>
<tr><td>Bulk density (Mg/m^3 or g/cm^3)</td><td>The soil bulk density (BD) expresses the ratio of the mass of solid particles to the total volume of the soil, $\rho b = M_S / V_T$. In moist bulk density determinations, the mass of the soil is the oven dry weight and the total volume of the soil is determined when the soil is at or near field capacity. Bulk density values should fall between 1.1 and 1.9 Mg/m^3.</td><td>BD was estimated using porosity (PO) data in Schulze (2007) for each Land Type.
PO = 1-BD/2.65</td></tr>
</table>

Table 1 continued.

Parameter name	Definition/description	Methodology/reasoning
Soil albedo (non-dimensional value between 0 and 1)	The ratio of the amount of solar radiation reflected by a body to the amount incident upon it, expressed as a fraction. The value for albedo should be reported when the soil is at or near field capacity.	Albedo values were assigned to each soil series in the Land Type Database of SA according to Table 2 below[1].
Texture of soil layer [optional]	This data is not processed by the model and the line may be left blank.	Assigned using clay classes given to each soil form in the Land Type Database of SA.
Clay content (% soil weight)	The percent of soil particles which are < 0.002 mm in equivalent diameter.	Clay content in the A-horizon was assigned using the average topsoil clay classes given to each soil form in the Land Type Database of SA. Clay content in the B-horizon was assigned to each soil series in the Land Type Database by adjusting the clay values of the A-horizon to clay-factors given in Table 3 below[1].
Silt content (% soil weight)	The percentage of soil particles which have an equivalent diameter between 0.05 and 0.002 mm.	Due to the lack of data, silt content for A and B horizons were assigned values between 10-22.5%, increasing with increase in clay as follows[1]: percentage of Land Type with <= 6% clay = 10% silt; 6.1 - 15% clay = 15% silt; 15.1 - 25% clay = 17.5% silt; 25.1 - 35% clay = 20% silt; 35.1 - 55% clay = 22.5% silt.
Sand content (% soil weight)	The percentage of soil particles which have a diameter between 2.0 and 0.05 mm.	Due to the lack of data, sand content for A and B horizons were assigned as follows: Sand = 100% – (%clay + %silt).
Rock fragment content (% soil weight)	The percent of the sample which has a particle diameter > 2 mm, i.e. the percent of the sample which does not pass through a 2 mm sieve.	Used agricultural restriction/rock (MB) classes in Land Type Database of SA as follows: MB0=0%; MB1=20%; MB2=50%; MB3=20%; MB4=100% (no soil).
Organic carbon content (% soil weight)	When defining by soil weight, the soil is the portion of the sample that passes through a 2 mm sieve.	A soil organic carbon map of SA of Schulze and Schütte (2020) (derived from soil profile data and Land Type Database) were used to assign average carbon values for A and B horizons per Land Type.
(K) factor in SI units t/ha per unit 'erosivity'	USLE equation soil erodibility described by Wischmeier and Smith (1978).	Using the SLEMSA model of Elwell (1976), erodibility units were established and used as a guide to the assignment of USLE (Wischmeier and Smith, 1978) K-factors to Land Types (Le Roux *et al.*, 2008).

1. Information from personal communication with:
Dr G. Paterson, Agricultural Research Council – Soil, Water and Climate: Pretoria, South Africa. 17 September 2013.
Prof. J. van Tol, Department of Soil, Crop and Climate Sciences, University of the Free State: Bloemfontein, South Africa. 20 January 2021.

Table 2: Description of soil albedo values given to Land type broad soil pattern codes.

Broad soil pattern	Description	Soil albedo (non-dimensional value between 0-1)	
		A-horizon	B-horizon
Aa	Freely drained, red and yellow apedal soils with humic topsoils comprise >40% of the Land Type	0.6	0.4
Ab	Freely drained, red and yellow, dystrophic/mesotrophic, apedal soils comprise >40% of the Land Type (yellow soils <10%)	0.4	0.4
Ac	Freely drained, red and yellow, dystrophic/mesotrophic, apedal soils comprise >40% of the Land Type (red and yellow soils each >10%)	0.4	0.4
Ad	Freely drained, red and yellow, dystrophic/mesotrophic, apedal soils comprise >40% of the Land Type (red soils comprise <10%)	0.3	0.3
Ae	Freely drained, red, eutrophic, apedal soils comprise >40% of the Land Type (yellow soils comprise <10%)	0.3	0.3
Af	Freely drained, red, eutrophic, apedal soils comprise >40% of the Land Type (yellow soils comprise <10%); with dunes	0.3	0.3
Ag	Freely drained, shallow (<300 mm deep), red, eutrophic, apedal soils comprise >40% of the Land Type (yellow soils comprise <10%)	0.3	0.3
Ah	Freely drained, red and yellow, eutrophic, apedal soils comprise >40% of the Land Type (red and yellow soils each comprise >10%)	0.2	0.2
Ai	Freely drained, yellow, eutrophic, apedal soils comprise >40% of the Land Type (red soils comprise <10%)	0.1	0.1
Ba	Red and yellow, dystrophic/mesotrophic, apedal soils with plinthic subsoils (plinthic soils comprise >10% of Land Type, red soils comprise >33% of Land Type)	0.4	0.4
Bb	Red and yellow, dystrophic/mesotrophic, apedal soils with plinthic subsoils (plinthic soils comprise >10% of Land Type, red soils comprise <33% of Land Type)	0.4	0.4
Bc	Red and yellow, eutrophic, apedal soils with plinthic subsoils (plinthic soils comprise >10% of Land Type, red soils comprise >33% of Land Type)	0.4	0.4
Bd	Red and yellow, eutrophic, apedal soils with plinthic subsoils (plinthic soils comprise>10% of Land Type, red soils comprise <33% of Land Type)	0.4	0.4
Ca	Land Type qualifies as Ba-Bd, but >10% occupied by upland duplex/margalitic soils	0.4	0.4
Da	Duplex soils (sandier topsoil abruptly overlying more clayey subsoil) comprise >50% of Land Type; >50% of duplex soils have red B horizons	0.4	0.6
Db	Duplex soils (sandier topsoil abruptly overlying more clayey subsoil) comprise >50% of Land Type; <50% of duplex soils have non-red B horizons	0.4	0.6
Dc	Either red or non-red duplex soils (sandier topsoil abruptly overlying more clayey subsoil) comprise >50% of Land Type; plus >10% occupied by black or red clays	0.4	0.6
Ea	Black or red clays comprise >50% of Land Type	0.7	0.7
Fa	Shallow soils (Mispah & Glenrosa forms) predominate; little or no lime in landscape	0.4	0.4
Fb	Shallow soils (Mispah & Glenrosa forms) predominate; usually lime in some of the bottomlands in landscape	0.4	0.4
Fc	Shallow soils (Mispah & Glenrosa forms) predominate; usually lime throughout much of landscape	0.4	0.4
Ga	Podzols occur (comprise >10% of Land Type); dominantly deep	0.7	0.7
Gb	Podzols occur (comprise >10% of Land Type); dominantly shallow	0.7	0.7
Ha	Deep grey sands dominant (comprise >80% of Land Type)	0.6	0.6
Hb	Deep grey sands sub dominant (comprise >20% of Land Type)	0.5	0.5
Ia	Deep alluvial soils comprise >60% of Land Type	0.3	0.3
Ib	Rock outcrops comprise >60% of Land Type	0.3	0.3
Ic	Rock outcrops comprise >80% of Land Type	0.3	0.3

Table 3: Clay content in the B-horizon, assigned to each soil series in the Land Type Database (according to clay factor values of the A-horizon).

Class	Description	Soil forms	Land Type %	B-horizon clay-factor change
S1	Soils with humic topsoil horizons	Ia, Ma, Kp, No	>60	1.1
S2	Freely drained, structureless soils	Hu, Cv, Gf, Sd, Oa	>60	1.2
S3	Red or yellow structureless soils with a plinthic horizon	Av, Gc, Bv, Pn	>60	1.2
S4	Excessively drained sandy soils	Sp, Ct, Vf, Du, Fw	>60	1*
S5	Dark clay soils which are not strongly swelling	Ar	>60	1.7
S6	Swelling clay soils	Bo, Ik, Tk	>40**	1.2
S7	Soils with a pedocutanic (blocky structured) horizon	Va, Sw	>60	1.7
S8	Imperfectly drained soils, often shallow and often with a plinthic horizon	We, Cf, Lo, Wa Kd	>60	1.3
S9	Podzols	Lt, Hh	>60	1.3
S10	Poorly drained dark clay soils which are not strongly swelling	Wo	>10**	1.5
S11	Poorly drained swelling clay soils	Rg	>60	1.2
S12	Dark clay soils, often shallow, on hard or weathering rock	My, Mw	>60	1*
S13	Lithosols (shallow soils on hard of weathering rock)	Ms, Gs	>60	1*
S14	Duplex soils (a sandy topsoil abruptly overlying a clayey, structured subsoil), often poorly drained	Es, Ss, Kd	>60	2
S15	Wetlands	Ch, Ka, Fw	>60	1*
S16	Non soil land classes	P, S, E, M, H**	>60	1*
S17	Rock	R**	>60	1*

*Clay content average given in Land Type data remains the same for A-horizon.
**Land Types where S-class consists less than 60%.

2.2.4 Weather statistics

Weather Generator (WGN) input files consists of several weather statistics needed by SWAT to generate representative daily climate data for simulated catchments in two instances: when the user specifies that simulated weather will be used or when measured data is missing. Table 4 indicates the SWAT parameters of a WGN file. WGN files were created by acquiring and interpreting climate data from ARC (2021) weather stations in different climate zones in SA. The completeness of climate data was the most important consideration for selecting weather stations. The station data of the ARC network that were used include precipitation, temperature, solar radiation, relative humidity and wind speed (ARC, 2021). Two sets of (12) WGN files were prepared covering the periods 1981-2000 and 2001-2020 respectively. Station 30673, however, did not have enough data to calculate statistics for the period 1981-

2000, and excluded the 1981-2000 dataset. Thus, the 1981-2000 WGN dataset consists of 11 individual WGNs.

The ARC weather network historically consisted of mechanical weather stations. These were replaced with automatic weather stations over time. Most of the stations were automated after the year 2000. Table 5 indicates the latitude and longitude and date of automation of the 12 weather stations. Figure 4 illustrates the location of the 12 ARC weather stations used to create the WGN files, superimposed over the rainfall erosivity factor (R) map of Le Roux *et al.*, 2008). Mechanical stations only measured sunshine hours and not solar radiation. Therefore, solar radiation data for the period 2001-2020 were used for the calculation of all WGN files. Furthermore, the mechanical stations measured wind run, which was converted to wind speed for the purposes of this project.

Table 4: SWAT parameters of a weather generator file.

Parameter code	Min value	Max value	Default value	Units	Definition
OID	na	na	na	na	Unique ID.
SUBBASIN	na	na	na	na	Subbasin ID
STATION	na	na	na	na	Weather Station name.
WLATITUDE	-90	90	15	[Degrees]	Latitude of weather station used to create statistical parameters.
WLONGITUDE	-180	180	20	[Degrees]	Longitude of weather station.
WELEV	0	5000	16	[m]	Elevation of weather station.
TMPMX	-30	50	0	[deg c]	Average maximum air temperature for month.
TMPMN	-40	40	1	[deg c]	Average minimum air temperature for month.
TMPSTDMX	0.1	100	2	[deg c]	Standard deviation for maximum air temperature in month.
TMPSTDMN	0.1	30	3	[deg c]	Standard deviation for minimum air temperature in month.
PCPMM	0	600	4	[mm/dd]	Average amount of precipitation falling in month.
PCPSTD	0.1	50	5	[mm/dd]	Standard deviation for daily precipitation in month.
PCPSKW	-50	20	6	na	Skew coefficient for daily precipitation in month.
PR_W1	0	0.95	7	[fraction]	Probability of a wet day following a dry day in the month.
PR_W2	0	0.95	8	[fraction]	Probability of a wet day following a wet day in the month.
PCPD	0	31	9	[days]	Average number of days of precipitation in month.
RAINHHMX	0	125	10	[mm]	Maximum 0.5 hour rainfall in entire period of record for month.
SOLARAV	0	750	11	[MJ/m^2-day]	Average daily solar radiation in month.
DEWPT	-50	25	12	[deg c]	Average dew point temperature in month.
WNDAV	0	100	13	[m/s]	Average wind speed in month.

Table 5: **12 ARC weather stations used to create WGN files.**

Station number	Latitude	Longitude	Elevation	Date of automation
30879	-29.20625	31.1548	651	2010/09/01
30673	-29.48298	27.13462	1500	2005/01/01
30142	-27.9576	24.8399	1180	1997/02/01
31004	-28.3142	28.70859	1693	2015/07/01
30420	-25.45458	30.97157	673	2000/05/01
30846	-34.26865	20.08052	280	2009/04/01
30608	-33.5142	26.82037	226	2003/10/01
30180	-28.46388	21.20541	798	1996/04/01
30731	-30.96861	21.98324	1327	2006/05/01
30835	-28.21915	32.48986	14	2008/10/01
30717	-31.03385	19.52388	800	2005/11/01
30093	-25.60398	28.35429	1168	1998/02/01

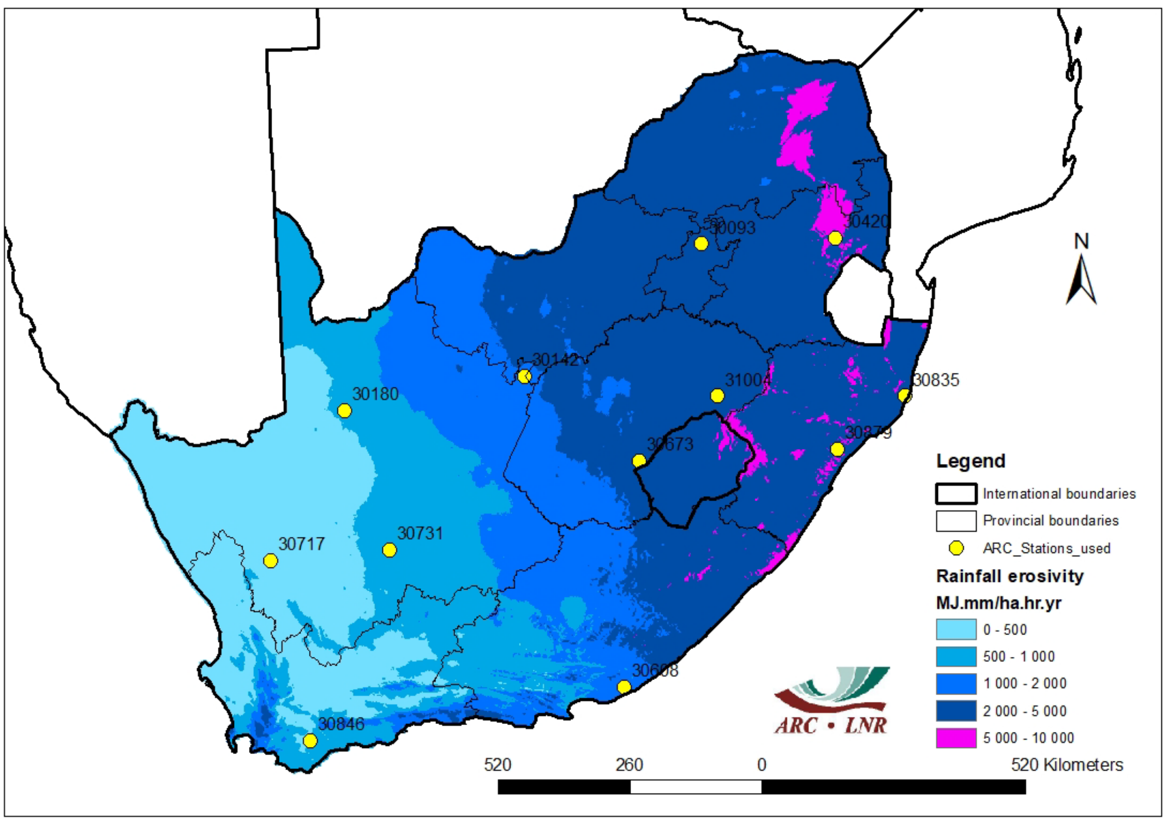


Figure 4: **Location of the 12 ARC weather stations used to create the WGN files, superimposed over the rainfall erosivity factor (R) map of Le Roux *et al.*, 2008).**

2.3 Application of baseline input data in four research catchments

Using the national baseline datasets, the next step was to apply ArcSWAT in four (previously simulated) research catchments including Middle Olifants, Lower Vaal, Mkabela and Tsitsa (see Figure 5). These catchments were simulated using the same weather data, over the same timeframes, as before. Various topographical, soil and land cover input datasets were used in the previous/original case studies; in many cases more detailed data (spatially and temporally) than the baseline input dataset utilized here. The original catchment data used are described in the "Model configuration, parameterization and calibration sections below. The reason for duplicating the application of ArcSWAT in these previously modelled catchments is to compare the results (flow and sediment yield) of the two different input datasets used in each catchment, as well as hydrological accuracy against measured streamflow data. Comparing the results and accuracies of the two input datasets (original input versus baseline input), allowed appraisal of the performance of the baseline input data.

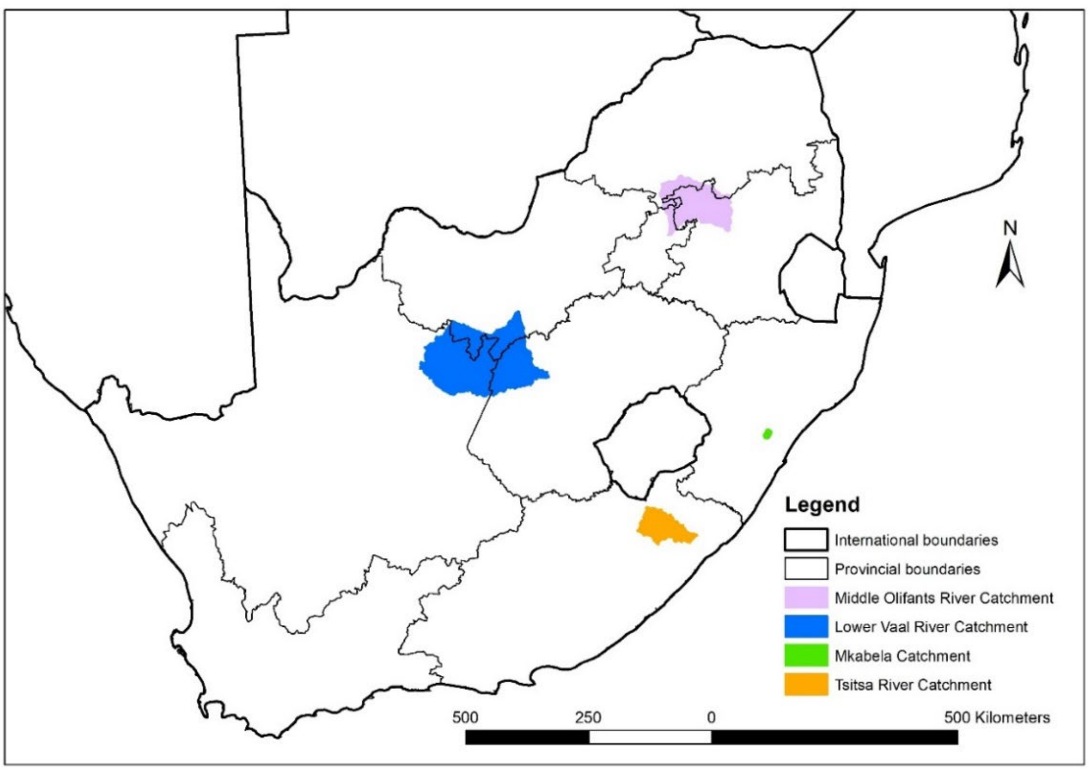


Figure 5: Location map of four research catchments where ArcSWAT is applied including Middle Olifants, Lower Vaal, Mkabela and Tsitsa.

2.3.1 SWAT application in the Middle Olifants River Catchment

Site description of the MORC

The MORC is the third most water stressed catchment in SA (Walter *et al.*, 2011) and covers an area of 22 550 km^2. The catchment is located in the area between Loskop Dam and Penge and has four main tributaries; the Selons, Moses, Elands and Mohlapitse Rivers (Singh and van Veelen, 2001) and includes the towns of Marble Hall, Groblersdal and Roedtan (DWS, 2011). The altitude varies from 800 m along the Olifants River to 1 900 m above sea level at the catchment margin. The area has a semi-arid climate and receives a mean annual rainfall of 500 mm (DWS, 2011). The primary water users in this area are urban and rural households as well as the mining industry and large-scale agriculture (Walter *et al.*, 2011). Extensive irrigation is, however, the main economic activity in this area (DWS, 2013) and the economy is characterized by intensive irrigated agriculture (particularly in the Marble Hall and Groblersdal areas), commercial dryland agriculture (in the Springbok flats) and some subsistence agriculture with platinum mining in the area as well (DWS, 2011).

The geology consists of various formations which chronologically belongs to the Vaalian Group formed 2600-1800 MY, including the Transvaal Sequence, the Bushveld Igneous Complex (BIC), as well as elements of the Pretoria Group (Council for Geoscience, 2007). The BIC is dominated by Rashoop Granophyre and Lebowa Granite Suite subdivisions, which is very important economically due to large deposits of major mineral deposits. Platinum, chrome and vanadium mining is prominent (DWS, 2013). Soils in the catchment vary significantly (Land Type Survey Staff, 1972-2006). Soils with a plinthic catena occur throughout the catchment, especially in low lying areas. In contrary, soils with minimal development, usually shallow on hard or weathering rock, occur in higher (mountainous) areas with steep slopes. Strongly structured cracking soils (vertic soils mainly dark coloured, dominated by swelling clays) occur in the northwest.

Model configuration, parameterization and calibration of the MORC

Topographic and drainage network data

The ASTER DEM (ASTER-GDEM, 2009) was selected to partition the catchment into sub-catchments due to its relatively high spatial resolution (15 m). The number of sub-catchment links or outlets was manually adjusted, representing all the relevant tributaries of the main river into 60 sub-catchments that are comparative in size, as well as to ensure that flow monitoring points spatially overlay with sub-catchment outlet points for calibration of model simulations with field measurements (see Figure 6).

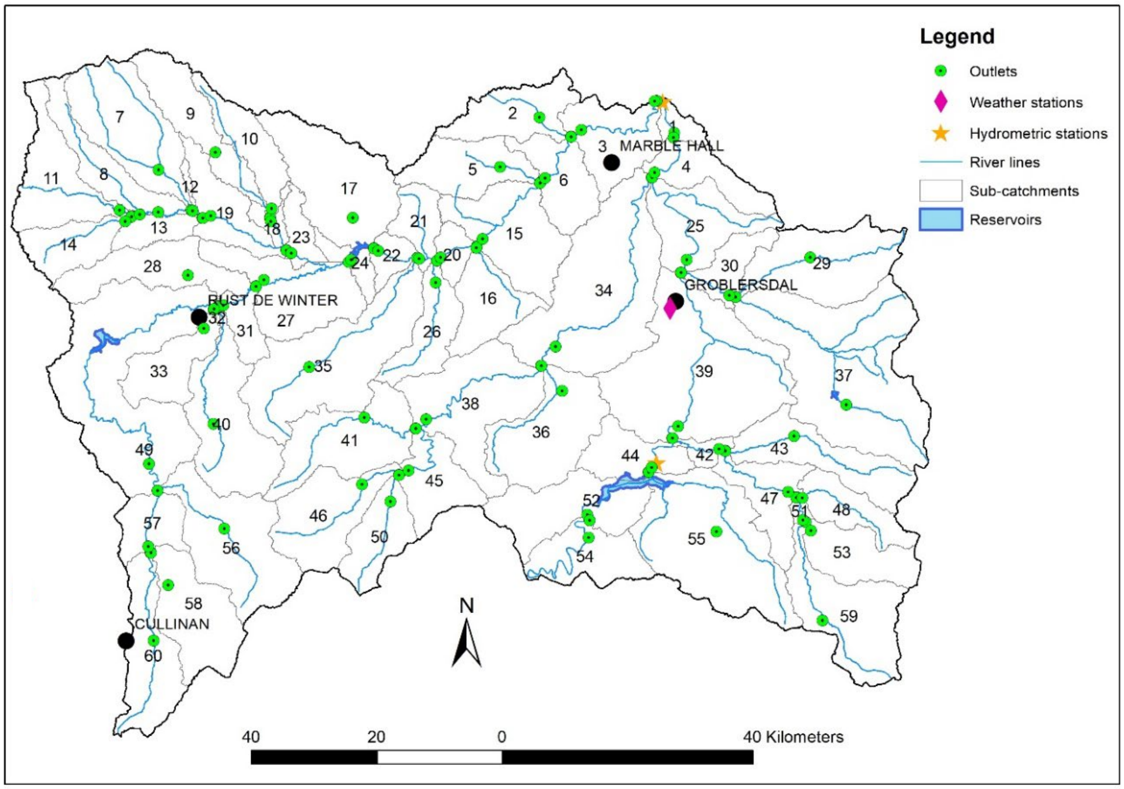


Figure 6: **The location of the MORC illustrating the 60 delineated sub-catchments, streams, outlets, reservoirs, hydrometric and weather station locations.**

Land use-cover and soil data

The land use map chosen for the catchment data model was the 2013/2014 South African National Land Cover dataset (SANLC, 2014) (see Figure 7a). The SANLC (2018) dataset was not yet available at the time the study was conducted. The majority of the catchment area is uncultivated (70%), with agricultural activity occupying 20% of the study area. Residential areas make up 5% of the land use and industrial activities account for less than 1%. The soil map (Figure 7b) used for the original catchment simulation was developed by the Land Type Database (Land Type Survey Staff, 1972-2006). A fair amount of soil data required by the SWAT model were available from the Land Type Database, however, some parameters had to derived from other sources. One of the sources used was the South African Atlas of Climatology and Agrohydrology (Schulze, 2007). Another source used was the SPAW (Soil-Plant-Air-Water) model pedotransfer functions (Saxton *et al.*, 2006). A breakdown of how the soil parameters were identified for use in the model is provided in Table 6.

The land area in the sub-catchments were partitioned into HRUs (Arnold *et al.*, 2012). These are areas of the landscape that have similar land uses, soils and slope within a sub-catchment. The three slope classes used where < 8%, 8-30%, and > 30%, which are based on the FAO classification of level to gently undulating (< 8%), rolling to hilly (8-30%) and steeply dissected to mountainous (> 30%) (Gyamfi *et al.*, 2016). Hydrologic response thresholds (HRTs) were set to 10%/10%/10% for land use, soils and slope respectively, for defining the HRUs. These thresholds ensure a unique combination of land use, soils and slope in the determination of HRUs by simplifying catchment processes (Gyamfi *et al.*, 2016).

Table 6: Method used to assign value to required SWAT soil characteristics.

Parameter name	Definition/description	Methodology/reasoning
Number of layers in the soil	---	One soil layer/horizon was incorporated into each soil component.
Depth from soil surface to bottom of layer (mm)	Depth of each individual soil layer.	Depth descriptions/classes in the Land Type Database (Land Type Survey Staff, 1972-2006) and Schulze (2007) were used to assign depth to each Land Type layer.
Maximum rooting depth of soil profile (mm)	If no depth is specified, the model assumes the roots can develop throughout the entire depth of the soil profile.	Depth descriptions/classes in the Land Type Database and Schulze (2007) were used to assign depth to each Land Type layer.
Soil Hydrologic Group (A,B,C,D)	The U.S. Natural Resource Conservation Service (NRCS) classifies soils into four hydrologic groups based on infiltration characteristics of the soils. In term of runoff potential, Soil Group A = low, B = moderately low, C = moderately high, D = high.	Used the hydrological classes given in Schulze (2007) for each Land Type.
Available water capacity of the soil layer (mm H_2O/mm soil)	The plant available water, also referred to as the available water capacity, is calculated by subtracting the fraction of water present at permanent wilting point from that present at field capacity, AWC = FC - WP where AWC is the plant available water content, FC is the water content at field capacity, and WP is the water content at permanent wilting point.	AWC = FC - WP given in Schulze (2007) for each Land Type.
Saturated hydraulic conductivity (mm/hr)	The saturated hydraulic conductivity, K_{sat}, relates soil water flow rate (flux density) to the hydraulic gradient and is a measure of the ease of water movement through the soil. K_{sat} is the reciprocal of the resistance of the soil matrix to water flow.	Derived from Soil-Plant-Air-Water (SPAW) model pedotransfer functions (Saxton *et al.*, 2006).

Table 6 continued.

Parameter name	Definition/description	Methodology/reasoning
Moist bulk density (Mg/m^3 or g/cm^3)	The soil bulk density expresses the ratio of the mass of solid particles to the total volume of the soil, $\rho b = M_S / V_T$. In moist bulk density determinations, the mass of the soil is the oven dry weight and the total volume of the soil is determined when the soil is at or near field capacity. Bulk density values should fall between 1.1 and 1.9 Mg/m^3.	Derived from Soil-Plant-Air-Water (SPAW) model pedotransfer functions (Saxton *et al.*, 2006).
Moist soil albedo (non-dimensional value between 0 and 1)	The ratio of the amount of solar radiation reflected by a body to the amount incident upon it, expressed as a fraction. The value for albedo should be reported when the soil is at or near field capacity.	Similar to Le Roux *et al.* (2015), moist soil albedo was assigned to the Land Type based on texture of the dominant soil as follows: Sands = 0.25; clays = 0.7; remaining textures = 0.5.
Texture of soil layer [optional]	This data is not processed by the model and the line may be left blank.	Values were provided in the Land Type Database. The average of the range of values provided was used.
Clay content (% soil weight)	The percent of soil particles which are < 0.002 mm in equivalent diameter.	Similar to Le Roux *et al.* (2015), silt content was assigned values between 10-22.5%, increasing with increase in clay as follows: percentage of Land Type with <= 6% clay = 10% silt; 6.1 - 15% clay = 15% silt; 15.1 - 25% clay = 17.5% silt; 25.1 - 35% clay = 20% silt; 35.1 - 55% clay = 22.5% silt.
Silt content (% soil weight)	The percentage of soil particles which have an equivalent diameter between 0.05 and 0.002 mm.	Similar to Le Roux *et al.* (2015), sand content was estimated as follows: Sand = 100% – (%clay + %silt + %rock + %carbon).
Sand content (% soil weight)	The percentage of soil particles which have a diameter between 2.0 and 0.05 mm.	Similar to Le Roux *et al.* (2015), the agricultural restriction/rock (MB) classes in for each Land Type was as follows: MB0=0%; MB1=20%; MB2=50%; MB3=20%; MB4=100% (no soil).
Rock fragment content (% soil weight)	The percent of the sample which has a particle diameter > 2 mm, i.e. the percent of the sample which does not pass through a 2 mm sieve.	Values were provided in the Land Type Database. The average of the range of values provided was used.
Organic carbon content (% soil weight)	When defining by soil weight, the soil is the portion of the sample that passes through a 2 mm sieve.	Similar to Le Roux *et al.* (2015), an unpublished carbon map of SA derived from soil profile and Land Type datasets was used to assign carbon values to each Land Type in the study area.
(K) factor in SI units t/ha per unit 'erosivity'	USLE equation soil erodibility described by Wischmeier and Smith (1978).	Used K-values given in Schulze (2007) for each Land Type.

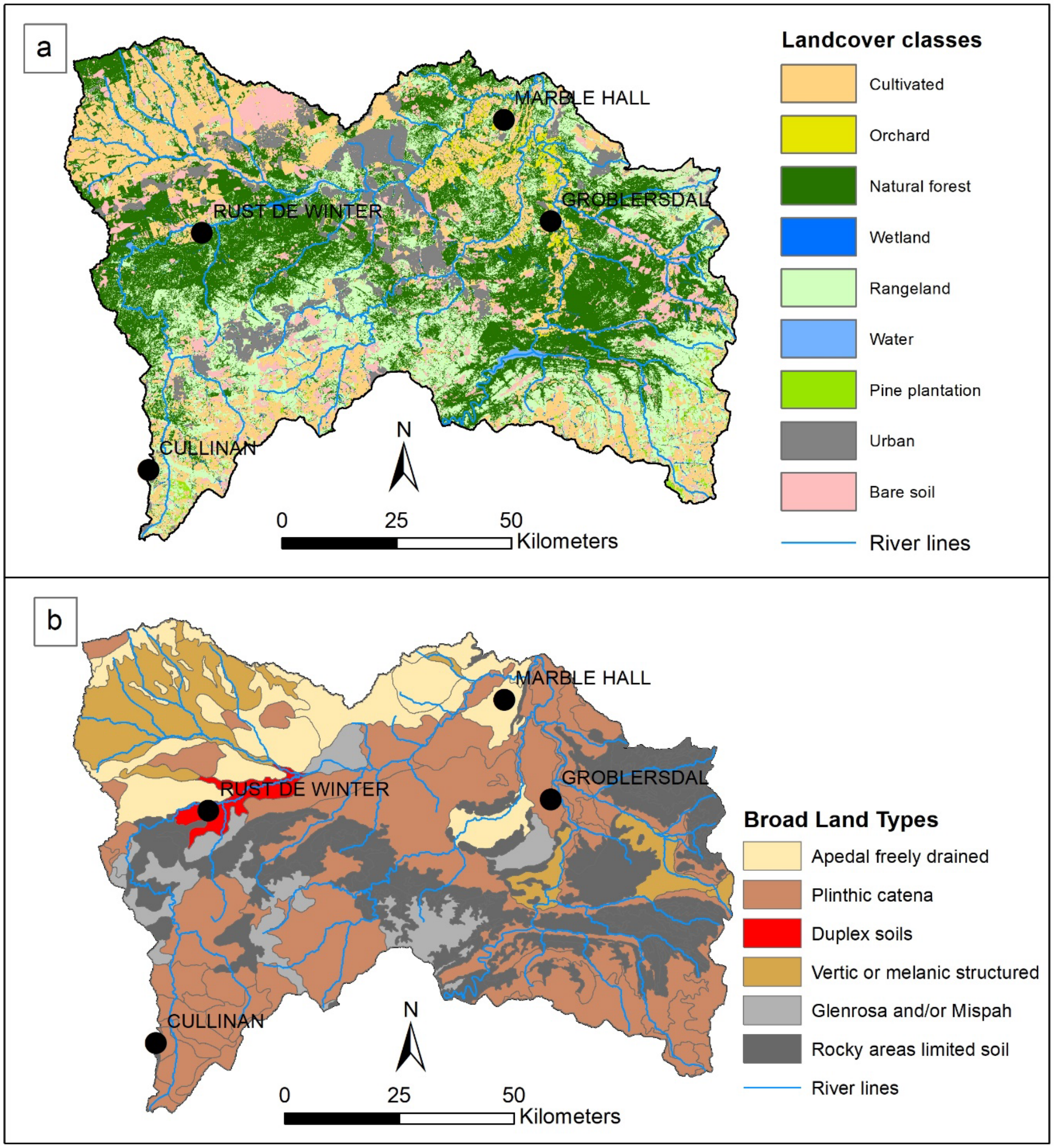


Figure 7: (a) Land cover map and (b) soil map of MORC.

Climate parameters

The climate data for the MORC was obtained from the ARC (2019), as well as the South African Weather Services (SAWS, 2019). Climate data were obtained from 1984 to 2015. There was insufficient data for solar radiation, wind speed and relative humidity for all the weather stations, however, the data that was available for these parameters was used to calculate statistical data for these parameters for the weather generator file. The weather generator input file contains the statistical data needed to generate representative daily climate data for the sub-catchments. The Loskop weather station had wind speed and relative humidity and one of the Marble Hall stations had solar radiation in addition to these parameters. The SWAT weather database (Essenfelder, 2016) was used to calculate these statistics.

Hydrological parameters

The reservoir file was edited to accommodate the Loskop Dam and the release data, which was supplied by DWS. The information was used to create a monthly reservoir outflow file, which summarized monthly outflow data from the reservoir on a monthly basis. This file contains the average daily flow rate for every month of the reservoirs operation (January to December). Inflow into Loskop Dam was calculated by adding the monthly volume from several rivers, which flow into the Loskop Dam, as outlined by Dabrowski *et al.* (2013). The DWS gauging stations used for this calculation were B1H002 (Spookspruit), B1H004 (Klipspruit), B1H010 (Olifants River), B1H015 (Klein Olifants River) and B2H015 (Wilge River).

Two point sources were added to the catchment, one in Marble Hall (sub-catchment 3) and one in Groblersdal (sub-catchment 41). Point source nutrient estimations were guided by the monitored values in the DWS database over a 9-year period (1999 to 2018). Information for only one wastewater treatment works (in Marble Hall) was available for this catchment and therefore the N and P average concentrations calculated were used for the Groblersdal wastewater treatment works as well. Flow was obtained from the DWS IRIS (Integrated Regulatory Information System) website. Information for the wastewater treatment works is shown in Table 7.

In addition, the source of irrigation water can be specified as a river reach (which can also indicate canal irrigation) reservoir, shallow aquifer, deep aquifer, or a location outside the catchment (Wei *et al.*, 2018). For this study the source of irrigation was set as the reservoir (Loskop Dam). An auto-irrigation operation (based on soil water content) was added to the management of the HRU's with agricultural land use and the source of the irrigation water was Loskop Dam.

Table 7: Wastewater treatment works location in the catchment, flow and nutrient concentrations.

Wastewater treatment plant	Sub-catchment location	Flow (m^3 day^{-1})	Outflow P Concentration (mg l^{-1})	Outflow N Concentration (mg l^{-1})
Marble Hall	3	1500	3.27	12
Groblersdal	41	5000	Used same value as Marble Hall	Used same value as Marble Hall

Management practices

The management file was edited in order to reflect the land use of the catchment in order to accommodate the agricultural land covered by intensively irrigated horticulture (mainly citrus with some vines) and agricultural land covered by maize. Developing a fertilization programme for a particular field is complex. The amount of nutrients already in the soil, target yield of the crop, and method of application are some of the factors that can all strongly influence how much fertiliser is applied at different stages. It was not possible to get management information specific to the study catchment and therefore more general practices were relied on. The following information was used to decide on N and P fertiliser rates for the important crops grown in the region. For maize, farmers were assumed to apply 15 kg N ha^{-1} and 4 kg P ha^{-1} for every ton of grain produced (Fertasa, 2007; ARC-GCI, 2013). We conservatively estimated an average maize yield of 10 t ha^{-1} under irrigation, so 150 kg N ha^{-1} and 40 kg P ha^{-1} was assumed to be broadcast at planting. Based on (Department of Agriculture Forestry and Fisheries, 2016), wheat was assumed to receive 160 kg N ha^{-1} and 18 kg P ha^{-1}, also broadcast at planting. For the orchard crops (mostly citrus), 200 kg N ha^{-1} (Dasberg, 1987) and 31 kg P ha^{-1} (Hume *et al.*, 1985) was assumed to be applied. The agricultural practices used for the catchment are summarised in Table 8.

Table 8: Agricultural practices for the MORC.

Agricultural practice	Crop		
	Maize	**Wheat**	**Orange trees**
Planting	November	June	September
Fertilization (N)	November (150 kg N ha^{-1})	June (160 kg N ha^{-1})	January (0.5479 kg N ha^{-1} continuous application)
Fertilization (P)	November (40 kg P ha^{-1})	June (18 kg P ha^{-1})	January (0.0849 kg P ha^{-1} continuous application)
Harvest	February	October	-
Tillage (disk plough)	February	October	-

Calibration and validation

Model simulation was conducted over a period of 30 years (1987 to 2015), preceded by a two-year (1985-1986) warm-up period to get the hydrological cycle fully operational. Model calibration and validation was performed for streamflow with observation data obtained from DWS river monitoring station B3H021. This gauge is named Olifants River North of Loskop Dam and the coordinates are 24°55'36.0"S, 29°23'21.9"E. The monitoring point is located in sub-catchment 3, which is downstream of Loskop Dam and just upstream of Flag Boshielo Dam in the study catchment. A temporal split-sample approach was used to split the observation data into two periods for calibration and validation. Data from 1987 to 2001 were used for calibration and data from 2002 to 2015 for validation. The SWAT model was calibrated and validated using the Parameter Estimator (SPE) algorithm, which is the successor of SUFI-2, available in SWAT Calibration Uncertainties Program (SWAT-CUP) premium version (Abbaspour et al., 2004). The SPE expresses uncertainties in the model output variables as the 95% probability distributions, commonly referred to as the 95% prediction probability uncertainty (95PPU) band (Abbaspour et al., 2007). The objective in SUFI-2 calibration is for the model results (95PPU) to capture most of the observations. Two statistical indices, namely p-factor and r-factor are used to quantify the 95PPU. P-factor is the percentage of the observed data bracketed by the 95PPU, whereas the r-factor is the thickness of the 95PPU bracket. A value greater than 0.7 for p-factor is suggested when modelling streamflow while a value closer to zero for r-factor is recommended. The selected parameters used for the calibration of the SWAT model in the LVRC are shown in Table 9. Successive iterations were performed until a reasonable coefficient of efficiency (NSE) of Nash and Sutcliffe (1970) was achieved, as well as the coefficient of determination (r^2). In addition, a per cent deviation method (Dv) of Martinec & Rango (1989) was used as a measure of goodness-of-fit between simulated and measured streamflow.

Table 9: Selected SWAT parameters and values used for the calibration of SWAT model hydrology in the MORC.

	Initial Value Range		Catchment Data Model	National Data Model
Parameters	Min	Max	Fitted Value	Fitted Value
R__CN2.mgt	-0.200	0.200	-0.182	-0.170
V__ESCO.hru	0.700	0.100	0.5494	0.259
R__SOL_K().sol	-0.800	0.800	-0.6128	-0.760
R__SOL_BD().sol	-0.600	0.600	0.4956	0.462
R__SOL_AWC().sol	-0.500	0.500	-0.053	0.355
V__ALPHA_BF.gw	0.000	1.000	0.905	0.645
V__GW_DELAY.gw	0.000	500.0	182.5	492.5
R__HRU_SLP.hru	-0.500	0.500	0.019	0.205
R__OV_N.hru	-0.500	0.500	-0.007	-0.265
R__SLSUBBSN.hru	-0.500	0.500	0.005	-0.435
V__GWQMN.gw	0.000	5000	1135	1125
V__REVAPMN.gw	0.000	1000	137.0	505.0
V__GW_REVAP.gw	0.020	0.100	0.026	0.086
V__EPCO.bsn	0.000	1.000	0.831	0.795
V__RCHRG_DP.gw	0.000	1.000	0.381	0.045
V__SHALLST.gw	0.000	5000	1885	3775
V__DEEPST.gw	0.000	10000	2070	3850

Parameters with V__ means existing parameter value is to be replaced by a given value.
Parameters with R__ means existing parameter is multiplied by (1 + a given value).

2.3.2 SWAT application in the Lower Vaal River Catchment

Site description of the LVRC

The study chose the LVRC in SA as study area, which is comprised of two quaternary sub-catchments labelled C91A and C91B by the DWS. It has a surface area of approximately 7 220 km^2. The area is drained by the Vaal River downstream of Bloemhof Dam with the main catchment outlet near the wall of Vaalharts Weir. Topography is relatively flat with no distinct features (DWS, 2015). The altitude varies from 1 183 m along the Vaal River to 1 431 m above sea level at the catchment margin. The area has a semi-arid climate and receives mean annual rainfall ranging from 200 to 600 mm. The area harbours a savannah type of vegetation which includes Kimberley Thornveld (SVk4) and Schmidtsdrif Thornveld (SVk6) (Mucina and Rutherford, 2006). The primary land uses are dryland crop farming and extensive livestock husbandry. Irrigated crop production is the dominant economic activity among farmers with water user rights, who are mostly located in the vicinity of the Vaal River. About 80% of water use in the Vaal River is for irrigation, 12% for urban and industrial activities, while the remaining 8% is for household supply and mining. Urban development is fragmented and is mainly located around major agricultural centres.

The main geological features of the study area include dark-grey mudrock, rhyolite with subordinate pyroclastic rocks, minor sandstone and light grey medium-grained biotite granite intruded by a network of dolerite sills, sheets and dykes. Conglomerate, grit, quartzite, sub-grey wacke, shale lenses, andesitic lavas and tuffs belonging to the Ventersdorp Supergroup (late Archaean-early Proterozoic sequences) also occur (Burger, 2013). These have given rise to freely drained sandy soils with low nutrient holding abilities, requiring fertiliser application to enable profitable crop production.

Model configuration, parameterization and calibration of the LVRC

Topographic and drainage network data

The catchment was partitioned into 27 sub-catchments using the SRTM DEM at 30 m resolution (see Figure 8). Each of the 27 sub-catchments were allowed to have its point source automatically generated. A reservoir was defined at the main catchment outlet (sub-catchment 24), representing the Vaalharts Weir outlet.

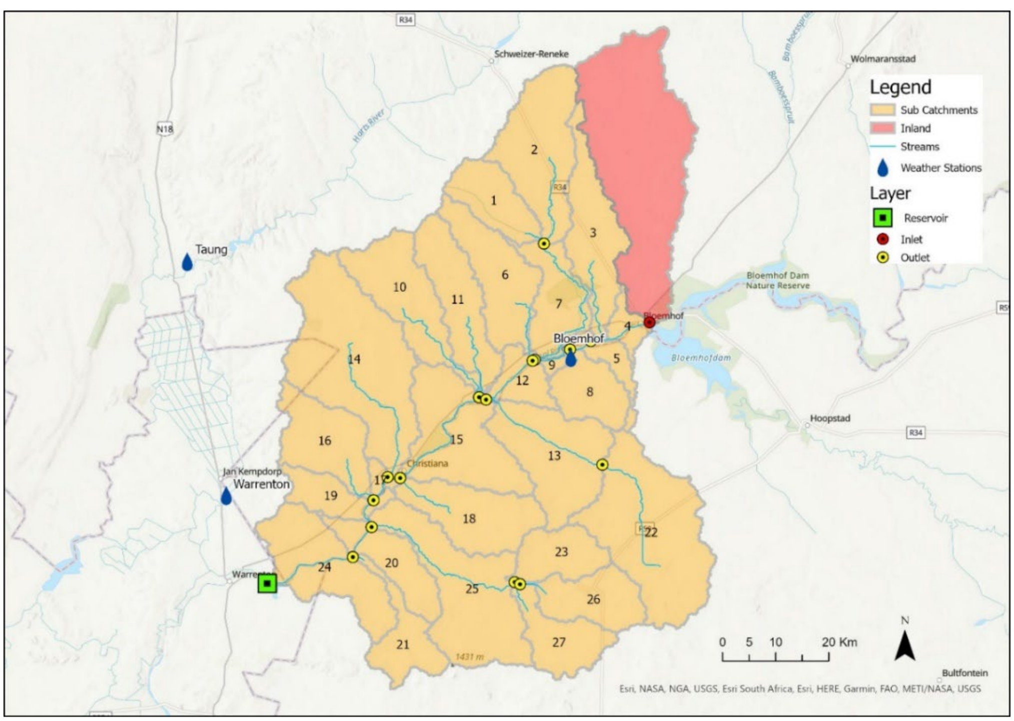


Figure 8: The location of the LVRC illustrating the 27 delineated sub-catchments, streams, outlets, reservoir and weather stations location.

Land use-cover and soil data

Land use data with seven classes (waterbodies [WATR], wetlands [WETN], barren land [BARR], cultivated – dryland [AGRL], cultivated – irrigated [AGRC/ POTA], rangeland [RNGB] and built-up areas [URBN]) were derived from 2018 Landsat imagery (Mararakanye *et al.*, 2021) (see Figure 9a). Soil data were obtained from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009) (see Table 10 and Figure 9b). HWSD comprised of a 30 cm topsoil and a 70 cm subsoil layer, based on combined existing regional and national soil information from different countries. The base layer used for HWSD in SA is the Land Type Database (Land Type Survey Staff, 1972-2006), available at a scale of 1: 250 000. Other soil information not included in the HWSD, such as moist bulk density, an available water capacity of the soil layer, saturated hydraulic conductivity, moist soil albedo, soil erodibility factor and electrical conductivity, were estimated based on pedotransfer functions using soil texture (Abbaspour *et al.*, 2019). Soil organic carbon contents of topsoil and subsoil layers were downloaded from the South African carbon sinks atlas (CSIR Smart Places, 2020). Subsequently, 259 HRUs were delineated in the catchment data model.

Table 10: Definition/description and methodology/reasoning used to assign soil parameter values to Land Types in LVRC.

Parameter name	Definition/description	Methodology/reasoning
Depth from soil surface to bottom of layer (mm)	Depth of each individual soil layer.	Schulze (2007) depth of the topsoil layer (*DEPAHO*) for each Land Type was used.
Maximum rooting depth of soil profile (mm)	If no depth is specified, the model assumes the roots can develop throughout the entire depth of the soil profile.	Schulze (2007) depth of the topsoil layer (*DEPAHO*) for each Land Type was used.
Soil Hydrologic Group (A,B,C,D)	The U.S. Natural Resource Conservation Service (NRCS) classifies soils into four hydrologic groups based on infiltration characteristics of the soils. In term of runoff potential, Soil Group A = low, B = moderately low, C = moderately high, D = high.	The soil hydrological groups given in Schulze (2007) for each Land Type were used.
Available water capacity of the soil layer (mm H_2O/mm soil)	The plant available water, also referred to as the available water capacity, is calculated by subtracting the fraction of water present at permanent wilting point from that present at field capacity, AWC = FC - WP where AWC is the plant available water content, FC is the water content at field capacity, and WP is the water content at permanent wilting point.	Schulze (2007) plant available water (*PAW*) given for each Land Type was used.

Table 10 continued.

Parameter name	Definition/description	Methodology/reasoning
Saturated hydraulic conductivity (mm/hr)	The saturated hydraulic conductivity, Ksat, relates soil water flow rate (flux density) to the hydraulic gradient and is a measure of the ease of water movement through the soil. Ksat is the reciprocal of the resistance of the soil matrix to water flow.	Estimation of hydraulic conductivity was based on the soil texture classes of the dominant soil in the Land Type, assuming 2.5% organic matter (OM) and no salinity, gravel or density adjustment (Saxton *et al.*, 2006) as follows: Sa = 108.1, L-Sa = 96.7, Sa-L = 50.3, L = 15.5, Si-L = 16.1, Si = 22, Sa-CL = 11.3, CL = 4.3, Si-C-L = 5.7, Si-C = 3.7, Sa-C = 1.4 and C = 1.1
Moist bulk density (Mg/m3 or g/cm3)	The soil bulk density expresses the ratio of the mass of solid particles to the total volume of the soil, ρb = MS /VT. In moist bulk density determinations, the mass of the soil is the oven dry weight and the total volume of the soil is determined when the soil is at or near field capacity. Bulk density values should fall between 1.1 and 1.9 Mg/m3.	Bulk density of the soil was obtained from Schulze (1995) approximations based on clay content as follows: 2 - 5% clay = 1.7, 6 - 15% clay = 1.6, 16 - 25% clay = 1.5, 26 - 32% clay = 1.4, 33 - 40% clay = 1.3
Moist soil albedo (non-dimensional value between 0 and 1)	The ratio of the amount of solar radiation reflected by a body to the amount incident upon it, expressed as a fraction. The value for albedo should be reported when the soil is at or near field capacity.	Moist soil albedo was assigned to the Land Type based on texture of the dominant soil as follows: Sands = 0.25; clays = 0.7; remaining textures = 0.5 (Le Roux *et al.*, 2015).
Organic carbon content (% soil weight)	When defining by soil weight, the soil is the portion of the sample that passes through a 2 mm sieve.	An unpublished carbon map of SA derived from soil profile and Land Type data was used to assign carbon values to each Land Type.
Clay content (% soil weight)	The percent of soil particles which are < 0.002 mm in equivalent diameter.	The maximum value of clay percentages range given in a Land Type Memoir for the dominant soil was used (Land Type Survey Staff, 1972-2006).
Silt content (% soil weight)	The percentage of soil particles which have an equivalent diameter between 0.05 and 0.002 mm.	Similar to Le Roux *et al.* (2015), silt content was assigned values between 10-22.5%, increasing with increase in clay as follows: percentage of Land Type with <= 6% clay = 10% silt; 6.1 - 15% clay = 15% silt; 15.1 - 25% clay = 17.5% silt; 25.1 - 35% clay = 20% silt; 35.1 - 55% clay = 22.5% silt.
Sand content (% soil weight)	The percentage of soil particles which have a diameter between 2.0 and 0.05 mm.	Similar to Le Roux *et al.* (2015), sand content was estimated as follows: Sand = 100% – (%clay + %silt + %rock + %carbon).
Rock fragment content (% soil weight)	The percent of the sample which has a particle diameter > 2 mm, i.e. the percent of the sample which does not pass through a 2 mm sieve.	The agricultural restriction/rock (MB) for each Land Type was as follows: MB0=0%; MB1=20%; MB2=50%; MB3=70%; MB4=100% (no soil).
(K) factor in SI units t/ha per unit 'erosivity'	USLE equation soil erodibility described by Wischmeier and Smith (1978).	Used K-values given in Schulze (2007) for each Land Type.

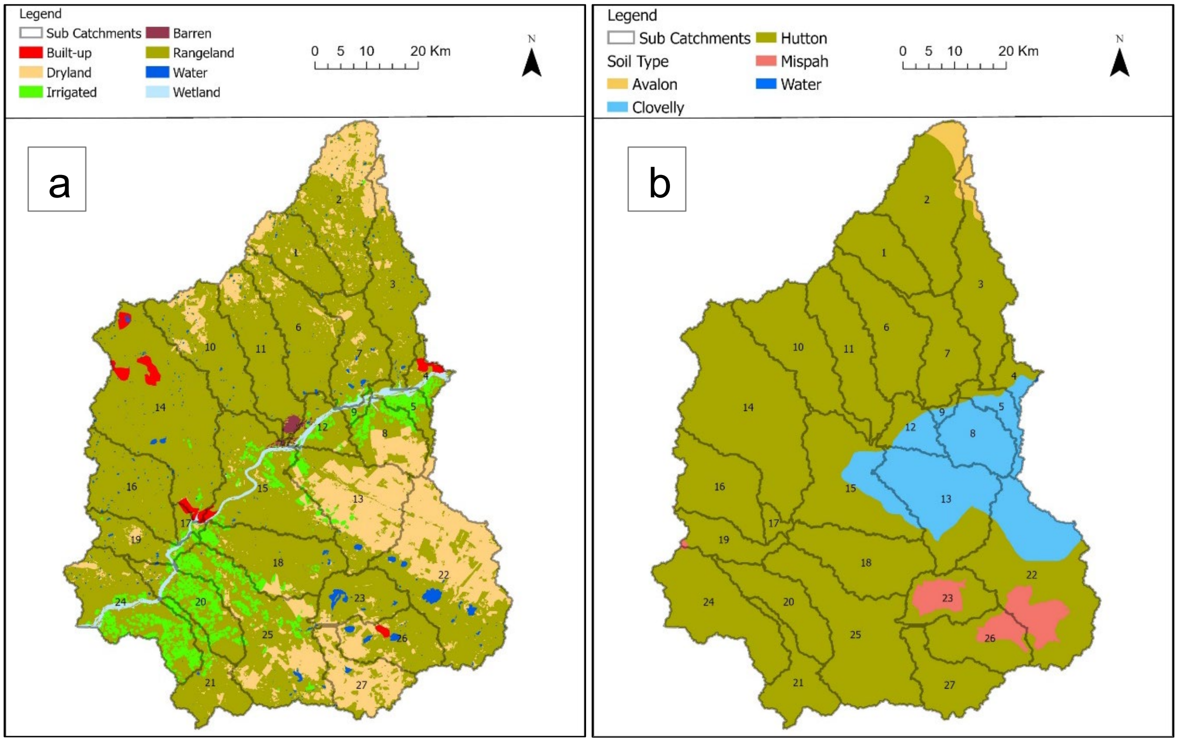


Figure 9: (a) Land cover map and (b) soil map of LVRC.

Climate parameters

Daily weather data for both models were acquired from the ARC (2019) and SAWS (2019) gauge stations at Bloemhof, Warrenton and Taung (see Figure 8). Gaps and missing values were filled with satellite-based Climate Forecast System Reanalysis (CFSR) weather data (Saha *et al.*, 2010) and the nearest neighbour interpolation method implemented in XLSTAT® for Microsoft Excel (Addinsoft, 2019).

Hydrological parameters

Flow contributions from several point or inlet sources in the lower Vaal River were incorporated into the SWAT model. Daily flow volume data from the Bloemhof Dam station were used to calculate the inlet input flow. At the same time, monthly nutrient loads were added as a point source to the most upstream sub-catchment, as described in Mararakanye *et al.* (2021). The Penman-Monteith equations were used to calculate potential evapotranspiration (PET). SWAT uses PET to estimate actual evapotranspiration (AET), considering soil moisture and crop development (Aouissi *et al.*, 2016). Point sources at Bloemhof Sewage Treatment Works (STWs), Christiana Wastewater Treatment Plant (WWTP) and Christiana STWs were constant input to the SWAT model. The average measured daily PO_4-P concentrations from DWS point sources database (http://ws.dwa.gov.za/IRIS) at Bloemhof STWs, Christiana WWTP and Christiana STWs were 11 mg/ℓ, 0.2 mg/ℓ and 3 mg/ℓ respectively. Average daily NO_3-N discharge from Christiana Aventura Vaal Spa point source was relatively small (0.6 mg/ℓ).

The Vaalharts Barrage characteristics included in the model were the surface area and the volume of water required to fill the reservoir (Table 11). Diversion of flow to the Vaalharts Irrigation Scheme was accounted for by converting daily flow measurements at station C9H018 into monthly volumes (m^3).

Table 11: Reservoir management operations parameters.

Parameter	Description	Value	Unit
RES_ESA	Reservoir surface area when the reservoir is filled to the emergency spillway	2 124	ha
RES_PSA	Reservoir surface area when the reservoir is filled to the principal spillway	2 119	ha
RES_EVOL	Volume of water stored in reservoir when filled to the emergency spillway	5 073	10^4 m^3
RES_PVOL	Volume of water stored in reservoir when filled to the principal spillway	5 068	10^4 m^3
RES_VOL	Initial reservoir volume	5 068	10^4 m^3
IFLOD1R	Beginning month of non-flood season	April	Month
IFLOD2R	Ending month of non-flood season	November	Month

Management practises

Agricultural practices such as plant growth, tillage, harvest, and fertilization are described in SWAT under the “Management” (.mgt) files. These practices affect the simulation of water balance, erosion, and pollutant load generation through the impacts of the plant growth cycle on evapotranspiration. Due to the lack of data on agricultural practices, assumptions had to be made in order to provide appropriate input to the SWAT model. The parameterization of agricultural management practices and crop rotation was described in Mararakanye *et al.* (2021) and summarized in Table 12.

Table 12: Crop rotation and management schedule for potato-onion-maize-grass under irrigation and maize-sunflower-maize-soybean under dryland used in SWAT (Mararakanye *et al.*, 2021).

		Operations date (dd-mm)					Fertiliser application date (dd-mm)			
Season	**Crop**	**Ripper**	**Harrow**	**Ridging**	**Planting**	**Irrigation (mm)**	**1st**	**2nd**	**3rd**	**Harvest / kill**
Irrigated crops										
1	Potato	30-Sep	15-Oct	30-Oct & 01-Dec	01-Nov	400	01-Nov	01-Dec	n/a	20-Feb
2	Onion	n/a	20-Apr	30-Apr	01-May	500	01-May	20-May	20-Jun	30-Sep
3	Maize	n/a	10-Oct	n/a	20-Oct	350	20-Oct	20-Nov	20-Dec	20-Apr
3-6	Grass	n/a	n/a	n/a	20-May	20	n/a	n/a	n/a	25-Sep
Dryland crops										
1	Maize	30-Sep	n/a	n/a	01-Dec	n/a	01-Dec	01-Jan	01-Feb	15-Jul
2	Sunf.	30-Sep	n/a	n/a	01-Jan	n/a	01-Jan	01-Feb	n/a	01-Jun
3	Maize	30-Sep	n/a	n/a	01-Dec	n/a	01-Dec	01-Jan	01-Feb	15-Jul
4	Soy.	30-Sep	n/a	n/a	20-Nov	n/a	n/a	n/a	n/a	01-Jun

Notes: Sunf. = Sunflower, Soy. = Soybean

Calibration and validation

After SWAT configuration, model calibration and validation were performed for streamflow at the catchment outlet with observation data obtained from the DWS river monitoring station. A temporal split-sample approach was used to split the observation data into two periods for calibration and validation. Data from 1980 to 2006 were used for calibration and data from 2007 to 2018 for validation, ensuring that most available data were used for calibration. The performance of the models was assessed before any calibration or validation was done. The SWAT model was calibrated and validated using the sequential uncertainty fitting 2 (SUFI-2) calibration optimization algorithm (Abbaspour *et al.*, 2004) available in SWAT Calibration Uncertainties Program (SWAT-CUP) version 5.1.6. The SUFI-2 expresses uncertainties in the model output variables as the 95% probability distributions, commonly referred to as the 95% prediction probability uncertainty (95PPU) band (Abbaspour *et al.*, 2007). The objective in SUFI-2 calibration is for the model results (95PPU) to capture most of the observations. Two statistical indices, namely p-factor and r-factor are used to quantify the 95PPU. P-factor is the percentage of the observed data bracketed by the 95PPU, whereas the r-factor is the thickness of the 95PPU bracket. A value greater than 0.7 for p-factor is suggested when modelling streamflow while a value closer to zero for r-factor is recommended. The selected parameters used for the calibration of the SWAT model in the LVRC are shown in Table 13.

Successive iterations were performed until a reasonable coefficient of efficiency (NSE) of Nash and Sutcliffe (1970) was achieved, as well as the coefficient of determination (r^2). A per cent deviation method (Dv) of Martinec & Rango (1989) was used as a measure of goodness-of-fit between simulated and measured streamflow.

Table 13: Selected SWAT parameters and values used for the calibration of SWAT model hydrology in the LVRC.

	Initial Value Range		Catchment Data Model	National Data Model
Parameters	**Min**	**Max**	**Fitted Value**	**Fitted Value**
r__CN2.mgt	-0.3	0.2	-0.29	-0.28
v__ESCO.hru	0	1	0.60	0.63
r__SOL_K().sol	-0.5	0.5	0.41	-0.42
r__SOL_BD().sol	-0.5	0.5	0.50	0.35
r__SOL_AWC().sol	-0.5	0.5	-0.44	0.32
v__ALPHA_BF.gw	0	1	0.86	0.53
v__GW_DELAY.gw	0	500	429.50	468.50
r__HRU_SLP.hru	-0.5	0.5	-0.01	-0.08
r__OV_N.hru	-0.5	0.5	-0.10	-0.42
r__SLSUBBSN.hru	-0.5	0.5	0.39	0.31
v__GWQMN.gw	0	5000	355	3715
v__REVAPMN.gw	0	1000	577	393
v__GW_REVAP.gw	0.02	0.2	0.16	0.07
v__EPCO.bsn	0	1	0.30	0.27
v__RCHRG_DP.gw	0	1	0.22	0.70
v__SHALLST.gw	0	5000	4165	4665
v__DEEPST.gw	0	10000	550	3450

Parameters with v__ means existing parameter value is to be replaced by a given value.
Parameters with r__ means existing parameter is multiplied by (1 + a given value).

2.3.3 SWAT application in the Mkabela Catchment

Site description of the MC

The MC lies between 29º 21' 12" and 29º 27' 16" south and 30º 36' 20" and 30º 41' 46" east in the KwaZulu-Natal Province of SA, northeast of the town Pietermaritzburg. Elevation ranges from 880 m at the catchment outlet in the southwest to 1 057 m upstream in the northeast of the catchment. The catchment area of 4 154 ha is drained by a tributary of the Mngeni River with a flow length of approximately 12.6 km from its source to the catchment outlet. Connectivity is influenced by a series of 9 farm dams and 5 wetlands along the axial valley, ranging between 0.6-10 and 5.4-22 ha, respectively. Landforms are complex, ranging from gently undulating footslopes and valley floors to very steep midslopes exceeding 20%. The climate is sub-humid with a mean annual rainfall of 825 mm of which around 80% is recorded in the summer season extending from October to April. The mean annual potential evaporation is 680 mm, as estimated by the Priestley and Taylor (1972) method in SWAT. July is the coolest month whereas February is the warmest month with mean minimum and maximum temperatures ranging from 6 to 21ºC and 17 to 28ºC, respectively. The catchment falls within the Savanna Biome (Mucina and Rutherford, 2006) but natural vegetation in the catchment has been replaced or modified by agricultural activities several decades ago. Most of the catchment is under sugarcane cultivation (3 100 ha or 75% of the catchment) with minority land uses including forestry (13%), pasture (8%) and a cabbage plot (3%).

The geology consists of sandstone of the Natal Group of the Cambrian Age and a relatively small pocket of Ecca sedimentary rocks in the north (Council for Geoscience, 2007). Soils vary from poorly drained clays predominately in the northern part of the catchment and areas with low relief (e.g. Westleigh form) to well drained sandy soils mainly in the southern part of the catchment in areas with high relief and steep slopes (e.g. Hutton form) (Land Type Survey Staff, 1972-2006). The major soil types occur in the central part of the catchment, including shallow sandy soils on steep and convex hillslopes with little water holding capacity (Cartef form occupying approximately 36% of the catchment) and deeper sandy soils on midslopes with soft or hard plinthic sub-horizons that is permeable to water (Glencoe and Avalon forms occupying approximately 20% of the catchment).

Model configuration, parameterization and calibration of the MC

Topographic and drainage network data

The MC was originally simulated by means of the AVSWAT-X model. AVSWAT-X is a graphical user interface for SWAT and ArcView® software extension (Di Luzio *et al*., 2004). First, topographic and drainage network data were prepared from a digital elevation model (DEM) with a grid cell resolution of 20 m (GISCOE, 2001). Automated routines in AVSWAT-X calculated the slope and divided the catchment into 19 sub-catchments from the DEM (see Figure 10). Appropriate contributing source areas and sub-catchment sizes had to be established by the user as percentage area of the entire catchment, i.e. 30%. The number of sub-catchment links or outlets was manually adjusted, representing all the relevant tributaries of the main river into 19 sub-catchments that are comparative in size, as well as to ensure that flow monitoring points spatially overlay with sub-catchment outlet points for calibration of model simulations with field measurements. Thus, each of the 19 sub-catchments consists of a channel with unique geometric properties not shown here including slope gradient, length and width. Manning's roughness coefficient was assigned to each segment in order to represent conditions observed in the field. Channel erosion parameters were set to default representing non-erosive channels due to the lack of data but also to eliminate channel erosion in simulations.

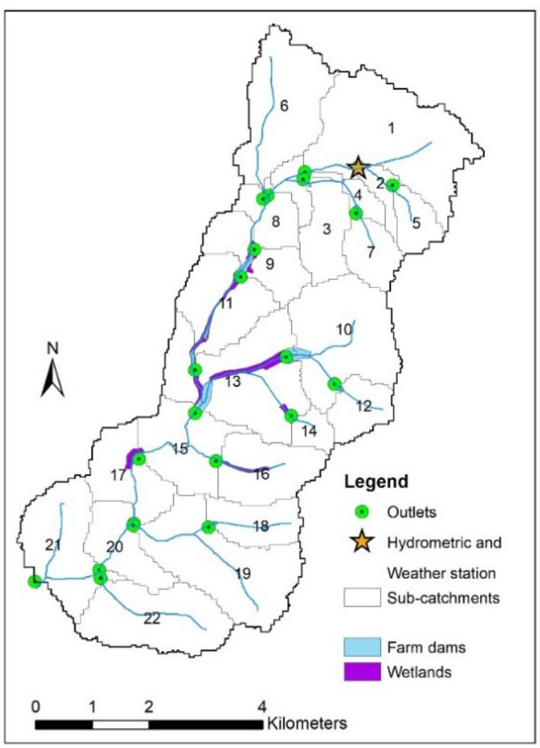


Figure 10: Sub-catchment boundaries, streams, outlets, hydrometric and weather station locations.

Land use-cover and soil data

A land cover map was digitized from SPOT 5 imagery acquired in 2006, followed by ground truthing (see Figure 11a). The land cover map was linked to a database in AVSWAT-X consisting of several plant growth parameters. An unpublished pedological soil map at a scale of 1:100 000 with textural profile descriptions for all major soils was used (Le Roux *et al.*, 2006) (see Figure 11b). The initial soil components (6 units) provided by Le Roux *et al*. (2006) were divided into smaller terrain units (18) using typical topographical algorithms of Evans (1979) and Schmidt *et al*. (2003). To account for soil variability with depth, up to three layers/horizons were incorporated into each soil component. Textural parameter values were assigned to each unit and layer according to the textural profile descriptions given by the soil map. Pedotransfer functions similar to van Tol *et al.* (2010) were used to generate the required hydraulic parameters, including available water capacity and saturated hydraulic conductivity (see Table 14). The overlay of land cover and soil maps created 130 hydrological response units (HRUs).

Table 14: Definition/description and methodology/reasoning used to assign soil parameter values in MC.

Parameter name	Definition/description	Methodology/reasoning
Number of layers in the soil	---	One soil layer/horizon was incorporated into each soil component.
Depth from soil surface to bottom of layer (mm)	Depth of each individual soil layer.	Depth descriptions/classes in the Land Type Database of SA (Land Type Survey Staff, 1972-2006) were used to assign depth to each Land Type in catchment.
Maximum rooting depth of soil profile (mm)	If no depth is specified, the model assumes the roots can develop throughout the entire depth of the soil profile.	As above.
Soil Hydrologic Group (A,B,C,D)	The U.S. Natural Resource Conservation Service (NRCS) classifies soils into four hydrologic groups based on infiltration characteristics of the soils. In term of runoff potential, Soil Group A = low, B = moderately low, C = moderately high, D = high.	HSG values (A, B, C, or D) were assigned to each dominant soil type within a given soil map unit or hillslope using the descriptions of Le Roux *et al.* (2006) and soil texture classes.
Available water capacity of the soil layer (mm H_2O/mm soil)	The plant available water, also referred to as the available water capacity, is calculated by subtracting the fraction of water present at permanent wilting point from that present at field capacity, AWC = FC - WP where AWC is the plant available water content, FC is the water content at field capacity, and WP is the water content at permanent wilting point.	Used the total profile available water given in Schulze (2007) for each Land Type.

Table 14 continued.

Parameter name	Definition/description	Methodology/reasoning
Saturated hydraulic conductivity (mm/hr)	The saturated hydraulic conductivity, K_{sat}, relates soil water flow rate (flux density) to the hydraulic gradient and is a measure of the ease of water movement through the soil. K_{sat} is the reciprocal of the resistance of the soil matrix to water flow.	Values were derived from SWAT look-up tables based on the soil texture classes of each soil series in the Land Type Database of SA – to spatially assign a conductivity value to each Land Type polygon, the values related to each soil series were weighted according to the area occupied by that soil within the polygon; therefore, the final values are an area weighted average for a Land Type.
Moist bulk density (Mg/m^3 or g/cm^3)	The soil bulk density expresses the ratio of the mass of solid particles to the total volume of the soil, $\rho b = M_S / V_T$. In moist bulk density determinations, the mass of the soil is the oven dry weight and the total volume of the soil is determined when the soil is at or near field capacity. Bulk density values should fall between 1.1 and 1.9 Mg/m^3.	Obtained from Le Roux *et al.* (2006).
Moist soil albedo (non-dimensional value between 0 and 1)	The ratio of the amount of solar radiation reflected by a body to the amount incident upon it, expressed as a fraction. The value for albedo should be reported when the soil is at or near field capacity.	Assigned according to individual soil colour characteristics.
Texture of soil layer [optional]	This data is not processed by the model and the line may be left blank.	Assigned using clay classes given to each soil form in the Land Type Database of SA.
Clay content (% soil weight)	The percent of soil particles which are < 0.002 mm in equivalent diameter.	Percentage clay for each soil map unit was obtained from Le Roux *et al.* (2006).
Silt content (% soil weight)	The percentage of soil particles which have an equivalent diameter between 0.05 and 0.002 mm.	The percentage particles smaller than 0.75 mm was used as an indication of percentage fine sand and silt (from Le Roux *et al.*, 2006).
Sand content (% soil weight)	The percentage of soil particles which have a diameter between 2.0 and 0.05 mm.	Sand = 100% – (%clay + %silt).
Rock fragment content (% soil weight)	The percent of the sample which has a particle diameter > 2 mm, i.e. the percent of the sample which does not pass through a 2 mm sieve.	Rock content values were set to 0 due to the lack of data.
Organic carbon content (% soil weight)	When defining by soil weight, the soil is the portion of the sample that passes through a 2 mm sieve.	Organic carbon content is available from the soil survey data of Le Roux *et al.* (2006).
(K) factor in SI units t/ha per unit 'erosivity'	USLE equation soil erodibility described by Wischmeier and Smith (1978).	Using the SLEMSA model of Elwell (1976), erodibility units were established and used as a guide to the assignment of USLE (Wischmeier and Smith, 1978) K-factors to Land Types (Le Roux *et al.*, 2008).

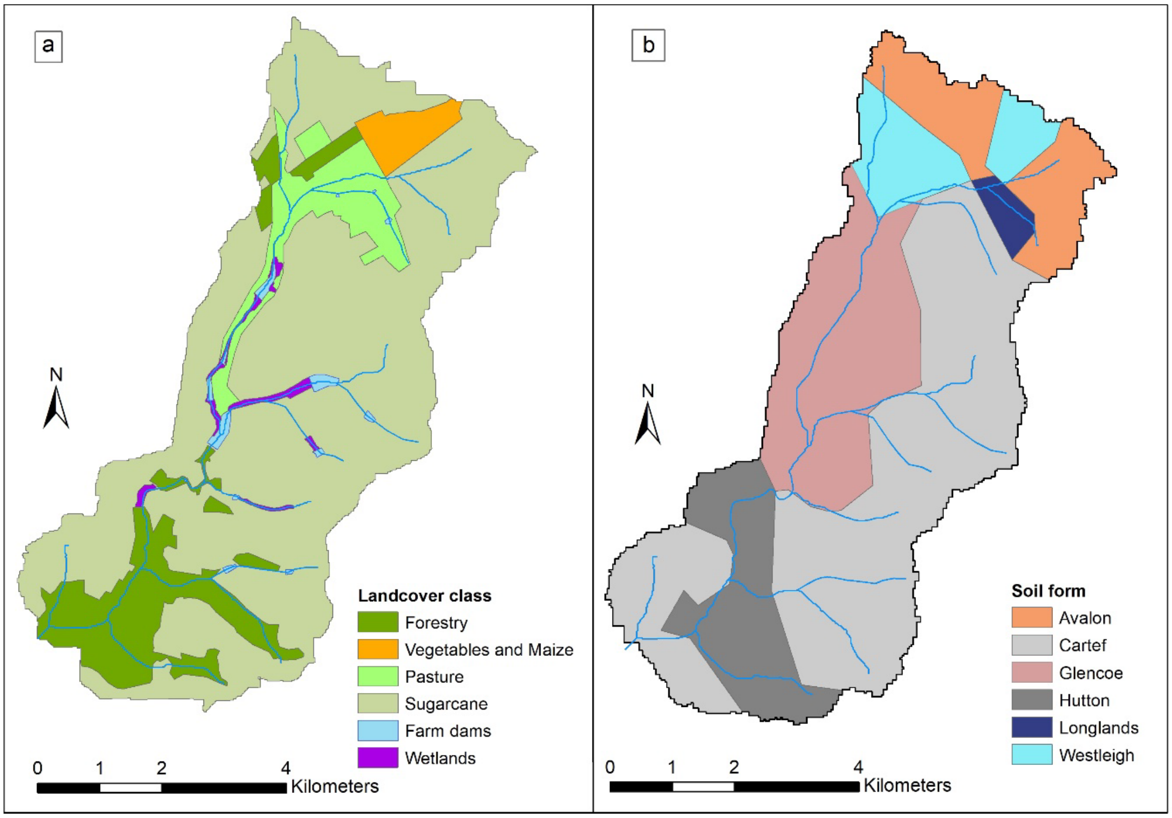


Figure 11: **(a) Land cover map and (b) soil map of MC.**

Climate parameters

AVSWAT-X also requires spatial data for several climate parameters including precipitation, temperature, solar radiation, relative humidity and wind speed. These were calculated from daily values over a 30-year period (1 January 1977 to 30 June 2008) from 4 stations within 2 kilometres or less of the catchment boundary (ARC, 2008). Since not all the stations have full records of the required parameters, incomplete records were patched with the most complete and closest stations. Finally, ancillary information regarding management practices in the catchment was incorporated including tillage operations, nutrient applications, irrigation scheduling and harvesting operations. Due to the lack of data on crop rotation systems and timing of agricultural operations, phenological plant development is based on daily accumulated heat units.

Hydrological parameters

In addition, 9 outlets were incorporated to represent outlets at the exit from 9 farm dams. AVSWAT-X also allows relatively small impoundments such as wetlands to receive loadings from a fraction of the sub-catchment area where it is located. The geographical distribution and extent of the farm dams and wetlands were digitized from SPOT 5 panchromatic sharpened images at 2.5 m resolution acquired in 2006. Table 15 contains parameter information obtained from Le Roux *et al.* (2009). The discretisation resulted in the definition of 19 sub-catchments that are joined by outlets and tributary channels branching off the main channel, including 9 farm dams and 5 wetlands along the axial valley (see Figure 10).

Table 15: Parameter information used to model each of the farm dams and wetlands in the MC (Le Roux *et al.*, 2009).

Sub-catchment	Dam area (ha)	Dam volume (m^3)	Wetland area (ha)	Wetland volume (m^3)
5	-	-	5.44	108 800
10	0.7	11 800	-	-
11	5.9	229 600	4.73	141 900
12	4.5	87 000	9.17	183 400
13	1.7	31 800	-	-
14	8.4	330 400	-	-
15	1.5	26 600	-	-
16	10.3	405 600	22.44	673 200
17	1.2	20 400	4.88	97 600
19	2.5	47 800	-	-

Management practises

Management practices data were not available at specific times, and parameter values were assigned to represent each management practice according to values provided in the SWAT database. Due to the lack of data on the crop rotation systems and timing of agricultural operations, phenological plant development is based on daily accumulated heat units. Management practices (tillage operations, nutrient applications, irrigation scheduling, and harvesting operations) were representing as follows:

- Specifying initial crop conditions prior to simulation;
- Specifying the application of nutrients to sugarcane and vegetables (the model can apply nutrients whenever the plant experiences a typical level of nitrogen stress);
- Specifying the application and source (deep aquiver in sub-catchment 1) of irrigation water applied to vegetables and pasture (the model can apply water as needed by the plant);
- Specifying harvesting operations for sugarcane (harvest only instead of harvest and kill), and forestry plantations (allows the plant to continue growing instead of frequent harvesting and killing).

Calibration and validation

Model simulation was conducted over a period of 2 years (1 July 2005 to 30 June 2008) preceded by a one-year model "warm-up" initialization period. A temporal split-sample approach was used to split the observation data into two periods for calibration and validation. Data from July 2005 to June 2006 were used for calibration and data from July 2007 to June 2008 for validation. Calibration and validation were restricted to measurements from an ISCO sampler and H-flume at the outlet of sub-catchment 8 (area of 96 ha). Calibration focused on the most sensitive model parameters similar to Tibebe and Bewket, 2010. The hydrological component was calibrated by modifying the curve number and base-flow coefficients. Model performance was improved by sequentially optimizing the coefficient of determination (r^2), as well as the widely used coefficient of efficiency (NSE) of Nash & Sutcliffe (1970). A per cent deviation method (Dv) of Martinec & Rango (1989) was used as a measure of goodness-of-fit between simulated and measured streamflow.

2.3.4 SWAT application in the Tsitsa River Catchment

Site description of the TRC

The Tsitsa River Catchment is located in the Eastern Cape Province of SA and is characterized by a steep landscape and erodible soils. It has a drainage area of 4 924 km^2 and lies between 30º 46' 58'' and 31º 28' 55'' south and 27º 55' 56'' and 29º 13' 47'' east. The Tsitsa River drains the Drakensberg Escarpment (approximately 2600 m.a.s.l.) and flows east into the Mzimvubu River (at approximately 200 m.a.s.l.) after a flow length of approximately 200 km. The climate is sub-humid with mean annual rainfall ranging from 625 mm in the lower plains to 1 327 mm in the mountains (ARC, 2012). The catchment falls mainly within the Grassland biome, with narrow bands of Bushveld along the river networks in the lower part of the catchment, as well as pockets of Afromontane Forest in fire protected ravines (Mucina and Rutherford, 2006). The main land use is extensive grazing with areas of pine and gum plantations and maize cultivation in the upper catchment.

The geology consists of a succession of sedimentary layers of the Triassic age, including Adelaide mudrock succeeded by mudstones of the Tarkastad, Molteno and Elliot Formations (Council for Geoscience, 2007). Mudstones are overlain by sandstone and siltstone of the Clarens Formation and capped by Drakensberg basaltic lava of the Jurassic age. Karoo dolerite sills and dykes are present in the sedimentary formations, leading to more resistant base level controls. Although soils in the catchment vary significantly, those from the mudstone parent material in the central part of the catchment are associated with duplex soils that are highly erodible with widespread gully erosion. Duplex soils have a marked increase in clay content from the topsoil to subsoil and an abrupt transition with respect to texture, structure and consistency (Land Type Survey Staff, 1972-2006). Soil forms that often have duplex properties include Katspruit, Kroonstad, Sterkspruit, Estcourt, and to a lesser extent Valsrivier, Swartland and Bonheim. These soils limit intrinsic permeability since water does not move readily into the subsurface matrix, which often leads to increased subsurface flow (van Tol *et al.*, 2013) causing tunnel and subsequent gully erosion. In the Tsitsa River Catchment, duplex soils often have prismacutanic subsoils that can easily be identified by the large structured prisms that are exposed on gully sidewalls or where the topsoil is completely eroded. Importantly, the subsurface matrix of duplex soils is often dispersive as a result of high sodium absorption (van Zijl *et al.*, 2014).

Model configuration, parameterization and calibration of the TRC

Topographic and drainage network data

Topographic and drainage network data were prepared from a 90 m SRTM DEM and the national 1:50 000 topographic maps with river lines (NGI, 2013). A total of 13 sub-catchments were delineated, representing all the relevant tributaries of the main river (see Figure 12).

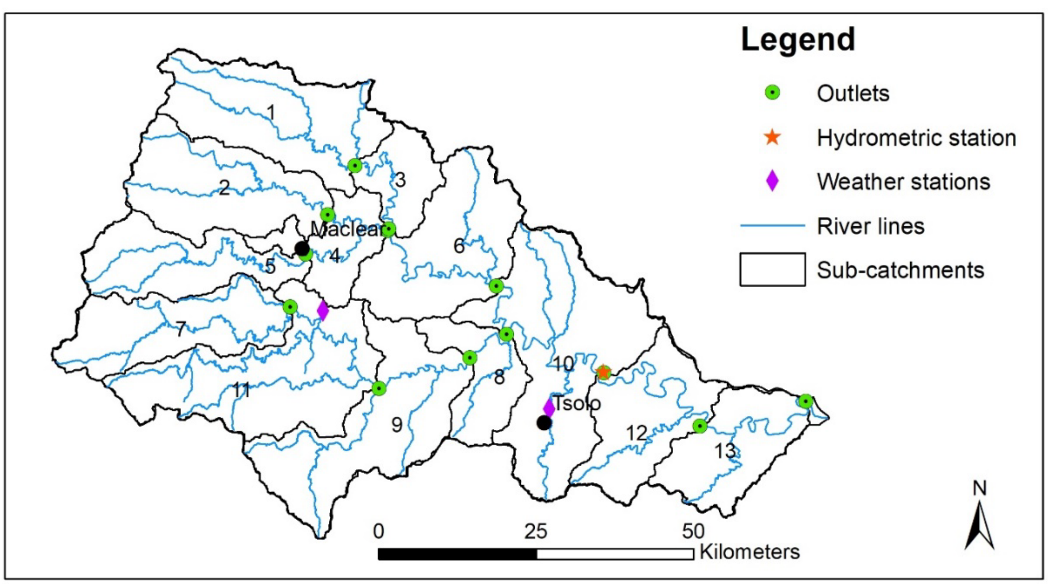


Figure 12: The TRC illustrating the 13 delineated sub-catchments, streams, outlets, hydrometric and weather station locations.

Land use-cover and soil data

A land cover map with 12 classes was created by means of unsupervised classification on SPOT 5 imagery acquired in 2011 (Figure 13a illustrates the extent of 5 most prominent land cover classes). The land cover map/classes were linked to the land cover types in the ArcSWAT database. In order to represent the variable soils in the catchment, textural and soil hydraulic parameter values were assigned to Land Types, a national soil, climate and terrain database usable at a scale of 1:250 000 (Land Type Survey Staff, 1972-2006). Pedotransfer functions similar to van Tol *et al.* (2010) were used to generate the required hydraulic parameters, including available water capacity and saturated hydraulic conductivity (see Table 16). Figure 13b illustrates the extent of the five most prominent soil types per sub-catchment in terms of soil erodibility (see Le Roux and Sumner, 2012).

Table 16: Definition/description and methodology/reasoning used to assign soil parameter values to Land Types in TRC.

Parameter name	Definition/description	Methodology/reasoning
Number of layers in the soil	---	One soil layer/horizon was incorporated into each soil component.
Depth from soil surface to bottom of layer (mm)	Depth of each individual soil layer.	Depth descriptions/classes in the Land Type Database of SA (Land Type Survey Staff, 1972-2006) were used to assign depth to each Land Type in catchment.
Maximum rooting depth of soil profile (mm)	If no depth is specified, the model assumes the roots can develop throughout the entire depth of the soil profile.	As above.
Soil Hydrologic Group (A,B,C,D)	The U.S. Natural Resource Conservation Service (NRCS) classifies soils into four hydrologic groups based on infiltration characteristics of the soils. In term of runoff potential, Soil Group A = low, B = moderately low, C = moderately high, D = high.	Used the hydrological classes given in Schulze (2007) for each Land Type.
Available water capacity of the soil layer (mm H_2O/mm soil)	The plant available water, also referred to as the available water capacity, is calculated by subtracting the fraction of water present at permanent wilting point from that present at field capacity, AWC = FC - WP where AWC is the plant available water content, FC is the water content at field capacity, and WP is the water content at permanent wilting point.	Used the total profile available water given in Schulze (2007) for each Land Type.
Saturated hydraulic conductivity (mm/hr)	The saturated hydraulic conductivity, Ksat, relates soil water flow rate (flux density) to the hydraulic gradient and is a measure of the ease of water movement through the soil. Ksat is the reciprocal of the resistance of the soil matrix to water flow.	Values were derived from SWAT look-up tables based on the soil texture classes of each soil series in the Land Type Database of SA – to spatially assign a conductivity value to each Land Type polygon, the values related to each soil series were weighted according to the area occupied by that soil within the polygon; the final values are an area weighted average for a Land Type.

Table 16 continued.

Parameter name	Definition/description	Methodology/reasoning
Moist bulk density (Mg/m^3 or g/cm^3)	The soil bulk density expresses the ratio of the mass of solid particles to the total volume of the soil, $\rho b = M_S / V_T$. In moist bulk density determinations, the mass of the soil is the oven dry weight and the total volume of the soil is determined when the soil is at or near field capacity. Bulk density values should fall between 1.1 and 1.9 Mg/m^3.	An average value of 1.6 g/cm^3 was assigned to all Land Types due to the lack of data at this scale.
Moist soil albedo (non-dimensional value between 0 and 1)	The ratio of the amount of solar radiation reflected by a body to the amount incident upon it, expressed as a fraction. The value for albedo should be reported when the soil is at or near field capacity.	Albedo values were assigned to each soil series in the Land Type Database of SA as follows: sands = 0.25 (coded as soil forms Ah, Ai, Ha and Hb); clays (soil forms coded Ea) = 0.7; remaining textures = 0.5.
Texture of soil layer [optional]	This data is not processed by the model and the line may be left blank.	Assigned using clay classes given to each soil form in the Land Type Database of SA.
Clay content (% soil weight)	The percent of soil particles which are < 0.002 mm in equivalent diameter.	Assigned using the average topsoil clay classes given to each soil form in the Land Type Database of SA.
Silt content (% soil weight)	The percentage of soil particles which have an equivalent diameter between 0.05 and 0.002 mm.	Due to the lack of data, silt content was assigned values between 10-22.5%, increasing with increase in clay as follows: percentage of Land Type with <= 6% clay = 10% silt; 6.1 - 15% clay = 15% silt; 15.1 - 25% clay = 17.5% silt; 25.1 - 35% clay = 20% silt; 35.1 - 55% clay = 22.5% silt.
Sand content (% soil weight)	The percentage of soil particles which have a diameter between 2.0 and 0.05 mm.	Sand = 100% – (%clay + %silt + %rock + %carbon).
Rock fragment content (% soil weight)	The percent of the sample which has a particle diameter > 2 mm, i.e. the percent of the sample which does not pass through a 2 mm sieve.	Used agricultural restriction/rock (MB) classes in Land Type Database of SA as follows: MB0=0%; MB1=20%; MB2=50%; MB3=20%; MB4=100% (no soil).
Organic carbon content (% soil weight)	When defining by soil weight, the soil is the portion of the sample that passes through a 2 mm sieve.	An unpublished Carbon map of SA (derived from soil profile data and Land Type Database) was used to assign carbon values to each Land Type in the catchment.
(K) factor in SI units t/ha per unit 'erosivity'	USLE equation soil erodibility described by Wischmeier and Smith (1978).	Using the SLEMSA model of Elwell (1976), erodibility units were established and used as a guide to the assignment of USLE (Wischmeier and Smith, 1978) K-factors to Land Types (Le Roux *et al.*, 2008).

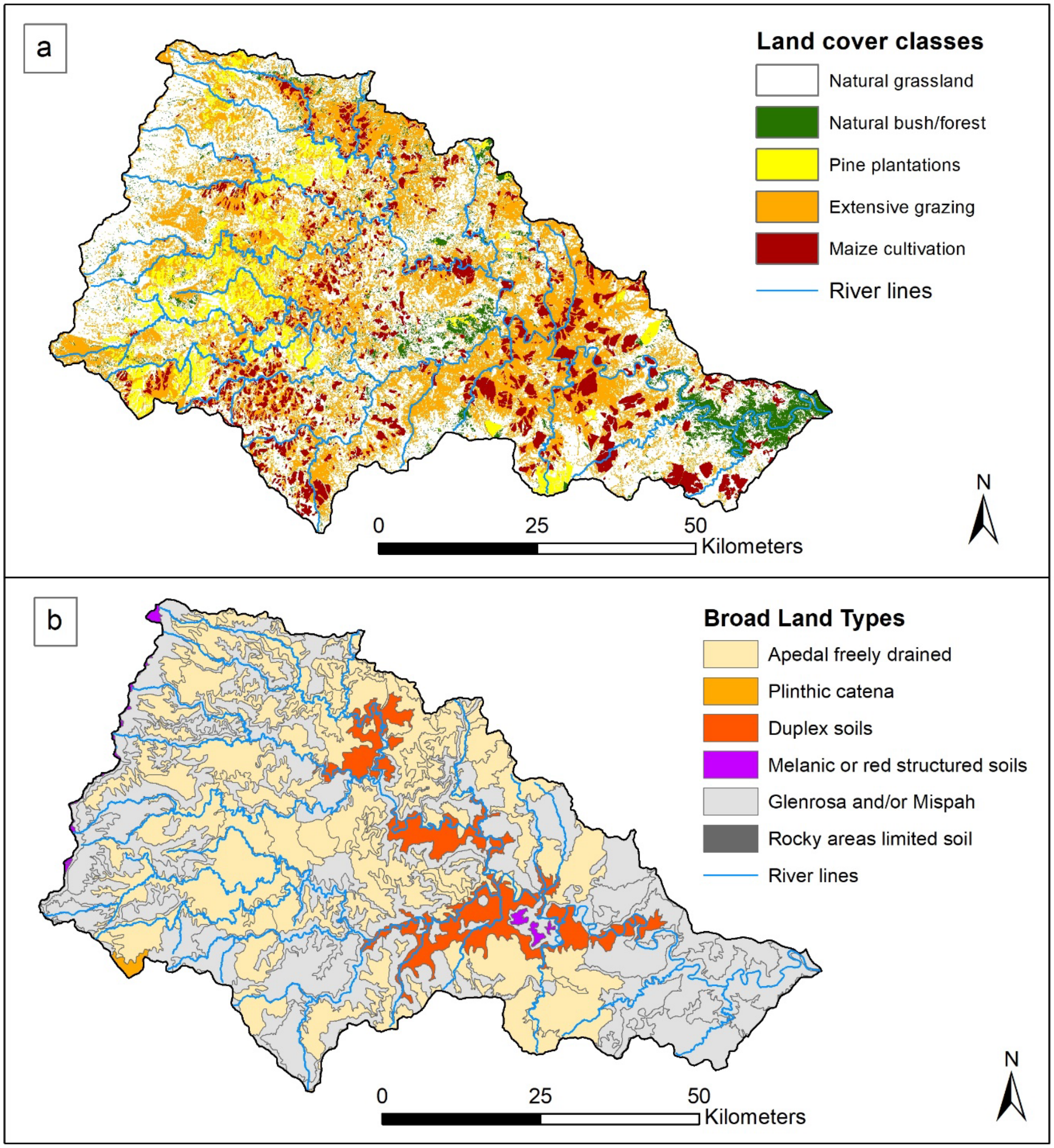


Figure 13: (a) Land cover map and (b) soil map of TRC.

Climate parameters

Daily precipitation and temperature values over a 30-year period (1978 to 2007) were retrieved from 2 meteorological stations located within the catchment, more specifically near Maclear and Tsolo (ARC, 2012).

Management practises

Management practices data were not available at specific times, and parameter values were assigned to represent each management practice according to values provided in the SWAT database. Due to the lack of data on the crop rotation systems and timing of agricultural operations, phenological plant development is based on daily accumulated heat units.

Calibration and validation

Model simulation was conducted over a period of 5 years (2008 to 2012) preceded by a one-year warm-up period to get the hydrological cycle fully operational. A temporal split-sample approach was used to split the observation data into two periods for calibration and validation. Data from 2008 to 2009 were used for calibration and data from 2010 to 2012 for validation. Calibration of ArcSWAT was performed on the hydrological part of the model on a monthly time-step by adjusting the curve numbers and base-flow coefficients as done in other studies (e.g. Tibebe & Bewket, 2011). Model performance was improved by sequentially optimizing the coefficient of determination (r^2), as well as the widely used coefficient of efficiency (NSE) of Nash & Sutcliffe (1970). A per cent deviation method (*Dv*) of Martinec & Rango (1989) was used as a measure of goodness-of-fit between simulated and measured streamflow. The next Section presents the results of modelling the four research catchments with the national baseline datasets.

3. RESULTS AND DISCUSSION

First, the national baseline datasets are presented in Section 3.1 as a series of maps and tables that are stored on a data portal system. Section 3.2 presents the results of modelling the four research catchments with the national baseline datasets including Middle Olifants, Lower Vaal, Mkabela and Tsitsa. Comparison of the results and accuracies of the two input datasets (original input versus baseline input), allows appraisal of the performance of the baseline input data.

3.1 Online Data Portal System for South Africa

The national input database to run the ArcSWAT model is stored in the Water Research Observatory data portal of WRC project: C2020/2021-00440 titled "Development and application of a big data platform to improve agricultural water resources management in South Africa" (see Figure 14). The URL is: https://www.waterresearchobservatory.org/data-and-resources/hydrological-data-and-modelling. The portal provides geo-spatial input datasets including:

- SWAT catchment outline data (tertiary and quaternary) including the hydrologically corrected SRTM DEM of SA at 90 m resolution (Weepener *et al.*, 2012);
- South African National Land Cover (SANLC, 2014; 2018; 2020) linked to SWAT land cover codes;
- Soil map with SWAT attribute data for each Land Type of SA (Land Type Survey Staff, 1972-2006);
- Weather statistics (WGN) files required as input by the model.

Appendix 1 lists the Metadata of the datasets, based on ISO 19115 Geospatial metadata standards.

Appendix 2 lists the post-graduate students whose studies are contributed to the project.

Appendix 3 indicates different forms of knowledge dissemination.

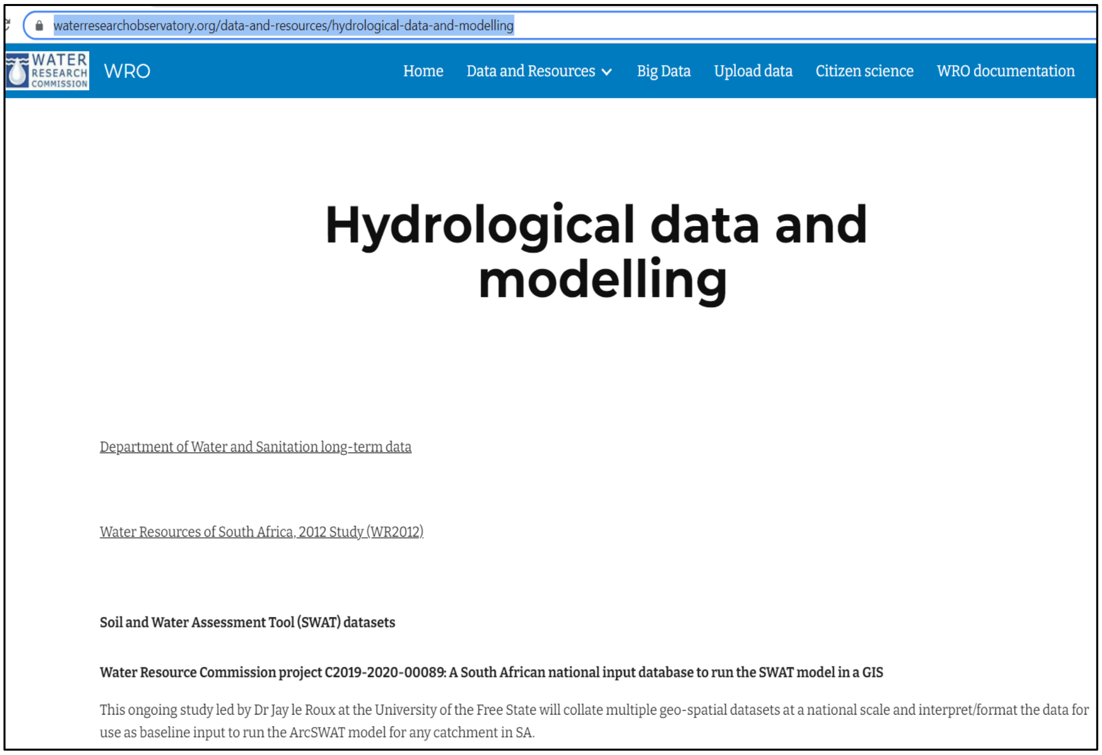


Figure 14: Title page of the national input database to run the ArcSWAT model (in a GIS) is stored in the Water Research Observatory data portal of WRC project: C2020/2021-00440.

3.1.1 SWAT catchment outline data

Catchment outlines were obtained and prepared from the hydrologically corrected 90 m SRTM DEM and derived products of Weepener *et al.* (2012). Catchment data are stored on the portal as: SWAT_SA_DEM_catchments including the following:

- SRTM_DEM_90m.zip – comprises the hydrologically corrected 90 m SRTM DEM of Weepener *et al.* (2012);
- Catchments_tertiary_buffer100m.zip – comprises the tertiary catchment shapefile with 100m buffer, needed by SWAT to identify the area of interest (see Figure 15);
- Catchments_quaternary_buffer100m.zip – comprises the quaternary catchments shapefile with 100m buffer, needed by SWAT to identify the area of interest (see Figure 16);
- Catchment_inlets_outlets.zip – comprises the tertiary and quaternary catchment outlets and inlets shapefile (see Figure 17);
- DWS_monitoringstations_locations.zip – comprises all the DWS monitoring stations locations shapefile (DWS, 2022b) (see Figure 18).

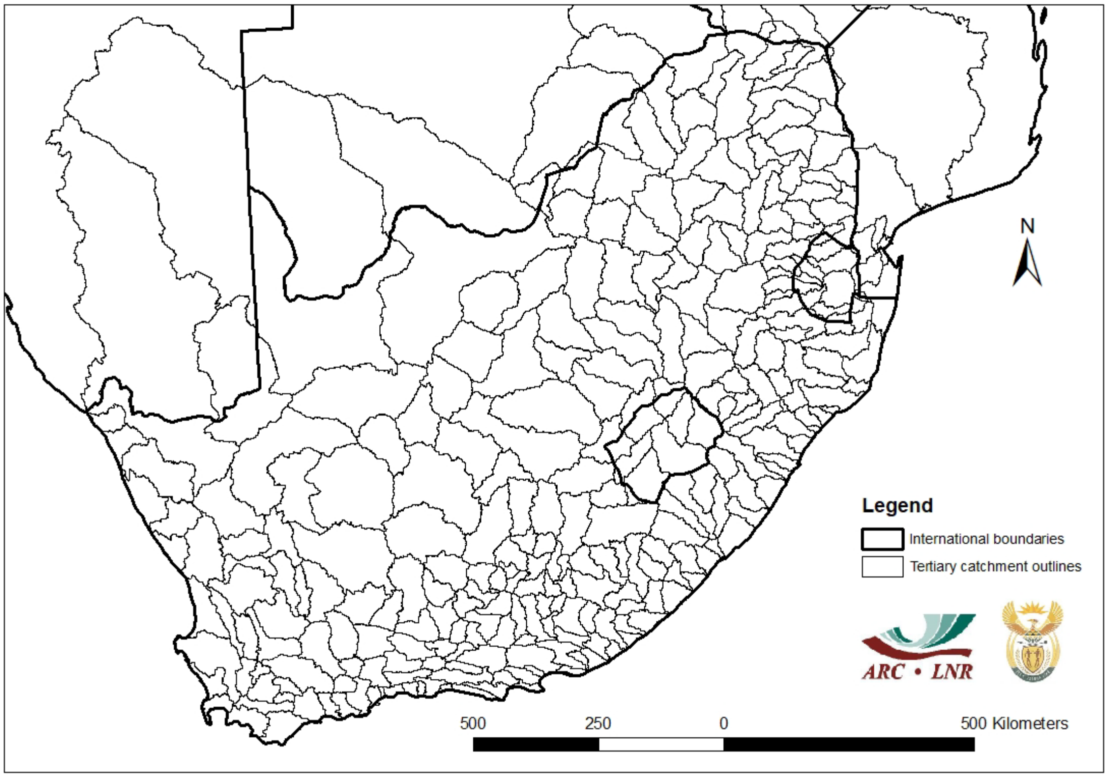


Figure 15: Tertiary catchment boundaries of SA (Weepener *et al.*, 2012).

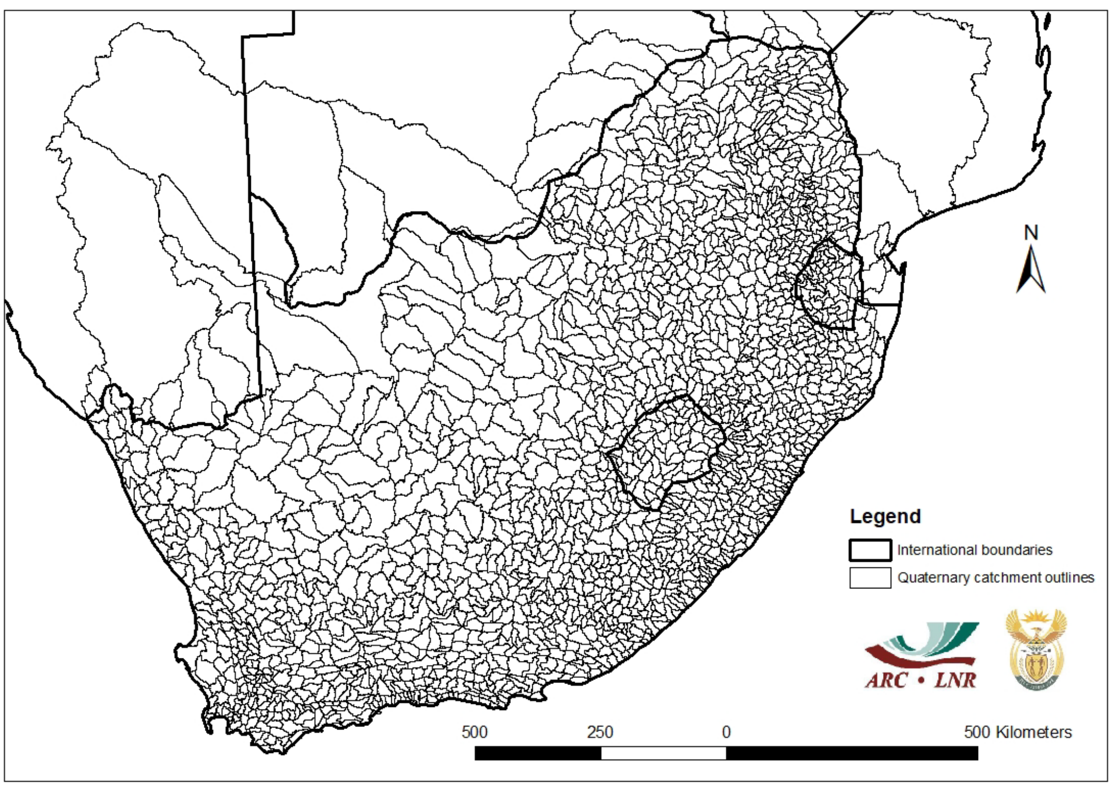


Figure 16: Quaternary catchment boundaries of SA (Weepener *et al.*, 2012).

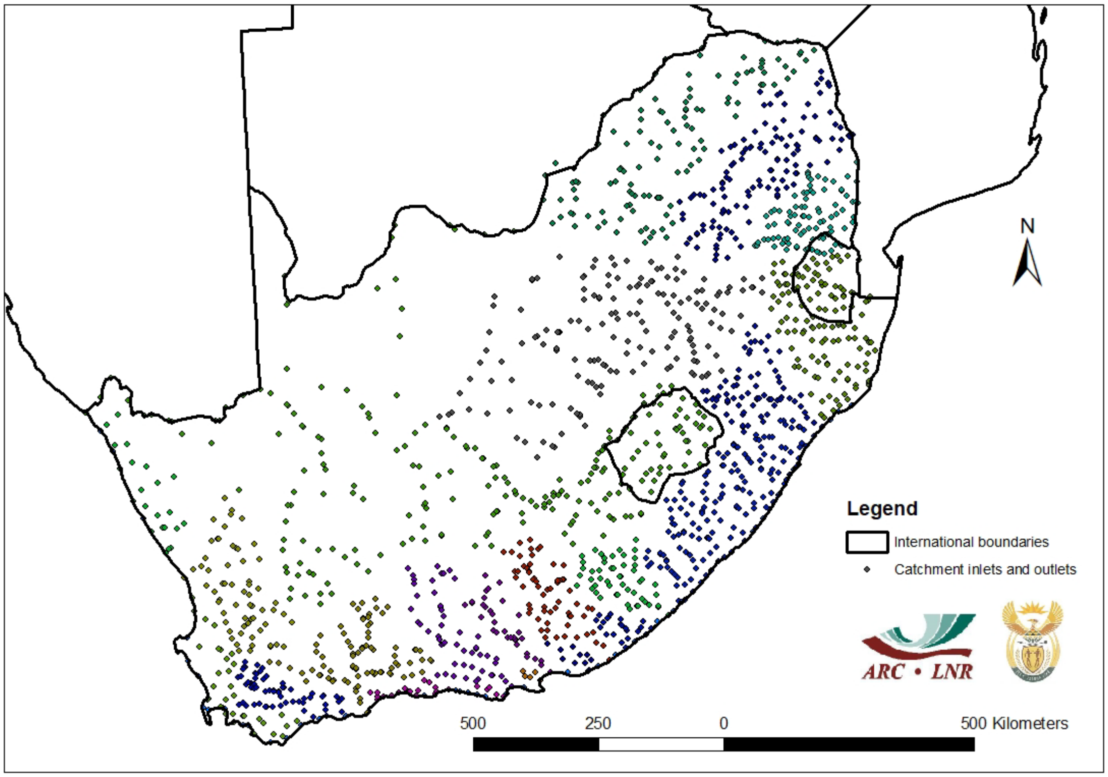


Figure 17: **Tertiary and Quaternary catchment inlet and outlet locations of SA (Weepener *et al.*, 2012).**

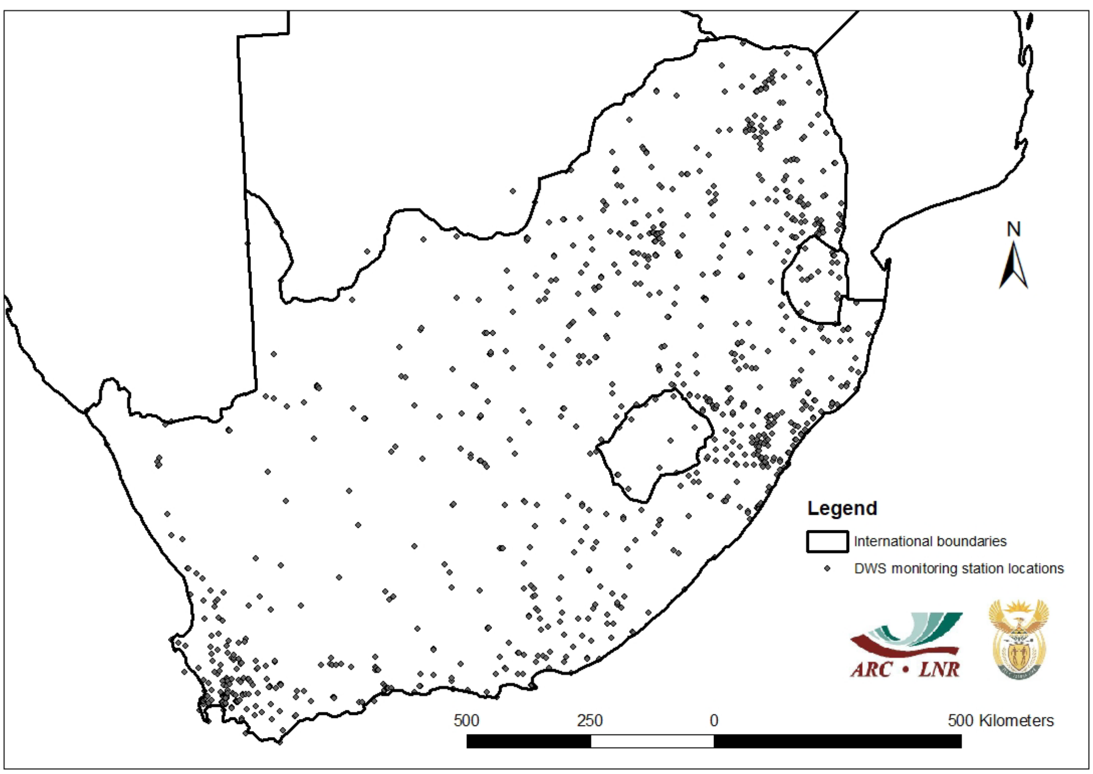


Figure 18: **DWS monitoring stations locations of SA (DWS, 2022b).**

3.1.2 SWAT land cover data

Table 17 lists the land cover types in the ArcSWAT database (excluding parameters) that were linked to the National Land Cover (2013/14 and 2018) maps of SA. Land cover data are stored on the portal as: SWAT_SA_landcover including the following:

- NLC2020_SWAT.zip – comprises the 2020 national land cover grid of SA, of which the grid values are reclassified to match the SWAT land cover codes;
- LU_landcover2020.txt – comprises the user lookup table that links abovementioned 2020 national land cover grid with the SWAT land cover codes;
- NLC2018_SWAT.zip – comprises the 2018 national land cover grid of SA, of which the grid values are reclassified to match the SWAT land cover codes;
- LU_landcover2018.txt – comprises the user lookup table that links abovementioned 2018 national land cover grid with the SWAT land cover codes;
- NLC2013_SWAT.zip – comprises the 2013/14 national land cover grid of SA, of which the grid values are reclassified to match the SWAT land cover codes;
- LU_landcover2013.txt – comprises the user lookup table that links abovementioned 2013 national land cover grid with the SWAT land cover codes.

Table 17: List of land cover types in the ArcSWAT database (excluding parameters) that was linked to the National Land Cover (2013/14, 2018 and 2020) maps of SA.

SWAT land cover Code*	SWAT land cover name*	2013 NLC type	2018 NLC type 2020 NLC type
Crop classes			
1	Agricultural Land-Generic	Cultivated comm fields (high) Cultivated comm fields (med) Cultivated comm fields (low)	Commercial annual crops rain-fed / dryland
2	Agricultural Land-Row Crops	Cultivated subsistence (high) Cultivated subsistence (med) Cultivated subsistence (low)	Subsistence / small-scale annual crops
3	Agricultural Land-Close-grown	Cultivated comm pivots (high) Cultivated comm pivots (med) Cultivated comm pivots (low)	Commercial annual crops pivot irrigated Commercial annual crops non-pivot irrigated
4	Orchard	Cultivated orchards (high) Cultivated orchards (med) Cultivated orchards (low)	Cultivated commercial permanent orchards
5	Hay		
6	Forest-Mixed	Thicket /Dense bush Shrubland fynbos	Contiguous low forest & thicket Low shrubland (fynbos)
7	Forest-Deciduous	Woodland/Open bush	Open woodland
8	Forest-Evergreen	Indigenous Forest	Contiguous (indigenous) forest Dense forest & woodland

Table 17 continued.

SWAT land cover Code*	SWAT land cover name*	2013 NLC type	2018 NLC type 2020 NLC type
9	Wetlands-Mixed		
10	Wetlands-Forested		Mangrove wetlands
11	Wetlands-Non-Forested	Wetlands	Herbaceous wetlands (currently mapped) Herbaceous wetlands (previously mapped)
15	Range-Grasses	Grassland Urban school and sports ground Urban sports and golf (low grass)	Sparsely wooded grassland Natural grassland Urban recreational fields (grass)
16	Range-Brush	Low shrubland Urban sports and golf (dense) Urban sports and golf (open)	Low shrubland (other) Urban recreational fields (tree) Urban recreational fields (bush)
17	Southwestern US (Arid) Range		Low shrubland (succulent karoo) Low shrubland (nama karoo)
18	Water	Water seasonal Water permanent Mines water seasonal Mines water permanent	Natural rivers Natural estuaries & lagoons Natural ocean & coastal Natural lakes Natural pans (flooded @ observation times) Artificial dams (including canals) Artificial sewage ponds Artificial flooded mine pits
26	Sugarcane	Cultivated cane pivot - crop Cultivated cane pivot - fallow Cultivated cane commercial - crop Cultivated cane commercial - fallow Cultivated cane emerging - crop Cultivated cane emerging - fallow	Cultivated commercial sugarcane pivot irrigated Cultivated commercial sugarcane non-pivot Cultivated emerging farmer sugarcane non-pivot
94	Pine	Plantations / Woodlots mature Plantation / Woodlots young Plantation / Woodlots clearfelled	Contiguous & dense plantation forest Open & sparse plantation forest Temporary unplanted (clear-felled) plantation
98	Vineyard	Cultivated vines (high) Cultivated vines (med) Cultivated vines (low)	Cultivated commercial permanent vines
114	Pineapple	Cultivated permanent pineapple	Cultivated commercial permanent pineapples
118	Barren	Mines 1 bare Mines 2 semi-bare Erosion (donga) Bare none vegetated Urban informal (low veg / grass) Urban informal (bare) Urban residential (bare) Urban smallholding (bare) Urban sports and golf (bare) Urban township (bare) Urban village (bare)	Natural rock surfaces Dry pans Eroded lands Sand dunes (terrestrial) Coastal sand & dunes Bare riverbed material Other bare Fallow land & old fields (trees) Fallow land & old fields (bush) Fallow land & old fields (grass) Fallow land & old fields (bare) Fallow land & old fields (low shrub) Residential informal (bare) Smallholdings (bare) Urban recreational fields (bare) Mines: extraction pits, quarries Mines: salt mines Mine: tailings and resource dumps Land-fills Fallow land & old fields (wetlands)

* SWAT land cover codes/names not corresponding to the NLC data were not used/linked. Thus, only SWAT land cover codes/names similar to NLC classes were linked.

Table 17 continued.

SWAT land cover Code*	SWAT land cover name*	2013 NLC type	2018 NLC type 2020 NLC type
Urban classes			
100	Residential-High Density	Urban built-up (low veg / grass)	Residential formal (tree)
200	Residential-Medium Density	Urban built-up (dense trees / bush) Urban built-up (open trees / bush) Urban built-up (bare) Urban township (dense trees / bush) Urban township (open trees / bush) Urban township (low veg / grass)	Residential formal (bush) Residential formal (low veg / grass) Residential formal (bare)
300	Residential-Med/Low Density	Urban residential (dense trees/bush) Urban residential (open trees/bush)	
400	Residential-Low Density	Urban informal (dense trees / bush) Urban informal (open trees / bush) Urban residential (low veg / grass) Urban smallholding (dense trees / bush) Urban smallholding (open trees / bush) Urban smallholding (low veg / grass) Urban village (dense trees / bush) Urban village (open trees / bush) Urban village (low veg / grass)	Residential informal (tree) Residential informal (bush) Residential informal (low veg / grass) Village scattered (bare & low veg/ grss combo) Village dense (bare & low veg / grss combo) Smallholdings (tree) Smallholdings (bush) Smallholdings (low veg / grass)
500	Commercial	Urban commercial	Commercial
600	Industrial	Urban industrial Mine buildings	Industrial Artificial sewage ponds Mines: surface infrastructure
700	Transportation		Roads & rails (major linear)

* SWAT land cover codes/names not corresponding to the NLC data were not used/linked. Thus, only SWAT land cover codes/names similar to NLC classes were linked.

3.1.3 SWAT user soil data

Tables 18 and 19 indicate parameter values assigned to a selection of Land Types of SA. Table 18 indicates the topsoil parameter values, whereas Table 19 indicates the subsoil (or second soil layer) parameter values of the selected Land Types. Soil data are stored on the portal as: SWAT_SA_usersoil including the following:

- Usersoil_SWAT.zip – comprises the Land Type shapefile of SA without soil attribute data;
- Usersoil.accdb – comprises a Microsoft Access database table with soil attribute data for each Land Type of SA;
- LU_usersoil.txt – comprises the user lookup table that links abovementioned Land Type grid with abovementioned Access database table with soil attribute data for each Land Type of SA.

Table 18: Topsoil parameter values assigned to the Land Types of SA (only a selection of Land Types are shown).

Land Type code	Hydrological Group (A, B, C, D)	Max soil depth (mm)	Layer 1 soil depth (mm)	Bulk density (g/cm³)	Available water (mm H2O/ mm soil)	Hydraulic conductivity (mm/hr)	Soil albedo (0-1)	Clay content (%)	Silt content (%)	Sand content (%)	Rock content (%)	Organic carbon content (%)	K factor (t/ha per unit 'erosivity')
Aa7	C	820	300	1.5	0.4	210	0.6	2.5	10	87.5	7.2	1.5	0.1582
Ac14	B	850	300	1.5	0.5	210	0.4	2.6	10	87.4	28.6	2	0.176175
Ae177	D	380	300	1.6	0.4	210	0.3	3.3	10	86.7	5.7	0.75	0.259375
Ag107	A	1100	300	1.5	0.5	210	0.3	4	10	86	9	0.05	0.1875
Ai37	C	800	280	1.5	0.4	210	0.1	1.8	10	88.2	12.4	1	0.3045
Bb107	C	820	300	1.5	0.4	210	0.4	2.1	10	87.9	7.2	0.75	0.276525
Bc47	C	510	290	1.5	0.4	210	0.4	1.3	10	88.7	17.8	0.5	0.20955
Ca127	B	600	300	1.6	0.3	210	0.4	2.7	10	87.3	16.8	0.75	0.48347
Da115	B	390	290	1.5	0.3	210	0.4	1.7	10	88.3	19.1	1	0.4928
Db114	A	650	300	1.4	0.5	210	0.4	8	15	77	5.2	1	0.46065
Dc30	C	320	250	1.5	0.4	210	0.4	1.5	10	88.5	14.3	1	0.41173
Ea101	C	270	220	1.5	0.4	210	0.7	4.8	10	85.2	24.8	2	0.248125
Fa1023	B	920	300	1.6	0.5	210	0.4	3.1	10	86.9	2.6	1	0.227325
Fb112	C	330	230	1.5	0.4	210	0.4	6.6	15	78.4	10.9	1	0.298275
Fc131	B	510	290	1.5	0.4	210	0.4	1.7	10	88.3	20.5	1	0.45474
Gb22	A	260	250	1.5	0.4	61	0.7	9.3	15	75.7	23.7	1	0.33225
Ha29	B	1160	300	1.4	0.5	210	0.6	3	10	87	9.8	0.75	0.179
Hb117	A	1120	280	1.4	0.5	210	0.5	2.3	10	87.7	0.5	0.25	0.3433
Ia103	C	320	210	1.5	0.4	210	0.3	3.8	10	86.2	0.8	1	0.2723
Ib116	B	1140	300	1.4	0.1	210	0.3	3.9	10	86.1	68.1	0.5	0.18456
Ic51	B	860	290	1.5	0.4	210	0.3	1.8	10	88.2	84.8	0.05	0.132

The default value of 0.5 was used for the fraction of porosity (void space) from which anions are excluded.
The default value of 0.5 was used for the crack volume potential of soil.
Electrical conductivity is not currently active.

Table 19: Subsoil parameter values assigned to the Land Types of SA (only a selection of Land Types are shown).

Land Type code	Hydrological Group (A, B, C, D)	Max soil depth (mm)	Layer 1 soil depth (mm)	Bulk density (g/cm^3)	Available water (mm H2O/ mm soil)	Hydraulic conductivity (mm/hr)	Soil albedo (0-1)	Clay content (%)	Silt content (%)	Sand content (%)	Rock content (%)	Organic carbon content (%)	K factor (t/ha per unit 'erosivity')
Aa7	C	820	820	1.5	0.4	210	0.4	2.5	10	87.5	7.2	0.75	0.1582
Ac14	B	850	850	1.5	0.5	210	0.4	2.6	10	87.4	28.6	1.25	0.176175
Ae177	D	380	380	1.5	0.3	210	0.3	4	10	86	5.7	0.75	0.259375
Ag107	A	1100	1100	1.4	0.5	210	0.3	4.8	10	85.2	9	0.05	0.1875
Ai37	C	800	800	1.6	0.3	210	0.1	1.8	10	88.2	12.4	0.05	0.3045
Bb107	C	820	820	1.5	0.4	210	0.4	2.1	10	87.9	7.2	0.05	0.276525
Bc47	C	510	510	1.5	0.4	210	0.4	1.3	10	88.7	17.8	0.05	0.20955
Ca127	B	600	600	1.5	0.3	210	0.4	2.7	10	87.3	16.8	0.25	0.48347
Da115	B	390	390	1.6	0.3	210	0.6	1.7	10	88.3	19.1	0.05	0.4928
Db114	A	650	650	1.5	0.4	13	0.6	16	17.5	66.5	5.2	0.05	0.46065
Dc30	C	320	320	1.5	0.4	210	0.6	1.5	10	88.5	14.3	0.05	0.41173
Ea101	C	270	270	1.5	0.4	210	0.7	4.8	10	85.2	24.8	1	0.248125
Fa1023	B	920	920	1.5	0.5	210	0.4	3.1	10	86.9	2.6	1	0.227325
Fb112	C	330	330	1.5	0.4	210	0.4	6.6	15	78.4	10.9	0.05	0.298275
Fc131	B	510	510	1.5	0.3	210	0.4	1.7	10	88.3	20.5	0.05	0.45474
Gb22	A	260	260	1.5	0.5	61	0.7	9.3	15	75.7	23.7	0.05	0.33225
Ha29	B	1160	1160	1.4	0.5	210	0.6	3	10	87	9.8	0.05	0.179
Hb117	A	1120	1120	1.4	0.5	210	0.5	2.3	10	87.7	0.5	0.05	0.3433
Ia103	C	320	320	1.5	0.4	210	0.3	4.6	10	85.4	0.8	1	0.2723
Ib116	B	1140	1140	1.4	0.2	210	0.3	3.9	10	86.1	68.1	0.05	0.18456
Ic51	B	860	860	1.5	0.4	210	0.3	1.8	10	88.2	84.8	0.05	0.132

The default value of 0.5 was used for the fraction of porosity (void space) from which anions are excluded.
The default value of 0.5 was used for the crack volume potential of soil.
Electrical conductivity is not currently active.

3.1.4 SWAT weather generator files

Tables 20 and 21 indicates two sets of (12) weather statistics (WGN) covering the periods 2001-2020 and 1981-2000, respectively. Both Tables provide the weather statistics for the month of January only (for practical purposes). Table 20 indicates the WGN statistics from 2001-2020 of 12 stations for the month of January, whereas Table 21 indicates the WGN statistics from 1981-2000 of 11 stations for the month of January. The station in Hobhouse (30673), did not have enough data to calculate statistics for the period 1981-2000, and were excluded in the 1981-2000 dataset. WGN files are stored on the portal as: SWAT_SA_wgn including the following:

- wgenuser1981-2000.accdb – comprises a Microsoft Access database table with 11 WGN files containing the WGN statistics from 1981-2000 of 11 stations;
- wgenuser2001-2020.accdb – comprises a Microsoft Access database table with 12 WGN files containing the WGN statistics from 2001-2020 of 12 stations.

SWAT users should select the WGN file closest to a respective study site (catchment). Using the national baseline datasets, the next step was to apply ArcSWAT in four (previously simulated) research catchments including Middle Olifants, Lower Vaal, Mkabela and Tsitsa.

Table 20: Weather statistics (WGN) from 1981-2000 of 12 stations for the month of January.

Station	Lat-itude	Long-itude	Elevation (m.a.s.l.)	Max tempe-rature (˚C)	Min tempe-rature (˚C)	Max tempe-rature SD (˚C)	Min tempe-rature SD (˚C)	Precipit-ation (mm/ month)	Precipit-ation SD (mm/ month	Precipit-ation Skp (mm/ month	Proba-bility wet day after dry	Proba-bility wet day after wet	Days of Precip-itation	Max 0.5 hour rain-fall (mm)	Daily solar radia-tion (MJ/ m^2-day)	Dew-point tempe-rature (˚C)	Wind speed (m/s)
Roodeplaat. Pretoria	-25.6	28.35	1168	30.26	16.98	3.2	1.9	120.66	11.09	6.79	0.29	0.53	11.55	12.12	21.88	61.73	0.91
Vaalharts, Jan Kemp-dorp	-27.95	24.83	1180	33.64	16.49	4	3.21	67.2	6.46	4.18	0.31	0.51	11.68	7.37	26.05	54.53	1.48
Upington, Proefplaas	-28.46	21.2	798	36.49	17.76	4.2	3.5	41.4	5.55	6.13	0.14	0.38	5.3	6.33	27.78	43.63	1.46
ITSG, Nelspruit	-25.45	30.97	673	28.89	18.29	6.75	4.09	148.33	11.67	3.89	0.35	0.59	14.45	14.36	20.39	67.56	0.8
AWS, Bathurst	-33.51	26.82	226	26.3	17.07	4.92	3.00	53.96	8.26	10.73	0.28	0.48	10.33	6.43	20.88	71.02	2.00
Hobhouse	-29.48	27.13	1500	29.64	14.8	3.82	2.46	89.57	6.83	4.24	0.3	0.63	13.93	8.24	24.92	57.66	1.39
Debrak, Loeriesfontein	-31.03	19.52	800	33.29	13.58	4.23	3.81	5.32	1.13	10.48	0.06	0.25	2.46	1.22	29.94	46.32	3.57
Carnarvon	-30.96	21.98	1327	32.44	15.63	3.13	2.95	25.36	3.59	5.97	0.11	0.34	4.23	4.23	26.05	42.22	3.36
St. Lucia	-28.21	32.48	14	30.62	20.98	3.45	1.96	363.26	52.87	7.92	0.26	0.63	12.05	26.94	19.18	76.11	1.38
Protem	-34.26	20.08	280	28.06	15.85	3.93	2.05	34.08	4.83	8.84	0.16	0.4	6.5	5.02	22.77	67.97	2.37
Stanger	-29.2	31.15	651	25.15	16.8	7.33	4.43	165.09	11.83	4.09	0.41	0.63	16	16.49	17.84	73.42	1.06
Kestell	-28.31	28.7	1693	26.48	13.37	3.56	2.08	153.68	11.5	3.53	0.31	0.54	12.2	15.44	24.93	64.61	1.61

Table 21: Weather statistics (WGN) from 2001-2020 of 11 stations for the month of January.

Station	Lat-itude	Long-itude	Elevation (m.a.s.l.)	Max tempe-rature (˚C)	Min tempe-rature (˚C)	Max tempe-rature SD (˚C)	Min tempe-rature SD (˚C)	Precipit-ation (mm/month)	Precipit-ation SD (mm/month	Precipit-ation Skp (mm/month	Proba-bility wet day after dry	Proba-bility wet day after wet	Days of Precip-itation	Max 0.5 hour rain-fall (mm)	Daily solar radia-tion (MJ/m^2-day)	Dew-point tempe-rature (˚C)	Wind speed (m/s)
Roodeplaat. Pretoria	-25.6	28.35	1168	29.63	16.87	2.96	1.82	121.37	9.96	4.73	0.28	0.52	11.68	12.79	21.88	59	1.49
Vaalharts, Jan Kemp-dorp	-27.95	24.83	1180	32.73	17.78	3.74	2.71	96.96	14.13	12.85	0.23	0.47	9.55	10.37	26.05	52.22	1.21
Upington, Proefplaas	-28.46	21.2	798	34.6	17.54	3.2	3.51	26.42	4.74	7.74	0.07	0.23	2.7	5.24	27.78	47.5	1.55
ITSG, Nelspruit	-25.45	30.97	673	29.66	18.91	3.35	1.75	130.87	10.86	4.29	0.29	0.52	11.85	13.96	20.39	65.26	1.02
AWS, Bathurst	-33.51	26.82	226	26.21	17.41	3.81	2.21	62.08	7.44	7.3	0.26	0.46	10.2	9.42	20.88	67.63	2.03
Debrak, Loeriesfontein	-31.03	19.52	800	30.56	13.1	4.11	3.56	8.65	2.29	10.41	0.01	0.5	1	1.63	29.94	45.73	1.05
Carnarvon	-30.96	21.98	1327	32.28	15.23	3.17	3.06	12.71	2	6.44	0.06	0.29	2.52	2.23	26.05	37.76	3.41
St. Lucia	-28.21	32.48	14	30.79	20.49	3.81	2.2	122.17	15.97	12.29	0.28	0.53	11.85	18.4	19.18	71.08	0.76
Protem	-34.26	20.08	280	27.28	15.3	3.78	2.01	28.55	4.61	9.02	0.12	0.33	4.9	4.98	22.77	64.8	2.87
Stanger	-29.2	31.15	651	25.78	17.35	4.1	2.35	174.95	11.88	3.52	0.39	0.64	16.36	15.33	17.84	74.88	1.17
Kestell	-28.31	28.7	1693	26.12	13.04	3.69	2	121.21	9.54	4.31	0.32	0.48	11.7	12.56	24.93	63.78	1.96

3.2 Comparison between catchment and national data model results

Four catchments were simulated with the national datasets but using the same weather data, over the same timeframes, as before. Various topographical, soil and land cover input datasets were used in the previous/original case studies; in many cases more detailed data (spatially and temporally) than the baseline input dataset utilized here. As mentioned above, the reason for duplicating the application of ArcSWAT in these previously modelled catchments is to compare the results (flow and sediment yield) of the two different input datasets used in each catchment, as well as hydrological accuracy against measured streamflow data. Comparison of the results and accuracies of the two input datasets (original input versus baseline input) allow appraisal of the performance of the baseline input data. From here on, ArcSWAT simulations with original catchment input data are referred to as 'catchment data model(s)', whereas ArcSWAT simulations with the national baseline data are referred to as 'national data model(s)'.

3.2.1 Comparison of data models in the Middle Olifants River Catchment

Streamflow simulation results

For simulation with catchment data, streamflow at the main catchment outlet ranges between 1.6 m^3/s in December 2002 to 347 m^3/s in March 2001 with an average of 13.8 m^3/s during the simulation period (1987-2015). Streamflow simulation with national data obtained similar results. The performance evaluation metrics (p and r-factor) for both models are shown in Table 22. Given that a value greater than 0.7 for p-factor is suggested when modelling streamflow indicate that the 95PPU band of both models captured insufficient observed data. The SWAT model was slightly superior with catchment data during calibration and validation. The catchment data model has slightly higher p-factor values (0.65 and 0.62) than the national data model (0.61 and 0.63) during calibration and validation periods, respectively. The r-factor values above 2 indicate that both the catchment and national data models have estimation error variances that are significantly larger than the variance of streamflow observations.

Table 22: MORC performance metrics obtained from monthly streamflow calibration and validation for each model from 1989-2015.

SWAT Model Name	Calibration					Validation				
	p	r	NSE	r^2	*Dv*	p	r	NSE	r^2	*Dv*
Catchment data model	0.65	2.11	0.10	0.40	47.0	0.62	3.06	-1.20	0.57	11.7
National data model	0.61	2.01	0.10	0.40	46.9	0.63	3.05	-0.88	0.30	23.5

Graphical comparisons of catchment and national data models of simulated mean monthly streamflow at the main catchment outlet are illustrated in Figure 19. Graphically, streamflow simulations with catchment and national data appear similar. Although both models seem to have over-predicted streamflow, they were able to identify all the major wet seasons. It is important to note that soil and land use-cover input data for both (catchment and national) data models were virtually the same; thus, no significant differences between the data were anticipated. Input data similarities are described in the Discussion Section below.

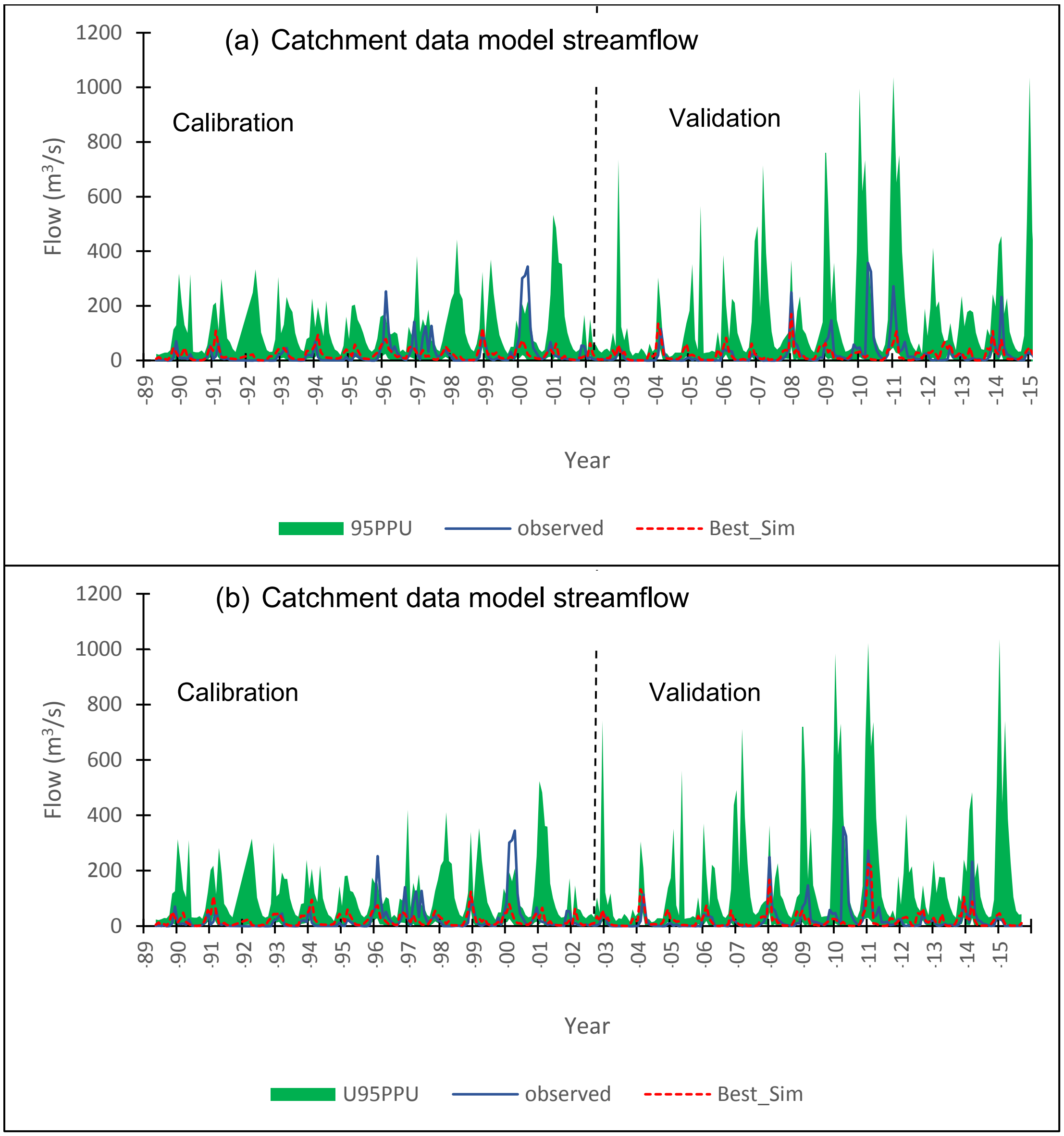


Figure 19: Graphical comparison of monthly streamflow (in m^3/s) at the hydrometric station with (a) catchment and (b) national data during the observation period (1989 to 2015) in the MORC.

Sediment simulation results

For simulation with catchment data, sediment load at the main catchment outlet ranges between 0 t in July 2002 to 36 350 t in March 2000 with an annual average load of 1 217 t/yr and a total of 25 410 041 t during the simulation period (1987-2015). Likewise, for simulation with national data, sediment load ranges from 0 t in June 2015 to 105 300 t in January 2008, with an average of 1 287 t/yr and a total of 26 881 760 t during the simulation period. The graph comparing average sediment load over time between SWAT simulations with catchment and national data is shown in Figure 21. Graphically, sediment load simulations with catchment and national data appear similar, with occasional steep peaks that can be associated with wetter periods.

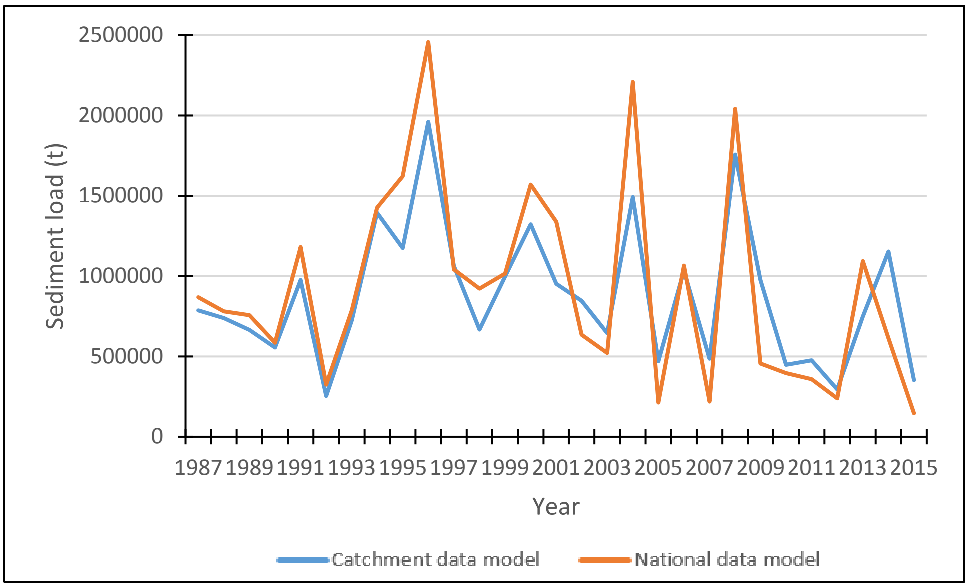


Figure 20: Graphical comparison of total annual sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1987 to 2015) in the MORC.

Figure 22 illustrates the inter-annual variation of the average monthly sediment load for SWAT model simulations with catchment and national data. There are similar trends between simulations obtained using the catchment and national data. Both simulations show that the sediment load is relatively high during high rainfall months (December to March). Low rainfall months extending from June to August have low sediment loads due to low or no rainfall during winter in the MORC.

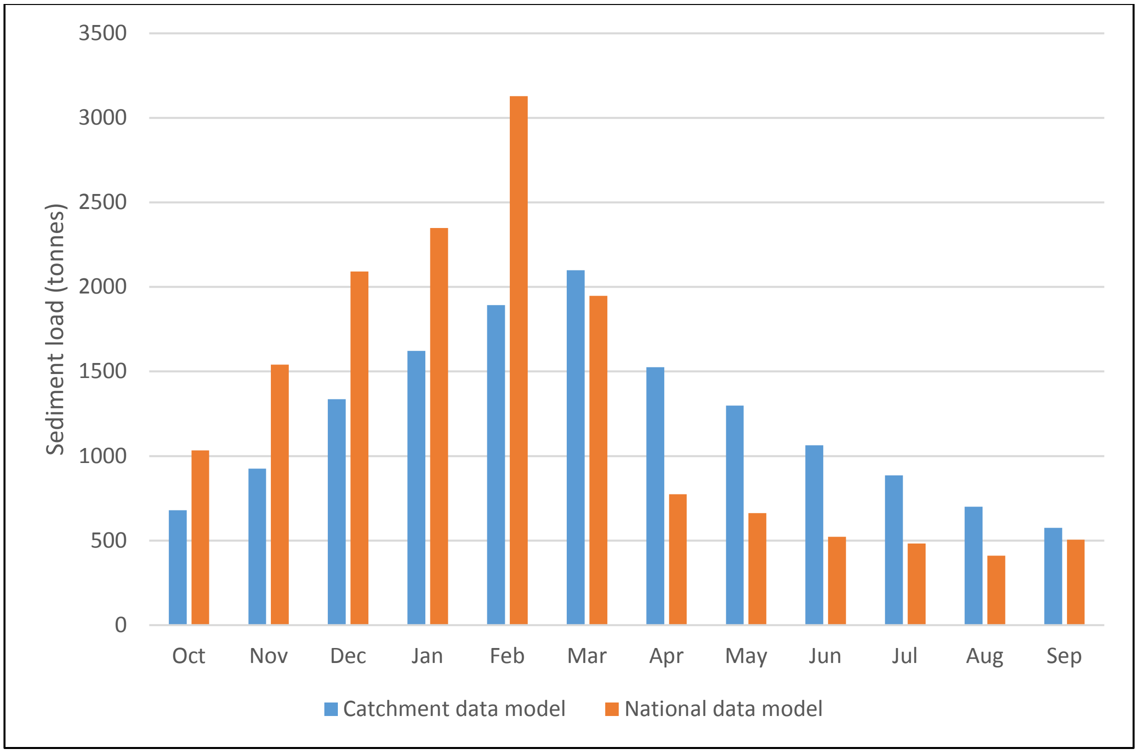


Figure 21: **Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1987 to 2008) in the MORC.**

To spatially illustrate sediment source areas, the average annual sediment yield (in t/ha/yr) for each sub-catchment is shown in Figure 23. Both models simulated similar average sediment yield in the MORC (1.18 and 1.25 t/ha/yr). Figure 23 illustrates the highest sediment yield occur in the same areas. Both models indicate that sub-catchments 3, 4, 20, 22, 24, 32, 42 and 51 have the highest sediment yield (>0.6 t/ha/yr). The spatial correlation in sediment yield between catchment and national data models are attributed to similar soil and land use-cover associations.

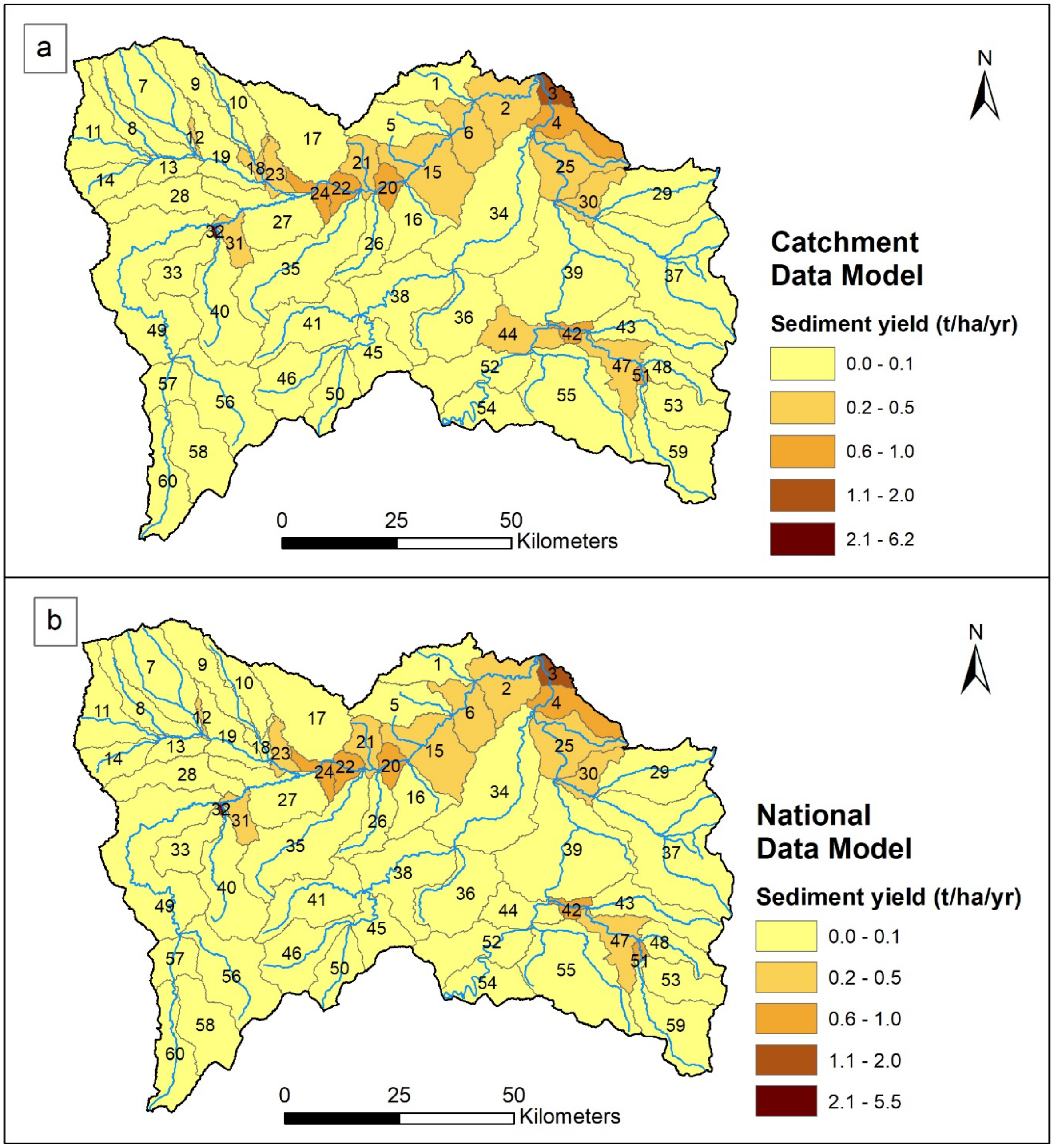


Figure 22: Spatial comparison of average annual sediment yield (in Kg/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the MORC.

Discussion of model differences

As mentioned above, in both catchment and national data models, soil data were obtained from the Land Type Database of SA (Land Type Survey Staff, 1972-2006) and the South African Atlas of Climatology and Agrohydrology (Schulze, 2007). Likewise, the land use-cover classes for both models were obtained from the South African National Land Cover datasets

(SANLC, 2014; 2018). For the catchment data model, land use-cover was obtained from the 2013/14 land cover dataset (SANLC, 2014), whereas for the national data model the land use-cover was obtained from the 2018 land cover dataset (SANLC, 2018). Despite the 5-year difference between these two datasets, the land cover classes in both datasets are spatially similar. According to the two land cover datasets, land cover remained virtually the same from 2013 to 2018 in the MORC. Both datasets indicate that more agricultural land use-cover classes (dominated by irrigated agriculture and orchards) occur in sub-catchments with the highest sediment yield (>0.6 t/ha/yr). It is postulated that the agricultural land use-cover classes mentioned above reduce the vegetation cover, especially due to overgrazing and during tillage operations. Poor vegetation cover, combined with the frequency of disturbance, accounts for relatively high sediment yield in these sub-catchments.

3.2.2 Comparison of data models in the Lower Vaal River Catchment

Streamflow simulation results

For simulation with catchment data, streamflow at the main catchment outlet ranges between 0.4 m^3/s in year 2007 to 1 576 m^3/s in 1996 with an average of 37.9 m^3/s during the simulation period (1980-2008). For simulation with national data, streamflow at the main catchment outlet ranges between 0.6 m^3/s in 2015 to 1 561 m^3/s in 1996 with an average of 33.9 m^3/s. Overall, the performance of both models were equally well during calibration (1980 to 2006) and validation (2007 to 2018). The SWAT model was slightly superior with catchment data during validation, as shown by the higher r^2 value (0.97) than the national data model (0.96) (Table 23). The catchment data model had low *Dv* values (11.3 and 17.3) during calibration and validation periods respectively, while the national data model had 20.8 and 25.4 *Dv* values for the same periods. A p-factor value of between 0.31 to 0.52 for both simulations indicated that the 95PPU band captured few of the observed data. An r-factor value of 0.11 and 0.12, however, showed a smaller uncertainty of the streamflow simulation. It is important to note that the soil input data for both catchment and national data models were derived from the Land Type Database of SA (Land Type Survey Staff, 1972-2006); thus, no significant differences between the data were anticipated. Although land use-cover data were obtained from different sources, with different classes, both datasets captured the main land use activities of the catchment. It is for these reasons that the performance of both models are similar.

Table 23: LVRC performance metrics obtained from monthly streamflow calibration and validation for each model from 1980-2018.

SWAT Model Name	Calibration					Validation				
	p	r	NSE	r^2	*Dv*	p	r	NSE	r^2	*Dv*
Catchment data model	0.52	0.11	0.94	0.94	11.3	0.31	0.11	0.97	0.97	17.3
National data model	0.49	0.12	0.94	0.94	20.8	0.34	0.12	0.97	0.96	25.4

Graphical comparisons of observed versus simulated mean monthly streamflow for simulations with catchment and national data are presented in Figure 24. Graphically, streamflow simulations with catchment and national data appear similar, with a good level of agreement between observed and simulated mean monthly streamflow, especially flow peaks. While both models seem to have identified and matched the major peak flow events and wet seasons well, the dry and/or base flow conditions does not appear to match the observed base flow condition well. Future research should focus on the improvements of base flow simulation.

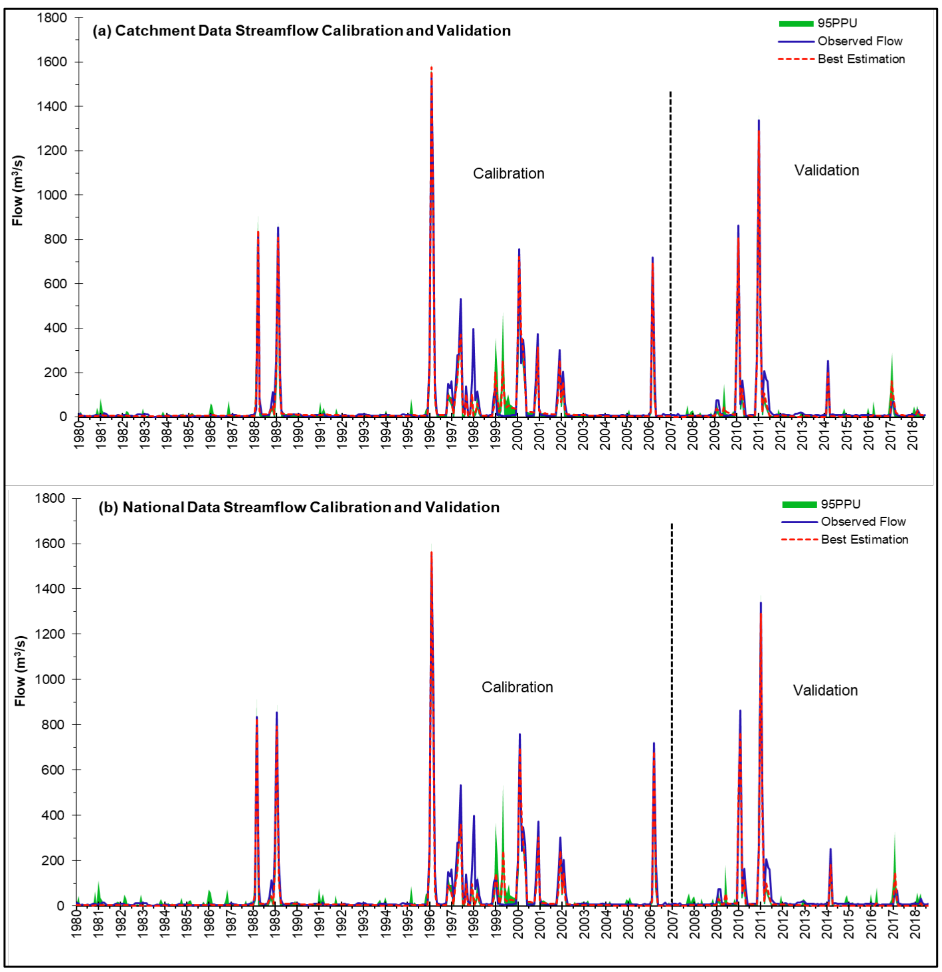


Figure 23: Graphical comparison of monthly streamflow (in m³/s) for SWAT simulations with (a) catchment and (b) national data during the observation period (1980 to 2018) in the LVRC.

Sediment simulation results

Sediment load at the main catchment outlet ranges between 206 t in 1987 to 96 086 t in 1988 for simulation with catchment data, with an annual average load of 12 812 t/yr and a total of 499 675 t during the simulation period (1980-2018). For simulation with national data, sediment load ranges from 149 t in 2015 to 89 225 t in 1988, with an average of 11 298 t/yr and a total of 440 606 t during the simulation period. The graph comparing average sediment load over time between SWAT simulations with catchment and national data is shown in Figure 25. SWAT simulations with catchment and national data show similar trends in sediment load estimations, with occasional steep peaks that can be associated with wetter periods. Sediment load estimation with national data was lower (440 606 t) than the simulation with catchment data. A notable exception was the year 2000, where simulation with national datasets had a higher peak (73 445 t compared to 26 885 t).

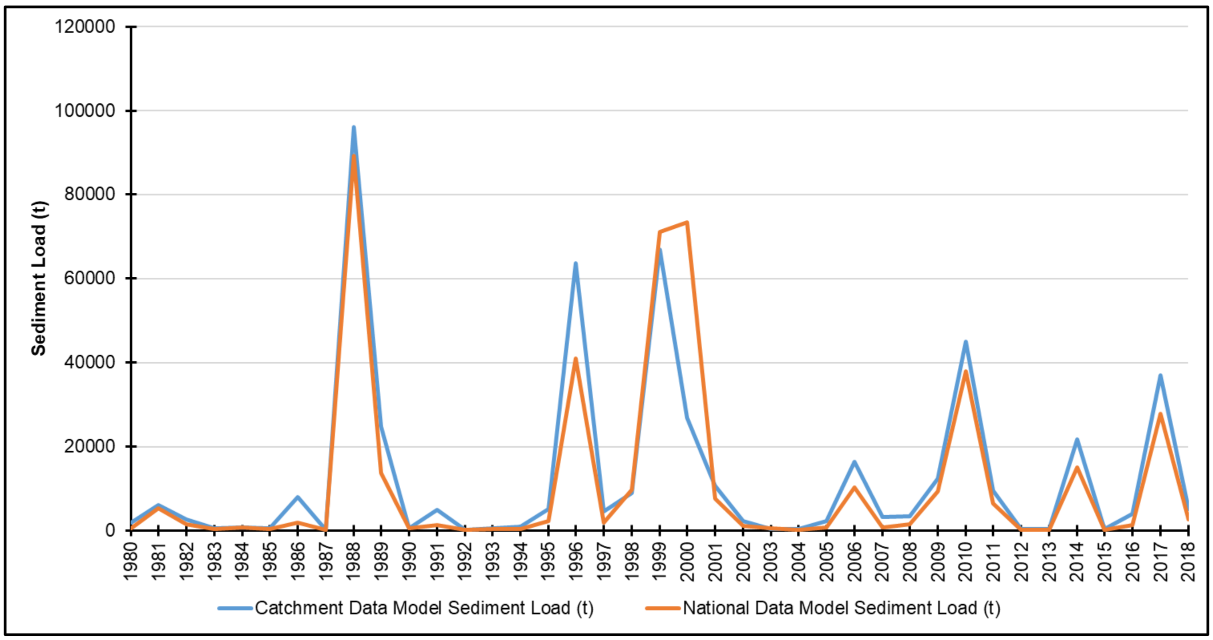


Figure 24: Graphical comparison of total annual sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1987 to 2018) in the LVRC.

Figure 26 illustrates the inter-annual variation of the average monthly sediment load for SWAT model simulations with catchment and national data. There are similar trends between simulations obtained using the catchment and national data. However, similar to Figure 25, sediment loads simulated with national data were lower than sediment loads of the catchment data model for most of the months. Both simulations show that the sediment load is relatively high during high rainfall months (January to March). Low rainfall months extending from June to August have low sediment loads due to low or no rainfall during winter in the LVRC.

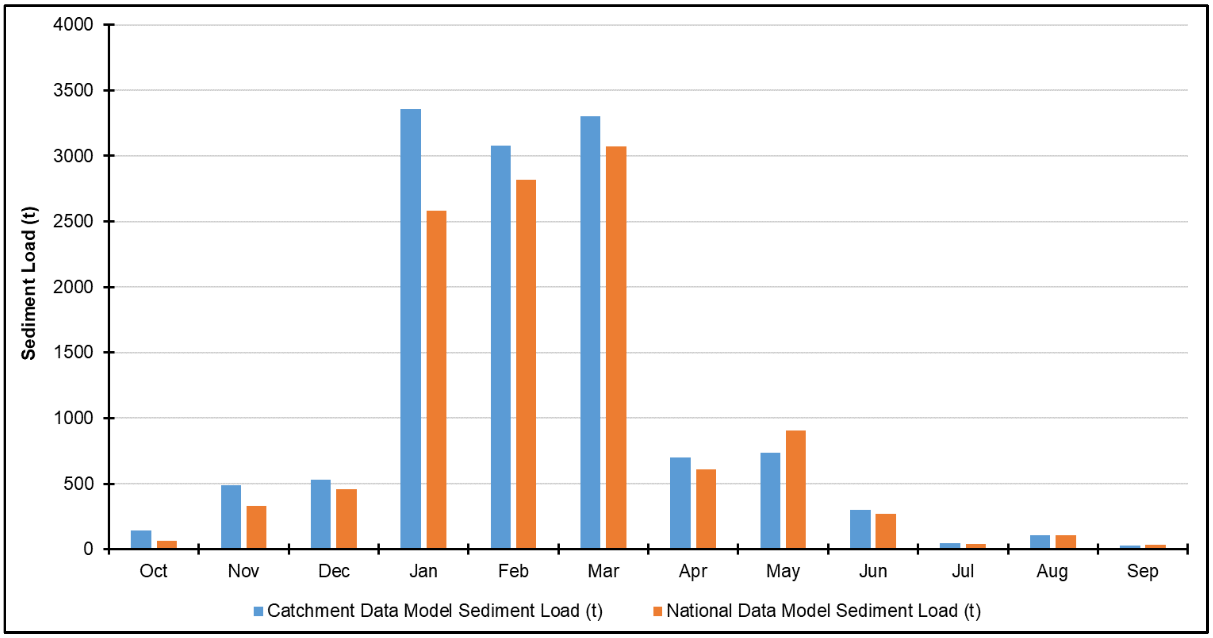


Figure 25: Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (1980-2018) in the LVRC.

To spatially illustrate sediment source areas, the average annual sediment yield (Kg/ha/yr) for each sub-catchment is shown in Figure 27. Some differences are noted between the catchment and national data models. Although the average sediment yield in the LVRC is similar (69 and 84 Kg/ha/yr), the catchment data model illustrates sub-catchments 2, 12, 13, 17 and 22 have the highest sediment yield (>100 Kg/ha/yr), whereas the national data model identifies 13, 17, 20, 21 and 25 as important source areas. The spatial differences in sediment yield between the catchment and national data models are attributed to land use-cover variances since the topography and soils in both models are similar. As mentioned above, soil data were obtained from the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009) which is derived from the Land Type Database of SA (Land Type Survey Staff, 1972-2006), similar to the national data model.

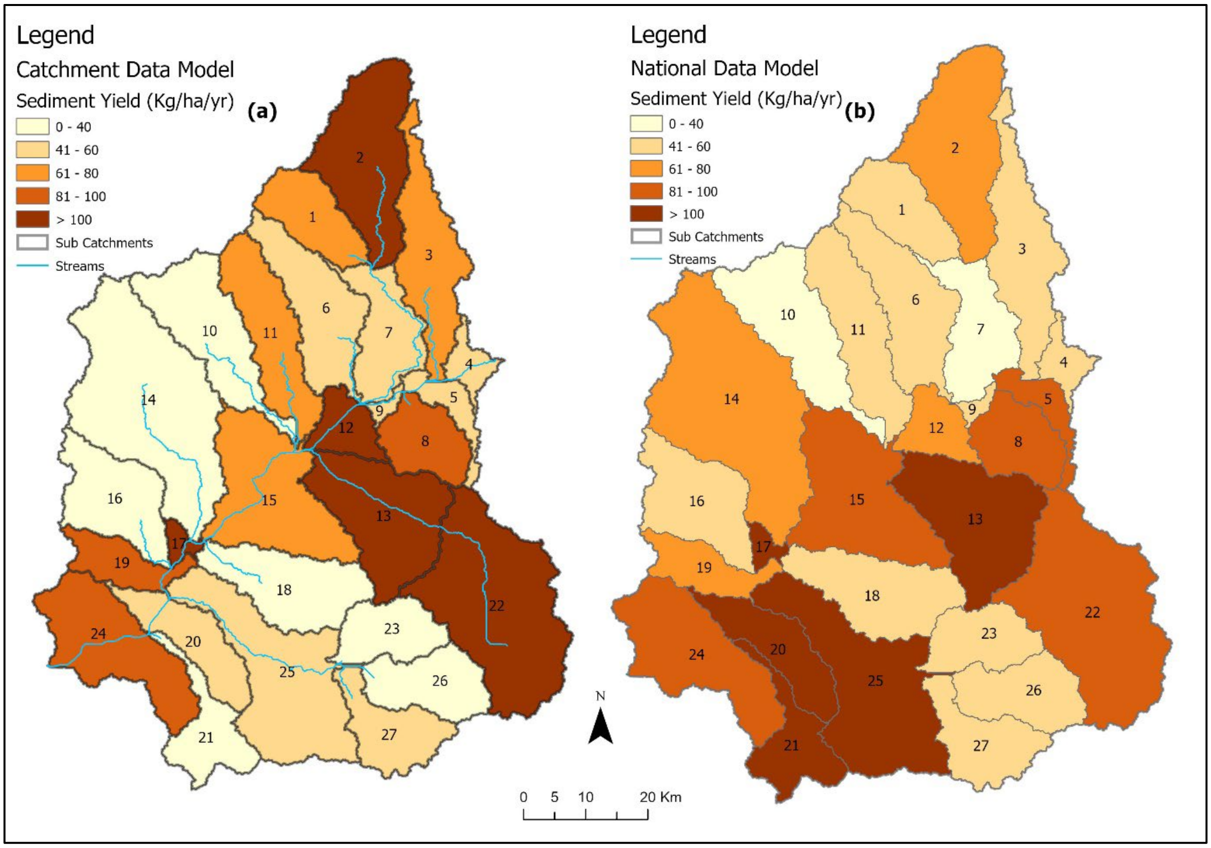


Figure 26: Spatial comparison of average annual sediment yield (in Kg/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the LVRC.

Discussion of model differences

In the national data model, more agricultural land use-cover classes (dominated by irrigated agriculture) occur in sub-catchments 20, 21 and 25 than in the catchment data model. In contrary, in the catchment data model, more agricultural land use-cover classes (dominated by dryland agriculture) occur in sub-catchments 2 and 22 than in the national data model. Sediment yield usually has an inverse relationship with vegetation cover (Le Roux *et al.*, 2013). The agricultural land use-cover classes mentioned above usually reduce the vegetation cover, especially during tillage operations. Poor vegetation cover, combined with the frequency of disturbance, accounts for relatively high sediment yield in these sub-catchments.

3.2.3 Comparison of data models in the Mkabela Catchment

Streamflow simulation results

For simulation with catchment data, streamflow at the main catchment outlet ranges between 0.02 m^3/s in September 2007 to 2.01 m^3/s in January 2006 with an average of 0.60 m^3/s during the simulation period (January 2006 to June 2008). For simulation with national data, streamflow at the main catchment outlet ranges between 0.02 m^3/s in in September 2007 to 1.36 m^3/s in April 2007 with an average of 0.51 m^3/s. Graphical comparisons of catchment and national data models of simulated mean monthly streamflow at the main catchment outlet are illustrated in Figure 28. Graphically, streamflow simulations with catchment and national data appear similar. The national data model has slightly lower peaks than the catchment data model. A possible reason for the discrepancy between the catchment and national data models is that observed data were inadequate to calibrate and validate streamflow at the main catchment outlet. Calibration and validation were restricted to measurements from an ISCO sampler and H-flume at the outlet of sub-catchment 2 (located upstream with a surface area of 96 ha) from August 2006 to March 2008.

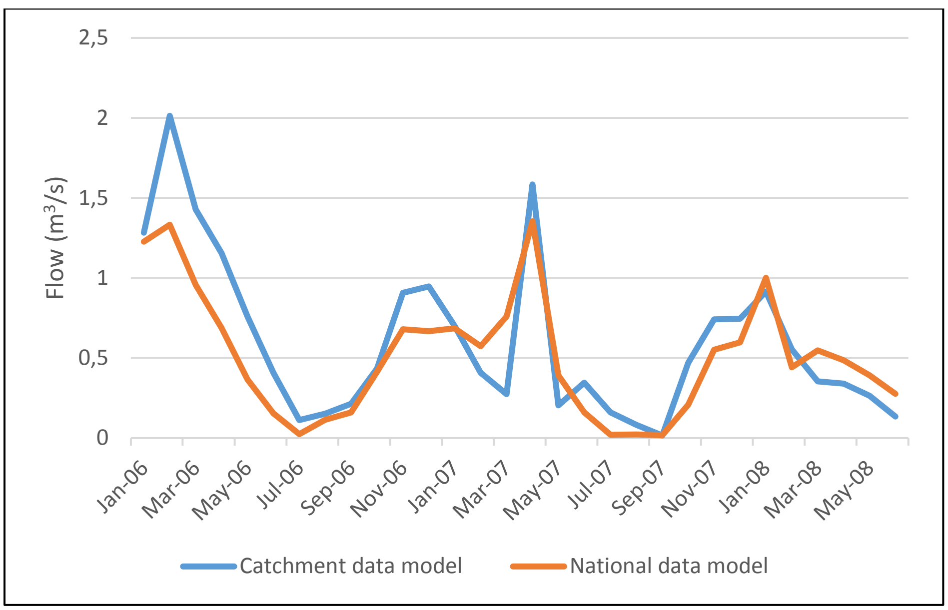


Figure 27: Graphical comparison of monthly streamflow (in m^3/s) at the main catchment outlet with catchment and national data during the observation period (January 2006 to June 2008) in the MC.

As mentioned above, the performance of the catchment and national data models were evaluated at the outlet of sub-catchment 2 using the coefficient of efficiency (NSE) of Nash and Sutcliffe (1970), the coefficient of determination (r^2), as well as the per cent deviation method (*Dv*) of Martinec and Rango (1989). The SWAT model was superior with catchment data during validation, as shown by the higher NSE, r^2 and *Dv* values (see Table 24). The catchment data model over-predicted streamflow by 1.61% as determined by *Dv*, the goodness of fit expressed by NSE was 60% and r^2 was 82%. The national data model over-predicted streamflow by 18.6% as determined by *Dv*, the goodness of fit expressed by NSE was 51% and r^2 was 71%. Graphical comparisons of catchment and national data models of simulated mean monthly streamflow at the outlet of sub-catchment 2 are illustrated in Figure 29a-b. The main reason for the superior performance of the catchment data model is probably due to soil data being obtained from different sources at different spatial scales. These differences are described in the Discussion Section below.

Table 24: MC performance metrics obtained from monthly streamflow calibration and validation for each model from January 2006 to June 2008.

SWAT Model Name	Calibration			Validation		
	NSE	r^2	*Dv*	NSE	r^2	*Dv*
Catchment data model	0.50	0.71	6.2	0.60	0.82	1.61
National data model	0.43	0.57	36.8	0.51	0.71	18.6

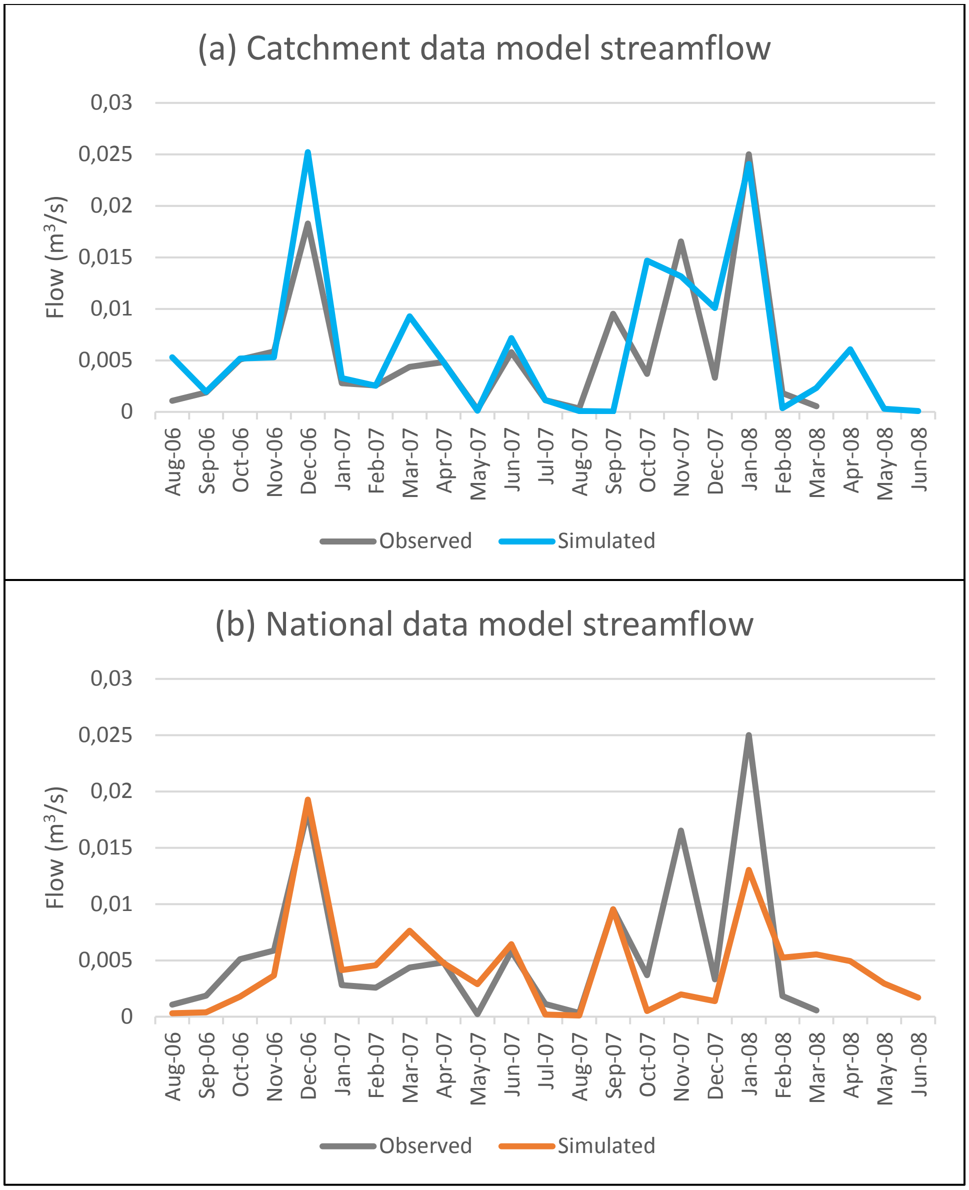


Figure 28: Graphical comparison of monthly streamflow (in m^3/s) at the H-flume outlet of sub-catchment 2 with (a) catchment and (b) national data during the observation period (August 2006 to March 2008) in the MC.

Sediment simulation results

Sediment load at the main catchment outlet ranges between 0.004 t in September 2008 to 49.84 t in February 2006 for simulation with catchment data, with an annual average load of 5.15 t/yr and a total load of 154.58 t during the simulation period (January 2006 to June 2008). For simulation with national data, sediment load ranges from 0 t in July 2006 to 40.96 t in February 2006, with an average of 5.25 t/yr and a total of 157.60 t during the simulation period. The graph comparing average sediment load over time between SWAT simulations with catchment and national data is shown in Figure 30. The catchment and national data models show similar trends (four peaks that can be associated with rainfall events).

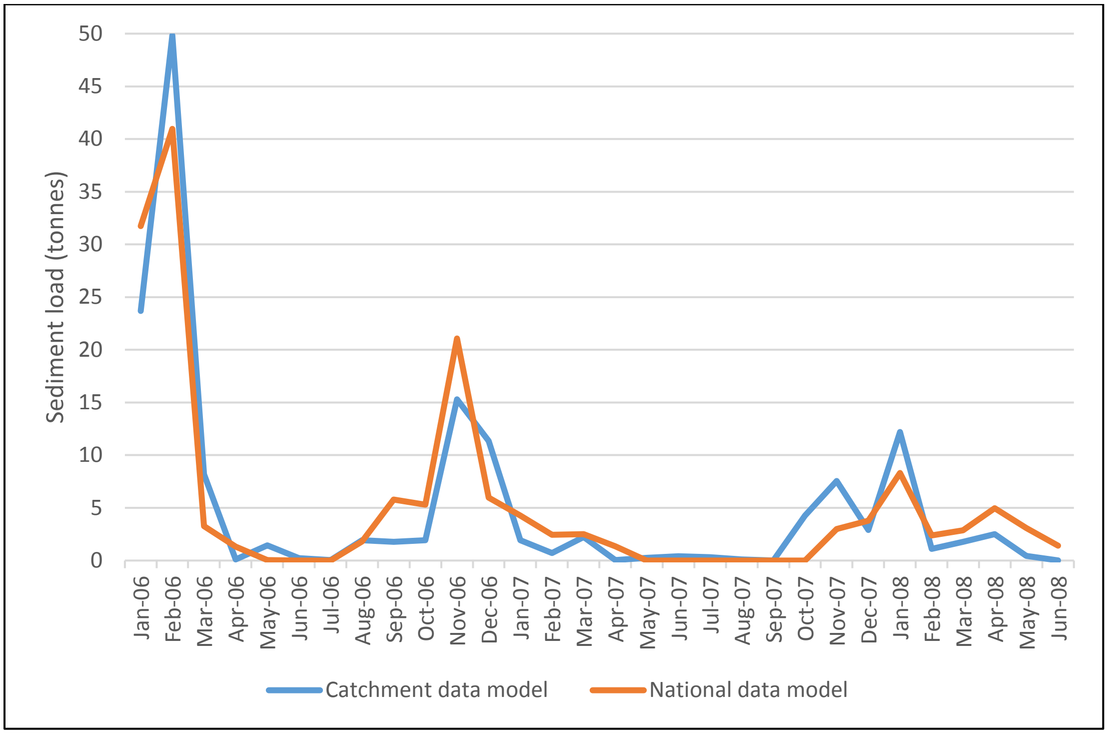


Figure 29: Graphical comparison of total annual sediment load (in metric) for SWAT simulations with catchment and national data during the observation period (January 2006 to June 2008) in the MC.

Figure 31 illustrates the inter-annual variation of the average monthly sediment load for SWAT model simulations with catchment and national data. There are similar trends between simulations obtained using the catchment and national data. Both simulations show that the sediment load is mainly high during the summer rainfall season (extending from October to April). Low rainfall months extending from May to August have low sediment loads due to low or no rainfall during winter in the MC.

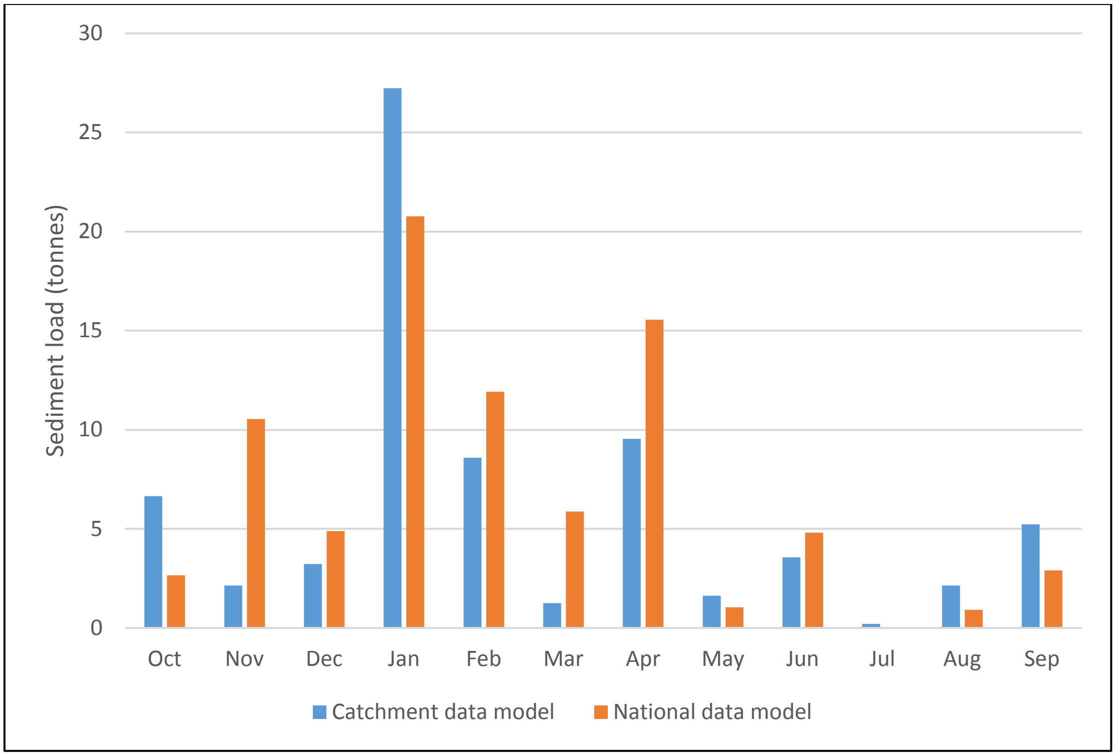


Figure 30: **Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (January 2006 to June 2008) in the MC.**

To spatially illustrate sediment source areas, the average annual sediment yield (Kg/ha/yr) for each sub-catchment is shown in Figure 32. Some differences are noted between the catchment and national data models. Although the average sediment yield of the catchment and national data models are similar (0.35 and 0.39 kg/ha/yr respectively), the catchment data model identifies sub-catchments 1, 3, 4 and 7 as important sources of sediment (>0.3 Kg/ha/yr), whereas the national data model identifies sub-catchments 2, 7, 20 and 21 as sediment source areas. In both catchment and national data models, the spatial extent of the main land use-cover remained the same from 2006 to 2018 in the MC. Therefore, the spatial differences in sediment yield between the catchment and national data models are attributed to differences in the spatial scale of the input data including more comprehensive model calibration.

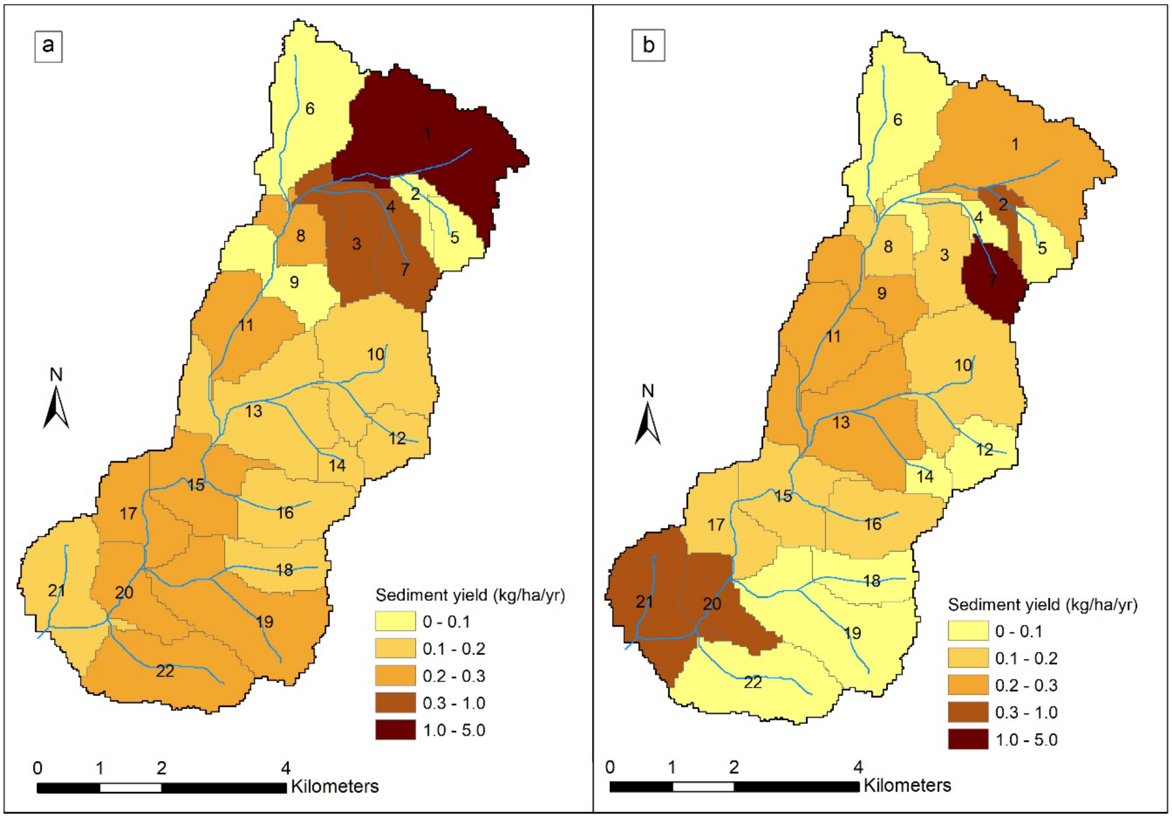


Figure 31: Spatial comparison of average annual sediment yield (in Kg/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the MC.

Discussion of model differences

The soil input data were obtained from different sources at different spatial scales (and detail). Soil input data of the catchment data model were obtained from a relatively detailed pedological soil map at a scale of 1:100,000 with textural profile descriptions for all major soils (Le Roux *et al.*, 2006). In contrary, soil input data for the national data model was derived from the Land Type Database of SA (Land Type Survey Staff, 1972-2006) and the South African Atlas of Climatology and Agrohydrology (Schulze, 2007) at a scale of 1:250,000. In addition, soil input data were improved with delineation of the initial soil components (6 units) into 18 smaller terrain units and invoking a three-layered soil system instead of two layers used in the national data model. However, ancillary soil information of the catchment data model was less critical compared to hydrological structures and land use-cover (Lorentz *et al.*, 2012).

Land use-cover input data were also obtained from different sources. For the catchment data model, land use-cover was obtained from SPOT 5 imagery acquired in 2006, whereas for the national data model the land use-cover was obtained from the 2018 land cover dataset (SANLC, 2018). Despite the 12-year difference between these two datasets, the land cover classes in both datasets are spatially similar. Both datasets captured the main land use activities of the catchment, which is dominated by sugarcane plantations. According to the two land cover datasets, the spatial extent of the main land use-cover (sugarcane) remained the same from 2006 to 2018 in the MC. The national data model, however, could not identify sub-catchment 1 as a sediment source as effectively as the catchment data model. According to the national data model, most sediment originates from (the surrounding) sugarcane fields, not the cabbage plot. Sediment generated on the relatively small cabbage plot is not spatially identified within the sub-catchment it is located. The main reason the national data model could not identify the vegetable plot in sub-catchment 1 as a sediment source is due to differences in the spatial scale of the input data including more comprehensive model calibration. The catchment data model adjusted the curve numbers and Manning's Roughness Coefficient. For example, to account for relatively poor vegetation cover and high runoff on the cabbage plot in sub-catchment 1, curve numbers and Manning's roughness Coefficient for cabbage were increased and decreased, respectively. The opposite adjustments were implemented to better account for low runoff for land use-cover with relatively good seasonal groundcover (e.g. sugarcane). These parameters were not adjusted or modified within the national data model. Despite this, the national data model appears to be efficient in representing farm dams as a series of storages where connectivity is reduced at the catchment scale.

3.2.4 Comparison of data models in Tsitsa River Catchment

Streamflow simulation results

For simulation with catchment data, streamflow at the main catchment outlet ranges between 0.001 m^3/s in September 2010 to 131.9 m^3/s in January 2011 with an average of 30.77 m^3/s during the simulation period (1980-2008). For simulation with national data, streamflow at the main catchment outlet ranges between 0 m^3/s in August 2010 to 127.3 m^3/s in in January 2011 with an average of 37.2 m^3/s. The catchment data model was superior compared to the national data model during validation, as shown by the higher NSE, r^2 and *Dv* values (see Table 25). The catchment data model over-predicted streamflow by 14.4% as determined by *Dv*, the goodness of fit expressed by NSE was 75% and r^2 was 88%. The national data model over-predicted streamflow by 29.2% as determined by *Dv*, the goodness of fit expressed by

NSE was 65% and r^2 was 87%. The main reason for the slightly superior performance of the catchment data model is due to differences between soil and land use-cover datasets. These differences are described in the Discussion Section below.

Table 25: TRC performance metrics obtained from monthly streamflow calibration and validation for each model from January 2008 to December 2012.

SWAT Model Name	Calibration			Validation		
	NSE	r^2	*Dv*	NSE	r^2	*Dv*
Catchment data model	0.55	0.47	18.7	0.75	0.88	14.4
National data model	0.43	0.61	34.3	0.65	0.87	29.2

Graphical comparisons of observed versus simulated mean monthly streamflow for simulations with catchment and national data are presented in Figure 33a-b. Graphically, streamflow simulations with catchment and national data appear similar, with a good level of agreement between observed and simulated mean monthly streamflow.

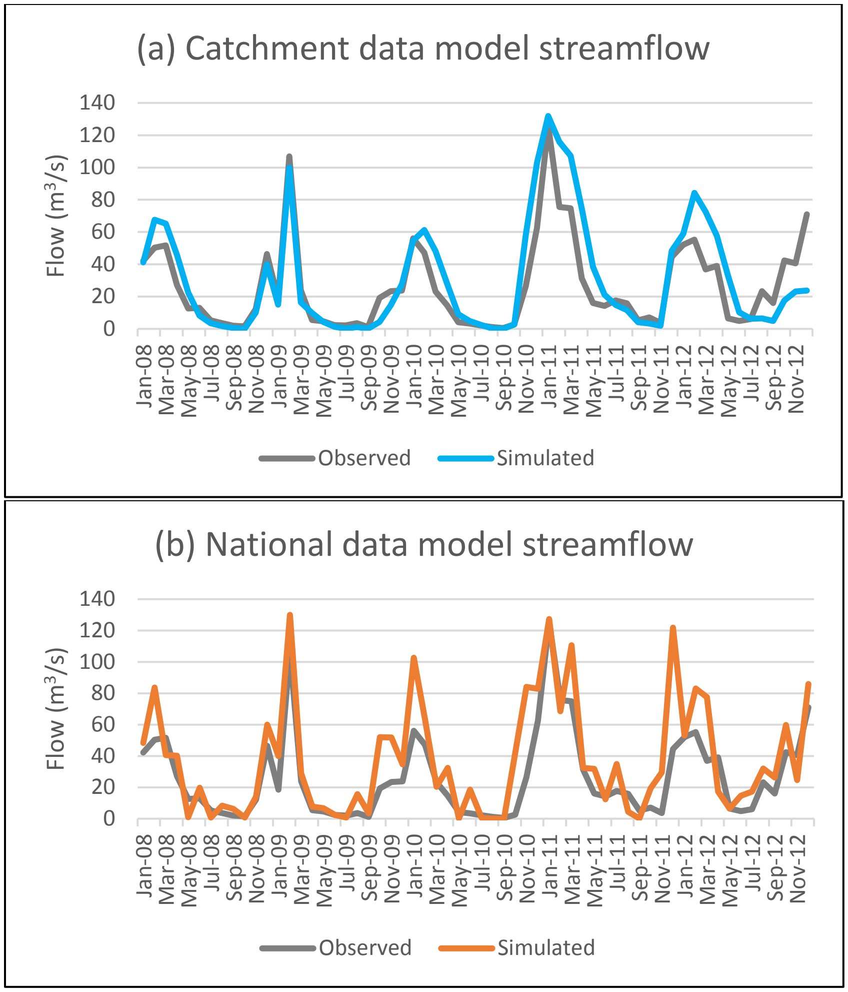


Figure 32: **Graphical comparison of monthly streamflow (in m^3/s) for SWAT simulations with (a) catchment and (b) national data during the observation period (20018-2012) in the TRC.**

Sediment simulation results

For simulation with catchment data, sediment load at the main catchment outlet ranges between 3.06 t in September 2010 to 167,100 t in January 2012 with an annual average load of 28 096 t/yr and a total load of 1.6 million t during the simulation period (2008-2012). For simulation with national data, sediment load at the main catchment outlet ranges between 1.37 t in August 2010 to 117,700 t in January 2012 with an average load of 23 812 t/yr and a total load of 1.4 million t during the simulation period (2008-2012). The graph comparing average sediment load over time between SWAT simulations with catchment and national data is shown in Figure 34. SWAT simulations with catchment and national data show similar trends in sediment load estimations, with occasional steep peaks that can be associated with wetter months.

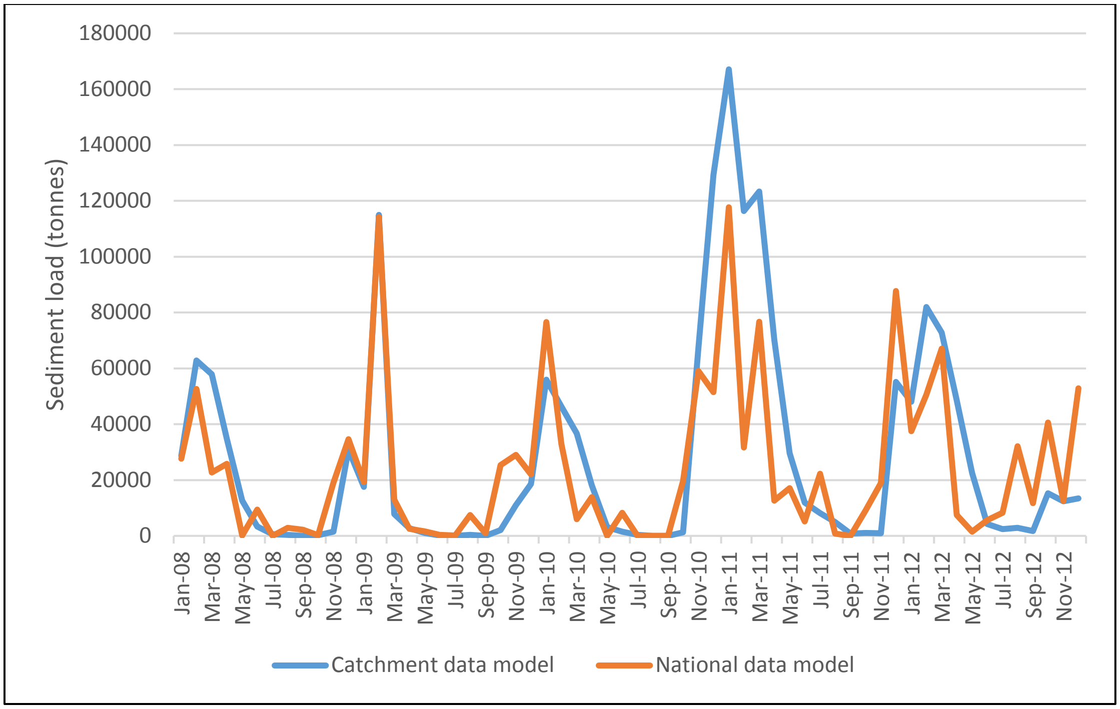


Figure 33: **Graphical comparison of total annual sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (20018-2012) in the TRC.**

Figure 35 illustrates the inter-annual variation of the average monthly sediment load for SWAT model simulations with catchment and national data. There are similar trends between simulations obtained using the catchment and national data. Both simulations show that the sediment load is mainly high during the summer rainfall season (extending from October to April). Low rainfall months (extending from May to August) have low sediment loads due to low or no rainfall during winter in the TRC.

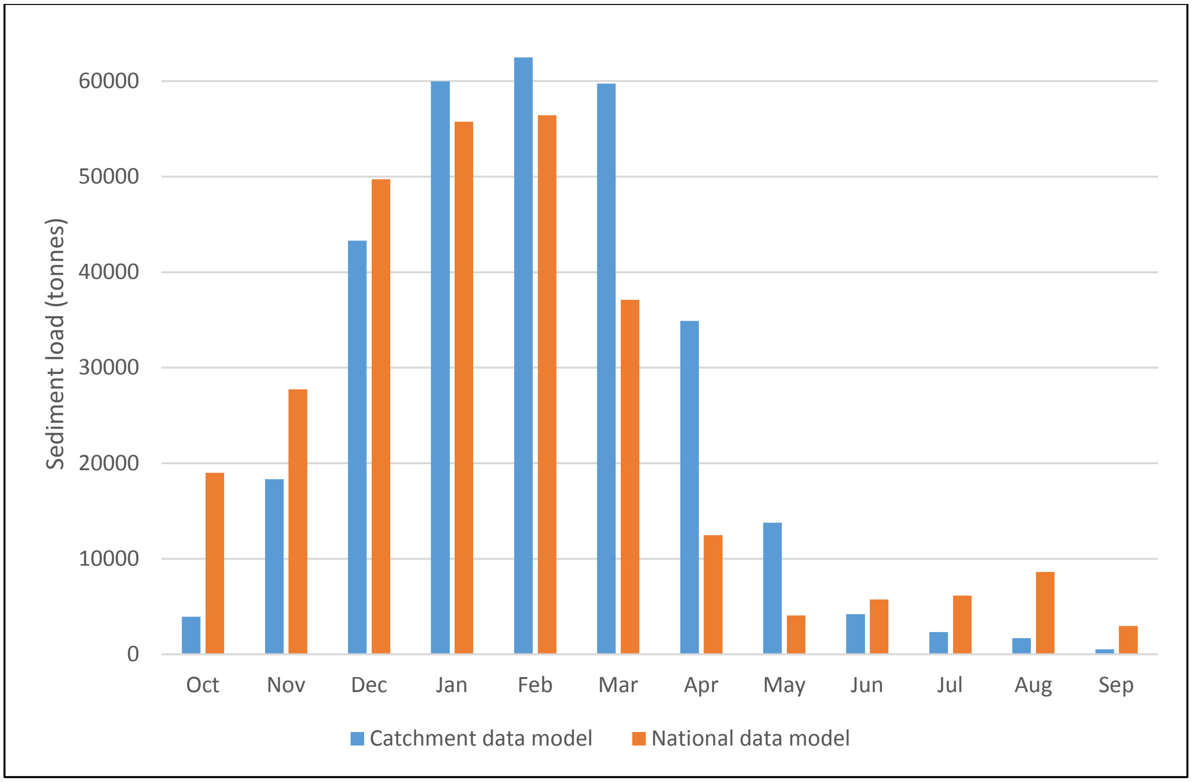


Figure 34: Graphical comparison of monthly average sediment load (in metric t) for SWAT simulations with catchment and national data during the observation period (20018-2012) in the TRC.

To spatially illustrate sediment source areas, the average annual sediment yield (t/ha/yr) for each sub-catchment is shown in Figure 36. Some differences are noted between the catchment and national data models. Although the average sediment yield of the catchment and national data models are similar (0.85 and 0.72 t/ha/yr respectively), the national data model identifies the lower half of the TCR (sub-catchments 6, 8, 9, 10 and 12) as important sediment source areas (>2 t/ha/yr), whereas the catchment data have moderate sediment yield values (1.0 to 2.0 t/ha/yr) throughout the catchment. The spatial differences in sediment yield between the catchment and national data models are attributed to land use-cover variances since the topography and soils in both models are similar.

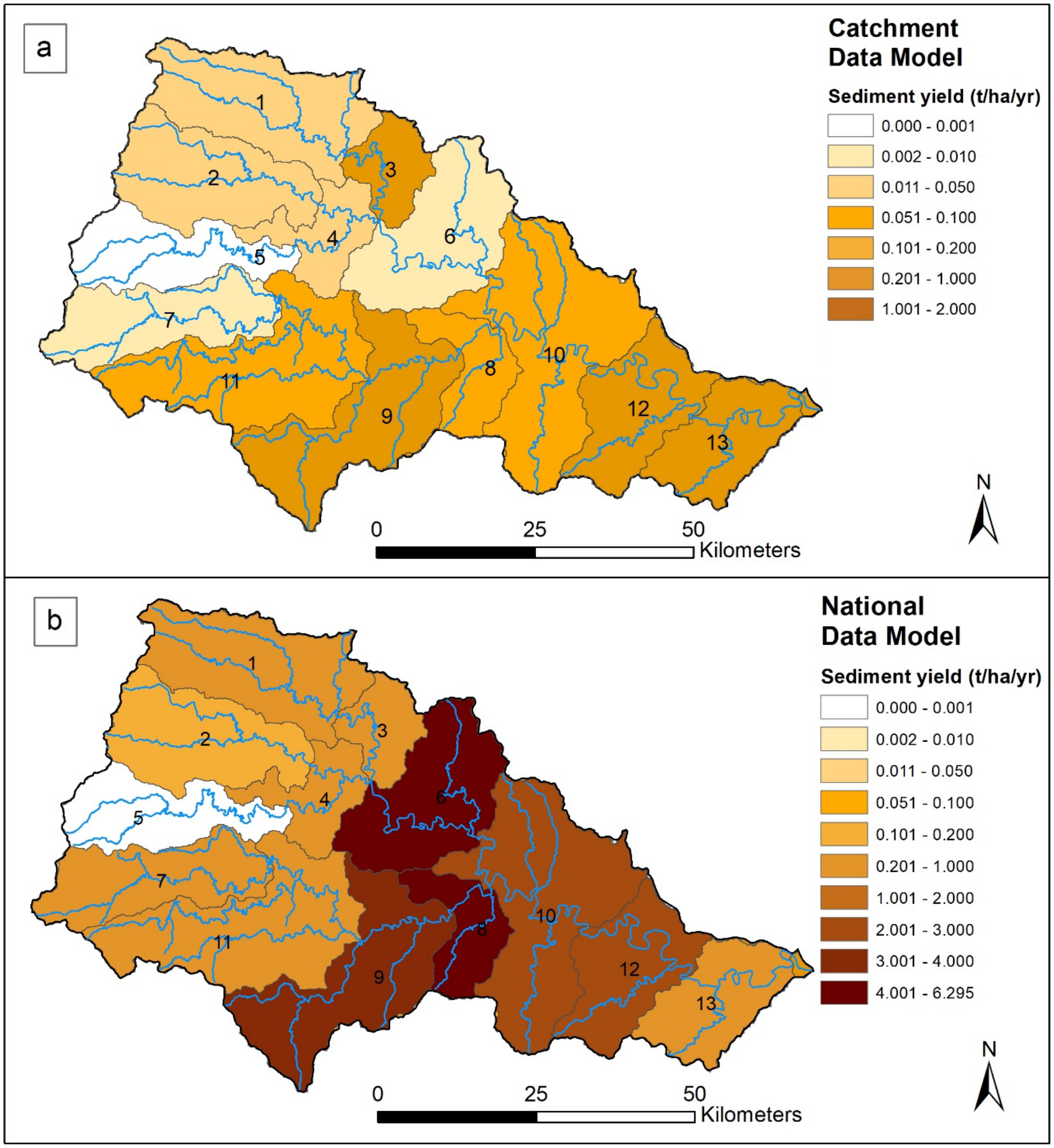


Figure 35: Spatial comparison of average annual sediment yield (in t/ha/yr) simulated by the SWAT model with (a) catchment data and (b) national data in the TRC.

Discussion of model differences

It is important to note that the soil input data for both catchment and national data models were derived from the Land Type Database of SA (Land Type Survey Staff, 1972-2006); thus, no significant differences between the data were anticipated. For both data models, similar methodology/reasoning was followed in the assignment of the required parameter values to

Land Types. In contrary, however, land use-cover classes were obtained from different sources. For the catchment data model, land use-cover was created by means of unsupervised classification on SPOT 5 imagery acquired in 2011, whereas for the national data model the land use-cover was obtained from the 2018 land cover dataset (SANLC, 2018). In the national data model, more barren land occur in the lower half of the TCR (sub-catchments 6, 8, 9, 10, 12 and 13) than in the catchment data model. Barren land is associated with no vegetation cover which accounts for relatively high sediment yield in these sub-catchments. Latter-mentioned sub-catchments are also associated with erodible dispersive soils which, in combination with overgrazing and abandonment of cropland, are the main factors causing severe erosion in the TRC (Le Roux, 2018). Despite these differences, the national baseline data appears to be an efficient input dataset for modelling streamflow and sediment dynamics in the TRC.

3.2.5 Overall comparison of catchment and national data models

Performance of the national data models was determined by comparing the hydrological accuracy against measured streamflow data. Figure 36 compares the performance metrics (NSE, r^2 and *Dv*) of the four catchments with (a) catchment and (b) national data. In each of the four case studies, catchment data models were slightly superior compared to the national data models, as shown by more accurate performance metrics. Between the four catchments, the LVRC performed the best, followed by the MC and then TRC, whereas the MORC performed the poorest. Model performance depends largely on the detail or quality of input data (Mararakanye *et al*., 2020; Glenday *et al*., 2021). It is postulated that differences in the performance between catchments was largely influenced by the quality of rainfall data, since the other input datasets (DEM, soil and land use-cover) were similar in all four catchments. For example, the LVRC utilized three weather stations distributed throughout the catchment whereas the MORC utilized only one weather station located in the east of the catchment. The use of only one weather station in such a large catchment (22 550 km^2) possibly caused errors in daily rainfall in other parts of the catchment and subsequent streamflow output. Furthermore, the LVRC is three times smaller (7 220 km^2) than the MORC. Although the MC also used one weather station, the MC is the smallest catchment (4 154 ha) of the four, with less probability of uneven rainfall distribution within the catchment.

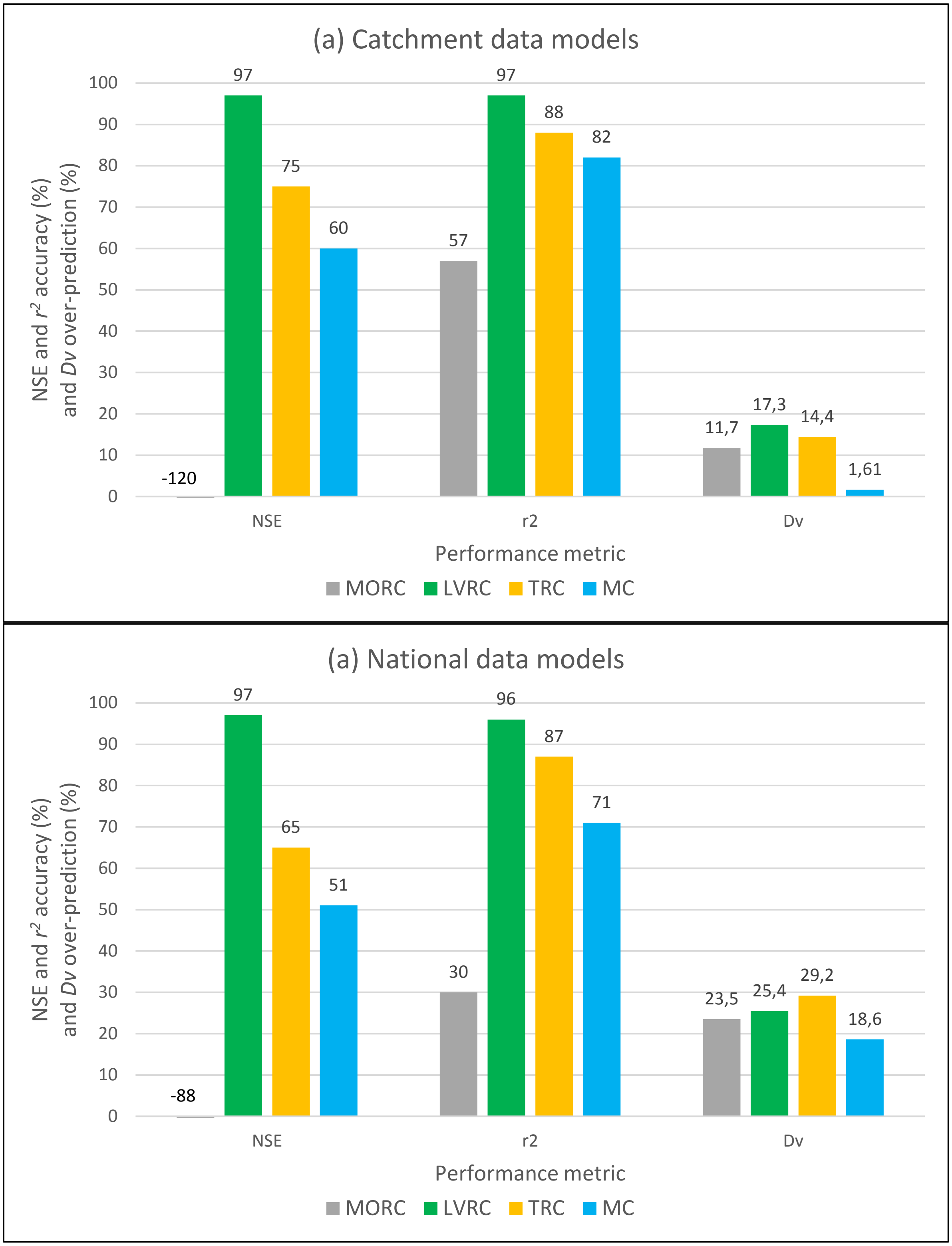


Figure 36: Comparison of the performance (NSE, r^2 and *Dv*) of the four catchments (MORC, LVRC, MC and TRC) with (a) catchment and (b) national data.

Performance of the national data models was further assessed by comparing streamflow and sediment outputs with previous modelled catchment data models. Graphically, in each of the four catchments, streamflow and sediment load of both data models appear similar. Spatial similarities and/or differences of sediment source areas are illustrated by means of sediment yield maps of the respective catchments. Except for the MORC, the catchment and national data models illustrate different sub-catchments as sediment source areas. Spatial differences in sediment yield between catchment and national data models are mainly attributed to land use-cover variances since the other input datasets (DEMs and soil input data) are in essence similar between data models. Each case study utilized DEMs with similar spatial resolutions (15-30 m). Furthermore, in each case study the soil input data for catchment and national data models were mostly derived from the Land Type Database of SA (Land Type Survey Staff, 1972-2006). Similar methodology/reasoning was followed in the assignment of the required parameter values to Land Types. Land use-cover classes, however, were obtained from different sources with different acquisition dates. In the TRC for example, land use-cover for the catchment data model was created by means of unsupervised classification on SPOT 5 imagery acquired in 2011, whereas for the national data model the land use-cover was obtained from the 2018 land cover dataset. In the national data model of the TRC, more barren land occur in the lower half of the catchment than in the catchment data model. Barren land is associated with no vegetation cover which accounts for relatively high sediment yield in these sub-catchments. Despite these differences, the national baseline data appears to be an efficient input dataset capable of modelling streamflow and sediment dynamics at a catchment scale.

4. CONCLUSION AND RECOMMENDATIONS

The study provides well-structured geo-spatial input datasets to set up and run SWAT in SA. The national input database is stored in the Water Research Observatory data portal: https://www.waterresearchobservatory.org/data-and-resources/hydrological-data-and-modelling. The portal provides geo-spatial input datasets including:

- SWAT catchment outline data (tertiary and quaternary) including the hydrologically corrected SRTM DEM of SA at 90 m resolution (Weepener *et al.*, 2012);
- South African National Land Cover (SANLC, 2014; 2018; 2020) linked to SWAT land cover codes;
- Soil map with SWAT attribute data for each Land Type of SA (Land Type Survey Staff, 1972-2006);
- Weather statistics (WGN) files required as input by the model.

The national input database is an important step forward in the application of hydrological modelling by assisting modellers to set-up and run the SWAT model anywhere in SA. One of the biggest challenges to set-up and run the SWAT model in SA is to obtain appropriate soil data. This study addressed this challenge by providing appropriate soil data for use in SWAT. The input datasets consist of more detailed and higher resolution soil data than the global datasets of Abbaspour *et al.* (2019). Another challenge is input data preparation and model set-up is a laborious task, especially due to the lack of appropriate and representative data. A large part of modelling effort goes into the construction of input datasets (Jetten *et al*., 2003; Glenday *et al*., 2021). Therefore, this database will save time with model set-up, as well as assist in the standardization of SWAT modelling efforts in SA. Although SWAT users could use the input data 'as is', it is recommended to supplement, improve and/or replace the input data with recent/sophisticated data, especially rainfall data. In addition, stream channel processes and hydrological structures need to be characterised, allowing deposition of excess sediment depending on the carrying capacity and/or sediment storages where connectivity is reduced (Chen and Mackay, 2004). Ancillary information regarding management practices in the catchment should also be incorporated including tillage operations, nutrient applications, irrigation scheduling and harvesting operations. Calibration of model simulations with measured data is essential by adjusting the most sensitive model parameters such as curve number and base-flow coefficients (Glenday *et al*., 2021).

Further refinement of the national baseline data will be possible given additional research, including the following. It is recommended to expand the database to include data input for cross-bordering catchments. For example, at a global scale, Abbaspour *et al*. (2019) prepared databases of soil, land use-cover, as well as weather databases that could serve as standard inputs in SWAT models. It is further recommended to update the input datasets continually with new data, especially land use-cover data (e.g. see Toucher *et al*., 2020). Climate data could be included such as the 50-year (1950-1999) climate data from the South African Atlas of Climatology and Agrohydrology (Schulze, 2007), which will be extended to include 2000-2020 by WRC project C20192020-00205 lead by Dr David Clark at UKZN. Daily climate data could be reformatted for use in SWAT. Another recommendation is to improve the soil input dataset by means of digital soil modelling (DSM) techniques (see e.g. Wahren *et al*., 2016; Van Tol *et al.,* 2020; van Zijl *et al.*, 2020). Further improvements can be achieved by means of a hydropedological approach in the simulation of soil water contents to obtain more accurate representation of the dominant hydrological processes in catchments (van Tol *et al.*, 2021; Harrison et al., 2022). Lastly, the underestimation of wetland water regimes in SWAT can be reduced by formatting existing wetland data (van Deventer *et al*., 2020) that reflect wetland structure and processes in SWAT models.

5. REFERENCES

Abbaspour KC, Johnson CA, van Genuchten MTH, 2004. Estimating uncertain flow and transport parameters using a Sequential Uncertainty Fitting procedure. *Vadose Zone Journal* **3**: 1340-1352. https://doi.org/10.2136/vzj2004.1340

Abbaspour KC, Vaghefi SA, Yang H, Srinivasan R. 2019. Global soil, land use, evapotranspiration, historical and future weather databases for SWAT Applications. *Scientific Data – Nature* **6**: 1-11. https://doi.org/10.1038/s41597-019-0282-4

Abbaspour KC, Yang J, Maximov I, Siber R, Bogner K, Mieleitner J, Zobrist J, Srinivasan R. 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology* **333**: 413-430. https://doi.org/10.1016/j.jhydrol.2006.09.014

Addinsoft, 2019. XLSTAT statistical and data analysis solution. Addinsoft. Long Island, New York.

Akoko G, Le TH, Gomi T, Kato T. 2021. A Review of SWAT Model Application in Africa. *Water* **13**(1313): 1-20.

ARC-GCI. 2013. *Agricultural Research Council Maize information guide*. Agricultural Research Council – Grain Crops Institute: Potchefstroom, South Africa. Available at: http://www.arc.agric.za/arc-gci/Documents/MIG2013/7290.pdf (Accessed: 15 April 2020).

ARC. 2008. *Agroclimatology Database,* Agricultural Research Council-Soil, Climate and water, 600 Belvedere, Pretoria 0083, South Africa.

ARC. 2012. *Agroclimatology Database,* Agricultural Research Council-Soil, Climate and water, 600 Belvedere, Pretoria 0083, South Africa.

ARC. 2019. *Agroclimatology Database,* Agricultural Research Council-Soil, Climate and water, 600 Belvedere, Pretoria 0083, South Africa.

ARC. 2021. *Agroclimatology Database,* Agricultural Research Council-Soil, Climate and water, 600 Belvedere, Pretoria 0083, South Africa.

Arnold JG, Kiniry JR, Srinivasan R, Williams EB, Haney EB, Neitsch SL. 2012. *SWAT: input/output file documentation, version 2012, report no. TR-439*. Texas Water Resources Institute: Texas USA. pp. 650. Available on website https://swat.tamu.edu/media/69296/swat-io-documentation-2012.pdf.

Arnold JG, Srinivasan R, Muttiah RS, Williams JR. 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association* **34**(1): 73-89.

Arnold JG, Williams JR Maidment DR. 1995. Continuous-time water and sediment-routing model for large basins. *Journal of Hydraulic Engineering* **121**(2): 171-183.

Aouissi J, Benabdallah S, Lili Chabaâne Z, Cudennec C, 2016. Evaluation of potential evapotranspiration assessment methods for hydrological modelling with SWAT—Application in data-scarce rural Tunisia. *Agricultural Water Management* **174**: 39-51. https://doi.org/10.1016/j.agwat.2016.03.004

ASTER-GDEM. 2009. *Advanced Spaceborne Thermal Emission and Reflection Radiometer – Global Digital Elevation Model*: Courtesy of the Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA). Available at: http://gdem.ersdac. jspacesystems.or.jp.

Burger M. 2013. 1*:1,000,000 Scale Geology for South Africa [WWW Document]. 11000000 Geol. Map*. URL https://www.arcgis.com/home/item.html?id=739c8b22b99b47bb81c2bed660d6c5de (accessed 5.15.20).

Chen E, Mackay DS. 2004. Effects of distribution-based parameter aggregation on a spatially distributed agricultural nonpoint source pollution model. *Journal of Hydrology* **295**: 211-224.

Council for Geoscience. 2007. *Geological data 1:250 000*. Council for Geoscience: Pretoria, South Africa.

CSIR Smart Places. 2020. Total soil organic carbon for 2018 in t/ha [Data set]. Department of Environment, Forestry and Fisheries: Pretoria, South Africa. https://doi.org/10.15493/DEA.CARBON.10000112

Elwell HA. 1976. *Natal Agricultural Research Bulletin No 7, Soil Loss Estimator for Southern Africa*. Department of Agricultural Technical Services: Natal, South Africa.

FERTASA. 2007. *The Fertilizer Society of South Africa Handbook*. The Fertilizer Society of South Africa: Pretoria, South Africa.

GISCOE. 2001. *GISCOE Digital Terrain Models*, *Unpublished*. GIMS: Midrand, South Africa.

Green WH, Ampt GA. 1911. Studies on soil physics, 1. The flow of air and water through soils. *Journal of Agricultural Sciences* **4**:11-24.

Dabrowski J, Oberholster P, Dabrowski J, Le Brasseur J, Gieskes J. 2013. Chemical characteristics and limnology of Loskop Dam on the Olifants River (South Africa), in light of recent fish and crocodile mortalities, *Water SA* **39**: 675-686.

De Vente J, Poesen J, Verstraeten G, Govers G, Vanmaercke M, Van Rompaey A, Arabkhedri M, Boix-Fayos C. 2013. Predicting soil erosion and sediment yield at regional scales: Where do we stand? *Earth-Science Reviews* **127**: 16-29. DOI: 10.1016/j.earscirev.2013.08.014

Du Plessis C, van Zijl G, Johan Van Tol J, Manyevere A. 2020. Machine learning digital soil mapping to inform gully erosion mitigation measures in the Eastern Cape, South Africa. Geoderma 368.

DWS. 2011. *Classification of Significant Water Resources in the Olifants Water Management Area (WMA 4): Management Classes of the Olifants WMA. Report No: RDM/WMA04/00/CON/CLA/0213*. Report compiled by Golder Associates Africa, Prime Africa and Retha Stassen: Department of Water and Sanitation: Pretoria, South Africa.

DWS. 2013. *Development and implementation of irrigation water management plans to improve water use efficiency in the Agricultural Sector. Loskop Irrigation Board Water Management Plan Final Report. Project No: WP 10276*: *Directorate Water Use Efficiency*. Department of Water and Sanitation: Pretoria, South Africa.

DWS. 2015. *Business Case for the Vaal River Catchment Management Agency*. Department of Water and Sanitation: Pretoria, South Africa.

DWS. 2022a. *Spatial Data & Application Portal*. Department of Water and Sanitation: Pretoria, South Africa. Available from: https://gia.dws.gov.za/portal/home. Accessed 03 March 2022.

DWS. 2022b. *Resource Quality Information Services Portal*. Department of Water and Sanitation: Pretoria, South Africa. Available from: *https://www.dws.gov.za/iwqs/gis_data*. Accessed 03 March 2022.

Evans IS. 1979. *An integrated system of terrain analysis and slope mapping*. Final report on grant DA-ERO-591-73-G0040. University of Durham: England. http://www.soi.city.ac.uk/~jwo/phd.

FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009. Harmonized World Soil Database (version 1.1).

Essenfelder AH. 2016. SWAT Weather Database: A Quick Guide. Version: v.0.16.06. doi:10.13140/RG.2.1.4329.1927.

Gassman PW, Reyes MR, Green CH, Arnold JG. 2007. The Soil and Water Assessment Tool: historical development, applications, and future research directions. *Transactions of the ASABE* **50**(4): 1211-1250.

Gassman PW, Sadeghi AM, Srinivasan R. 2014. Applications of the SWAT Model Special Section: Overview and Insights. *Journal of Environmental Quality* **43**: 1-8.

Glenday J, Gokool S, Gwapedza D, Holden P, Rebelo A, Tanner J, Jumbi F, Metho P. 2021. Critical catchment model inter-comparison and model use guidance development. *WRC report K5/2927*. Water Research Commission: Pretoria, South Africa.

Görgens, A.H.M., Lorentz, S.A., Van der Laan, M., Jovanovic, N.Z., Matthews, N., Annandale J., Grové, B., Le Roux, J.J., 2012: *Modelling Agricultural Non-Point Source Pollution and Economic-Environmental Trade-Offs of Pollution Control Measures: A Project Overview*, *WRC report TT 516/12*. Water Research Commission: Pretoria, South Africa. ISBN 978-1-4312-0240-0.

Guzha AC, Rufino MC, Okoth S. Jacobs S, Nóbrega RLB. 2018. Impacts of land use and land cover change on surface runoff, discharge and low flows: Evidence from East Africa. *Journal of Hydrology: Regional Studies* **15**: 49-67.

Gyamfi C, Ndambuki J, Salim W. 2016. Simulation of Sediment Yield in a Semi-Arid River Basin under Changing Land Use: An Integrated Approach of Hydrologic Modelling and Principal Component Analysis, *Sustainability* **8**(11).

Harrison RL, van Tol J, Toucher ML. 2022. Using hydropedological characteristics to improve modelling accuracy in Afromontane catchments. *Journal of Hydrology: Regional Studies* **39**(100986).

Hill TR, Scott-Shaw BC, Gillham JS, Dickey M, Duncan GE, Everson CS, Everson TM, Zuma K, Birkett CK. 2019*. Assessing the impact of erosion and sediment yield from different land uses in farming and forestry systems and their effect on water resources in selected catchments of South Africa. WRC Report No. TT 788/19*. Water Research Commission, Pretoria.

Hume LJ, Healy WB, Tama K, Hosking WJ, Manarangi A, Reynolds J. 1985. Responses of citrus (Citrus sinensis) to nitrogen-phosphorus-potassium (NPK) fertiliser on 2 soils of Rarotonga, Cook Islands 2. Effects of NPK fertiliser rate, soil properties, and leaf nutrient levels on yield and tree size. *New Zealand Journal of Agricultural Research* **28**(4): 487-495.

Jetten V, Govers G, Hessel R. 2003. Erosion models: quality of spatial patterns. *Hydrological Processes* **17**: 887-900.

Kirkby MJ, Jones RJA, Irvine B, Gobin A, Govers G, Cerdan O, Van Rompaey AJJ, Le Bissonnais Y, Daroussin J, King D, Montanarella L, Grimm M, Vieillefont V, Puigdefabregas J, Boer M, Kosmas C, Yassoglou N, Tsara M, Mantel S, van Lynden GWJ, Huting J. 2004. *Pan-European Soil Erosion Risk Assessment: The PESERA Map. Vers. 1 Oct. 2003. Explanation of Special Publication Ispra 2004 No. 73 (S.P.I.04.73). European Soil Bureau Report 16 (EUR 21176)*. Office for Official Publications of the European Communities: Luxembourg.

Land Type Survey Staff (1972-2006*). Land Types of South Africa: Maps (69 sheets) and Memoirs (39 books).* Agricultural Research Council - Soil, Climate and Water, Pretoria.

Le Roux JJ. 2018. Sediment yield potential in South Africa's only large river Network without a dam: implications for water resource management. *Land Degradation and Development* **29**(3): 765-775. https://0-doi.org.wam.seals.ac.za/10.1002/ldr.2753

Le Roux JJ, Barker CH, Weepener HL, Van den Berg EC, Pretorius SN, 2015: *Sediment yield modelling in the Mzimvubu River Catchment, WRC Report No. 2243/1/15.* Water Research Commission, Pretoria. ISBN 978-1-4312-0667-4.

Le Roux JJ, Germishuyse T, Lorentz S. 2009. *Case studies and scenarios of NPS control measures and mitigation of downstream pollution: Using the Soil and Water Assessment Tool, KwaZulu-Natal. ISCW Report No. GW/A/2009/04.* ARC: Pretoria, South Africa.

Le Roux JJ, Morgenthal TL, Malherbe J, Sumner PD, Pretorius DJ. 2008. Water erosion prediction at a national scale for South Africa. *Water SA* **34**(3): 305-314.

Le Roux JJ, Sumner PD. 2012. Factors controlling gully development: comparing continuous and discontinuous gullies. *Land Degradation & Development* **23**: 440-449. DOI: 10.1002/ldr.1083

Le Roux JJ, Sumner PD, Lorentz SA, Germishuyse T. 2013. Connectivity Aspects in Sediment Migration Modelling Using the Soil and Water Assessment Tool. *Geosciences* **3**(1): 1-12.

Le Roux PAL, Fraenkel CH, Bothma CB, Gutter JH, Du Preez CC. 2006. *Soil survey report: Mkabela catchment*. Department of Soil, Crop and Climate Sciences, University of the Free State: Bloemfontein, South Africa.

Lenhart T, Van Rompaey A, Steegen A, Fohrer N, Frede H, Govers G. 2005. Considering spatial distribution and deposition of sediment in lumped and semi-distributed models. *Hydrological Processes* **19**: 785-794.

Lorentz SA, Kollongei J, Snyman N, Berry SR, Jackson W, Ngaleka K, Pretorius JJ, Clark D, Thornton-Dibb S, le Roux JJ, Germishuyse T, Görgens AHM. 2012. *Modelling Nutrient and Sediment Dynamics at the Catchment Scale*. *WRC report 1516/3/12*. Water Research Commission: Pretoria, South Africa.

Lorentz SA, Miller J, Lechler P, Mackin G, Lord M, Kollongei JK, Pretorius J, Ngeleka K, Zondi N, Le Roux JJ. 2011. *Definition of process zones and connectivity in catchment scale NPS processes, WRC report 1808/1/11*. Water Research Commission: Pretoria, South Africa.

Mararakanye N, Le Roux JJ, Franke AC. 2020. Using satellite-based weather data as input to SWAT in a data poor catchment. *Physics and Chemistry of the Earth* **117**: 102871.

Mararakanye N, Le Roux JJ, Franke AC. 2021. Long-term water quality assessments under changing land use in a large semi-arid catchment in South Africa. *Science of The Total Environment* 151670. https://doi.org/10.1016/j.scitotenv.2021.151670

Martinec J, Rango A. 1989. Merits of statistical criteria for the performance of hydrological models. *Water Resources Bulletin* **25**: 421-432.

Mishra A, Froebrich J, Gassman PW. 2007. Evaluation of the SWAT model for assessing sediment control structures in a small watershed in India. *Transactions of the ASABE* **50**(2): 469-478.

Mucina L, Rutherford MC. 2006. *The vegetation of South Africa, Lesotho and Swaziland*. *Strelitzia* **19**. South African National Biodiversity Institute, Pretoria, South Africa.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models. *Journal of Hydrology* **10**: 282-290.

Neitsch SL, Arnold JG, Kiniry JR, Williams JR. 2011. *Soil and Water Assessment Tool Theoretical Documentation, Version 2009*. USDA Agriculture Research Service, Blackland Research Center, Texas Agricultural Experiment Station: Texas, USA.

NGI. 2013. *What is ITIS Portal*. Chief Directorate – National Geo-spatial Information: Capetown, South Africa. Available from: http://www.ngi.gov.za/index.php/online-shop/what-is-itis-portal. Accessed 14 October 2013.

Overton DE. 1966. Muskingum flood routing of upland streamflow. *Journal of Hydrology* **4**: 185-200.

Priestley CHB, Taylor RJ. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review* **100**:81-92.

Saha S, Moorthi S, Pan HL, Wu X, Wang Jiande, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang Jun Hou YT, Chuang H, Juang HMH, Sela J, Iredell M, Treadon R, Kleist D, Van Delst P, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, van den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm JK, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou CZ, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M, 2010. The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc*. **91**: 1015-1058. https://doi.org/10.1175/2010BAMS3001.1

SANLC. 2014. National Land Cover Data of South Africa 1990 and 2013-14 Department of Environmental Affairs: Pretoria, South Africa. Available from https://www.environment.gov.za/projectsprogrammes/egis_landcover_datasets

SANLC. 2018. National Land Cover Data of South Africa 2018 Department of Environmental Affairs: Pretoria, South Africa. Available from https://www.environment.gov.za/projectsprogrammes/egis_landcover_datasets

SANLC. 2020. National Land Cover Data of South Africa 2020. Department of Environmental Affairs: Pretoria, South Africa. Available from https://www.environment.gov.za/projectsprogrammes/egis_landcover_datasets

SAWS. 2019. South African Weather Service Data Portal. Unpublished, South African Weather Service: Pretoria, South Africa. Available from https://www.weathersa.co.za/

Saxton K, Willey P, Rawls W. 2006. Field and Pond Hydrologic Analyses with the SPAW Model. In: ASABE Annual International Meeting, 9-12 July 2006, Portland, Oregon.

Schaap MG, Leij FJ, van Genuchten MT. 2001. ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* **251**(3-4): 163-176. https://doi.org/10.1016/S0022-1694(01)00466-8.

Schmidt J, Evans IS, Brinkmann J. 2003. Comparison of polynomial models for land surface curvature calculation. *International Journal of Geographical Information Science* **17**(8): 797-814.

Schulze RE. 1995. *Hydrology and Agrohydrology: A text to accompany the ACRU 3.00 Agrohydrological Modelling System, WRC report no TT69/95*. Water Research Commission: Pretoria, South Africa.

Schulze RE. 2007. *South African Atlas of Climatology and Agrohydrology, WRC Report 1489/1/06*. Water Research Commission: Pretoria, South Africa.

Schulze RE, Schütte S. 2020. Mapping soil organic carbon at a terrain unit resolution across South Africa. *Geoderma* **373**. https://doi.org/10.1016/j.geoderma.2020.114447.

Scott-Shaw BC, Hill TR, Gillham JS. 2020. Calibration of a modelling approach for sediment yield in a wattle plantation, KwaZulu-Natal, South Africa. *Water SA* **46**(2) 171-181.

Singh A, van Veelen M. 2001. *Olifants River Ecological Water Requirements Assessment: Ecological Reserve Report. Project number: P9104. DWAF report number: PB-000-00-5299*. Available at: https://www.environment.gov.za/sites/default/files/docs/olifant_ecological_reserve_report.pdf.

Srinivasan R, Ramanarayanan TS, Arnold JG, Bednarz ST. 1998. Large area hydrologic modeling and assessment part II: model application. *Journal of the American Water Resources Association* **34**: 91-101.

Srinivasan R, Zhang X, Arnold J. 2010. SWAT ungauged: hydrological budget and crop yield predictions in the upper Mississippi river basin. *Transactions of the ASABE* **53**(5): 1533-1546.

Tibebe D, Bewket W. 2011. Surface runoff and soil erosion estimation using the SWAT model in the Keleta watershed, Ethiopia. *Land Degradation and Development* **22**(6): 551-564.

Toucher MI, Ramjeawon M, Mcnamara MA, Rouget M, Bulcock H, Kunz RP, Moonsamy J, Mengistu M, t. Naidoo, Vather T, Aldworth TA. 2020. Resetting the baseline land cover against which streamflow reduction activities and the hydrological impacts of land use change are assessed. *WRC Report No. 2437/1/19.* Water Research Commission, Pretoria. ISBN 978-0-6392-0110-8.

USDA SCS. 1972. *National Engineering Handbook, Section 4 Hydrology*. USDA Agricultural Conservation Service: USA.

van der Laan M, Franke A. 2019: *Quantifying and managing agricultural nitrogen and phosphorus pollution from field to catchment scale. WRC Report No. TT 792/19*. Water Research Commission, Pretoria.

van Deventer H. van Niekerk L, Adams J, Dinala MK, Gangat R, Lamberth SJ, Lötter M, Mbona N, MacKay F, Nel JL, Ramjukadh C, Skowno A, Weerts SP. 2020. National Wetland Map 5: An improved spatial extent and representation of inland aquatic and estuarine ecosystems in South Africa. *Water SA* **46**(1) 66-79.

Vanmaercke M, Poesen J, Verstraeten G, De Vente J, Ocakoglu F. 2011. Sediment yield in Europe: Spatial patterns and scale dependency. *Geomorphology* **130**(3-4): 142-161.

van Tol JJ, Bieger K, Arnold JG. 2021. A hydropedological approach to simulate streamflow and soil water contents with SWAT+. *Hydrological Processes* **35**(6): 1-14.

van Tol JJ, Le Roux PAL, Hensley M. 2010. Soil indicators of hillslope hydrology in the Bedford catchments, South Africa. *South African Journal of Plant and Soil* **27**(3): 242-251.

van Tol JJ, Le Roux PAL, Lorentz SA, Hensley M. 2013. Hydropedological Classification of South African Hillslopes. *Vadose Zone Journal* **12**(4).

van Tol JJ, van Zijl GM, Julich S. 2020. Importance of Detailed Soil Information for Hydrological Modelling in an Urbanized Environment. *Hydrology* **7**(34): 1-15.

van Zijl GM, Ellis F, Rozanov A. 2014. Understanding the combined effect of soil properties on gully erosion using quantile regression. *South African Journal of Plant and Soil* **31**(3): 163-172. https://doi.org/10.1080/02571862.2014.944228

Van Zijl GM, Van Tol JJ, Bouwer D, Lorentz S, le Roux PAL. 2020. Combining Historical Remote Sensing, Digital Soil Mapping and Hydrological Modelling to Produce Solutions for Infrastructure Damage in Cosmo City, South Africa. *Remote Sensing* **12**(433).

Van Zijl GM, Van Tol JJ, Riddell ES. 2016. Digital Soil Mapping for Hydrological Modelling. In: Zhang GL., Brus D., Liu F., Song XD., Lagacherie P. (eds) Digital Soil Mapping Across Paradigms, Scales and Boundaries. Springer Environmental Science and Engineering. Springer, Singapore.

Wahren FT, Julich S, Nunes JP, Gonzalez-Pelayo O, Hawtree D, Feger K, Keizer JJ. 2016. Combining digital soil mapping and hydrological modeling in a data scarce watershed in north-central Portugal. *Geoderma* **264**: 350-362.

Walter T, Kloos J, Tsegai D. 2011. Options for improving water use efficiency under worsening scarcity: evidence from the middle olifants sub-basin in South Africa. *Water SA* **37**: 357-370.

Wang X, Yang W, Melesse AM. 2009. Using Hydrologic Equivalent Wetland Concept within SWAT to estimate streamflow in watersheds with numerous wetlands. *Transactions of the ASABE* **51**(1): 55-72.

Weepener HL, Van den Berg HM, Metz M, Hamandawana H. 2012. The development of a hydrologically improved Digital Elevation Model and derived products for South Africa based on the SRTM DEM. WRC report 1908/1/11. Water Research Commission: Pretoria, South Africa. ISBN 978-1-4312-0217-1.

Wei X, Bailey RT, Tasdighi A. 2018. Using the SWAT Model in Intensively Managed Irrigated Watersheds: Model Modification and Application. *Journal of Hydrologic Engineering* **23**(10): 04018044.

Williams JR. 1975. Sediment-yield prediction with universal equation using runoff energy factor. In *Present and Prospective Technology for predicting sediment yield and sources. ARS.S-40*, US Gov. Print. Office: Washington, DC; 244-252.

Wischmeier WH, Smith DD. 1978. *Predicting Rainfall Erosion Losses, a Guide to Conservation Planning. USDA Agriculture Handbook No. 537*. USDA: Washington DC, USA.

APPENDIX 1: METADATA FOR ARCSWAT BASELINE DATA

Table 1. SWAT catchment outline data

Metadata type	Digital elevation model	Catchment outlines		Catchment inlets and outlet locations	DWS monitoring station locations
Title	SRTM_DEM_90m	Catchments_tertiary_ buffer100m	Catchments_quaternary_ buffer100m	Catchment_inlets_outlets	DWS_monitoringstations_ locations
Date created	2012	2018	2018	2022	2022
Purpose	Hydrologically improved DEM for SA	Tertiary catchment shapefile with 100m buffer to identify area of interest	Quaternary catchment shapefile with 100m buffer to identify area of interest	Tertiary and quaternary catchment outlets and inlets locations	DWS monitoring stations locations
Description of base datasets	Global SRTM DEM https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission	Tertiary catchment map available at DWS Spatial Data & Application Portal https://gia.dws.gov.za/portal/home	Quaternary catchment map available at DWS Spatial Data & Application Portal https://gia.dws.gov.za/portal/home	Flow paths derived from flow accumulation of hydrologically improved DEM for SA (where both entering and exiting flow paths overlay with catchment boundaries)	DWS monitoring stations locations available at DWS Resource Quality Information Services Portal https://www.dws.gov.za/iwqs/gis_data/
Limitations and assumptions	Overall accuracy 90%	Outlines based on Hydrologically improved DEM for SA of Weepener *et al.* (2012)	Outlines based on Hydrologically improved DEM for SA of Weepener *et al.* (2012)	Point locations based on Hydrologically improved DEM for SA of Weepener *et al.* (2012)	N/A
Place keywords	National	National	National	National	National
Citation	WRC report 1908/1/11, Weepener *et al.* (2012)	DWS (2022a)	DWS (2022a)	N/A	DWS (2022b)
Spatial representation	Format: GRID Resolution: 90 m Spatial Reference: WGS_1984_Albers	Format: Shapefile Spatial Reference: WGS_1984_Albers	Format: Shapefile Spatial Reference: WGS_1984_Albers	Format: Shapefile Spatial Reference: WGS_1984_Albers	Format: Shapefile Spatial Reference: WGS_1984_Albers

Table 2. South African National Land Cover linked to SWAT land cover codes

Metadata type	National land cover map 2013/14	National land cover map 2018	National land cover map 2020
Title	NLC2013_SWAT	NLC2018_SWAT	NLC2020_SWAT
Date created	2013-2014	2018	2020
Purpose	2013/14 national land cover grid of SA, of which the grid values are reclassified to match the SWAT land cover codes.	2018 national land cover grid of SA, of which the grid values are reclassified to match the SWAT land cover codes.	2020 national land cover grid of SA, of which the grid values are reclassified to match the SWAT land cover codes.
Description of base datasets	The South African National Land-Cover 2013/14 dataset with 72 land cover classes has been generated from 30 meter multi-seasonal Landsat 8 imagery. Available from https://www.environment.gov.za/ projectsprogrammes/egis_landcover_datasets	The South African National Land-Cover 2018 dataset with 73 land cover classes has been generated from 20 meter multi-seasonal Sentinel 2 satellite imagery. Available from https://www.environment.gov.za/ projectsprogrammes/egis_landcover_datasets	The South African National Land-Cover 2020 dataset with 73 land cover classes has been generated from 20 meter multi-seasonal Sentinel 2 satellite imagery. Available from https://www.environment.gov.za/ projectsprogrammes/egis_landcover_datasets
Limitations, assumptions	Overall map accuracy 82.53%	Overall map accuracy 91.32%	Overall map accuracy 85.47%
Place keywords	National	National	National
Citation	SANLC (2014)	SANLC (2018)	SANLC (2020)
Spatial representation	Format: GRID Resolution: 30 m Spatial Reference: WGS_1984_Albers	Format: GRID Resolution: 20 m Spatial Reference: WGS_1984_Albers	Format: GRID Resolution: 20 m Spatial Reference: WGS_1984_Albers

LU_landcover2020.txt – comprises the user lookup table that links abovementioned 2020 national land cover grid with the SWAT land cover codes;
LU_landcover2018.txt – comprises the user lookup table that links abovementioned 2018 national land cover grid with the SWAT land cover codes;
LU_landcover2013.txt – comprises the user lookup table that links abovementioned 2013 national land cover grid with the SWAT land cover codes.

Table 3. Soil map with SWAT attribute data for each Land Type of SA

Metadata type	Soil map (without soil attribute data)	Soil database table (with soil attribute data)
Title	Usersoil_SWAT	Usersoil.accdb
Date created	2022	2022
Purpose	Land Type shapefile of SA illustrating locations of land type (soil combinations).	Microsoft Access database table with soil attribute data for each of the 16556 Land Types of SA.
Description of base datasets	Digital map (1:250 000 scale) and soil inventory database that reflects the dominant soils in each land type by percentage.	Soil attribute data include textural and soil hydraulic parameter values for two soil layers.
Limitations and assumptions	Data cannot indicate exactly where within each Land Type the various soils occur. It is thus not a substitute for a detailed soil map.	Several assumptions (pedotransfer functions see Table 1 of report) had to be made to assign soil parameter values to Land Types at a national scale.
Place keywords	National	National
Citation	Land Type Survey Staff (1972-2006)	This report
Spatial representation	Format: Shapefile Spatial Reference: WGS_1984_Albers	Format: Microsoft Access database table

LU_usersoil.txt – comprises the user lookup table that links abovementioned Land Type grid with abovementioned Access database table with soil attribute data for each Land Type of SA.

Table 4. Weather statistics (WGN) files required as input by the model

Metadata type	Weather statistics (WGN) 1981-2000	Weather statistics (WGN) 2001-2020
Title	wgenuser1981-2000	wgenuser2001-2020
Date created	2021	2021
Purpose	Weather statistics database table with 11 WGN files containing the weather generator (WGN) statistics from 1981-2000 of 11 stations.	Weather statistics database table with 12 WGN files containing the WGN statistics from 2001-2020 of 12 stations.
	WGN input files consists of several weather statistics needed by SWAT to generate representative daily climate data for simulated catchments in two instances: when the user specifies that simulated weather will be used or when measured data is missing.	
Description of base datasets	Agroclimatology Database, Agricultural Research Council-Soil, Climate and water including precipitation, temperature, solar radiation, relative humidity and wind speed.	
Limitations and assumptions	Daily weather data are not provided. SWAT users have to obtain their own daily climate data. 12 WGN files were created by acquiring and interpreting climate data from ARC (2021) weather stations in different climate zones in SA. The completeness of climate data was the most important consideration for selecting weather stations. The station in Hobhouse (30673), did not have enough data to calculate statistics for the period 1981-2000, and were excluded in the 1981-2000 dataset. Mechanical stations only measured sunshine hours and not solar radiation. Therefore, solar radiation data for the period 2001-2020 were used for the calculation of all WGN files. Furthermore, the mechanical stations measured wind run, which was converted to wind speed for the purposes of this project.	
Place keywords	National	National
Citation	ARC (2021)	ARC (2021)
Spatial representation	Format: Microsoft Access database table	Format: Microsoft Access database table

APPENDIX 2: CAPACITY DEVELOPMENT

Several forms of capacity building took place because of the project including post-graduate students, university staff and institutional development. A national input database to run the SWAT model (in a GIS) will assist researchers and students to set up and run the SWAT model anywhere in SA. Students whose post-graduate studies are contributed to the project are shown in Table 1 below.

Table 1. Students and qualifications.

	Name	University	Degree	Student No.	Student ID	Gender	Race	Start	End
1	Aimee Coraline Thomson	UP	Hons Soil Science	16015925	9612030040086	Female	White	2020	2021
2	Marike Stander	UFS	PhD Geography	2003089741	8406050130088	Female	White	2019	2023
3	Ndifelani Mararakanye	UFS	PhD Geography	2016444883	7810105514082	Male	African	2017	2022
4	Christiaan Schutte	UP	PhD Water Resource Management	10063006	9103045027086	Male	White	2021	2024
5	Leushantha Mudaly	UP	PhD Soil Science	12108597	8509040218085	Female	Indian	2017	2023

APPENDIX 3: KNOWLEDGE DISSEMINATION

Knowledge dissemination workshop

An extended RG Meeting was planned in which the objectives of a Dissemination Workshop be included in the Agenda for the Final RG Meeting on the 28th of February 2022. However, the final meeting was moved to the 1st of March 2022 and the Dissemination Workshop was postponed to April 2022 (date TBC). Dr Le Roux will present on the datasets and how they can be used to run the SWAT model and sharing his screen and have a live demonstration using the datasets in ArcSWAT in one catchment.

Presentation at national conferences

The project outcomes will be presented:

- 12-14 September 2022 at the SSAG biennial conference at the University of Pretoria.
- January 2023 at the Combined conference at the University of the Free State.

Publication in peer-reviewed journal

A paper will be submitted to Water SA.