EXSMET: A Model for Soil Moisture and Evapotranspiration Determination in Fine Detail over the SADC Region

Exporting PyTOPKAPI and HYLARSMET over SADC including RSA with extended spatial and computational capacity of Soil Moisture and Evapotranspiration for flood and drought monitoring

Report to the
Water Research Commission

by

Geoffrey Pegram and Scott Sinclair
Pegram and Associates (Pty) Ltd.

WRC Report No. 2312/1/18
ISBN 978-0-6392-0014-9

June 2018

Obtainable from
Water Research Commission
Private Bag X03
Gezina, 0031

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

Printed in the Republic of South Africa

Executive Summary

What is this project about? The subtitle says it succinctly: Exporting PyTOPKAPI and HYLARSMET over SADC, including RSA, with extended spatial and computational capacity of Soil Moisture and Evapotranspiration for flood and drought monitoring. Timeous and routine monitoring of the spatial distribution of Soil Moisture and Evapotranspiration over a large region in fine detail has great value for coping with two weather extremes: Flash Floods and Droughts. The current state of soil moisture conditions has a major impact on the runoff response of a catchment to heavy rainfall; and monitoring the wetness of the soil in detail over large regions, without having to laboriously take expensive samples, is a bonus for agricultural managers who need to understand the status of crop growth potential. This is particularly relevant in the southern and central African countries in the SADC region which contribute importantly to the food basket of Africa.

Why did we do this work? This proposed project had two thrusts. The first was to extend coverage of near real-time Soil Moisture (SM) and Evapotranspiration (ET) monitoring over the SADC region using the existing methodology of operating the PyTOPKAPI hydrological model in Land Surface Modelling mode (as proven in RSA), which we called HYLARSMET in the completed WRC project K5/2024. This outreach project is likely to be beneficial in many ways, not only technically and socially, but also to help cement other forms of collaboration within the region.

The second thrust was to develop powerful computing techniques that will make it feasible to drive the modelling procedure for the vastly increased number of cells required to cover the SADC region. The same techniques can be applied to improve the spatial detail of distributed hydrological modelling at fine scale in RSA, as piloted in the WRC project K5/2024, to complement the South African Flash Flood Guidance system initiative. With a contribution from us, this was undertaken by Mr Eugene Poolman of SAWS in WRC project K5/2068. The improved computational speed will also add convenience when performing ensemble simulations to assess the impacts of climate change scenarios on hydrological variables at fine scale, hence the link with Prof Bruce Hewitson's CSAG group at UCT. We joined this group to exploit our strong link through their WRC project 'Managing limits in skill for seasonal climate forecasting'.

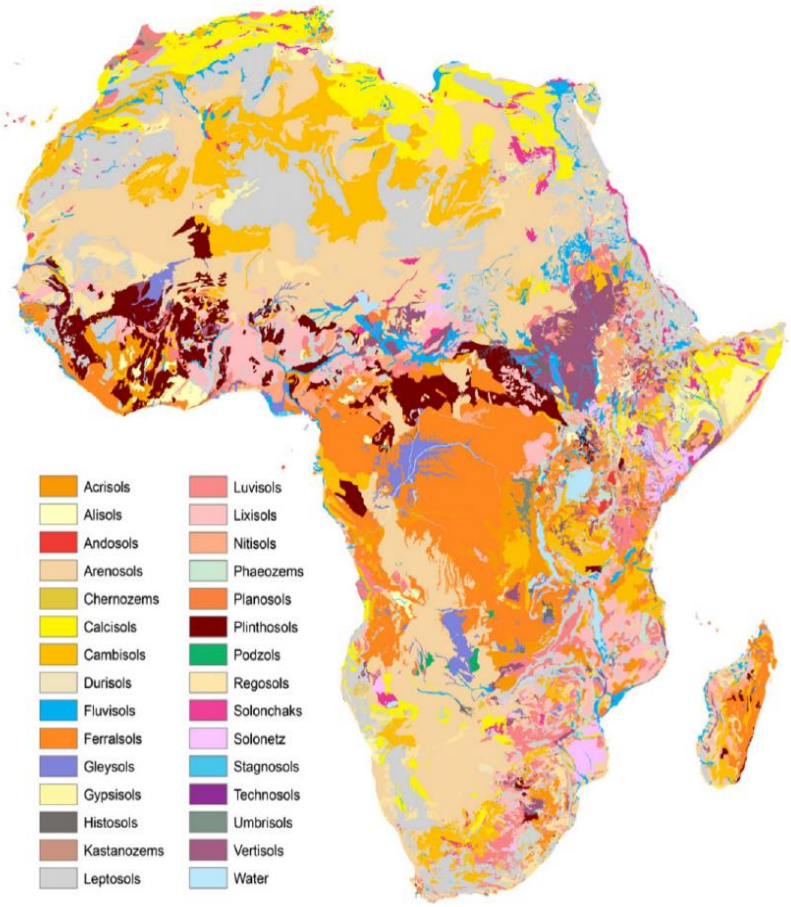
This report details answers to the following important research questions. How did we:

- ascertain what rainfall data are available in the SADC countries outside our borders and perform checks on their suitability for modelling
- obtain suitable ground cover and soil maps over the whole SADC region (e.g. FAO and others), for comparison with those already available in RSA
- exploit the sensitivity calculations performed on the RSA data-set under HYLARSMET
- determine the best way to compare FAO sets of ground-based data with our RSA sets

- determine if there are better alternative rainfall inputs to TRMM, for near real-time precipitation data input. If not, exploit bias adjustment of the TRMM product
- exploit the Soil Moisture estimates of the European Space Agency's SMOS [Soil Moisture & Ocean Salinity] mission when ready for use in model inter-comparisons
- determine how to best cope with the uncertainties associated with input parameters and forcing variables [TRMM in particular] when computing ensembles of historical and forecast data streams
- devise means of increasing computing capacity and the speed of calculations by refining key parts of the code and employing parallel [or High Performance] computing power
- determine the best ways of cold-starting calculations [model initialization] for both gauged and ungauged catchments?

The body of the report relates the methodology in as visual a way as possible – there are 107 Figures enhancing the message of the text. In the main, they illustrate the complexity of the problem we faced and solved. We have included almost no mathematics in the report – all computations were performed on computers using the Python language. For technically competent readers, we have given the details of the sites where the data and the programs we have devised can be accessed as open source material, at the end of this Executive Summary.

As an example of complexity and the necessity of very large data-sets, the following image from Section 2 illustrates the data handling problem well. The detail is at 1 km resolution, as stated in the report: 'Soil Mapping Units (SMUs) … have been gridded to a resolution of 30 arc-seconds (nominally about 1 km)' – Dewitte et al., 2013, noting that Africa's area is 30.37 million sq km.



Harmonised soil map of Africa (after Dewitte et al., 2013)

There are several major original research highlights that come out of the project. Our focus in EXSMET was to develop and utilise calculations at isolated 1 km pixels in 0.125° (about 12 km) squares over SADC, driven by hydrological modelling through our hydrological model PyTOPKAPI, leading to Soil Moisture assessments. We took advantage of gains made in an improved understanding of Soil Moisture and surface water hydrological processes through the completed WRC project HYLARSMET, which was blessed by the availability of better hydrological data, especially the various forms of more accurate remotely-sensed data with better precipitation coverage, like TRMM, recently replaced by GPM. Throughout this project, strategic partnerships with international expertise in climate, soil water and surface water hydrological research flourished. Hydrological tools that were developed in the past under HYLARSMET were upgraded, re-developed and accelerated by alternative tools that are suited to current data availability, improved knowledge and the recent technological advances in hydrological modelling using the PyTOPKAPI model.

In this EXSMET project, the loss of information due to continued deterioration of hydrological gauging processes and other installed earth measurement devices were managed through the intensive use of new data sources from remote sensing, coupled with the limited earth-based measurements.

In conclusion, the major outcomes we achieved here are that (i) we devised techniques to model Soil Moisture and Evapotranspiration over SADC in detail and (ii) model hydrological runoff on ungauged catchments over RSA, both with estimated levels of confidence, using accelerated parallel computing algorithm designs, which were especially devised for this project.

<<<<<>>>>>

Innovation:

This is probably the first WRC project to tackle Hydrometeorological problems outside RSA. The four-year study covered all SADC countries, modelling Evapotranspiration and Soil Moisture at 3-hour intervals over a 5-year period. The methodologies developed for modelling hydrological responses to precipitation and evaporation were applied over a very large area at nearly 50 000 sites. This massive computing task demanded the development of new tools for high speed parallel computing which were devised, archived and made free-ware. In addition, with poor meteorological information available outside data-rich RSA, it was mandatory that remote sensing tools should be exploited. A large amount of effort was made to ensure that the appropriate data sources were downloaded, compared, sorted and chosen.

Capacity building – postgraduate students:

Graduated:

MSc Eng: Simon Malose Ngoepe (2017) 'Conditioning Tropical Rainfall Measuring Mission (TRMM) data using ground-based rainfall data'

MSc Hydrology: Thigesh Vather (2016): Thesis entitled: 'Comparison between satellite based and cosmic ray probe soil moisture measurements: A case study in the Cathedral Peak Catchment'. Supervisor Prof Colin Everson in a joint WRC project with Pegram & Associates – K5/2323: 'The Validation of the Variables (Evaporation and Soil Moisture) in Hydrometeorological Models: Phase II, Application of Cosmic Ray Probes for Soil Water Measurement'

Currently being supervised:

[PhD] Stephanie Landman: Topic: 'High resolution ensemble numerical weather prediction forecasts for applications in the field of hydrology'

[MSc Eng, co-supervised] Tadiwanashe Gutsa: Utility of radar rainfall estimates in catchment rainfall analysis

Knowledge dissemination during this 4-year project from April 2014:

Papers Published in International Journals:

1. Bárdossy, A., and G. G. S. Pegram (2017). Combination of radar and daily precipitation data to estimate meaningful sub-daily point precipitation extremes, *Journal of Hydrology*, Volume 544, January 2017, pp 397-406
2. Sugimoto, Takayuki, András Bárdossy, Geoffrey G.S. Pegram, and Johannes Cullmann (2016), Investigation of hydrological time series using copulas for detecting catchment characteristics and anthropogenic impacts, *Hydrol. Earth Syst. Sci.*, 20, 2705-2720, doi:10.5194/hess-20-2705-2016.
3. Bárdossy, A., and G. G. S. Pegram (2016), Space-time conditional disaggregation of precipitation at high resolution via simulation, *Water Resour. Res.*, *52*, 920-937, doi:10.1002/2015WR018037.
4. Bárdossy, András, Geoffrey Pegram, Scott Sinclair, Justin Pringle and Derek Stretch, (2015), Circulation patterns identified by spatial rainfall and ocean wave fields in Southern Africa, *Frontiers in Environmental Science*, April, Volume3, Article 31
5. Bárdossy, András and Geoff Pegram (2014), Infilling missing precipitation records – A comparison of a new copula-based method with other techniques, *Journal of Hydrology*, Vol. 519, pp. 1162-1170
6. Yeboah Gyasi-Agyei and Geoffrey Pegram (2014) Interpolation of daily rainfall networks using simulated radar fields for realistic hydrological modelling of spatial rain field ensembles, *Journal of Hydrology*, Vol. 519, pp. 777-791

Books (Chapters in):

Geoff Pegram, Damien Raynaud, Eric Sprokkereef, Martin Ebel, Silke Rademacher, Jonas Olsson, Cristina Alionte-Eklund, Barbro Johansson, Göran Lindström and Henrik Spångmyr (2016). Present and Future Requirements for Using and Communicating Hydro-Meteorological Ensemble Prediction Systems for Short-, Medium-, and Long-Term Applications, in Q. Duan et al. (eds.), Handbook of Hydrometeorological Ensemble Forecasting, Springer-Verlag Berlin Heidelberg, DOI 10.1007/978-3-642-40457-3_39-1, pp. 4-20. ISBN 978-3-642-40457-3.

Presentations at International symposia & conferences:

1. Pegram, Geoff, András Bárdossy and Scott Sinclair (2017). Combination of radar and daily precipitation data to estimate meaningful sub-daily point precipitation extremes. *European Geosciences Union General Assembly*, Vienna
2. Pegram, Geoff, András Bárdossy & Scott Sinclair (2016). Meaningful Quantile adjustment of TRMM/GPM daily rainfall estimates. *European Geosciences Union General Assembly*, Vienna
3. Bárdossy, András and Geoffrey Pegram (2016). Space-time disaggregation of precipitation using daily precipitation and radar observations. *European Geosciences Union General Assembly*, Vienna
4. Bárdossy, András and Geoffrey Pegram (2015). Spatio-temporal statistical model for the optimal combination of precipitation measured at different timescales for estimating

unobserved point values and disaggregating to finer timescales, *European Geosciences Union General Assembly*, Vienna

5. Pegram, Geoff, András Bárdossy, Scott Sinclair, & Simon Ngoepe (2015). Accounting for uncertainty in the repair of rain gauge records and their spatial interpolation, Eur*opean Geosciences Union General Assembly*, Vienna
6. Sinclair, Scott, Geoff Pegram, Michael Mengistu (2015). Validating HYLARSMET: a Hydrologically Consistent Land Surface Model for Soil Moisture and Evapotranspiration Modelling over Southern Africa using Remote Sensing and Meteorological Data, Eur*opean Geosciences Union General Assembly*, Vienna
7. Pegram, Geoffrey and Yeboah Gyasi-Agyei (2014), Ensembles of spatial daily rainfall interpolations of raingauge networks using conditionally merged simulated radar fields, for hydrological modelling, AOGS, Sapporo, Japan
8. Pegram, Geoffrey and András Bárdossy (2014), Modelling daily rain-gauge network measurement responses under a changing climate, *Dooge-Nash International Symposium*, Dublin.
9. Pegram, Geoffrey and Yeboah Gyasi-Agyei (2014), A spatial daily rainfall model for interpolation of raingauge networks using artificial radar fields, for realistic hydrological modelling, *European Geosciences Union General Assembly*, Vienna
10. Bárdossy, András and Geoffrey Pegram (2014), Uncertainty estimation and validation of data for hydrological applications, *European Geosciences Union General Assembly,* Vienna
11. Pegram, Geoffrey and András Bárdossy (2014), Infilling and interpolation of precipitation at different temporal scales in South Africa, European Geosciences Union General Assembly, Vienna

Presentations at National Conferences and Work-shops:

1. Pegram, Geoffrey, Scott Sinclair and András Bárdossy (2016), Space-time modelling of precipitation using pluviometer, daily precipitation and radar data for updating IDF and ARF relationships, looking to the future, *18th SANCIAHS National Hydrology Symposium*, University of KwaZulu-Natal, September.
2. Pegram G.G.S., A. Bardossy and S. Sinclair (2016), Meaningful QQ adjustment of TRMM/GPM daily rainfall estimates, *18th SANCIAHS National Hydrology Symposium*, University of KwaZulu-Natal, September.
3. Sinclair, Scott, Geoff Pegram and Andras Bardossy (2016), New methods of infilling Southern African raingauge records enhanced by Annual, Monthly and Daily Precipitation estimates tagged with uncertainty, *18th SANCIAHS National Hydrology Symposium*, University of KwaZulu-Natal, September.
4. Pegram, Geoffrey, Scott Sinclair and Simon Ngoepe (2015). Comparing interpolated daily gauge rainfall over RSA with TRMM to determine possible bias correction for hydrological applications, 31st Conference of the South African Society for Atmospheric Science, Hennops River, RSA
5. Sinclair S., M. Mengistu, G.G.S. Pegram and C Everson (2014). HYLARSMET validation, *17th SANCIAHS National Hydrology Symposium*, University of the Western Cape, Bellville, September
6. Pegram G.G.S., S. Sinclair and A. Bardossy (2014). Accounting for uncertainty in the repair of daily rain gauge records, *17th SANCIAHS National Hydrology Symposium*, University of the Western Cape, Bellville, September

WRC Reports published since 2013:

1. Pegram GGS, Scott Sinclair and András Bárdossy (2016). New methods of infilling Southern African raingauge records enhanced by Annual, Monthly and Daily Precipitation estimates tagged with uncertainty. *Water Research Commission,* WRC Research Report No. 2241/1/15, ISBN 978-1-4312-0758-9, March
2. Poolman E, E de Coning, E Becker, GGS Pegram, S Sinclair, N Kroese (2015), Improvement of Early Preparedness and Early Warning Systems for Extreme Climatic Events – Flood Warnings, *Water Research Commission*, WRC Report No. 2068/1/15 ISBN 978-1-4312-0664-3
3. Pegram GGS, Scott Sinclair and András Bárdossy (2013). Modelling daily rain-gauge network measurement responses under changing climate scenarios, *Water Research Commission,* WRC Research Report No. 1964/1/13 ISBN 978-1-4312-0476-2, November.
4. Sinclair, S and GGS Pegram (2013**)** HYLARSMET: A Hydrologically Consistent Land Surface Model for Soil Moisture and Evapotranspiration Modelling over Southern Africa using Remote Sensing and Meteorological Data, *Water Research Commission,* WRC Research Report No. 2024/1/12, ISBN 978-1-4312-0392-5, March.

Archives:

It is important for the technical reader to know where to access the routines and data-sets developed in this report. These can be found in: PyTOPKAPI v0.4.0. Zenodo.

http://doi.org/10.5281/zenodo.820640

and in

https://github.com/sahg/PyTOPKAPI

Acknowledgements:

We gratefully acknowledge the financial support for this project by the Water Research Commission.

We also acknowledge the assistance of the Reference Group (and their willing alternatives) for their support and contributions.

They are:

Mr W Nomquphu	:	WRC (Chairperson)
Ms T Chetty	:	UKZN-CWRR
Prof C Everson	:	UKZN-CWRR
Dr E Kapangaziwiri	:	CSIR
Dr C Lennard	:	UCT
Dr J Mwenge Kahinda	:	CSIR
Dr J Malherbe	:	ARC-ISCW
Mr T Newby	:	NEOSS
Dr E Poolman	:	SAWS
Dr J van Tol	:	UFH
Mr A Sparks	:	AURECON

In particular, we are indebted to the Chairperson, Mr Wandile Nomquphu for his wisdom, support and guidance.

-----ooOoo----

Table of Contents

-----ooOoo----

List of Tables

-----ooOoo----

List of Figures

-----ooOoo----

List of Acronyms

ARC-ISCW	Agricultural Research Council – Institute of Soil, Climate and Water
BSD	BSD licenses are a family of permissive free software licenses imposing minimal restrictions on the use and redistribution of covered software
CATDS	Centre Aval de Traitement des Données
CDF	Cumulative Distribution Function
CGIAR	Consultative Group on International Agricultural Research
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CMORPH	Climate Prediction Centre MORPHING
CPC	Climate Prediction Centre
CPU	Central Processing Unit
DAG	Directed Acyclic Graph
DEM	Digital Elevation Model
DIS	Digital Information System
DPR	Dual Precipitation Radar
DWS	Department of Water and Sanitation
EAS	European Space Agency
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA	European ReAnalysis
ERAI	ERA-Interim
ET	Evapotranspiration
FAO	Food and Agriculture Organisation of United Nations
FEWS	Famine Early Warning System
GHCN	Global Historical Climatology Network
GIS	Geographic Information System
GLC-SHARE	Global Land Cover-SHARE
GLCC	Global Land Cover Characterization
GMI	GPM Microwave Imager
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Mission
GPU	Graphics Processing Units
GSMaP	Global Satellite Mapping of Precipitation
HWSD	Harmonized World Soil Database
HydroSHEDS	Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales
IGBP	International Geosphere-Biosphere Programme
IMERG	Integrated Multi-satellitE Retrievals for GPM
IR	Infrared
LEO	Low Earth Orbiting

LSM	Land Surface Model
MAP	Mean Annual Precipitation
MIRAS	Microwave Imaging Radiometer using the Aperture Synthesis
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NetCDF	Network Common Data Form
NRL	US Naval Research Laboratory
ODE	Ordinary Differential Equation
OPeNDAP	Open-source Project for a Network Data Access Protocol
PMW	Passive Microwave
PREC-L	PRECipitation REConstruction over Land
QMORPH	Quick version of CMORPH
RFE	Rainfall Estimate Algorithm
RFI	Radio Frequency Interference
RSA	Republic of South Africa
SADC	Southern African Development Community
SAFFG	South African Flash Flood Guidance
SAHG	Satellite Applications and Hydrology Group at UKZN
SAWS	South African Weather Service
SM	Soil Moisture
SMOS	Soil Moisture and Ocean Salinity Mission
SRTM	Shuttle Radar Topography Mission
SSI	Soil Saturation Index
TMI	TRMM Microwave Imager
TMPA	TRMM Multisensor Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
TRMM3B42	TRMM Multi-Satellite Precipitation Analysis
TRMM3B42-RT	TRMM Real-Time Multi-Satellite Precipitation Analysis
UKZN	University of KwaZulu-Natal
UMD	University of Maryland
USDA	United States Department of Agriculture
USGS	United States Geological Survey
WFD	WATCH Forcing Data
WFDEI	WATCH Forcing Dataset ERA Interim
WMO	World Meteorological Organisation
WRC	Water Research Commission

-----ooOoo----

List of Symbols

ψ_b	bubbling pressure
n_0	overland surface roughness
λ	pore size distribution index
θ_r	residual soil moisture content
θ_s	saturated moisture content
K_s	soil conductivity
L	soil depth
n_o	surface roughness coefficient (Manning)
$S_f = \tan\beta$	tangent of the angle of the surface slope (Manning)

-----ooOoo----

Introduction

Timeous and routine monitoring of the spatial distribution of Soil Moisture and Evapotranspiration over a large region in fine detail has great value for coping with two weather extremes: Flash Floods and Droughts. The prevailing state of soil moisture conditions has a major impact on the runoff response of a catchment to heavy rainfall. Monitoring the wetness of the soil in detail over large regions, without having to laboriously take expensive samples, is a bonus for agricultural managers who need to understand the status of crop growth potential. This is particularly relevant in the Southern and central African countries in the SADC region which contribute importantly to the food basket of Africa.

This research project had two thrusts. The first was to extend coverage of near real-time Soil Moisture (SM) and Evapotranspiration (ET) monitoring over the SADC region using the existing methodology of operating the PyTOPKAPI hydrological model in Land Surface Modelling mode (as proven in RSA), which we called HYLARSMET in the completed WRC project K5/2024 (Sinclair and Pegram, 2013b). This outreach project EXSMET is likely to be beneficial in many ways, not only technically and socially, but also to help cement other forms of collaboration within the region. The second thrust was to develop powerful computing techniques that make it feasible to drive the modelling procedure for the vastly increased number of cells required to cover the SADC region. The same techniques can be applied to improve the spatial detail of distributed hydrological modelling at fine scale in RSA, as piloted in the WRC project K5/2024, to complement the SAFFG initiative undertaken by Mr Eugene Poolman of SAWS in WRC project K5/2068, assisted by us in Pegram & Associates.

Notwithstanding the above, the core focus of this proposal was to build on the momentum of HYLARSMET's development of PyTOPKAPI as both a distributed hydrological model (Vischel et al., 2008; Sinclair and Pegram, 2010) and a Land Surface Model (Sinclair and Pegram, 2013a), in two earlier WRC projects K5/1683 and K5/2024, completed respectively in March 2010 and March 2013. The major thrusts in the HYLARSMET project were:

(1) Search for precipitation products possibly useful for EXSMET
(2) Obtain updated soil and land use data from ARC-ISCW and WRC/UKZN;
(3) Add a surface infiltration layer to TOPKAPI to allow flash flood runoff to be modelled;
(4) Improve rainfall and ET forcing after conducting an exhaustive sensitivity analysis;
(5) Set up hydrologically consistent implementations of HYLARSMET on several SAWS SAFFG target catchments, computing soil moisture at fine scale;
(6) Maintain the HYLARSMET tools for the community which included: an open source version of the PyTOPKAPI model, bug fixes, extensions to the model, website maintenance, user interaction, and a successful workshop.

To enable the extension of PyTOPKAPI into areas outside of South Africa, it was necessary to obtain suitable soil property data-sets over the region. Comparisons of hydrological response (which includes ET and SM behaviour) to different data-sets, were made to determine how sensitive are the estimates to these differences. The comparisons were made between our RSA data-sets and the global ones we obtained. Crucial to the transfer of information to the SADC region outside RSA was an estimation of the difference in responses of ET, SM and hydrology in the two sets. In the

HYLARSMET project K5/2024, the variable to which the SM modelling was found to be the most sensitive (and over which we have some control in terms of bias adjustment) is the rainfall input. In the SADC region outside RSA, the gauge rainfall records are sparser, by an order of magnitude, than those in RSA. We had to rely on the Tropical Rainfall Measuring Mission [TRMM] to provide us with rainfall estimates. However, we found in the HYLARSMET project that TRMM is biased, depending on topography and location. To tackle the problem, a postgraduate student involved with this project, Simon Ngoepe of DWS, concentrated on TRMM bias adjustment in his MSc research – he was awarded his degree in 2017.

In summary, this EXSMET proposal focused on developing and utilising integrated hydrological approaches in surface water and Soil Moisture assessments, water resource planning and management. It took advantage of gains made in improved understanding of Soil Moisture and surface water hydrological processes through the completed project HYLARSMET, as well as the availability of better hydrological data, especially the various forms of remotely-sensed data with better coverage, like TRMM. Hydrological software that was developed in the past under HYLARSMET was adapted, re-developed and accelerated by tools that are suited to current data availability, improved knowledge and the recent technological advances in hydrological modelling using the PyTOPKAPI model. In this EXSMET project, the continued deterioration of hydrological gauging processes and other installed earth measurement devices were managed through the intensive use of new data sources from remote sensing coupled with the limited earth-based measurements. Finally, the major outcomes were that we were able to (i) model Soil Moisture over SADC in detail (approximately 49 000 isolated pixels centred in 0.125° squares) and (ii) model hydrological runoff on ungauged catchments over RSA, both with estimated levels of confidence.

We consider that it is important for the technical reader to know where to access the routines and data-sets developed in this report. These can be found in: PyTOPKAPI v0.4.0. Zenodo. http://doi.org/10.5281/zenodo.820640 (Sinclair, et al. 2017, June 28) and in https://github.com/sahg/PyTOPKAPI

A good bias correction of TRMM, and other carefully chosen remote sensing products, was of paramount importance if we were to get the hydrology right. To achieve this, we partially relied on the output of our parallel WRC project K5/2241: originally named 'Revision of the Mean Annual Precipitation (MAP) estimates over Southern Africa' (Pegram et al., 2016) dedicated to gauge rainfall interpolation and TRMM bias correction.

In the Sections which follow, the Deliverables in the project are outlined, then our interpretation and results of the research are expanded upon in detail towards the end.

-----ooOoo----

1. Summary of precipitation products possibly useful for EXSMET

This first Section describes the work done under Deliverable 1: ***'Ascertain what rainfall data are available in the SADC countries outside our borders and perform checks on their suitability for modelling'***. In Table 1.1 are the details of the 23 Products/Data that are available at monthly down to daily and sub-daily time-steps, which were considered as possible candidates for testing in EXSMET. It was decided that all data-sets must be publicly accessible (preferably via the internet without onerous administrative restrictions, e.g. licensing.)

Table 1.1: Summary of precipitation products possibly useful for EXSMET

Name	Description	Spatial res.	Temporal res.	Start date	End date	Update freq.	Latency
CHARM	African NWP estimate constrined with GHCN gauge data	0.1°	Daily	1961	1996	Unknown	NA
CHIRPS	Quasi-global gridded precip.	0.05°	6 hourly, daily, monthly	1981-01	2014-06	Pentad?	2 days?
CHOMPS	Global daily rainfall from passive microwave	0.25°	Daily	2000	2007	Unknown	NA
CHOMPS-noA	Global daily rainfall from passive microwave (no AMSU)	0.25°	Daily	1998	2007	Unknown	NA
CMORPH	Global gridded satellite precip.	8km (equator)	30 min	2002-12	Present	30 mins?	18 hours
CRU TS	Global grids of monthly mean precip	0.5°	Monthly	1901	2012	Infreq.	NA
ERA40	Model reanalysis global	40 km	Hourly?	1957-09	2002-08	Never	NA
ERAI	Model reanalysis global	11 km	Hourly?	1979	Current	12 hr	2 month
ERAI \Land	Model reanalysis global - Water balance closure	11 km	Hourly?	1979	2010	Unknown	NA
CPC FEWS RFE	Africa and selected regions gridded satellite + gauge precip	0.1°	Daily	2001-01	Current	Daily	24 hr?
GHCN	Global station precip.	Point	Daily	1850	2014	Pentad?	Pentad?
GPCC	Global precip timeseries and climatology	0.5/1.0/2.5°	Monthly	1900	2010	Infreq.	NA
GPCP daily	Global gridded precip timeseries	1.0°	Daily	1996-01	2014-02	Monthly	Monthly?
GPCP monthly	Global gridded precip	2.5°	Monthly	1979-01	2013-08		
GSMaP	Global satellite derived gridded precip	0.1°	1 hr	????	Current	Hourly	NRT
NRL Monterey	Global satellite derived gridded precip	0.1°	3 hourly	2001-04	Current	3 hourly	NRT?
PERSIANN	Global gridded precip	0.25°	3, 6 hr	2000-03	2014-02	Infreq.	NA
PREC/L	Global gridded precip	0.5/1.0/2.5°	Monthly	1948-01	2014-06	Monthly?	NA
Princeton precip. dataset	Global hydro forcing	1.0°	3 hourly	1948	2008	Infreq.	NA
QMORPH	Global gridded satellite precip	8km, 0.25°	30 mins	-30 days	Current	30 mins?	3 hours
TMPA 3B42	Global merged microwave and IR - gauge adjusted	0.25°	3 hr	2000-03	Current	Monthly	Monthly?
TMPA 3B42-RT	Global merged microwave and IR	0.25°	3 hr	2000-03	Current	3 hr	+-8 hours
UDelaware precip climatology	Global precip timeseries and climatology	0.5°	Monthly	1900	2010	Infreq.	NA

The following paragraphs list of 12 (out of the 23) possibly useful products, selected from the full set of 23 in Table 1.1, based purely on whether they are currently available: CHIRPS, CMORPH, ERAI, CPC FEWS RFE, GHCN [Global station precipitation, Point Daily 1850 to 2014], GPCP, GSMaP, NRL, PREC/L, QMORPH, TRMM 3B42 and TRMM 3B42-RT.

CHIRPS
http://chg.geog.ucsb.edu/data/chirps/

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS, Funk et al., 2014) is a 30+ year quasi-global rainfall dataset. Spanning 50°S-50°N (and all longitudes), starting in 1981 to near-present. CHIRPS incorporates 0.05° resolution satellite imagery with in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring. As of May 1st, 2014, when the project was initiated, version 1.8 of CHIRPS was complete and available to the public.

CMORPH
http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html

CMORPH – CPC MORPHing technique (Joyce et al., 2004) – produces global precipitation analyses at very high spatial and temporal resolution. This technique uses precipitation estimates that have been derived from low orbiter satellite microwave observations exclusively, and whose features are transported via spatial propagation information that is obtained entirely from geostationary

satellite IR data. We incorporated precipitation estimates derived from the passive microwaves aboard the DMSP 13, 14 & 15 (SSM/I), the NOAA-15, 16, 17 & 18 (AMSU-B), and AMSR-E and TMI aboard NASA's Aqua and TRMM spacecraft, respectively. These estimates are generated by algorithms of Ferraro (1997) for SSM/I, Ferraro et al. (2000) for AMSU-B and Kummerow et al. (2001) for TMI. Note that this technique is not a precipitation estimation algorithm but a means by which estimates from existing microwave rainfall algorithms can be combined. Therefore, this method is extremely flexible such that any precipitation estimates from any microwave satellite source can be incorporated.

With regard to spatial resolution, although the precipitation estimates are available on a grid with a spacing of 8 km (at the equator), the resolution of the individual satellite-derived estimates is coarser than that – more on the order of 12 x 15 km or so. The finer 'resolution' is obtained via interpolation. In effect, IR data are used as a means to transport the microwave-derived precipitation features during periods when microwave data are not available at a location. Propagation vector matrices are produced by computing spatial lag correlations on successive images of geostationary satellite IR which are then used to propagate the microwave derived precipitation estimates. This process governs the movement of the precipitation features only. At a given location, the shape and intensity of the precipitation features in the intervening half hour periods between microwave scans are determined by performing a time-weighting interpolation between microwave-derived features that have been propagated forward in time from the previous microwave observation and those that have been propagated backward in time from the following microwave scan. We refer to this latter step as 'morphing' of the features.

Combined precipitation estimates are available from low earth orbiting microwave satellite platforms using geostationary IR data to provide motion fields describing the movement of 57 rainfall estimates between observations. 3 hourly data at 0.25° spatial resolution is available in a custom binary file format via FTP. There is a time lag of around 18 hours. The product is also available in 8 km and 30-minute resolution.

QMORPH

http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html

Similar to CMORPH, but only uses forward advection (i.e. forecasts). Combines precipitation estimates from low earth orbiting microwave platforms using geostationary IR data to provide motion fields describing the movement of rainfall between observations. Past 24 hours available in a custom binary file format via FTP, at 0.25 and 8 km spatial resolutions and 30-minute temporal resolution.

ERAI

http://www.ecmwf.int/en/research/climate-reanalysis/era-interim

ERA-Interim (Dee et al., 2011) is a global atmospheric reanalysis from 1979, continuously updated in real time. The data assimilation system used to produce ERA-Interim is based on a 2006 release of the IFS (Cy31r2). The system includes a 4-dimensional variational analysis (4D-Var) with a 12-hour analysis window. The spatial resolution of the data set is approximately 125 km (T255 spectral) on 60 vertical levels from the surface up to 0.1 hPa.

CPC FEWS RFE

http://www.cpc.ncep.noaa.gov/products/fews/rfe.shtml

As of January 1, 2001, RFE version 2.0 has been implemented by NOAA's Climate Prediction Center. Created by Ping-Ping Xie, this replaces RFE 1.0, the previous rainfall estimation algorithm that was operational from 1995 through 2000 (Herman et al., 1997). RFE 2.0 uses additional techniques to better estimate precipitation while continuing the use of cloud top temperature and station rainfall data that formed the basis of RFE 1.0. Meteosat 7 geostationary satellite infrared data is acquired in 30-minute intervals, and areas depicting cloud top temperatures of less than 235K are used to estimate convective rainfall. WMO Global Telecommunication System (GTS) data taken from 1000 stations provide accurate rainfall totals, and are assumed to be the true rainfall near each station. RFE 1.0 used an interpolation method to combine Meteosat and GTS data for daily precipitation estimates, and warm cloud information was included to obtain dekadal (sic – i.e. 10-day total) estimates. The two new satellite rainfall estimation instruments that are incorporated into RFE 2.0 are the Special Sensor Microwave/Imager (SSM/I) on board Defense Meteorological Satellite Program satellites, and the Advanced Microwave Sounding Unit (AMSU). Both estimates are acquired at 6-hour intervals and have a resolution of 0.25 degrees. RFE 2.0 obtains the final daily rainfall estimation using a two-part merging process, then sums daily totals to produce dekadal estimates. All satellite data is first combined using a maximum likelihood estimation method, and then GTS station data is used to remove bias. Warm cloud precipitation estimates are not included in RFE 2.0.

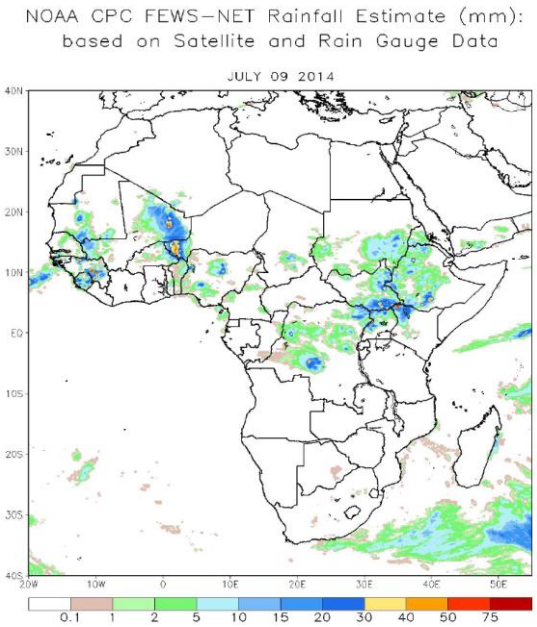


Figure 1.1: CPC FEWS daily RFE rainfall product for Africa on July 9 2014

GHCN

http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/

The Global Historical Climatology Network Daily (Menne et al., 2012; Durre et al., 2010) is an integrated database of daily climate summaries from land surface stations across the globe. Like its

monthly counterpart (GHCN-Monthly), GHCN-Daily is composed of daily climate records from numerous sources that have been integrated and subjected to a common suite of quality assurance reviews.

GHCN-Daily now contains records from over 75000 stations in 180 countries and territories. Numerous daily variables are provided, including maximum and minimum temperature, total daily precipitation, snowfall, and snow depth; however, about two thirds of 5 the stations report precipitation only. Both the record length and period of record vary by station and cover intervals ranging from less than year to more than 175 years.

The dataset is regularly reconstructed (usually every weekend) from its 20-plus data source components to ensure that GHCN-Daily is generally synchronous with its growing list of constituent sources. During this process, quality assurance checks are applied to the full dataset. On most weekdays, GHCN-Daily station data are updated when possible from a variety of data streams, which also undergo a suite of quality checks. Our interest is confined to cover SADC as shown in Figure 1.2.

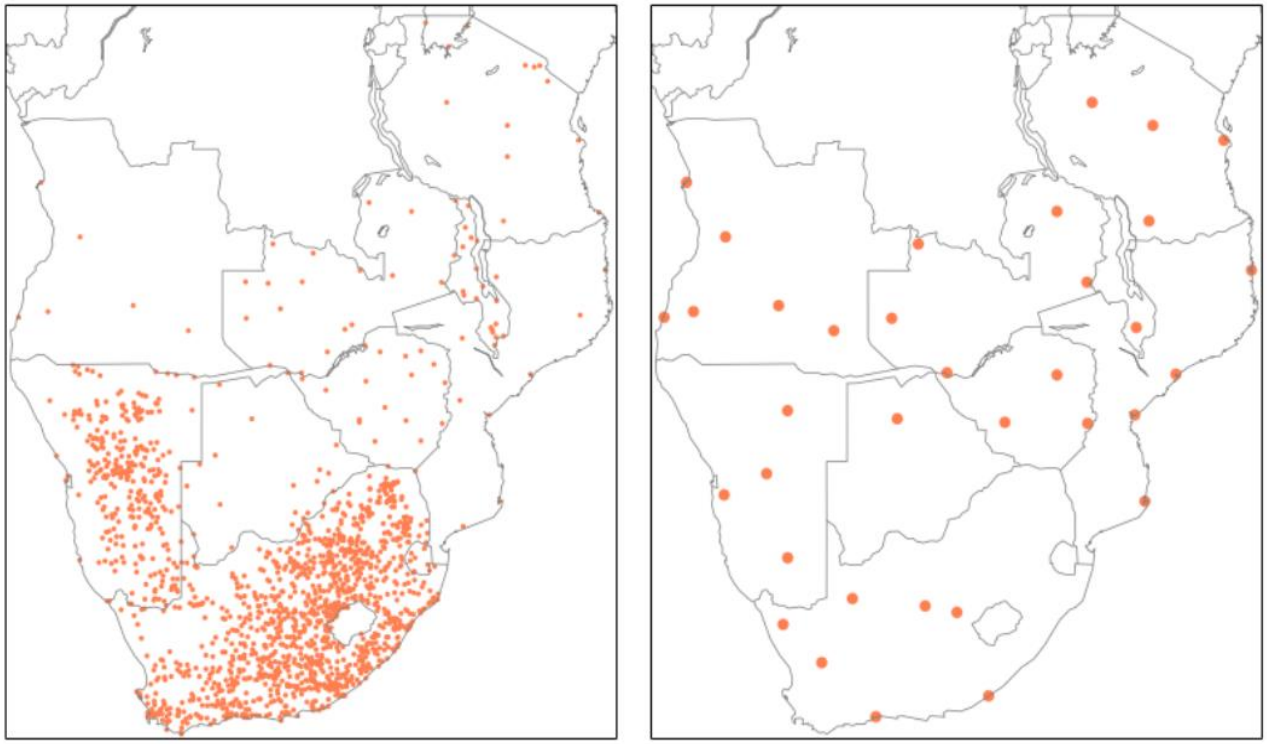


Figure 1.2: Left panel – The complete set of rain gauges available in the GHCN database for SADC countries, focused on the continental landmass. Right panel – The subset of SADC gauges that are likely to provide updated information via the WMOGTS system.

Figure 1.3 shows the availability of these data over time.

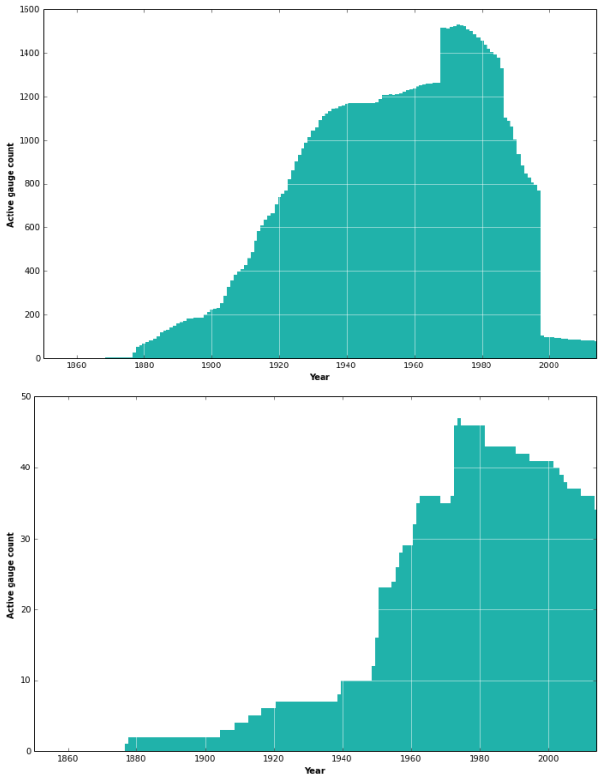


Figure 1.3: Top panel – The complete number of active SADC rain gauges available in the GHCN database. Bottom panel – The subset of SADC gauges that are likely to provide updated information via the WMO GTS system.

GPCP

http://precip.gsfc.nasa.gov/

The Global Precipitation Climatology Project (GPCP) was established by the World Climate Research Program (WCRP) and subsequently attached to the Global Energy and Water Exchange program (GEWEX) to address the problem of quantifying the distribution of precipitation around the globe over many years. The general approach is to combine the precipitation information available from each of several sources into a final merged product, taking advantage of the strengths of each data type. The passive microwave estimates are based on Special Sensor Microwave/Imager (SSMI) and Special Sensor Microwave Imager/Sounder (SSMIS) data from the series of Defense Meteorological Satellite Program (DMSP, United States) satellites that fly in sun-synchronous low-earth orbits at 6 a.m./p.m. The infrared (IR) precipitation estimates are computed primarily from geostationary satellites (United States, Europe, Japan), and secondarily from NOAA-series polar-orbiting satellites (United States). Additional low-Earth orbit estimates include Atmospheric Infrared Sounder (AIRS data from the NASA Aqua), and Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder (TOVS) and Outgoing Longwave Radiation (OLR) Precipitation Index (OPI) data from the NOAA series satellites. The precipitation gauge data are assembled and analysed by the Global Precipitation Climatology Centre (GPCC) of the Deutscher Wetterdienst.

The GPCP has promoted the development of an analysis procedure for blending the various estimates together to produce the necessary global gridded precipitation fields. The currently operational procedure is described in Huffman et al. (2001), Adler et al. (2003) and Huffman et al.

(2009), and has been used to produce the GPCP Version 2.2 Combined Precipitation Data Set, covering the period January 1979 through the present (with some delay). The primary product in the Version 2.2 dataset is a combined observation-only dataset, that is, a gridded analysis based on gauge measurements and satellite estimates of precipitation. There is a total of 27 fields in the data set providing information from the individual and intermediate estimates, including estimates of RMS random error.

GSMaP
http://sharaku.eorc.jaxa.jp/GSMaP/

The Global Satellite Mapping of Precipitation algorithm is a combination of the CMORPH technique and Kalman filter (Kubota et al., 2007). The IR data are used as a means to move the precipitation estimates from microwave observation during periods when microwave data are not available at a location in this study. The microwave sensors which we use are TRMM/TMI, Aqua/AMSR-E, and DMSP/SSMI (F13, 14, 15) for the GSMaP-MVK product; in addition to these, AMSU-B's are included in the GSMaPMVK+ product. The technique to have high resolution global precipitation map uses the Kalman filter to compute the estimates of the current surface rainfall rates at each 0.1° pixel of the infrared brightness temperature by the GEO-IR satellites. The filter predicts the precipitation rate from the microwave radiometer and its propagated product based on the IR data, and then refines the prediction based on the relationship between the IR brightness temperature and surface rainfall rate. The rain rates from the passive microwave radiometer are generated as suggested by Aonashi and Liu (2000).

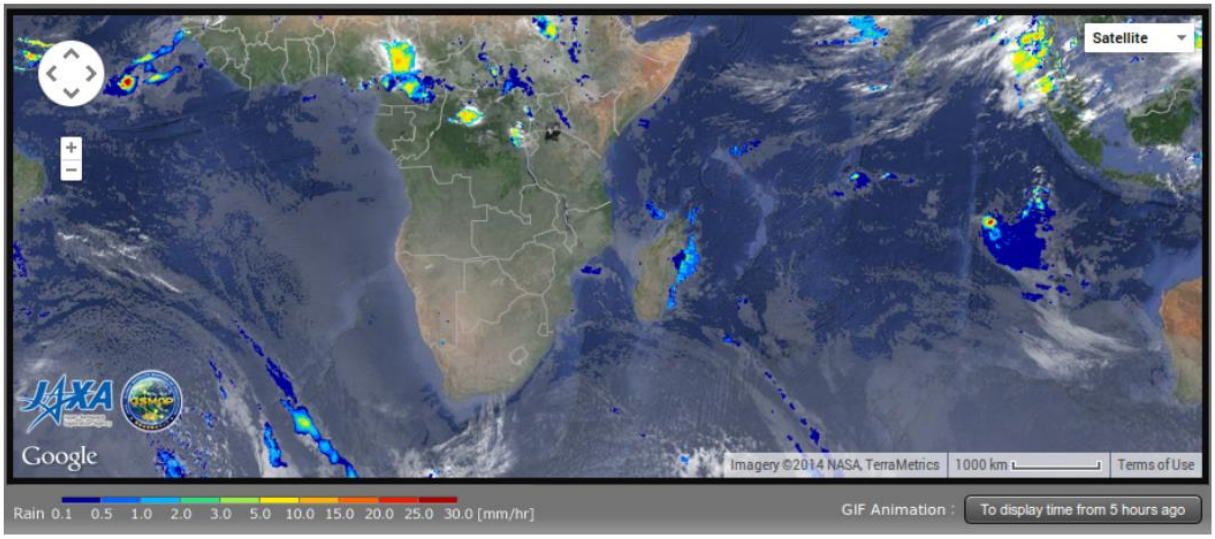


Figure 1.4: GSMaP rainfall product for Africa.

The blended satellite IR and passive Microwave rainfall estimates, provide both instantaneous and accumulated rainfall at 0.25° spatial resolution and 6-hourly time-steps. It's not clear from the website how one can access the data, nor what the details of the algorithm are (there is no reference to published work, but the contact person appears to be Joe Turk who we know and have met, and who is well published and respected in the field of satellite rainfall estimation).

NRL Monterey
http://www.nrlmry.navy.mil/sat-bin/rain.cgi

The US Naval Research Laboratory (NRL) blended satellite technique is based upon area-dependent statistical relationships derived from a precise, near real-time ensemble of collocated passive microwave (PMW) and infrared (IR) pixels from any or all low Earth-orbiting (LEO) and geostationary satellites, respectively, as their individual orbits and sensor scan patterns continuously intersect in space and observation time.

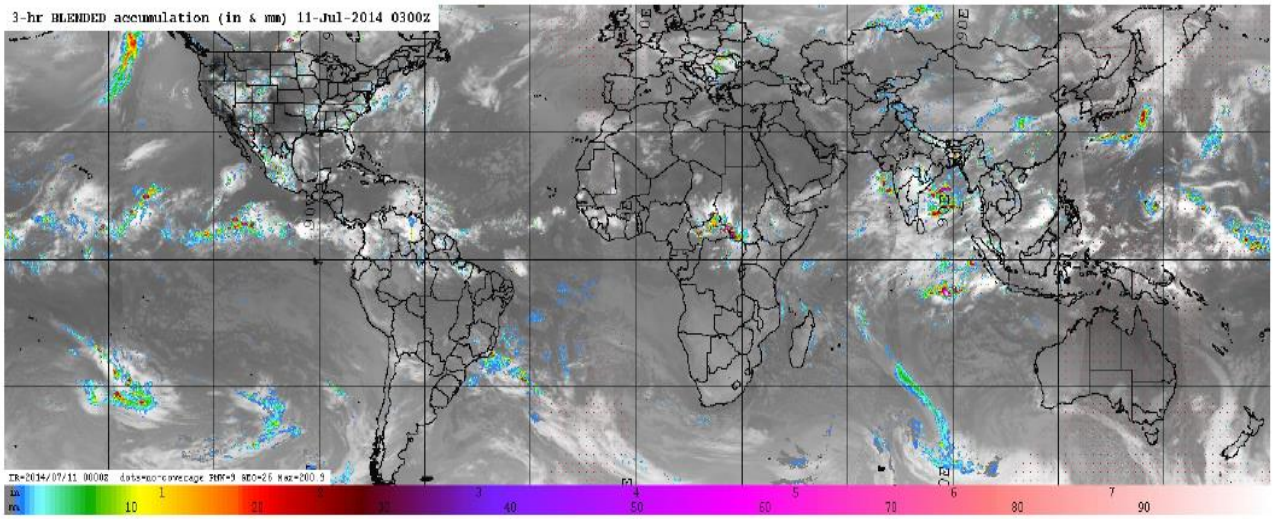


Figure 1.5: NRL Monterey 3 hour blended global rainfall rate product.

PREC-L
http://www.esrl.noaa.gov/psd/data/gridded/data.precl.html

PREC-L is NOAA's PRECipitation REConstruction over Land (Chen et al., 2002) product. The monthly data set consists of three files containing monthly averaged precipitation totals. Precipitation is available at 3 spatial resolutions. The global analyses are defined by interpolation of gauge observations over land. Gauge observations are from over 17 000 stations collected in the Global Historical Climatology Network (GHCN) version 2 (see above) and the Climate Anomaly Monitoring System (CAMS) datasets. An Optimal Interpolation technique was used to assimilate the observations. The mean distribution and annual cycle of precipitation observed in the PREC/L showed good agreement with those in several published gauge-based datasets, and the anomaly patterns associated with ENSO resemble those found in previous studies. All maps are stored using a landmask [MATLAB term] as smoothing puts values well over the oceans. We stored all values where the values were below 0. One may decide to further mask the data using their land file and a different cutoff.

TRMM 3B42
http://daac.gsfc.nasa.gov/precipitation/TRMM_README/TRMM_3B42_readme.shtml

TRMM Multi-Satellite Precipitation Analysis (Huffman et al., 2007), combined microwave and microwave/gauge calibrated IR estimates of precipitation rate. Spatial resolution is 0.25° and temporal resolution is 3 hourly. Estimates are post-processed on a Monthly basis, with the previous month's data typically available within a few days of month end. The data are available in HDF format and can be ordered via the TRMM Data Search and Order System, this process requires manual intervention, but could be partially automated.

TRMM 3B42-RT
ftp://trmmopen.gsfc.nasa.gov/pub/merged

TRMM Real-Time Multi-Satellite Precipitation Analysis (Huffman et al., 2007), combined microwave and microwave calibrated IR estimates of precipitation rate. Spatial resolution is 0.25° and the temporal resolution is 3 hourly. Estimates are available within 6-7 hours of observation times and can be downloaded via FTP in a custom binary file format.

1.1 Summary of decisions

At this early stage of the project we examined the following products for appropriateness in our study:

CHIRPS,
CMORPH,
ERA-I,
CPC FEWS RFE,
GHCN [Global station precip. Point Daily 1850 to 2014],
GPCP,
GSMaP,
NRL,
PREC/L,
QMORPH,
TRMM 3B42,
TRMM 3B42-RT

These were chosen because they are historically up-to-date and have promise of persevering into the future. This is important if we are to perform consistently reliable downscaling estimates of ground-level rainfall at the daily and monthly scale, as measured by gauges, as noisy as these may be. Thus, in the spirit of the description of this first deliverable, we ascertained what rainfall data are available in the SADC countries outside our borders and performed checks on their suitability for modelling, in particular for downscaling the appropriate remote sensing products. The difficult part now turned to how to determine the best way to export the downscaling techniques that we developed over data-rich RSA, to the data-poor countries north of us.

-----ooOoo----

2. The FAO maps used to determine parameters for PyTOPKAPI modelling

In order to extend the current HYLARSMET implementation in South Africa into the rest of SADC north of our borders, it was necessary to obtain suitable estimates for the PyTOPKAPI static parameters in these regions. The 2nd Deliverable was to: ***'Obtain suitable ground cover and soil maps over the whole SADC region (e.g. FAO and others), for comparison with those already available in RSA'.*** Therefore, in this Section we summarize our investigations into suitable alternatives to the detailed information available in South Africa. In subsequent Subsections we perform an extensive investigation of the differences between the South African and regional data sets.

In Subsection 2.1, we introduce the static parameters and techniques used to estimate them in South Africa. In Subsection 2.2 we mention a few alternatives to the Digital Elevation Model (DEM) used in our South African work. Subsection 2.3 summarises a variety of options for obtaining land cover information in SADC countries, while soil properties are reviewed in Subsection 2.4. Interim conclusions are presented in Subsection 2.5.

2.1 Estimation of the HYLARSMET parameters

Each PyTOPKAPI model cell in the HYLARSMET product (Pegram et al., 2010; Sinclair and Pegram, 2013b) is 1 km square in plan and is as deep as the active soil layer (approximately 1 m in South Africa, depending on location). The forcing variables driving the cell water content are rainfall and evapotranspiration, the latter depending on the soil water content of the cell as well as atmospheric variables. The parameters required to model the flow of water into and out of the cell are physically based properties:

- Soil store: local slope of cell, depth, saturated soil moisture, residual soil moisture and soil conductivity
- Overland store: local slope of cell and surface roughness.
- Infiltration layer: bubbling pressure and pore size distribution index, both parameters of the Green-Ampt (1911) Infiltration model.

2.1.1 Local slopes

The local slopes for each pixel were derived from CGIAR's SRTM DEM product (Jarvis et al., 2008) using raster analysis tools in GRASS GIS.

2.1.2 Overland surface roughness

In the concurrent implementation of HYLARSMET, we derive overland surface roughness (n_0) from GLCC v1.2 (see Figure 2.1) using an interpretation of the classes in line with table 12.2.1 of Maidment (1993), see table 2.1 in this Section.

2.1.3 Soil depth

For HYLARSMET in South Africa, we derived soil depth (L) maps by resampling and projecting the soil depths reported in Schulze et al. (2008).

2.1.4 Saturated soil moisture content

For HYLARSMET in South Africa, we derived saturated moisture content (θ_s) maps by resampling and projecting the total porosity maps reported in Schulze et al. (2008).

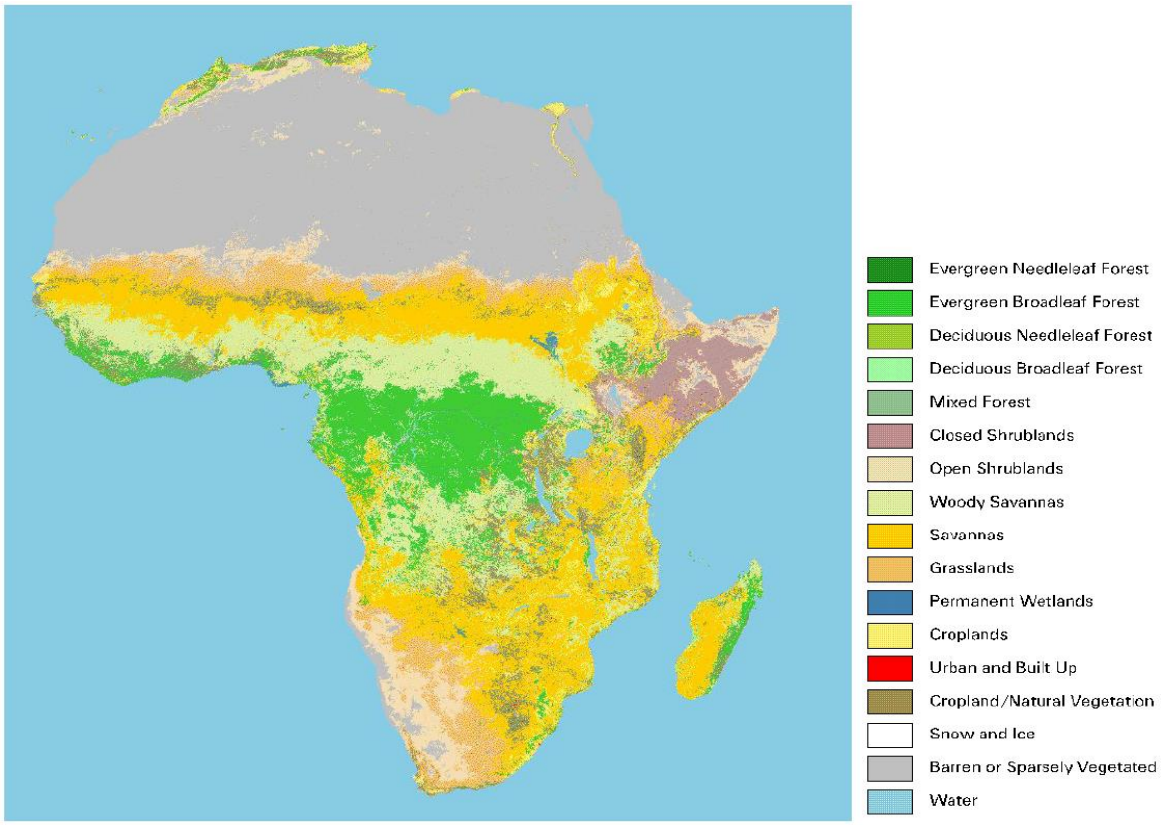


Figure 2.1: USGS GLCC v1.2 land cover product for the African continent

2.1.5 Residual soil moisture content

We obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The residual soil moisture content (θ_r) values are then the mean values for the relevant texture classes reported in table 5.3.2 of (Maidment, 1993) and in cases where more than one soil texture class is represented, an average has been taken.

2.1.6 Saturated hydraulic conductivity

We obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The soil conductivity (K_s) values were then estimated using table 5.5.5 of (Maidment, 1993) and in cases where more than one soil texture class is represented, the average of the values reported for each class is used.

2.1.7 Bubbling pressure

We obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The bubbling pressure (ψ_b) values were then estimated using table 2 of Rawls et al. (1982) and in

cases where more than one soil texture class is represented, the average of the values reported for each class is used.

Table 2.1: Lookup table of Manning roughness coefficients for USGS GLCC v1.2 land cover classes. Note that coefficients have only been assigned for classes that appear in South Africa, therefore some additions were required for application in SADC. Missing values are represented as -999.

GLCC class	Roughness co-efficient (n_o)
1	0.025
2	0.035
3	-999
4	-999
5	0.040
6	0.070
7	0.050
8	0.100
9	-999
10	0.070
11	0.100
12	-999
13	0.100
14	-999
15	0.100
16	-999
17	-999
18	-999
19	0.035
20	-999
21	-999
22	-999
23	-999
24	-999
-999	-999

2.1.8 Pore size distribution index

We obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The pore size distribution index (λ) values were then estimated using table 2 of Rawls et al. (1982) and in cases where more than one soil texture class is represented, the average of the values reported for each class is used.

Summary of 2.1

The local slopes for each cell in the SADC region can be calculated from the SRTM DEM as we did for South Africa. Possible alternative elevation models are listed in Subsection 2.2. Values of n_o for SADC can be derived from GLCC v1.2 exactly as they have been for the South African implementation of HYLARSMET. However, there are some more modern alternatives that were investigated for possible improvements (these are introduced in Subsection 2.3).

Since L and θ_s were derived from Schulze et al. (2008), we needed to use different data sets to calculate these parameters in the rest of the SADC region. L was the most challenging parameter, as there is not much information outside RSA borders. On the other hand, an estimate of θ_s can be derived from soil texture information and the tables in Rawls et al. (1982) and Maidment (1993) – see Figure 2.2. These two parameters are of particular importance, as the PyTOPKAPI model is most sensitive to errors in these parameters (see Sinclair and Pegram, 2013a).

The remaining parameters (θ_r, K_s, ψ_b and λ) can be estimated from soil texture information available in Figure 2.2, the result is that our focus was to identify sources of soil depth and texture information as a priority (the possibilities that we were aware of are presented in Subsection 2.4). In addition, we endeavoured to choose the 'best' data sets for all other parameters where there are several alternatives to choose from.

TABLE 2. HYDROLOGIC SOIL PROPERTIES CLASSIFIED BY SOIL TEXTURE

Texture class	Sample size	Total porosity (θ), cm³/cm³	Residual saturation (θ_r) cm³/cm³	Effective porosity (θ_e), cm³/cm³	Bubbling pressure (ψ b) Arithmetic, cm	Bubbling pressure (ψ b) Geometric,† cm	Pore size distribution (λ) Arithmetic	Pore size distribution (λ) Geometric†	Water retained at −0.33 bar tension, cm³/cm³	Water retained at −15 bar tension, cm³/cm³	Saturated Hydraulic Conductivity‡ (K_s) cm/h
Sand	762	0.437** (0.374–0.500)	0.020 (0.001–0.039)	0.417 (0.354–0.480)	15.98 (0.24–31.72)	7.26 (1.36–38.74)	0.694 (0.298–1.090)	0.592 (0.334–1.051)	0.091 (0.018–0.164)	0.033 (0.007–0.059)	21.00
Loamy sand	338	0.437 (0.368–0.506)	0.035 (0.003–0.067)	0.401 (0.329–0.473)	20.58 (0.0–45.20)	8.69 (1.80–41.85)	0.553 (0.234–0.872)	0.474 (0.271–0.827)	0.125 (0.060–0.190)	0.055 (0.019–0.091)	6.11
Sandy loam	666	0.453 (0.351–0.555)	0.041 (0.0–0.106)	0.412 (0.283–0.541)	30.20 (0.0–64.01)	14.66 (3.45–62.24)	0.378 (0.140–0.616)	0.322 (0.186–0.558)	0.207 (0.126–0.288)	0.095 (0.031–0.159)	2.59
Loam	383	0.463 (0.375–0.551)	0.027 (0.0–0.074)	0.434 (0.334–0.534)	40.12 (0.0–100.3)	11.15 (1.63–76.40)	0.252 (0.086–0.418)	0.220 (0.137–0.355)	0.270 (0.195–0.345)	0.117 (0.069–0.165)	1.32
Silt loam	1206	0.501 (0.420–0.582)	0.015 (0.0–0.058)	0.486 (0.394–0.578)	50.87 (0.0–109.4)	20.76 (3.58–120.4)	0.234 (0.105–0.363)	0.211 (0.136–0.326)	0.330 (0.258–0.402)	0.133 (0.078–0.188)	0.68
Sandy clay loam	498	0.398 (0.332–0.464)	0.068 (0.0–0.137)	0.330 (0.235–0.425)	59.41 (0.0–123.4)	28.08 (5.57–141.5)	0.319 (0.079–0.559)	0.250 (0.125–0.502)	0.255 (0.186–0.324)	0.148 (0.085–0.211)	0.43
Clay loam	366	0.464 (0.409–0.519)	0.075 (0.0–0.174)	0.390 (0.279–0.501)	56.43 (0.0–124.3)	25.89 (5.80–115.7)	0.242 (0.070–0.414)	0.194 (0.100–0.377)	0.318 (0.250–0.386)	0.197 (0.115–0.279)	0.23
Silty clay loam	689	0.471 (0.418–0.524)	0.040 (0.0–0.118)	0.432 (0.347–0.517)	70.33 (0.0–143.9)	32.56 (6.68–158.7)	0.177 (0.039–0.315)	0.151 (0.090–0.253)	0.366 (0.304–0.428)	0.208 (0.138–0.278)	0.15
Sandy clay	45	0.430 (0.370–0.490)	0.109 (0.0–0.205)	0.321 (0.207–0.435)	79.48 (0.0–179.1)	29.17 (4.96–171.6)	0.223 (0.048–0.398)	0.168 (0.078–0.364)	0.339 (0.245–0.433)	0.239 (0.162–0.316)	0.12
Silty clay	127	0.479 (0.425–0.533)	0.056 (0.0–0.136)	0.423 (0.334–0.512)	76.54 (0.0–159.6)	34.19 (7.04–166.2)	0.150 (0.040–0.260)	0.127 (0.074–0.219)	0.387 (0.332–0.442)	0.250 (0.193–0.307)	0.09
Clay	291	0.475 (0.427–0.523)	0.090 (0.0–0.195)	0.385 (0.269–0.501)	85.60 (0.0–176.1)	37.30 (7.43–187.2)	0.165 (0.037–0.293)	0.131 (0.068–0.253)	0.396 (0.326–0.466)	0.272 (0.208–0.336)	0.06

* First line is the mean value
Second line is ± one standard deviation about the mean
† Antilog of the log mean
‡ Obtained from Fig. 2

Figure 2.2: The table from Rawls et al. (1982) used to estimate soil properties from texture information.

2.2 Digital Elevation Models – Alternatives to CGIAR SRTM

Besides the SRTM DEM of Jarvis et al. (2008), in this Section we provide a list of other potential sources for elevation models:

- HydroSHEDS – see USGS https://hydrosheds.cr.usgs.gov/data.php
- opentopography.org
- African GIS database (B_odis, 2009)
- GeoNet and PyGeoNet – https://sites.google.com/site/geonethome/home -

From these sources, we can extract characteristics from high resolution topography data. We investigate these and others when we determine the best way to compare FAO data-sets with our RSA sets of ground-based data, in Section 4.

2.3 Land Cover – Alternatives to GLCC v1.2

There is a fairly large variety of different land cover classifications which could be used as alternatives to the GLCC v1.2, and several of these are listed in this Subsection. In general, these classification schemes are quite different from each other (e.g. see Figure 2.3). However, we showed in Sinclair and Pegram (2013a) that HYLARSMET is fairly insensitive to n_o, so we did not spend a great deal of time trying to make the 'best' selection amongst the options. In any event, we elected to keep using GLCC v1.2 as a matter of convenience.

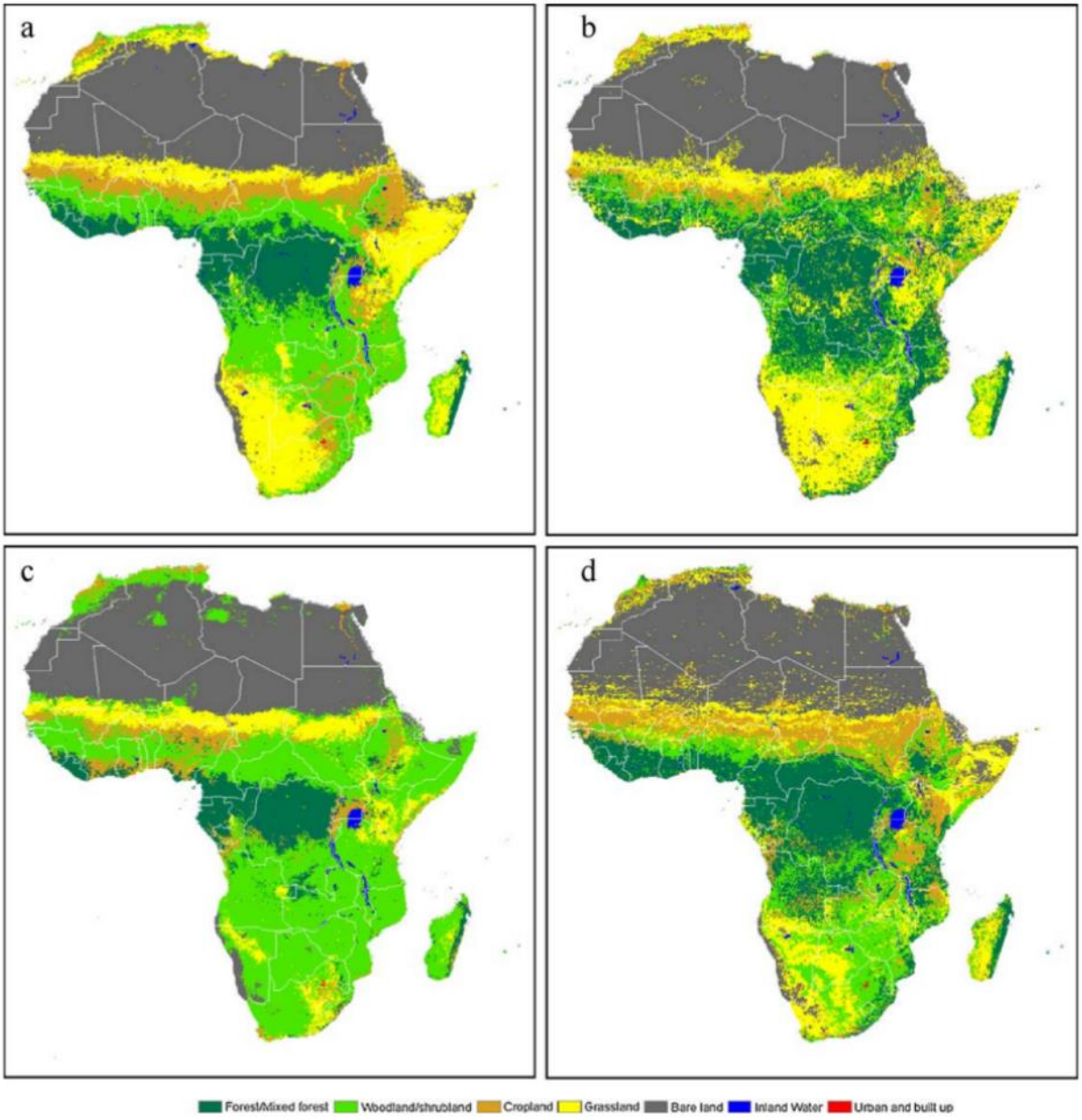


Figure 2.3: Comparison of (a) GLC2000, (b) GLOBCOVER, (c) MODIS and (d) ECOCLIMAP-II in Africa (after Tchuenté et al., 2011). Also compare with GLCC in Figure 2.1.

IGBP DIS

The International Geosphere Biosphere Program global land cover of Loveland et al. (2000). This is the alternative/new name for GLCC v1.2. More detail is available at: http://edc2.usgs.gov/glcc/glcc.php.

UMD

The University of Maryland global land cover of Hansen et al. (2000). This is a 1 km resolution global data set based on data from the Advanced Very High-Resolution Radiometer (AVHRR) on board the MODIS satellite. The approach used to create the set, is an unsupervised classification using decision trees.

GLC 2000

The Land Cover Classification System/GLC 2000 Bartholomé and Belward (2005). A global land cover map based on observations from the VEGETATION sensor on-board SPOT 4. The spatial resolution is 1 km and the classification has been regionally tuned by experts.

MODIS LC 1

The MODIS land cover type map of Friedl et al. (2002). This is the official MODIS product, also at 1 km resolution, and based on AVHRR. However, the methodology employed is a supervised classification scheme, based on a database of training sites, to guide an unsupervised procedure.

ECOCLIMAP 1 & 2

The CNRM ECOCLIMAP of Masson et al. (2003). This product provides initialization parameters for SVAT land surface models based on the tiling approach. It is also based on AVHRR at 1 km resolution, but provides many more variables in addition to land cover classes.

GLOBCOVER

Global land cover maps derived from the MERIS instrument on board ENVISAT (Defourny et al., 2006). The GLOBCOVER maps are designed to be frequently updated to allow for land cover/use change studies.

Global Land Cover-SHARE (GLC-SHARE)

http://www.glcn.org/databases/lc_glcshare_en.jsp –

'A new FAO land cover product: the Global Land Cover-SHARE (GLC-SHARE) has been just launched. It brings global land cover data under one roof for the first time and represents the most-reliable global view of planetary land cover assembled to-date.' See Figure 2.4 following:

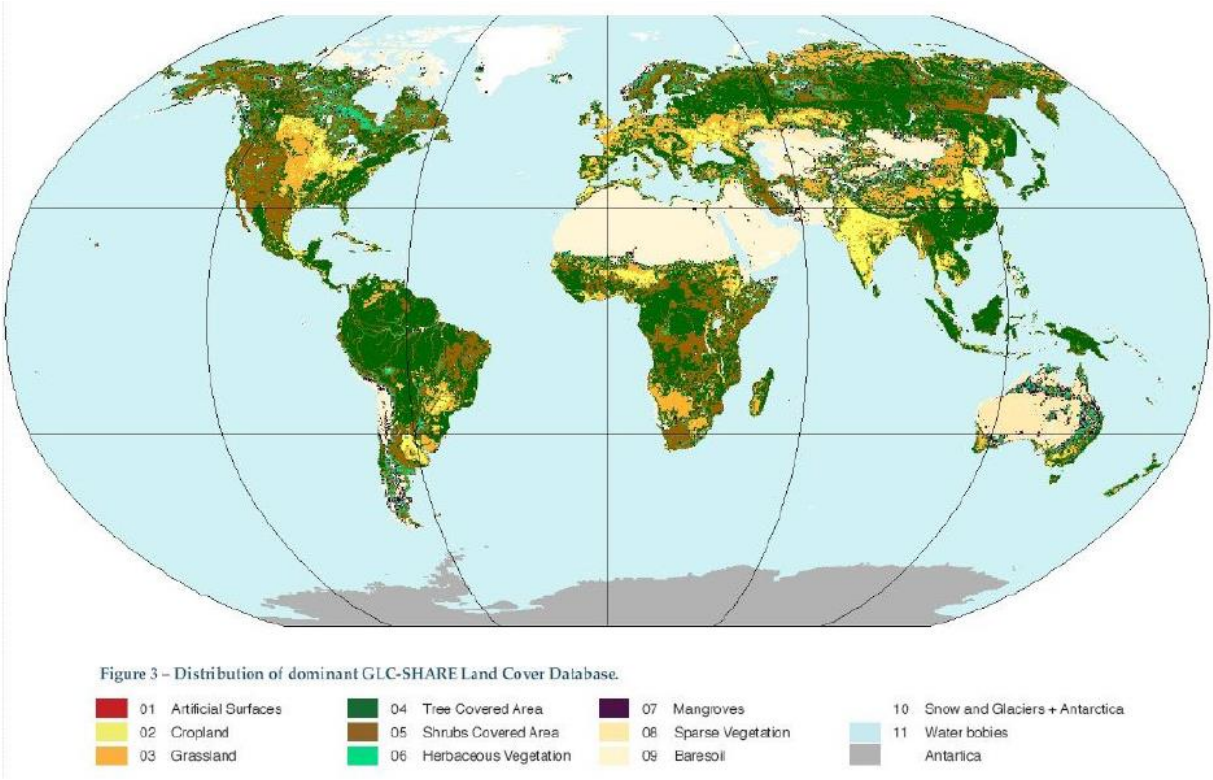


Figure 2.4: GLC-SHARE global land cover database

2.4 HWSD: Soil properties – Alternatives to WR2005 & Agro-Atlas

Harmonized World Soil Database: http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/

> 'The HWSD is composed of a raster image file and a linked attribute database. The raster database consists of 21600 rows and 43200 columns, of which 221 million grid cells cover the globe's land territory. Each grid cell in the database is linked to commonly used soil parameters, namely, organic carbon, pH, water storage capacity, soil depth, cation exchange capacity of the soil and the clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class, and granulometry. HWSD allows soil compositions to be displayed or queried in terms of user-selected soil parameters.'

This set was used in the development of the 'African GIS database' (Bódis, 2009) produced to allow the LISFLOOD model (Burek et al., 2013) to operate in the African continent. We rely heavily on the HWSD in order to derive the required soil parameters for SADC (despite some known issues with the data set).

Soil Atlas of Africa

The HWSD has a variety of discrepancies that have recently been corrected for the African continent during the production of the Soil Atlas of Africa (see Dewitte et al. (2013), selected paragraphs quoted verbatim below; see Figure 2.5 following).

'The new soil map of Africa represents an important contribution to the future sustainable use of soil resources of the continent. Together with the Soil Atlas of Africa it will raise awareness about the importance of soils in the support of an increasing population and threatened environment. The soil map and associated database also have the potential to enhance global studies on climate change, food production and land degradation for example. The explanation of the decisions that were made to produce the map will be useful to others who are attempting to harmonise legacy soil data sources to provide a usable information base.

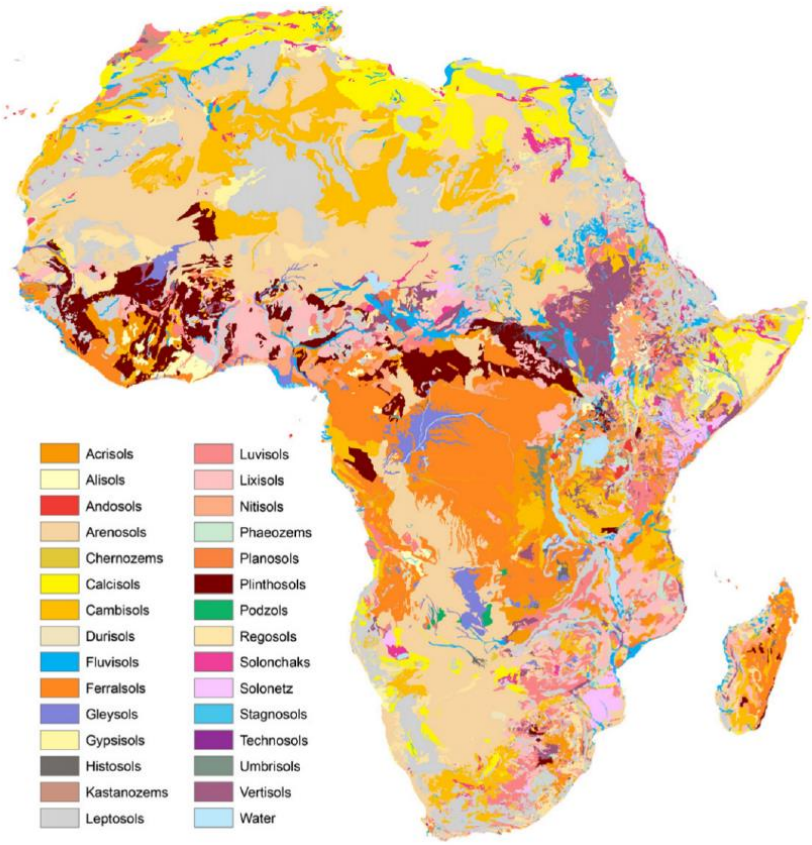


Figure 2.5: Harmonised soil map of Africa (after Dewitte et al., 2013)

'The Soil Atlas of Africa Project utilised the large body of legacy soil information for Africa collected over the last 60 years. The resulting harmonised soil map and database demonstrate the value of applying modern spatial analytical techniques to historic soil data to produce what is undoubtedly the best current soil information base for the African continent. Initially it is expected to satisfy the soaring demand for up-to-date and relevant soil data at international level in addition to the Africa Soil Information Service (AfSIS), which constitutes the African part http://www.africasoils.net of the GlobalSoilMap.net project (Sanchez et al., 2009). However, the resulting map highlights the need for applying new mapping techniques and collecting new data in Africa to meet 21st century soil information needs.

'The new map is at the heart of the Soil Atlas of Africa (Jones et al., 2013), displayed in a series of map sheets at the scale 1:3 M, constituting some forty per cent of the Atlas pages. The published Atlas, the Soil Map of Africa that it contains and the corresponding datasets (modified map and associated modified HWSD) are available for downloading free of charge

from the portals of the European Commission Joint Research Centre SOIL Action http://eusoils.jrc.ec.europa.eu/.'

The African soil atlas does not yet provide information on soil properties – only a description of soil types (see Figure 2.4). Unfortunately, the planned updates of 2016 were not released in time for use in this project.

SoilGrids1km Global Soil Information Based on Automated Mapping

SoilGrids1km is an interpolated set of soil parameters based on a large number of soil samples (shown in Figure 2.6), which have been interpolated using several regression covariates – including the HWSD (Hengl et al., 2014). This is an interesting project and a potentially useful data source for soil parameters in the SADC region.

The International Satellite Land Surface Climatology Project, Initiative II

http://daac.ornl.gov/ISLSCP_II/islscpii.shtml

'The International Satellite Land Surface Climatology Project, Initiative II (ISLSCP II) Project was part of the Global Energy and Water Experiment (GEWEX) and was responsible for addressing land- atmosphere interactions, process modelling, data retrieval algorithms, field experiment design and execution, and the development of global data sets. The ISLSCP II data set collection contains about 50 comprehensive data sets over the 10-year period from 1986 through 1995 focused on land cover, hydrometeorology, radiation, and soils. The ISLSCP II data were acquired from a number of U.S. and international agencies, universities, and institutions, then co-registered to equal-angle grids of one, one-half, and one-quarter degree resolution and reformatted into a common ASCII format.

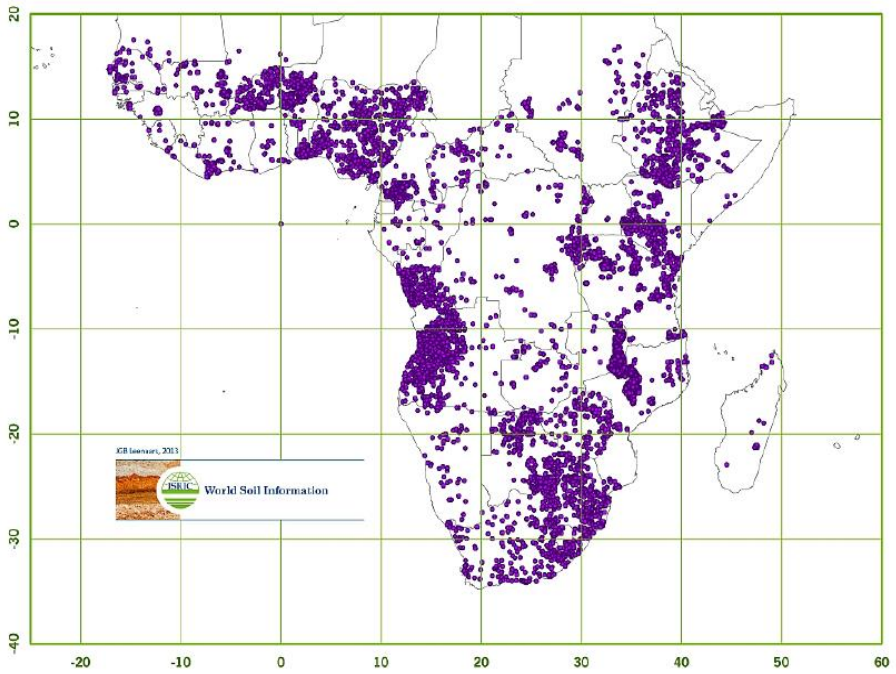


Figure 2.6: SoilGrid1km profile locations for Africa (after Hengl et al., 2014)

The ORNL DAAC ISLSCP II Data archive includes data products from the following categories: Ancillary data; Carbon; Hydrology and Soils; Near Surface Meteorology; Radiation and Clouds; Snow, Sea Ice, and Oceans; Socioeconomics; Vegetation'

GlobalSoilMap.net

http://globalsoilmap.net/

> 'There is a need for accurate, up-to-date and spatially referenced soil information. This need has been expressed by the modelling community, land managers, policy developers and decision makers. The need coincides with an enormous leap in technologies that allow for improvements in accurately collecting and predicting soil properties.
>
> A global consortium has been formed to make a new digital soil map of the world using state-of-the-art and emerging technologies. This new global soil map will predict soil properties at fine spatial resolution (100 m). These maps will be supplemented by interpretation and functionality options to support improved decisions for a range of global issues such as food production and hunger eradication, climate change, and environmental degradation. This is an initiative of the Digital Soil Mapping Working Group of the International Union of Soil Sciences IUSS.
>
> The project was officially launched on 17th February 2009, New York, USA.'

Unfortunately, there is no public activity on the website after early 2012, and we were not able to source any datasets.

2.5 Summary of Section 2

We conducted a review of both global and Africa focused data sets which allowed us to assemble the basic building blocks for an implementation of HYLARSMET in the SADC region.

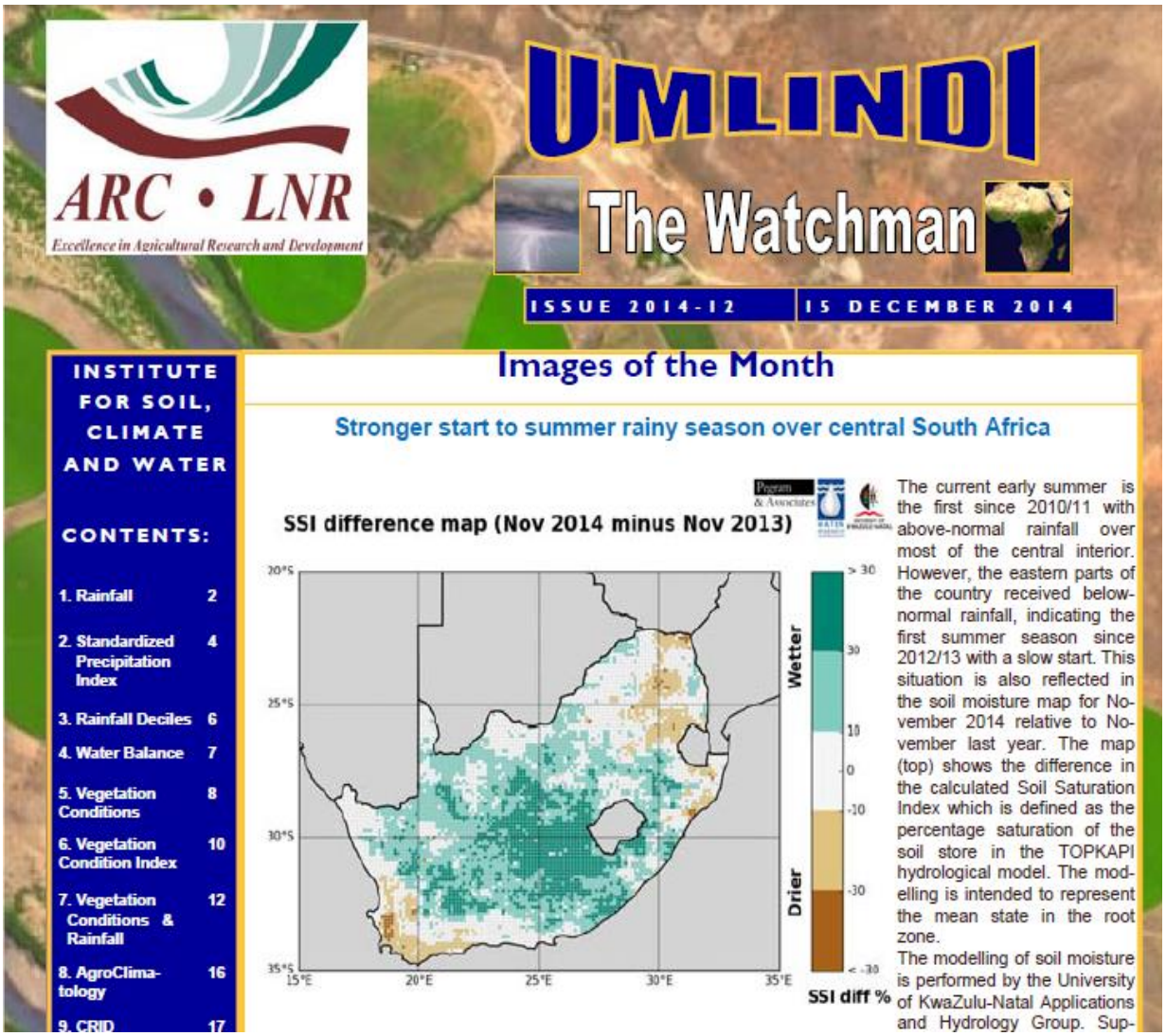
The slope and overland roughness parameters are derived using the same data sets as we have used in the South African implementation (SRTM Jarvis et al., 2008; GLCC v1.2). Of necessity, the soil physical parameters were derived from alternative data sets, but these are freely available and used the same techniques as we did for South Africa (primarily based on lookup from soil texture classes). The main difficulties were in obtaining soil depth information; we were reduced to using the (mostly) uniform soil depths presented in the HWSD.

As a result of this review, we were confident that we could proceed to the next stage of the project and carry out a comparative analysis to assess the effect of using the alternative parameter sets in modelling soil moisture using the HYLARSMET system.

-----ooOoo----

3. The UMLINDI newsletter highlighted the usefulness of the RSA product

The following image is a clip from the front page of [at the time] the most recent copy of Umlindi, the ARC-ISCW newsletter on monthly climate variables. We were delighted to have been given pride of place on the first page of the new issue in December 2014.



The text in the right column is legible. The second paragraph follows, highlighting the modelling procedure developed by Pegram and Associates. Note that SSI stands for Soil Saturation Index.

> 'The modelling of soil moisture is performed by the University of KwaZulu-Natal Satellite Applications and Hydrology Group. Supported by the WMO, the system and algorithms developed by the UKZN have been replicated at the ARC-ISCW, where the developing archive will be utilized in the expansion of the suite of drought monitoring products provided in near-real time. The SSI maps will be published in the newsletter in future. The wet conditions reflected over the central parts started by the end of October, with several rainfall events resulting in extremely wet conditions for the month over parts of the central interior and the far western areas.'

It bears expanding on this short description to put it in context. In the contract proposal, we made the following statement – please note in particular the highlighted text.

> ' ... the World Meteorological Organisation (WMO) held a workshop at ARCISCW in Pretoria on 29-30 May 2012 entitled 'STAKEHOLDER MEETING ON AGROMET AND SOIL MOISTURE APPLICATIONS FOR SOUTH AFRICA', adding considerably to the value of this initiative. It was led by Robert Stefanski, Chief, Agricultural Meteorology Division, WMO. Attendees beside members of ARC-ISCW, WMO and Pegram and Associates, included Prof Roland Schulze of UKZN, Mr Eugene Poolman of SAWS and Dr Stephan Steyn of UFS. Why this information is germane to this proposal, is that WMO want to perform validation (if possible) and model inter-comparison (certainly) of a high-resolution soil moisture product based on MODIS, developed by Prof. John Qu and his team at the George Mason University in the United States. The product will be distributed via the WMO's WAMIS (World Agro-Meteorological Information Service) and they are pursing possibilities for validation/intercomparison over SADC countries including RSA. We in Pegram and Associates and ARC-ISCW suggested that PyTOPKAPI is the correct tool to perform the model inter-comparison. This was endorsed by those present and it was agreed by the workshop attendees that a working copy of the HYLARSMET system should be installed and housed at ARC and that in addition, WMO would financially support Mr J Malherbe (ARC-ISCW) and Dr S Sinclair (P&A) to achieve the transfer of the HYLARSMET model code. They will take this opportunity to produce a user manual for installing and using the HYLARSMET system, in order to make it more transportable and 'friendly'. The transfer will provide valuable information on challenges and pitfalls that will greatly benefit this proposed EXSMET project. This means that the HYLARSMET system will be deployed to perform routine SM monitoring for agricultural planning and drought monitoring. Incidentally, the attendees asked for a renewed effort for deployment of Soil Moisture probes and both SAWS and ARC-ISCW agreed to move ahead with that, which will help with validation of the modelled estimates.'

During August, September, October and November of 2014 Dr Sinclair made 4 trips from Durban to Pretoria, of from 2 to 4 days long, to install this code and make it run 'on foreign soil' as it were. We are immensely proud of this achievement and grateful to the WMO for support and to ARC (in particular, Dr Johan Malherbe) for making the successful transplant possible.

It is still running today in 2017.

-----ooOoo----

4. Gathering appropriate FAO data-sets guided by our RSA ground-based data

Deliverable 4 was described as: ***'Determine the best way to compare FAO data-sets with our RSA sets of ground-based data'.*** We prepared the way for applying HYLARSMET (WRC project K5/2024) in the SADC region by testing the effect of using the best available replacements to the detailed hydraulic property data-sets available in South Africa. As explained in Section 2, this is necessary because the data-sets we used to derive the soil properties in HYLARSMET that are not available outside South Africa.

In Subsection 2.4 we introduced the HWSD (Harmonized World Soil Database), expanded in Subsection 4.2, and the soil characteristics used to develop the soil hydraulic properties needed for the PyTOPKAPI model (originally developed in WRC project K5/1683 from TOPKAPI and made open source in HYLARSMET). In Subsection 4.2, we describe the calculation of the proportional change in the relevant soil properties and use the sensitivity results reported in Sinclair and Pegram (2013) to evaluate the relative impact of the changes at each location in the HYLARSMET grid. Subsection 4.3 contains a summary of our findings and conclusions.

4.1. HWSD – the Harmonized World Soil Database

In this Subsection we describe the derivation of soil texture characteristics required to obtain hydraulic properties for PyTOPKAPI in the SADC region, based on the Harmonized World Soil Database (HWSD v1.2.1; FAO/IIASA/ISRIC/ISS-CAS/JRC (2012)).

The HWSD is currently the most complete and useful SADC-wide soil database containing sufficient information to allow the derivation of the soil properties needed to run the PyTOPKAPI model. The HWSD has been used as a base data-set by many land surface modellers (e.g. see Bódis, 2009; DeLannoy et al., 2014). There are a few alternative data-sets (e.g. Shangguan et al., 2014; Hengl et al., 2014; Jones et al., 2013) that we investigated, but these are either not widely used, or miss the soil texture information we require, so they are only mentioned for completeness.

The HWSD data-set is provided in the form of a 2 Gigabyte global raster, which specifies a soil mapping unit ID for each grid cell. The spatial resolution of the raster is approximately 1 km (30 arc seconds by 30 arc seconds) and consists of 43 200 rows by 21 600 columns of 16-bit unsigned integer values. Each mapping unit ID is associated with up to 9 soil units, resulting in a one-to-many relationship between the designated mapping unit ID and the mix of soil properties contained therein. Finally, each individual soil unit in a given mapping unit is defined according to its top- and sub-soil parameters (of which there are 48 in total per soil unit). Figure 4.1 shows an NDVI (Normalized Difference Vegetation Index) image at the same scale as the data-base.

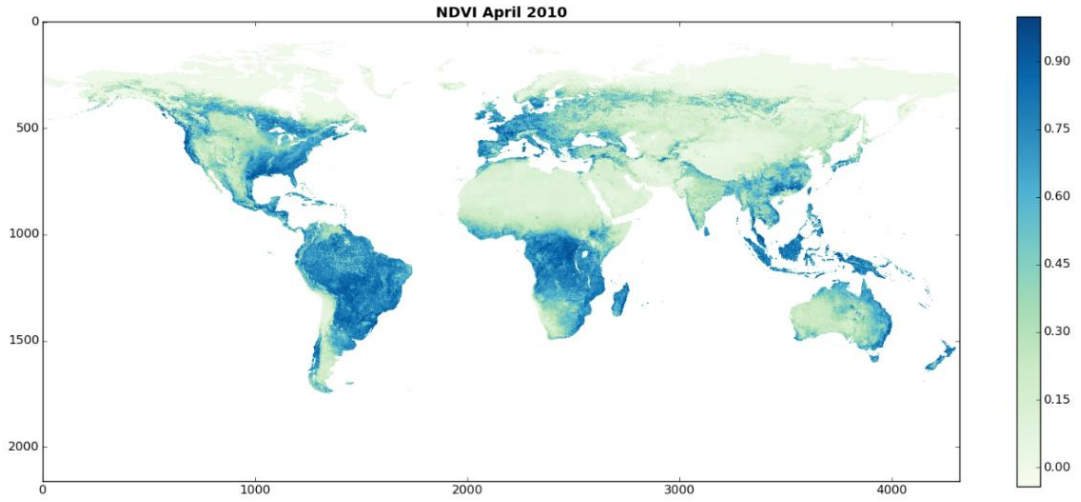


Figure 4.1: Global NDVI in April 2010.

Given this information, it is clear that the HWSD data structure is relatively convoluted and it was necessary to develop custom software to sample the database at the locations required. In addition, we spent a considerable amount of time developing a robust HWSD reader in Python, to ensure that the coherence of the data was maintained, while it was being parsed for specific applications in chosen regions. This extensive but invisible effort has added considerable value to the data resource and is the major contribution to the deliverable.

The properties sampled from the HWSD, for our purposes of developing soil parameters, are the sand, clay and organic carbon contents for top- and sub-soil respectively. The method used to convert these to hydraulic soil properties is described in Subsection 4.2. Here we present the sampled top- and sub-soil textural data from the HWSD that are used to derive the hydraulic properties. Figure 4.2 shows the sampled soil properties for the southern portion of SADC.

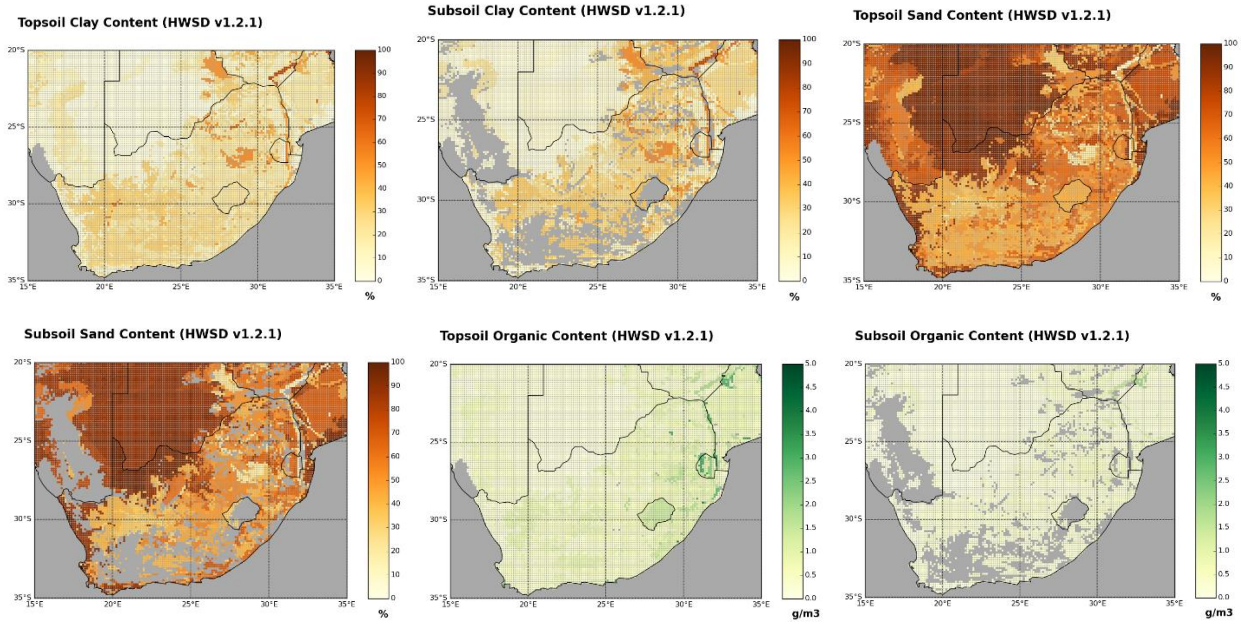


Figure 4.2: Top- and sub-soil texture and organic carbon sampled from the HWSD on an extended HYLARSMET grid (0.125°). Missing data are shown in grey. Of interest are the considerable areas of missing sub-soil properties; these are regions with shallow soils (see the HWSD reference depth map in Figure 4.4).

4.2. Derivation and comparison of soil hydraulic properties

In this Subsection we present the resulting changes in soil properties in South Africa where we have available detailed data-sets. For most of the parameters we follow the methodology of DeLannoy et al. (2014) to convert the Harmonized World Soil Database (HWSD v1.2.1; FAO/IIASA/ISRIC/ISS-CAS/JRC (2012)) texture classes to soil hydraulic properties.

TABLE 2. HYDROLOGIC SOIL PROPERTIES CLASSIFIED BY SOIL TEXTURE

Texture class	Sample size	Total porosity (θ_t), cm^3/cm^3	Residual saturation (θ_r) cm^3/cm^3	Effective porosity (θ_e), cm^3/cm^3	Bubbling pressure (ψ_b)		Pore size distribution (λ)		Water retained at -0.33 bar tension, cm^3/cm^3	Water retained at -15 bar tension, cm^3/cm^3	Saturated Hydraulic Conductivity‡ (K_s) cm/h
					Arithmetic, cm	Geometric,† cm	Arithmetic	Geometric†			
Sand	762	0.437**	0.020	0.417	15.98	7.26	0.694	0.592	0.091	0.033	21.00
		(0.374 0.500)	(0.001 0.039)	(0.354–0.480)	(0.24–31.72)	(1.36–38.74)	(0.298–1.090)	(0.334–1.051)	(0.018–0.164)	(0.007–0.059)	
Loamy sand	338	0.437	0.035	0.401	20.58	8.69	0.553	0.474	0.125	0.055	6.11
		(0.368–0.506)	(0.003 0.067)	(0.329–0.473)	(0.0 45.20)	(1.80–41.85)	(0.234–0.872)	(0.271 0.827)	(0.060–0.190)	(0.019–0.091)	
Sandy loam	666	0.453	0.041	0.412	30.20	14.66	0.378	0.322	0.207	0.095	2.59
		(0.351–0.555)	(0.0–0.106)	(0.283–0.541)	(0.0–64.01)	(3.45–62.24)	(0.140–0.616)	(0.186–0.558)	(0.126–0.288)	(0.031–0.159)	
Loam	383	0.463	0.027	0.434	40.12	11.15	0.252	0.220	0.270	0.117	1.32
		(0.375–0.551)	(0.0–0.074)	(0.334 0.534)	(0.0 100.3)	(1.63–76.40)	(0.086–0.418)	(0.137–0.355)	(0.195–0.345)	(0.069–0.165)	
Silt loam	1206	0.501	0.015	0.486	50.87	20.76	0.234	0.211	0.330	0.133	0.68
		(0.420–0.582)	(0.0–0.058)	(0.394–0.578)	(0.0–109.4)	(3.58–120.4)	(0.105–0.363)	(0.136 0.326)	(0.258–0.402)	(0.078–0.188)	
Sandy clay loam	498	0.398	0.068	0.330	59.41	28.08	0.319	0.250	0.255	0.148	0.43
		(0.332–0.464)	(0.0 0.137)	(0.235 0.425)	(0.0 123.4)	(5.57–141.5)	(0.079–0.559)	(0.125–0.502)	(0.186–0.324)	(0.085–0.211)	
Clay loam	366	0.464	0.075	0.390	56.43	25.89	0.242	0.194	0.318	0.197	0.23
		(0.409 0.519)	(0.0 0.174)	(0.279–0.501)	(0.0–124.3)	(5.80–115.7)	(0.070–0.414)	(0.100–0.377)	(0.250–0.386)	(0.115–0.279)	
Silty clay loam	689	0.471	0.040	0.432	70.33	32.56	0.177	0.151	0.366	0.208	0.15
		(0.418 0.524)	(0.0 0.118)	(0.347–0.517)	(0.0 143.9)	(6.68–158.7)	(0.039–0.315)	(0.090–0.253)	(0.304 0.428)	(0.138–0.278)	
Sandy clay	45	0.430	0.109	0.321	79.48	29.17	0.223	0.168	0.339	0.239	0.12
		(0.370 0.490)	(0.0 0.205)	(0.207–0.435)	(0.0 179.1)	(4.96–171.6)	(0.048 0.398)	(0.078–0.364)	(0.245 0.433	(0.162–0.316)	
Silty clay	127	0.479	0.056	0.423	76.54	34.19	0.150	0.127	0.387	0.250	0.09
		(0.425 0.533)	(0.0–0.136)	(0.334–0.512)	(0.0 159.6)	(7.04 166.2)	(0.040–0.260)	(0.074–0.219)	(0.332 0.442)	(0.193–0.307)	
Clay	291	0.475	0.090	0.385	85.60	37.30	0.165	0.131	0.396	0.272	0.06
		(0.427 0.523)	(0.0 0.195)	(0.269 0.501)	(0.0 176.1)	(7.43 187.2)	(0.037–0.293)	(0.068–0.253)	(0.326 0.466)	(0.208 0.336)	

* First line is the mean value
Second line is + one standard deviation about the mean
† Antilog of the log mean
‡ Obtained from Fig. 2

Table 4.1: The table from Rawls et al. (1982) used to estimate soil properties from texture information, rotated for readability. Unfortunately, the original document's text is not very clear.

However, the residual moisture content θ_r, bubbling pressure ψ_b and pore size distribution λ are calculated as before, using the table of Rawls et al. (1982) (see Table 4.1, repeated here from Figure 2.2 for convenience). The main difference from our RSA work is that for SADC we are using the new texture maps from the HWSD.

The basic procedure for converting soil texture to hydraulic properties can be explained by referring to the soil texture triangles shown in Figure 4.4.3.

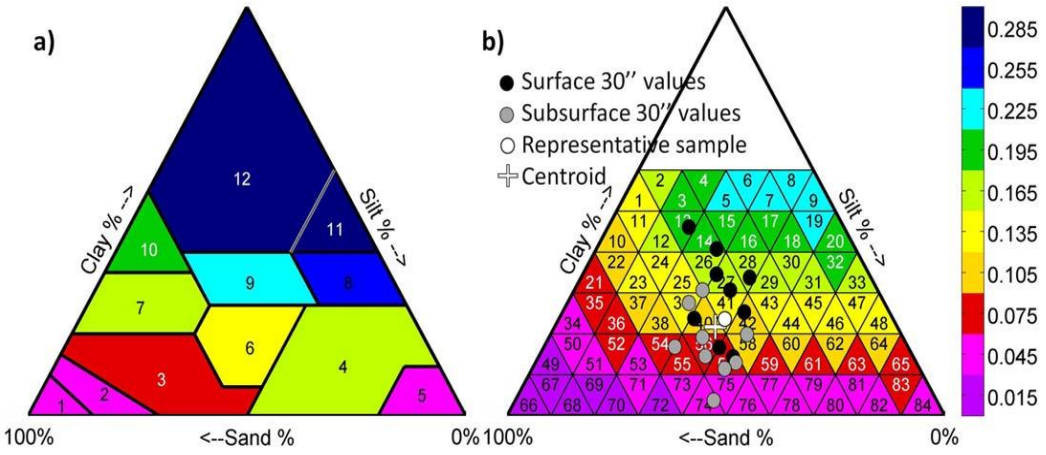


Figure 4.3: Soil texture triangles a) the 12 class USDA triangle used for defining the HYLARSMET parameters b) the 253-class triangle developed by DeLannoy et al. (2014) – their Figure repeated here

Figure 4.4.3a shows the USDA soil texture triangle, and the location of a given soil into a particular class can be determined using the properties shown in Figure 4.2 – the relevant hydraulic properties are then obtained using the soil texture data shown in Table 4.1. The method adopted by DeLannoy et al. (2014) subdivides the triangle into smaller regions and also stratifies the properties according to broad ranges of organic carbon content. The result is 253 unique soil types, with their associated properties determined by a look-up table which we extracted from the paper of DeLannoy et al. (2014). This set of activities was coded into the HWSD reader to enable semi-automatic calculations to be made.

There are 3 ways of obtaining soil properties which we examine in this Subsection in the following Subsections: 4.2.1 directly from HWSD; 4.2.2 combining HWSD texture and DeLannoy's triangles; 4.2.3 combining HWSD and Rawls' Table. We make inter-comparisons between them in maps of the region.

4.2.1 Properties obtained directly from the HWSD

Soil depth

For HYLARSMET in South Africa, we previously derived soil depth (L) maps by resampling and projecting the soil depths reported in Schulze et al. (2008). The top left panel in Figure 4.4.4 shows the depths used in the HYLARSMET analyses.

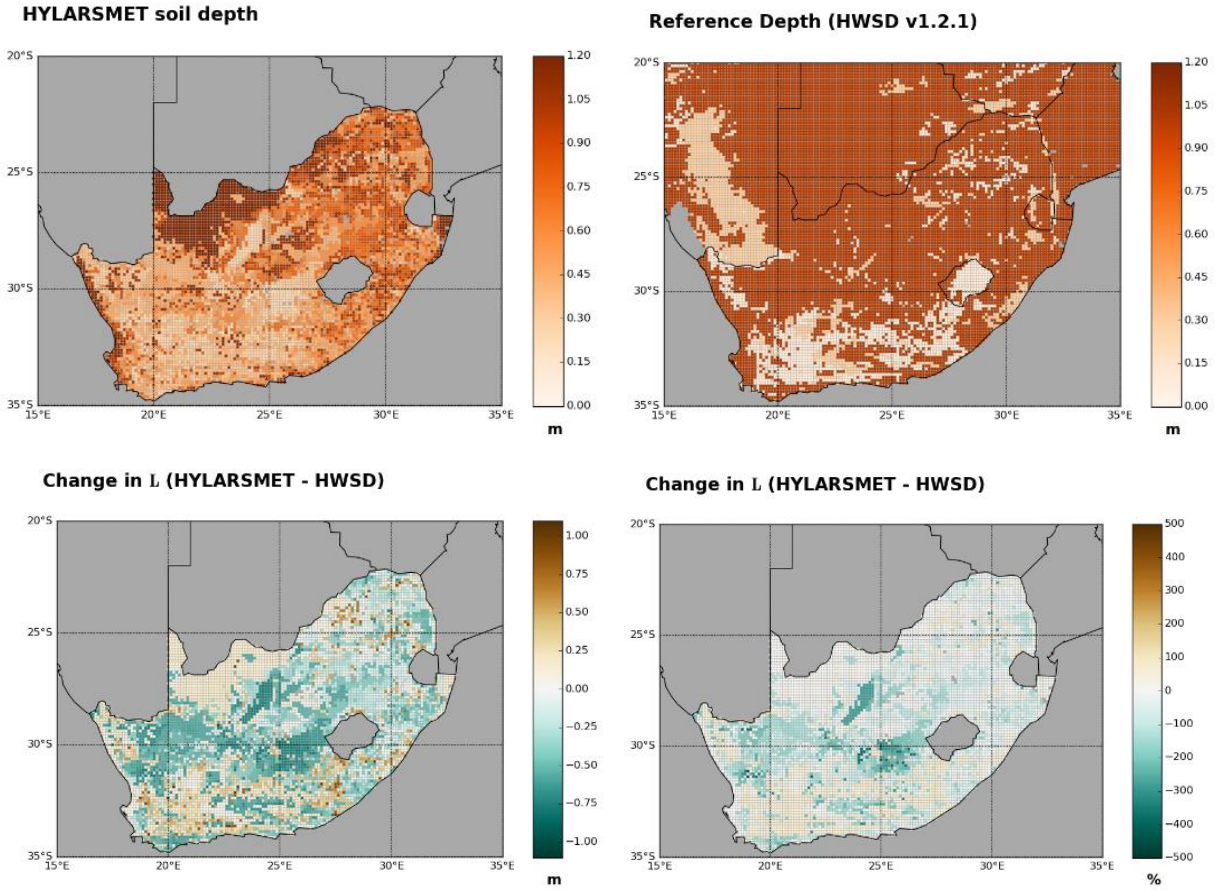


Figure 4.4: Depth of the soil layer currently used in the HYLARSMET modelling analyses (top left), HWSD reference depth map (top right), difference between HYLARSMET and HWSD (bottom left), and the proportional change introduced by using the HWSD texture maps (bottom right)

The only depth information in the HWSD is a property called Reference Depth, and it is uniformly 1 metre across the globe, except for areas with shallow and rocky soil where a depth of either 0.1 m or 0.3 m is defined. The top right panel of Figure 4.4.4 shows their distribution of reference depth in Southern Africa. The bottom left-hand panel in Figure 4.4.4 shows the change in soil depth (m) as a result of using the HWSD information compared to Schulze et al. (2008). The bottom right-hand panel in Figure 4.4.4 shows the *percentage* change in soil depth as a proportion of the original HYLARSMET depths based on Schulze et al. (2008).

4.2.2 Properties obtained by combining HWSD texture and the information from DeLannoy et al.

Saturated soil moisture content

For HYLARSMET in South Africa, we derived saturated moisture content (θ_s) maps by resampling and projecting the total porosity maps reported in Schulze et al. (2008). The top left panel in Figure 4.4.5 shows the saturated moisture content used in the HYLARSMET analyses.

To develop θ_s based on the HWSD texture and organic content information, we extracted the look-up tables from DeLannoy et al. (2014) and calculated the reported value of θ_s. The new map of θ_s is shown in the top right panel of Figure 4.4.5.

The bottom left-hand panel in Figure 4.4.5 shows the difference in saturated moisture content θ_s introduced by using the HWSD information.

The bottom right-hand panel in Figure 4.4.5 shows the percentage change in saturated moisture content, as a proportion of the original HYLARSMET θ_s based on Schulze et al. (2008).

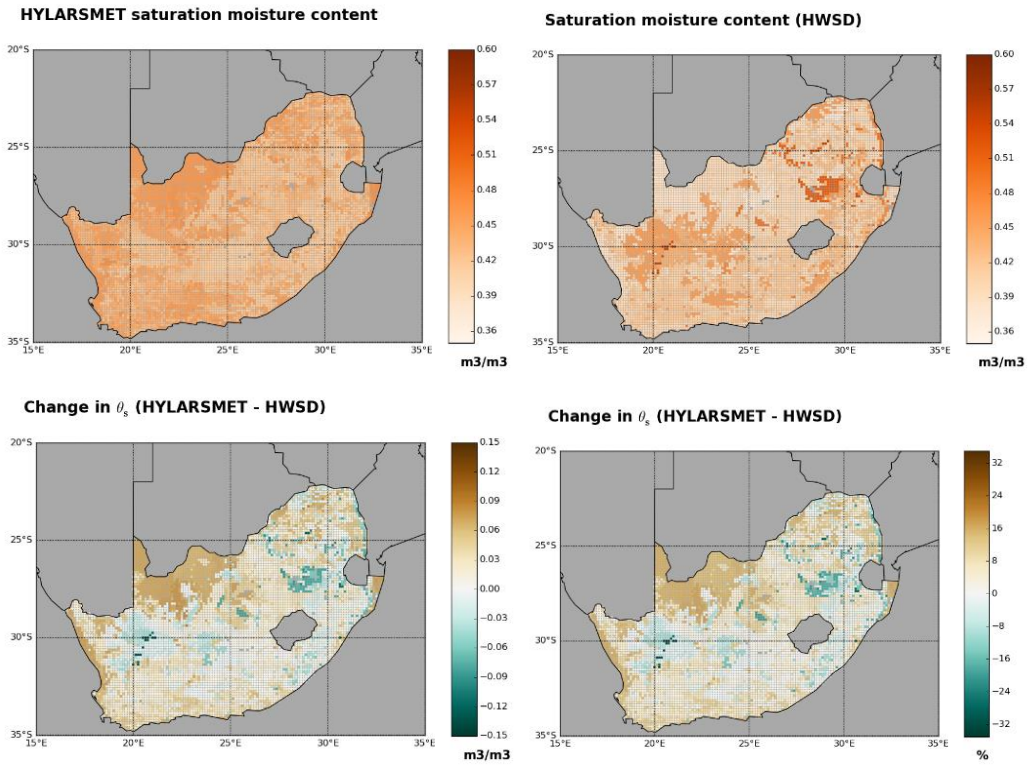


Figure 4.5: Saturation moisture content θ_s currently used in the HYLARSMET modelling analyses (top left), θ_s estimated from the HWSD (top right), difference map (bottom left), and the proportional change introduced by using the HWSD texture maps (bottom right).

Saturated hydraulic conductivity

For HYLARSMET in South Africa, we obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The soil conductivity (K_s) values were then estimated using table 5.5.5 of (Maidment, 1993) and in cases where more than one soil texture class is represented, the average of the values reported for each class is used. A map of the K_s values used in HYLARSMET is shown in the top left panel of Figure 4.4.6.

To develop maps of K_s based on the HWSD texture and organic content information, we used the look-up tables extracted from DeLannoy et al. (2014) and obtained the reported value of K_s. The new map of K_s is shown in the top right panel of Figure 4.4.6.

The bottom left-hand panel in Figure 4.4.6 shows the difference in K_s, when using the HWSD compared to the Middleton and Bailey (2009) distribution.

The bottom right-hand panel in Figure 4.4.6 shows the percentage change in saturated hydraulic conductivity, as a proportion of the original HYLARSMET K_s based on Maidment (1993).

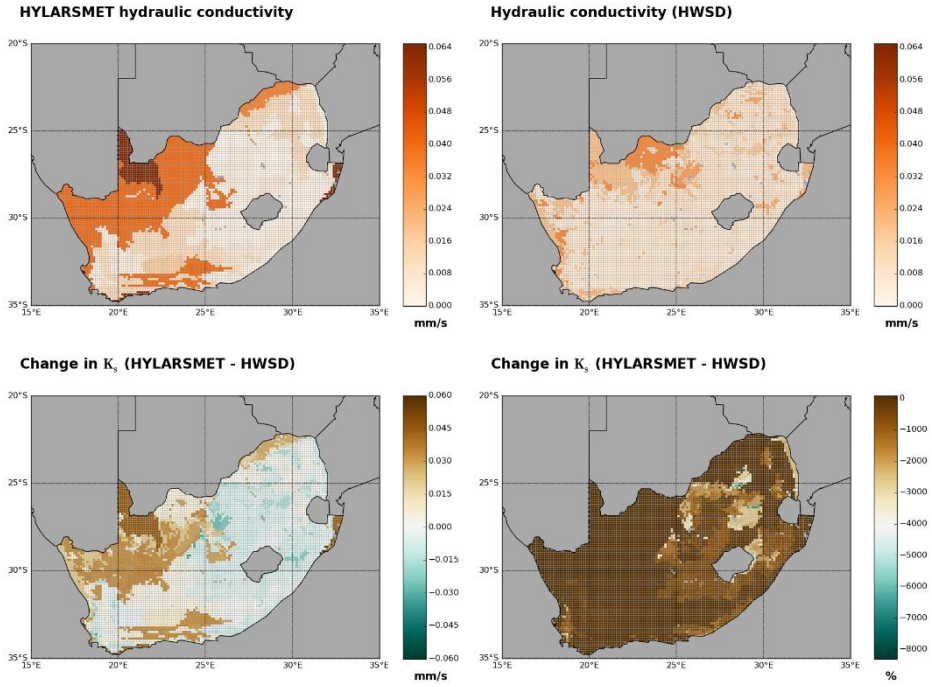


Figure 4.6: Saturated hydraulic conductivity K_S currently used in the HYLARSMET modelling analyses (top left), revised K_S map based on the HWSD texture information (top right), difference map (bottom left), and the proportional change introduced by using the HWSD texture maps (bottom right).

4.2.3 Properties obtained by combining HWSD texture and Table 4.1 (Rawls et al., 1982)

Residual soil moisture content

To develop maps of residual moisture content θ_r for HYLARSMET, we obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The θ_r values are then the mean values for the relevant texture classes reported in table 5.3.2 of (Maidment, 1993) and in cases where more than one soil texture class is represented, an average has been taken. The top left panel of Figure 4.4.7 shows the map of values used in HYLARSMET.

To develop new maps of θ_r based on the HWSD texture and organic content information, we used the same look-up tables extracted from (Maidment, 1993) and obtained the reported value of θ_r. The new map of θ_r is shown in the top right panel of Figure 4.4.7. The bottom left-hand panel in Figure 4.4.7 shows the change in θ_r, as a result of using the HWSD information. The bottom right-hand panel in Figure 4.4.7 shows the percentage change in θ_r, as a proportion of the original HYLARSMET θ_r based on the spatial distributions of Middleton and Bailey (2009).

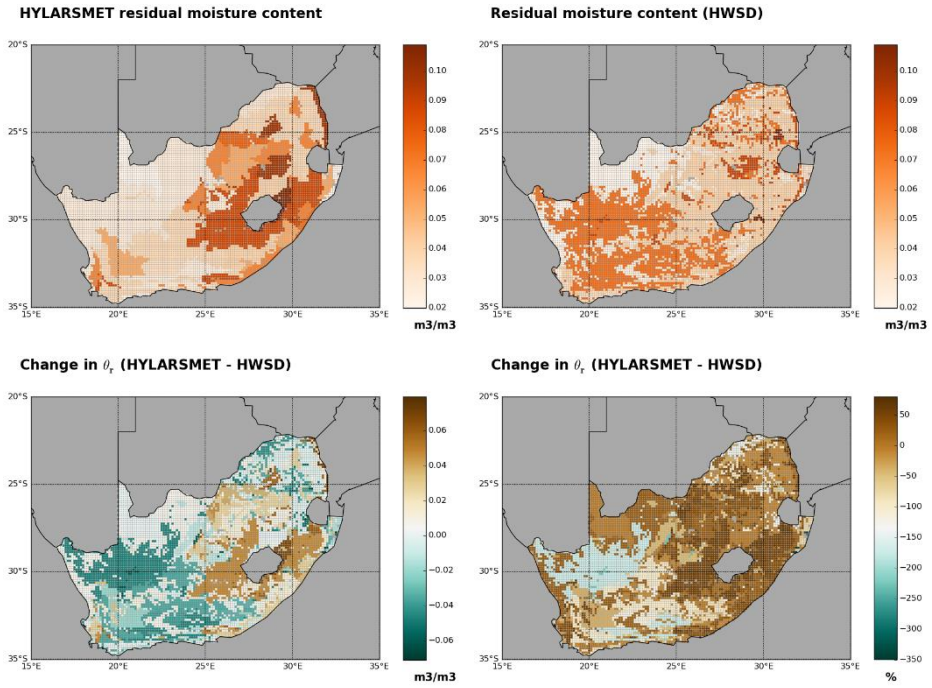


Figure 4.7: Residual moisture content θ_r currently used in the HYLARSMET modelling analyses (top left), revised map based on the HWSD (top right), difference map (bottom left), and the proportional change introduced by using the HWSD texture maps (bottom right).

Bubbling pressure

In developing HYLARSMET, we obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The bubbling pressure (ψ_b) values were then estimated using the values in Table 4.1 and in cases where more than one soil texture class is represented, the average of the values reported for each class is used. The HYLARSMET ψ_b distribution is shown in the top left panel of Figure 4.4.8.

To develop revised maps of ψ_b based on the HWSD texture and organic content information, we used the information in Table 4.1 and obtained the reported value of ψ_b. The new map of ψ_b is shown in the top right panel of Figure 4.4.8.

The bottom left-hand panel in Figure 4.4.8 shows the difference in ψ_b introduced as a result of using the HWSD data.

The bottom right-hand panel in Figure 4.4.8 shows the percentage change in ψ_b, as a proportion of the original HYLARSMET ψ_b based on the spatial distributions of Middleton and Bailey (2009).

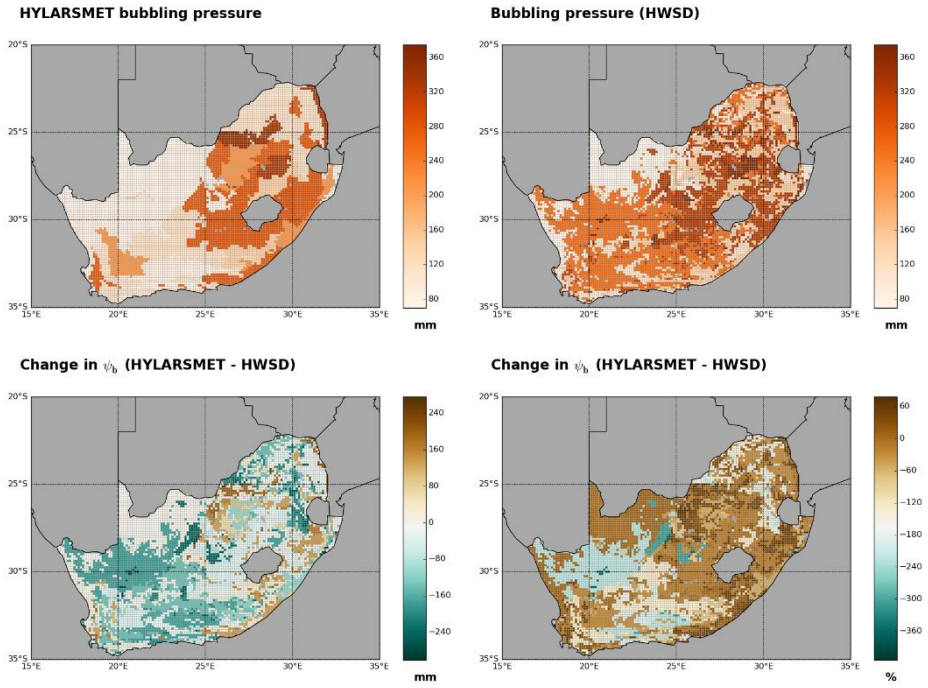


Figure 4.8: Bubbling pressure ψ_b currently used in the HYLARSMET modelling analyses (top left), the updated map based on the HWSD information (top right), difference map (bottom left), and the proportional change introduced by using the HWSD texture maps (bottom right).

Pore size distribution index

We obtained soil texture classifications from the WR2005 soils map (Middleton and Bailey, 2009). The pore size distribution index (λ) values were then estimated using Table 4.1 and in cases where more than one soil texture class is represented, the average of the values reported for each class is used. The HYLARSMET λ distribution is shown in the top left panel of Figure 4.9.

To develop revised maps of λ based on the HWSD texture and organic content information, we used the same look-up tables extracted from Table 4.1 and obtained the reported value of λ. The new map of λ is shown in the top right panel of Figure 4.4.9.

The bottom panel in Figure 4.4.9 shows the percentage change in λ, as a proportion of the original HYLARSMET λ based on the spatial distributions of Middleton and Bailey (2009).

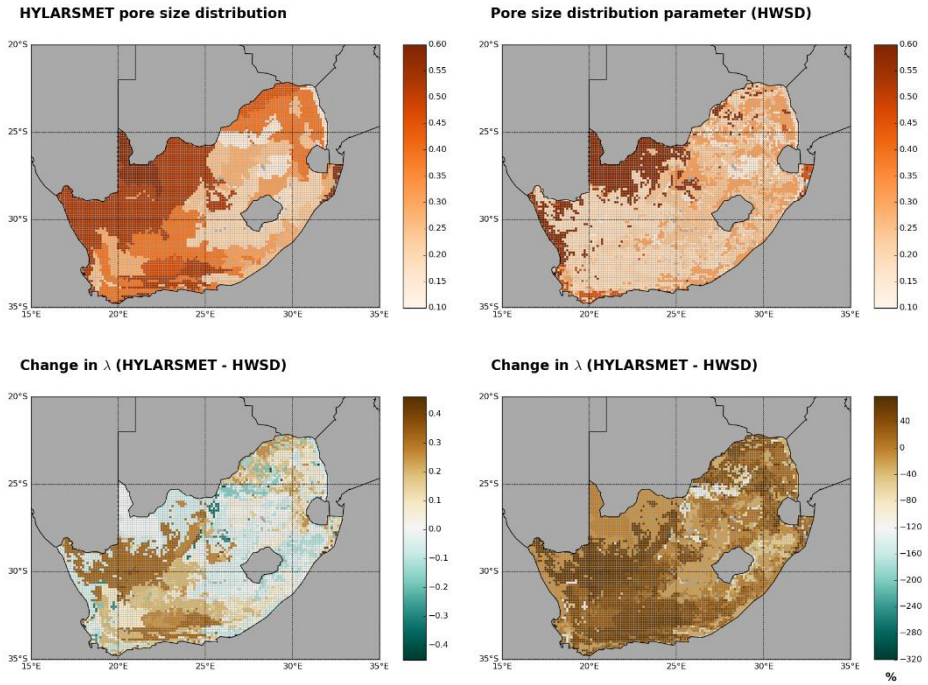


Figure 4.9: Pore size distribution parameter λ currently used in the HYLARSMET modelling analyses (top left), the HWSD based map (top right), difference between HYLARSMET and HWSD λ (bottom left), and the proportional change introduced by using the HWSD texture maps (bottom right).

4.3 Summary of Section 4

As reported in the heading of this Subsection we have sought then adopted a global set of soil properties from the Harmonized World Soil Database and then compared the parameters, thence derived for PyTOPKAPI, to the ones we developed in HYLARSMET, the WRC project K5/2024.

We

- sifted through a large range of products before we identified HWSD as a suitable data-source to cover SADC with the parameters we required at a fine enough scale
- developed a suite of code to access texture information in HWSD, in order to extract the information pertinent to the comparison
- identified and implemented a revised method of soil hydraulic parameter determination from DeLannoy et al. (2014)
- compared the derived parameters against our PyTOPKAPI ones and found that the data range is reasonable in most areas of South Africa, where we made the tests. We found that there are problems of mismatch and differences due to a different areal distribution of soil types, especially on the boundaries between our soil units/clusters.
- unfortunately, we do not have the same detailed data-sets over SADC as we do over RSA, so for the region outside our borders, we will have to use what is readily available. Regrettably, without ground-truthing, we are not completely sure which data-set is correct.

-----ooOoo----

5. Selecting best-match alternative remotely sensed rainfall data to TRMM

The intent of Deliverable 5 was to: ***'Determine if there are better alternative rainfall inputs to TRMM [such as the Global Precipitation Mission], for near real-time precipitation data input, if not exploit bias adjustment of the TRMM product'***, and to research the options for replacing the remotely sensed rainfall driving the HYLARSMET soil moisture modelling system. This tool was TRMM 3B42RT (part of a suite of rainfall products known as the TRMM Multisensor Precipitation Analysis – TMPA). Such replacement was only intended to take place if we were able to find a justifiably better alternative. However, this task took on a new importance with the decommissioning of the TRMM satellite's precipitation radar on 7 October 2014, and the TRMM Microwave Imager (TMI) on 8 April 2015, as the TRMM satellite was due to re-enter the Earth's atmosphere (Huffman, 2015). Luckily the TMPA rainfall products were scheduled to continue until mid-2017 (even though they no longer included inputs from TRMM); this gave us time to make a clean transition to the best usable alternative.

We start in what follows by describing two sub-daily rainfall products that are freely available with low update latency, based on data collected by the instruments flown on-board the Global Precipitation Mission (GPM) core observatory and other satellite-based instruments (Subsection 5.1). The choice was made primarily on the basis that these two products are both easily available in near real time and are being actively developed under the activities of the GPM.

In Subsection 5.2 we describe the data-sets collected and the analysis carried out to compare the new rainfall products with the TRMM3B42-RT product. Finally, in Subsection 5.3 we describe a preliminary choice of the best replacement product going forward.

5.1 GPM based rainfall products

In this Subsection we present a description of the GPM core observatory (Subsection 5.1.1) and then proceed to provide an overview of the two GPM derived near real time rainfall products, IMERG (Subsection 5.1.2), and GSMaP (Subsection 5.1.3). The intention is to give an overview of the instrumentation carried aboard the GPM core, and the relevant characteristics of the derived rainfall products.

5.1.1 The GPM core observatory and constellation

The GPM core observatory is the successor to TRMM and carries two instruments with improved capabilities over the decommissioned TRMM instruments. These instruments are:

1) the **GPM Microwave Imager (GMI)**, which is a replacement for the TRMM Microwave Imager (TMI). The GMI has 13 channels in the 10-183 GHz range, compared to 5 channels in the range 10-86 GHz of the TMI. The GMI measures the intensity of upwelling radiation from the Earth's surface and atmosphere, which is sensitive to the presence of moisture (e.g. rainfall, snow) at microwave frequencies.

2) the **Dual Precipitation Radar (DPR)**, which is a dual frequency cross-track scanning radar. The DPR scans at two frequencies 13.6 GHz and 35.5 GHz and is a replacement for the TRMM Precipitation Radar (PR), which was also a cross-track scanner at a single frequency, 13.8 GHz.

Figure 5.1 shows a schematic of the GPM core components, while the different scanning strategies for the GMI and DPR are illustrated in Figure 5.2.

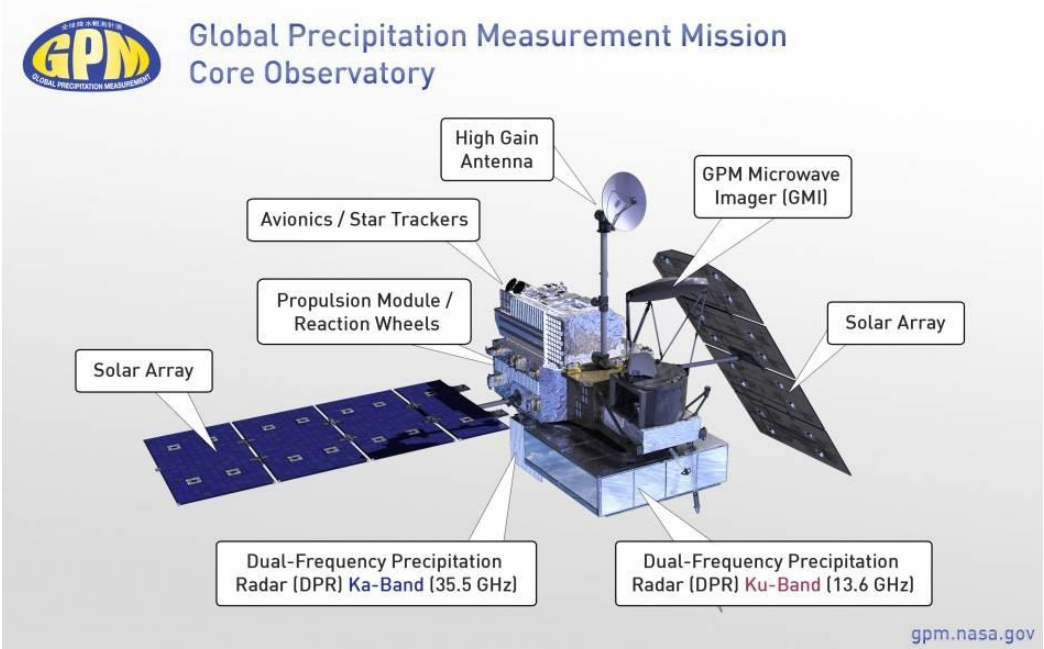


Figure 5.1: A diagram of the GPM core observatory, showing instrumentation. Of note are the GPM Microwave Imager (GMI) and the Dual Precipitation Radar (DPR).

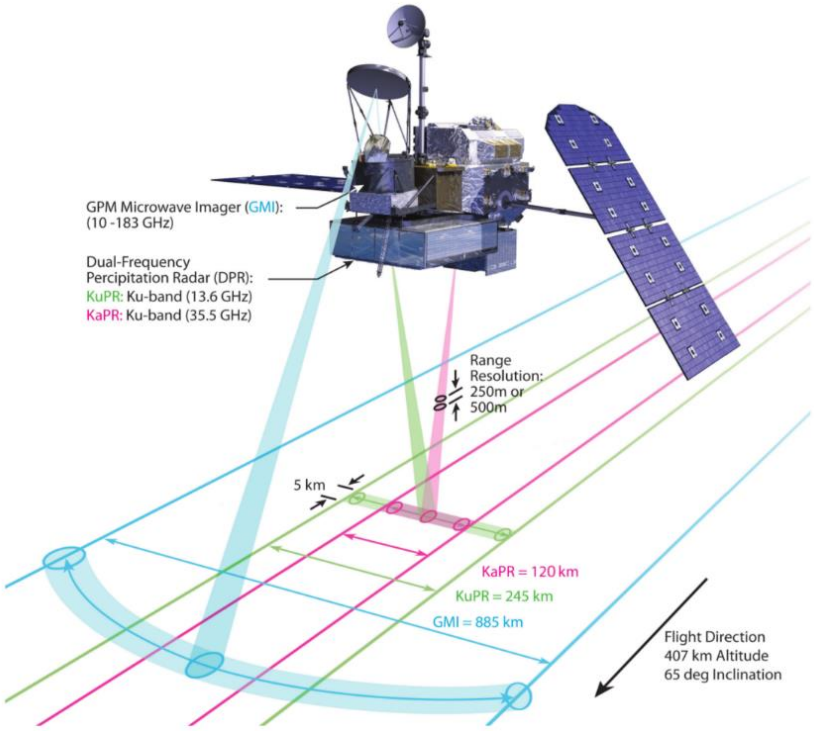


Figure 5.2: An illustration of the scanning geometry for the GPM core observatory. The DPR scans cross-track and the GMI scans an arc forward of the satellite. The ground footprints also have different spatial resolutions, and the DPR samples at a range of vertical intervals.

The GPM core will also have an extended orbit relative to TRMM, allowing it to observe a wider range of latitudes. Figure 5.3 is an illustration of three successive TRMM orbits in yellow, and the same for GPM core in blue. The wider coverage of the GPM core is clearly evident.

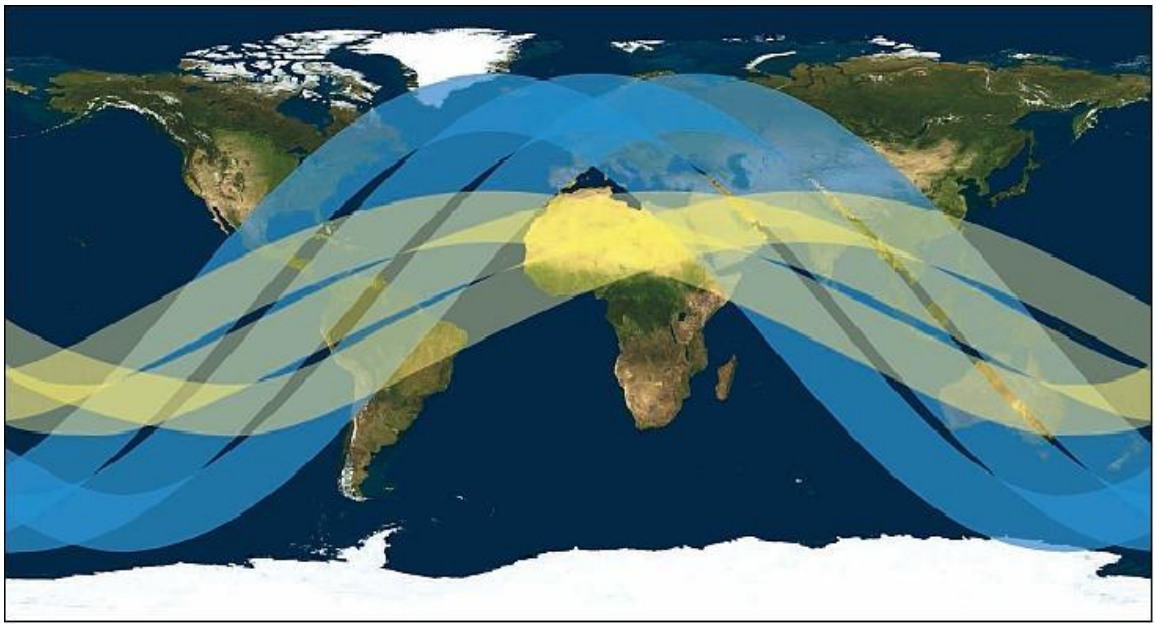


Figure 5.3: An illustration of three successive TRMM orbits in yellow, and the same for GPM core in blue. The wider coverage of latitudes by the GPM core is clearly shown.

The GPM core is the central satellite platform of the GPM initiative, which includes a constellation of low earth-orbiting satellites that carry microwave imagers and/or sounders capable of observing precipitation (see Figure 5.4 for an overview). The goal is to combine the relatively accurate and high spatial resolution, but intermittent microwave derived rainfall estimates, with the more indirect IR based estimates from geostationary platforms, which have better time and space coverage but coarser spatial resolution.

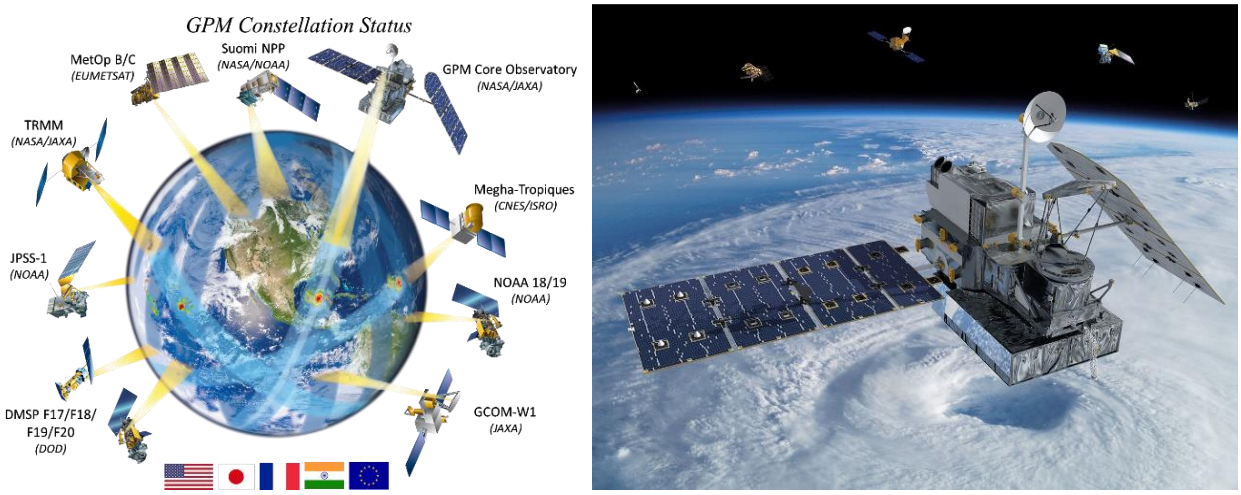


Figure 5.4: An illustration of the GPM constellation, showing the satellite platforms carrying suitable microwave instrumentation and the partner states/organizations involved in this global initiative.

5.1.2 IMERG rainfall products

In this Subsection we describe the IMERG rainfall product, which is NASA's successor to the TMPA suite of rainfall products based on TRMM (Huffman et al., 2007). The IMERG acronym stands for 'Integrated Multi-satellitE Retrievals for GPM' and is the NASA algorithm that provides the near-realtime multi-satellite precipitation product for the U.S. GPM team.

IMERG Half Hourly (HH) product has three variants that are relevant to this investigation, their basic characteristics are outlined in table 5.1. The space-time resolution is the same for each (0.1° grid at 30-minute intervals – a great improvement compared to TRMM 3B42RT's 0.25° grid at 3-hour intervals) – and the inputs and algorithms are similar. The main difference is in the latency at which the products become available. The IMERG HH Early and Late products are essentially identical, with the difference being that the Late run product may contain more microwave derived rainfall estimates (depending on the latency of the processed microwave rainfall products). The IMERG HH Final product includes an adjustment based on rain gauge observations (although the rain gauge network available over Africa is extremely sparse and includes only a small percentage of the complete gauge network). The 30-minute time-step is determined by the time resolution of the global combined IR precipitation field derived from geostationary satellite data.

Table 5.1: Basic characteristics of the IMERG Half-Hourly rainfall variants.

Product name	**IMERG HH Early**
Source	http://pmm.nasa.gov/data-access/downloads/gpm
Spatial resolution	0.1°
Temporal resolution	0.5 hr
Start date	2014-03-12 (but only most recent 5 months available)
End date	Present
Latency	6 hrs
Product name	**IMERG HH Late**
Source	http://pmm.nasa.gov/data-access/downloads/gpm
Spatial resolution	0.1°
Temporal resolution	0.5 hr
Start date	2014-03-12 (but only most recent 6 months available)
End date	Present
Latency	18 hrs
Product name	**IMERG HH Final**
Source	http://pmm.nasa.gov/data-access/downloads/gpm
Spatial resolution	0.1°
Temporal resolution	0.5 hr
Start date	2014-03-12
End date	Present
Latency	4 months

The IMERG data files contain a number of fields for each observation time-slot, and these are illustrated in Figures 5.5 through 5.13. A description of each field is given in the figure captions.

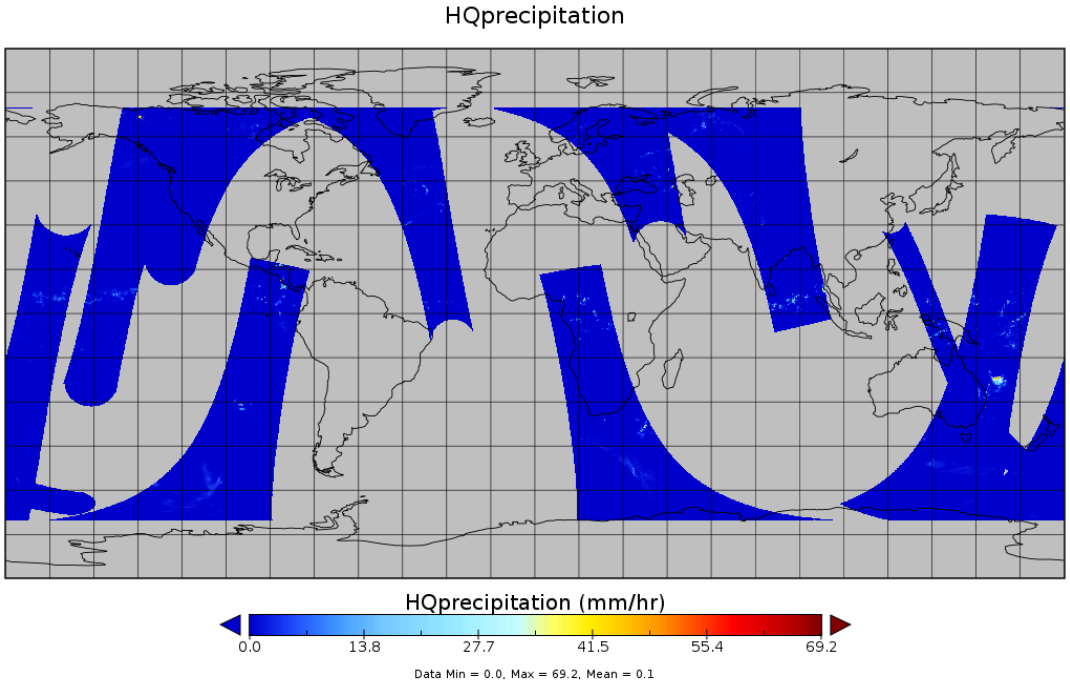


Figure 5.5: The IMERG HH HQprecipitation field for the time slot 2015-09-03 12:00. The HQprecipitation field is a blend of rainfall estimates derived from microwave instruments flown aboard low earth-orbiting (LEO) satellites (see the table in Figure 5.15 for the list of LEO estimates used in the GSMaP rainfall product). IMERG uses roughly the same selection. In this image we can identify several different instruments simultaneously sampling precipitation in the 30-minute interval after noon.

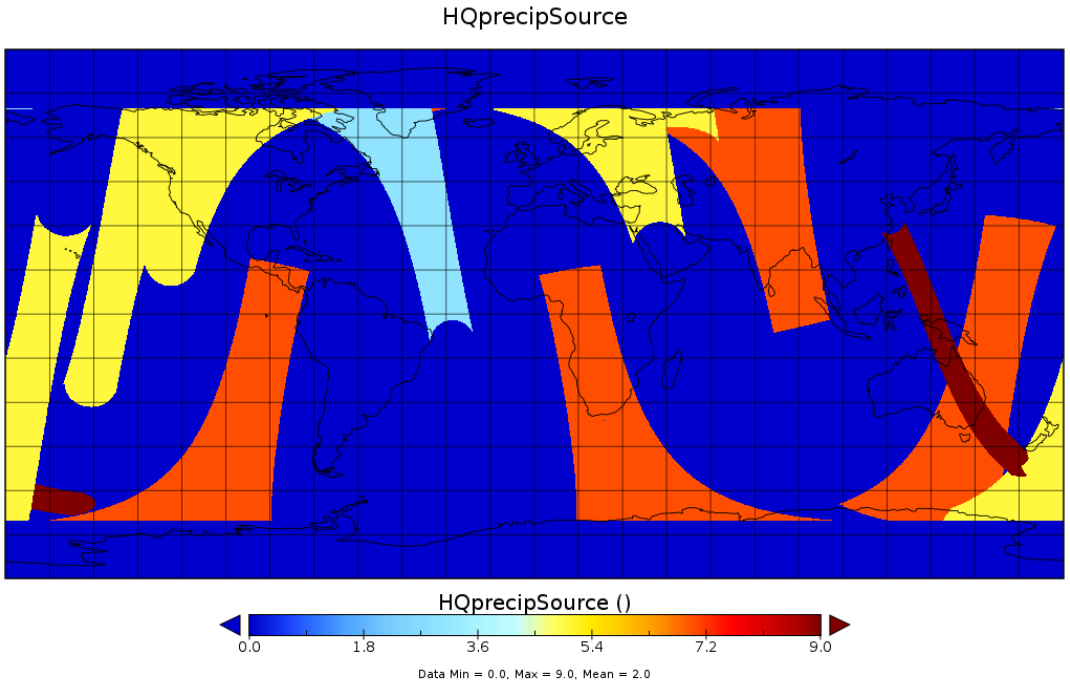


Figure 5.6: The IMERG HH HQprecipSource field for the time slot 2015-09-03 12:00. The HQprecipSource field contains integer keys that define which microwave product was used in each grid box to derive the HQprecipitation field shown in Figure 5.5.

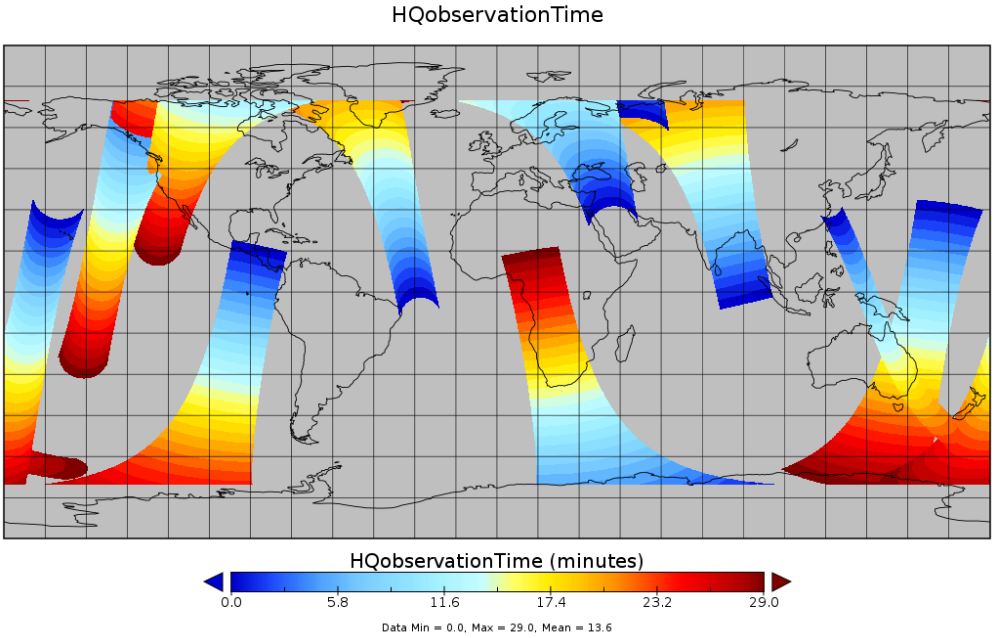


Figure 5.7: The IMERG HH HQobservationTime field for the time slot 2015-09-03 12:00. The HQobservationTime field shows the observation time for each microwave product used to derive the HQprecipitation field shown in Figure 5.5. The observation times are reported as an offset from the start time of each time-slot, starting blue and finishing red in the swaths.

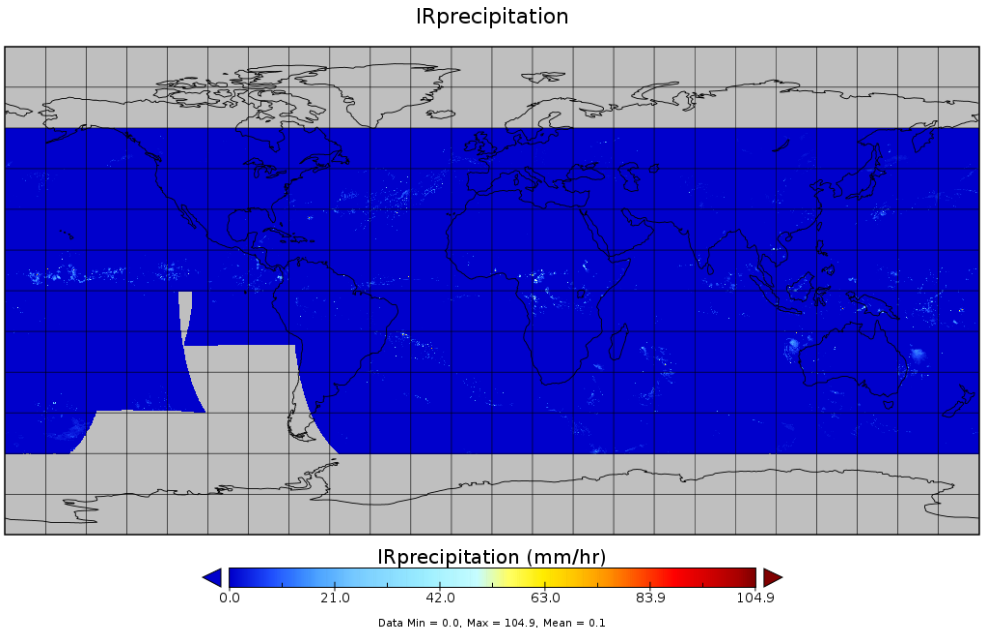


Figure 5.8: The IMERG HH IRprecipitation field for the 30-minute time slot on 2015-09-03, starting at 12:00. The IRprecipitation field is a global precipitation field derived from all available IR instruments aboard geostationary satellites. This IR field is combined with the HQprecipitation field shown in Figure 5.5 to produce the uncalibrated and calibrated precipitation fields shown in Figures 5.10 and 5.11.

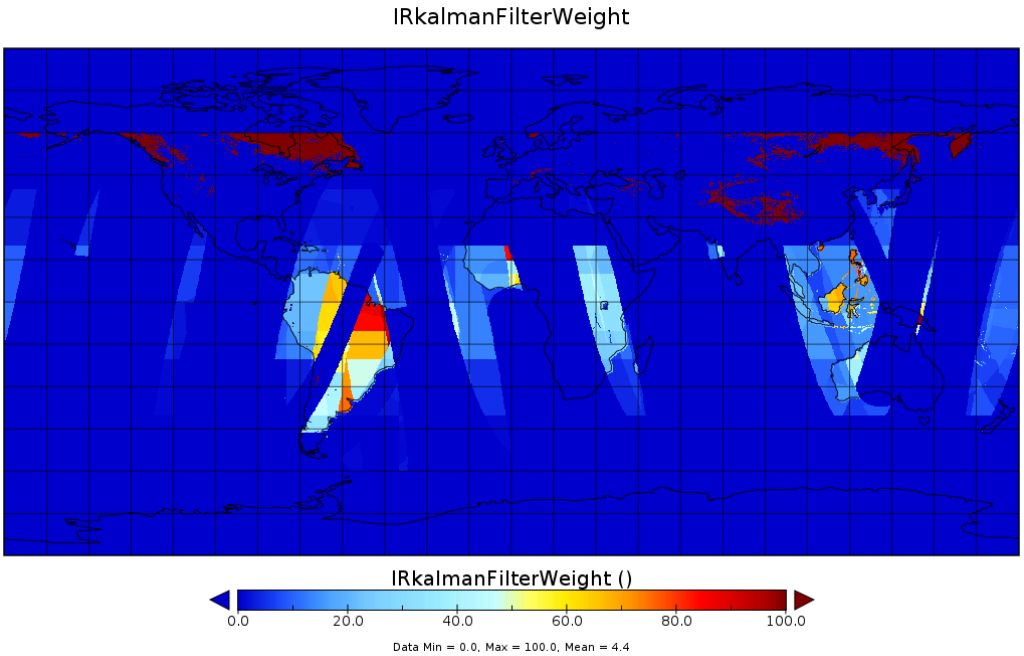


Figure 5.9: The IMERG HH IRkalmanFilterWeight field for the time slot 2015-09-03 12:00. The IRkalmanFilterWeight field gives the filter weights used to adjust the IR based precipitation estimates towards the microwave HQ precipitation estimates.

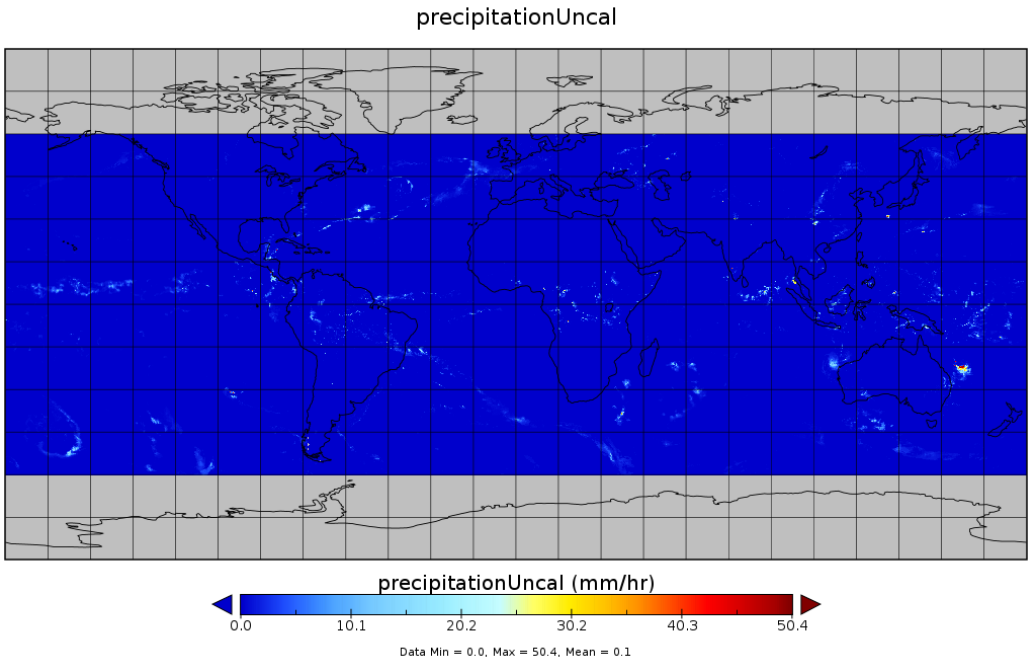


Figure 5.10: The IMERG HH precipitationUncal field for the time slot 2015-09-03 12:00. The precipitationUncal field is the combination of the IR based and microwave HQ precipitation estimates.

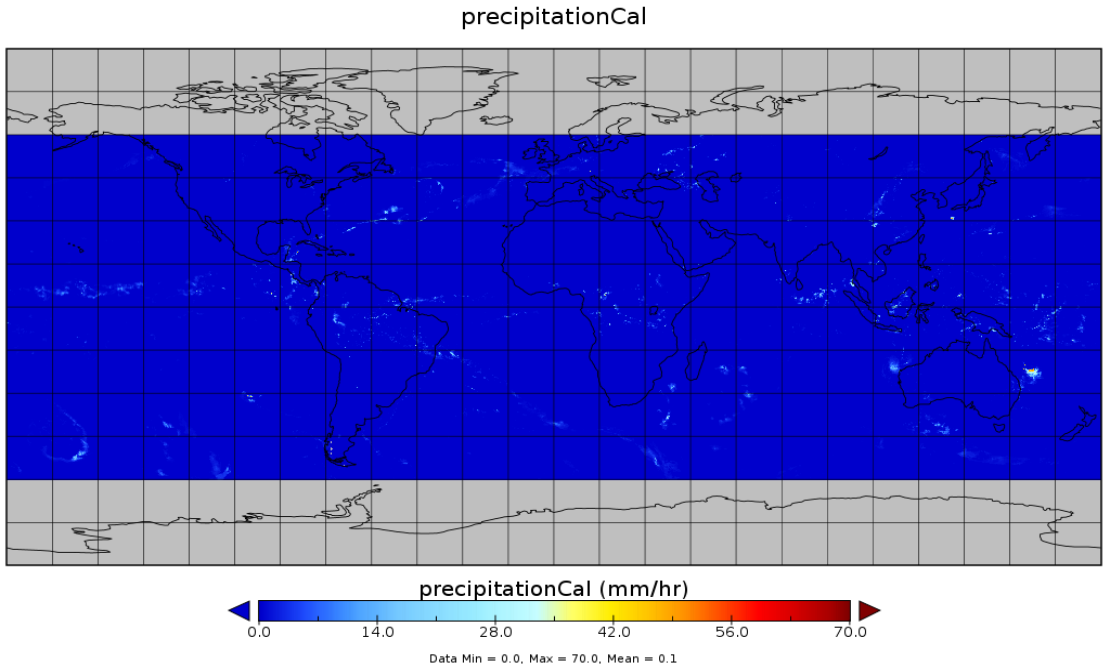


Figure 5.11: The IMERG HH precipitationCal field for the time slot 2015-09-03 12:00. The precipitationCal field is the combination of the IR based and microwave HQ precipitation estimates shown in Figure 5.10, but now adjusted to match monthly rain gauge distributions.

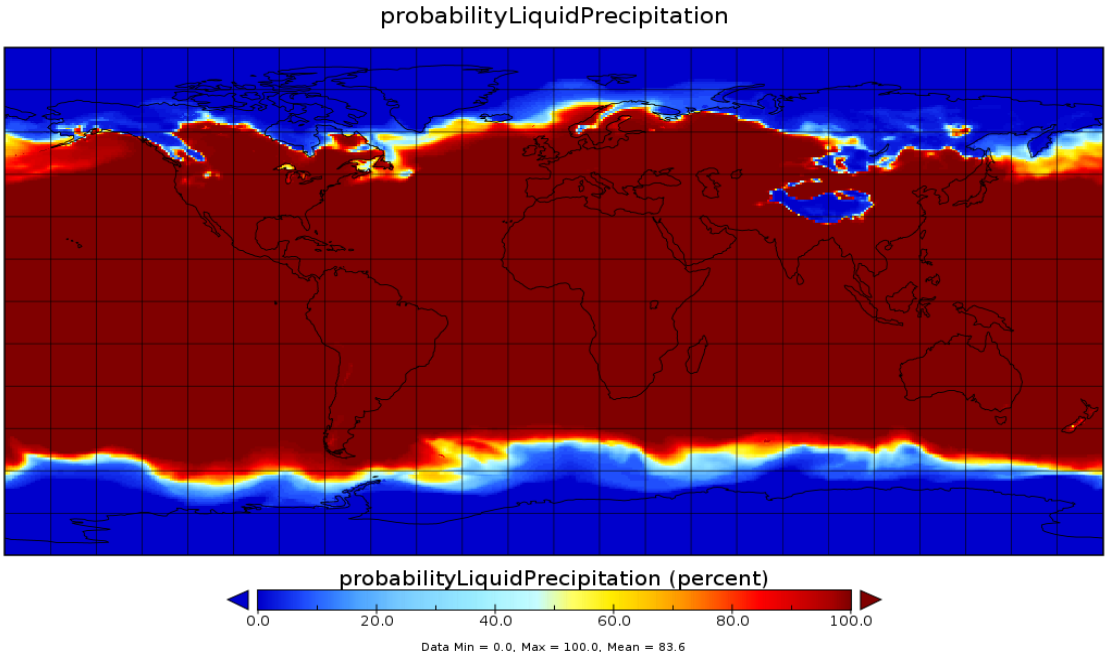


Figure 5.12: The IMERG HH probabilityLiquidPrecipitation field for the time slot 2015-09-03 12:00. The probabilityLiquidPrecipitation field defines the probability of observed precipitation being in liquid form rather than ice, snow or hail.

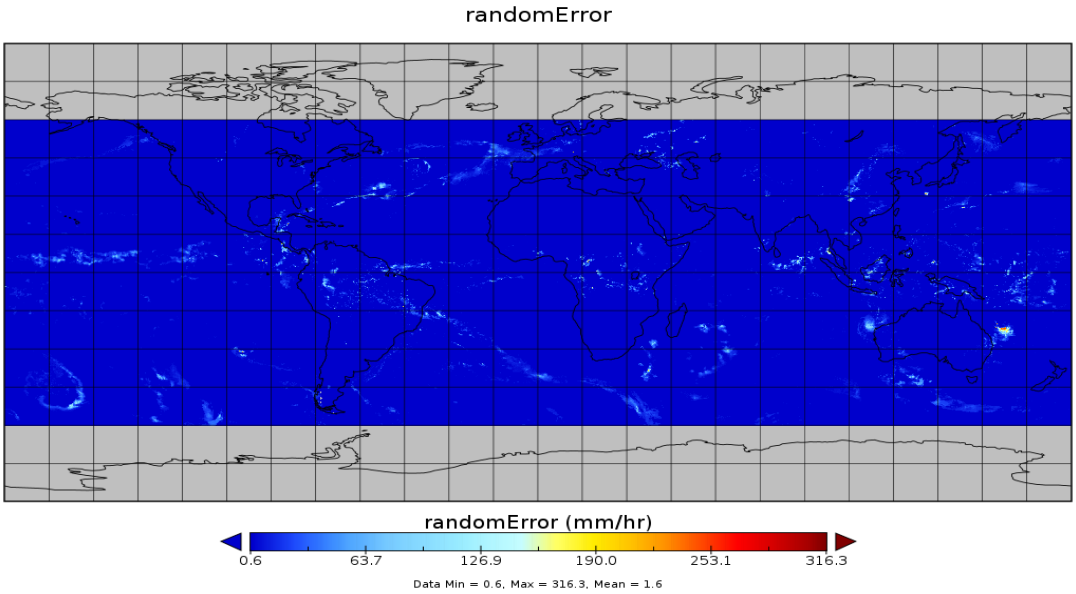


Figure 5.13: The IMERG HH randomError field for the time slot 2015-09-03 12:00. The randomError field defines the Kalman filter error variance in mm/hr.

5.1.3 GSMaP rainfall product

Table 5.2: Basic characteristics of the GSMaP hourly rainfall product.

Product name	**GSMaP NRT**
Source	http://http://sharaku.eorc.jaxa.jp/GSMaP/
Spatial resolution	0.1°
Temporal resolution	1 hr
Start date	2008-08-12
End date	Present
Latency	6 hrs

A full technical description of the GSMaP rainfall product is given in Kubota et al. (2007), and Tian et al. (2010). According to their website:

> 'We offer hourly global rainfall maps in near real time (about four hours after observation) using the combined MW-IR algorithm with GPM-Core GMI, TRMM TMI, GCOM-W AMSR2, DMSP series SSMIS, NOAA series AMSU, MetOp series AMSU and Geostationary IR data. Background cloud images are globally merged IR data produced by NOAA Climate Prediction Center (CPC), using IR data observed by JMA's MTSAT satellite, NOAA's GOES satellites and EUMETSAT's Meteosat satellites.'

The data files are available for download in raw binary grid which must be decoded and resampled for our analysis (see Subsection 5.2). The basic characteristics are given in table 5.2, while Figure 5.14 shows a sample rainfall map. Figure 5.15, which contains their Table 2, shows the level of complication of the blend of microwave rainfall retrievals used in this product.

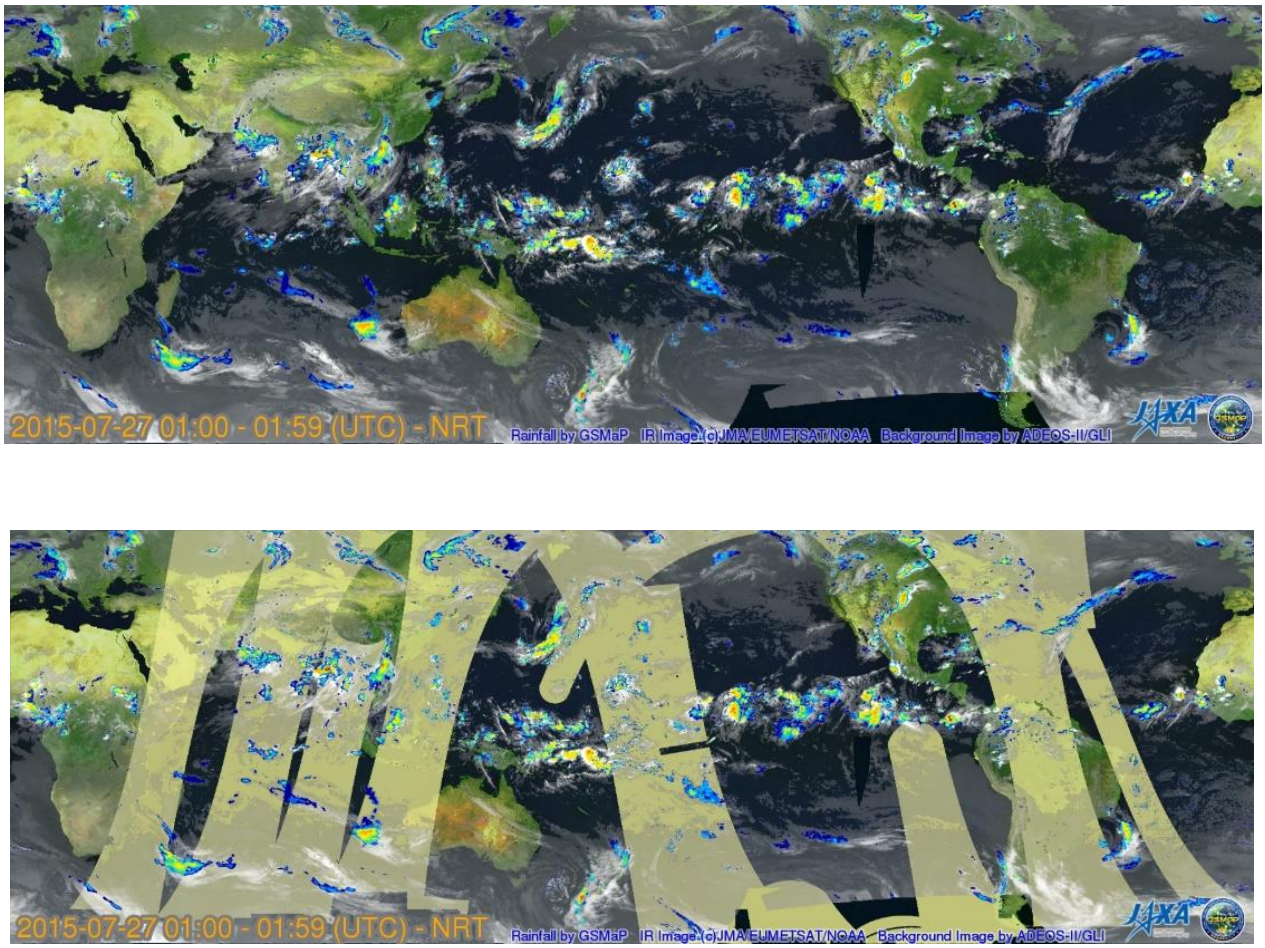


Figure 5.14: Example rainfall from the GSMaP product for a 1-hour time-slot on 2015-07-27 starting at 01:00. The lower image shows the overlays of the microwave overpasses in pale yellow.

Table 2 Summary of input low Earth orbit data

Satellite	Height (km)	Instrument	Category	frequency (GHz)	Note
GPM Core	407	GMI	imager	10.7, 19.4, 21.3, 37, 85.5, 166, 183.31±3, 183.31±7	Introduced into GSMaP_NRT since 3 Sep. 2014
TRMM	402	TMI	imager	10, 19, 21, 37, 85	
AQUA	705	AMSR-E	imager	7, 10, 19, 24, 37, 89	Not operational since 4 Oct. 2011
GCOM-W1	705	AMSR2	imager	7, 10, 19, 24, 37, 89	Introduced into GSMaP_NRT system since 1 Jul. 2013
DMSP-F13	833	SSM/I	imager	19, 22, 37, 85	Not operational since 18 Nov. 2009
DMSP-F14	833	SSM/I			Not operational since 24 Aug. 2008
DMSP-F15	833	SSM/I			Only rain over the ocean has been used since Aug. 2006. Not used since 3 Sep. 2014
DMSP-F16	833	SSMIS	imager/ sounder	19.4, 22.2, 37, 91.7, 60-63, 50-59, 150, 183.31±1, 183.31±3, 183.31±7	Introduced into GSMaP_NRT system since 11 Jun. 2010
DMSP-F17	850	SSMIS			
DMSP-F18	850	SSMIS			Introduced into GSMaP_NRT system since 1 Jul. 2013
DMSP-F19	850	SSMIS			Introduced into GSMaP_NRT system since 25 Mar, 2015
NOAA-N18	870	AMSU-A/ MHS	sounder	23.8-89.1 (AMSU-A), 89, 157, 183.311±3, 183.311±5, 190.311 (MHS)	Introduced into GSMaP_NRT since 3 Sep. 2014
NOAA-N19					Introduced into GSMaP_NRT system since 1 Aug. 2011
MetOp-A	817				
MetOp-B					Introduced into GSMaP_NRT since 3 Sep. 2014

Figure 5.15: Details of low earth orbit microwave observations incorporated in the GSMaP product (Table 2 from the GSMaP user guide).

5.2 Analysis and results

In this Subsection we describe the data-sets used and present some comparisons with the currently used TRMM 3B42RT rainfall product.

We downloaded several hundred Gigabytes of raw data; each set was in a different file format. This exercise took over two weeks to achieve and we then had to develop tools to read and manipulate each set because they were each differently organised.

The comparative analysis was carried out for the 13-month period 2014-04-01 through 2015-04-30. The ending date was decided by the most recent available IMERG HH Final rainfall products. More recent data are available for the IMERG HH Early and Late products; however, a full archive of these products is not available (at least we were unable to locate such an archive on any of the public NASA data portals). The IMERG HH Early/Late products are only available in a rolling window covering the previous 5 months (see Subsection 5.1.2), so we elected to carry out these early comparisons using the longer overlapping record offered by the IMERG HH Final product.

In order to make meaningful comparisons during the analysis period, we spatially resampled each of the rainfall products from their native grid resolution (of 0.1° and 0.25° respectively) onto a 0.125° grid, matching the HYLARSMET grid. Each of these resampled grids was stored in a Network Common Data Form (NetCDF) file at its native time resolution for further analysis.

5.2.1 Spatial comparisons

To get a feeling for the general behaviour of each product we started by analysing the spatial properties of the rainfall estimates. Figures 16 to 18 show the accumulated rainfall total for the 13-month analysis period 2014-04-01 to 2015-04-30 for each product. There is a very general agreement between the products with an increasing rainfall gradient from relatively dry over the western parts and the eastern portion of South Africa being wetter. This is as expected, so we are comfortable that the gross rainfall properties are reasonable. All of the products exhibit much higher rainfall over the ocean off the east coast when compared to the rainfall amounts estimated over the nearby coastal areas.

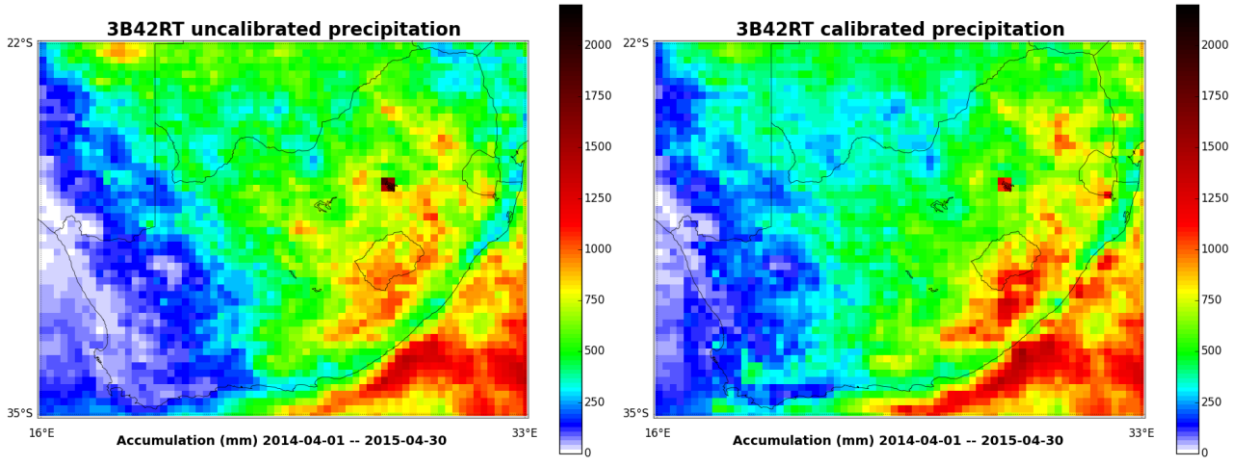


Figure 5.16: Total rainfall accumulation for the 13-month period 2014-04-01 to 2015-04-30 from the two 3B42RT combined fields. The precipitation field is the one we currently use for the HYLARSMET soil moisture modelling effort. Note the increased rainfall over the southern Cape mountains in the 'calibrated' precipitation field compared to the 'uncalibrated' version.

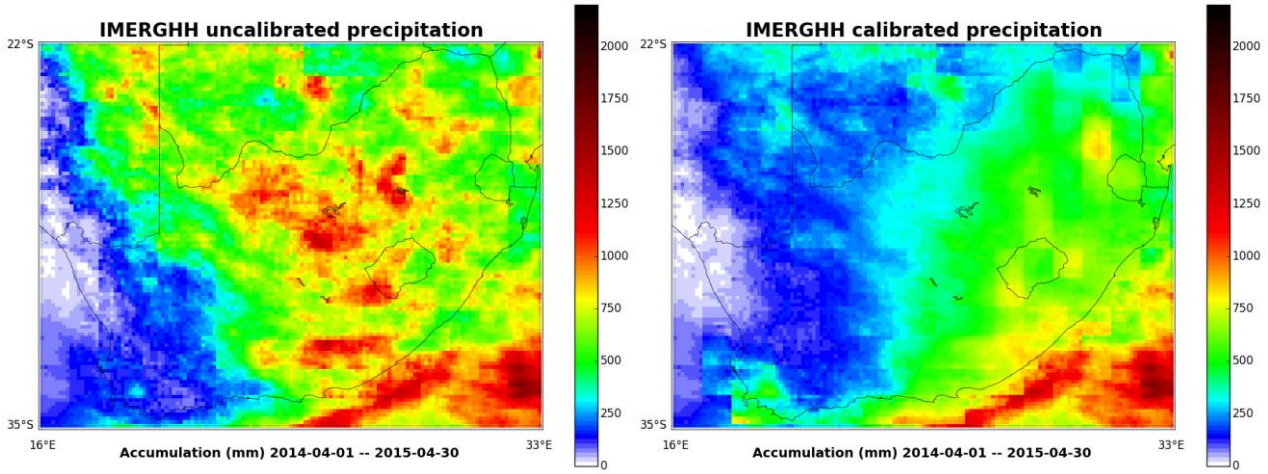


Figure 5.17: Total rainfall accumulation for the 13-month period 2014-04-01 to 2015-04-30 from the two IMERG combined fields. Note the increased rainfall over the southern Cape mountains in the 'calibrated' precipitation field compared to the uncalibrated version (this behaviour is similar to that shown in Figure 5.16. The rainfall over the interior is also significantly reduced by applying the 'calibrations'.

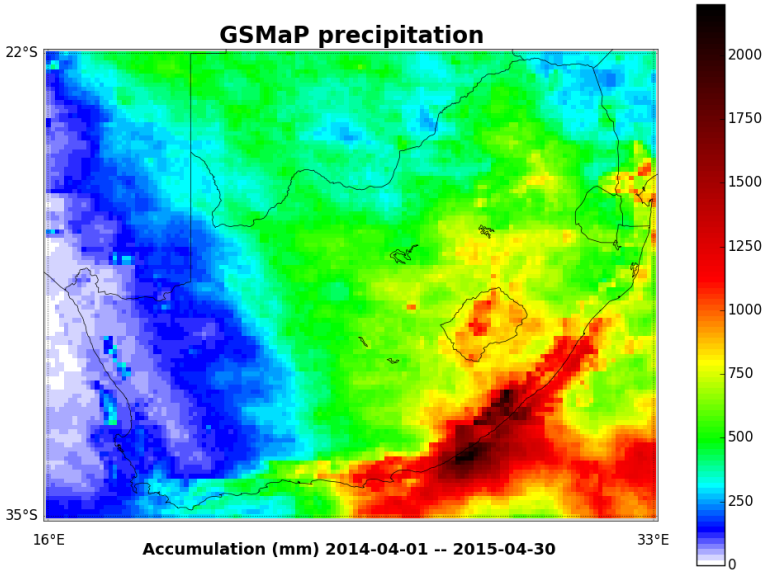


Figure 5.18: Total rainfall accumulation for the 13-month period 2014-04-01 to 2015-04-30 from GSMaP.

Figure 5.19 shows the proportion of missing time slots in the record for each grid box. TRMM 3B42RT and IMERG HH both have a low proportion of missing data, because they fall back on the IR rainfall field when there are problems with the microwave derived estimates. The IMERG map is smoother, but there are occasional problems with 3B42RT when the algorithm missed data for both microwave and IR (it is unclear why this occurred, but special processing was required in the analysis and use of the product as a result, which we undertook). The GSMaP product has data missing mainly due to failed microwave retrievals. The architects of this product do not fill in these missing retrievals with IR data as done for 3B42RT and IMERG, but instead report the missing value as a microwave retrieval failure. Nevertheless, even though the GSMaP missing proportion is much larger than the other two, it is still well below 10% during the analysis period.

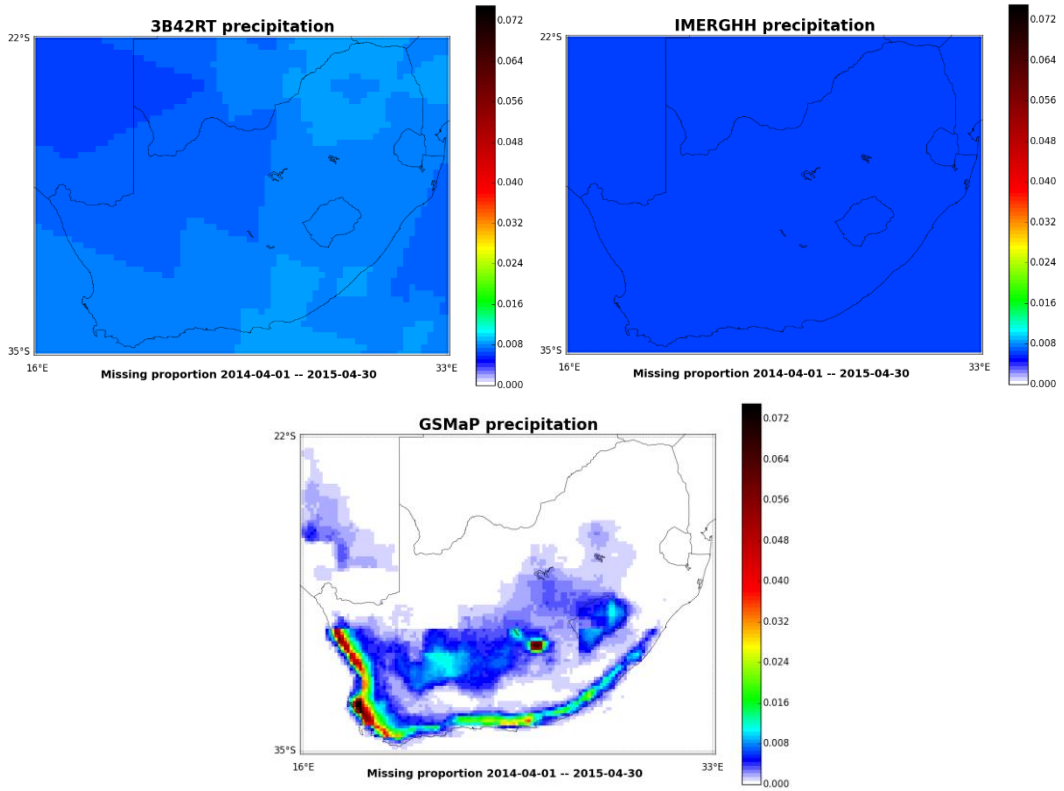


Figure 5.19: Maps comparing the proportion of missing data for each of the products during the 13-month analysis period. Although the GSMaP product appears to have more missing data, it is worth noting that the maximum missing is still less than 10% of the observation times.

Figures 5.20 to 5.22 map the maximum estimated rainfall rates during the analysis period for each product. In general, the 3B42RT maximum rates are far lower than those of the other products. This is mostly due to the larger grid box (0.25°) over which rainfall rates are averaged compared to the grid box for IMERG and GSMaP (0.1°). The sampling intervals also vary between products; 3-hourly for 3B42RT, 1 hourly for GSMaP, and half hourly for IMERG. The GSMaP product shows anomalous high maximum rainfall rates over the coastal regions, which is a concern.

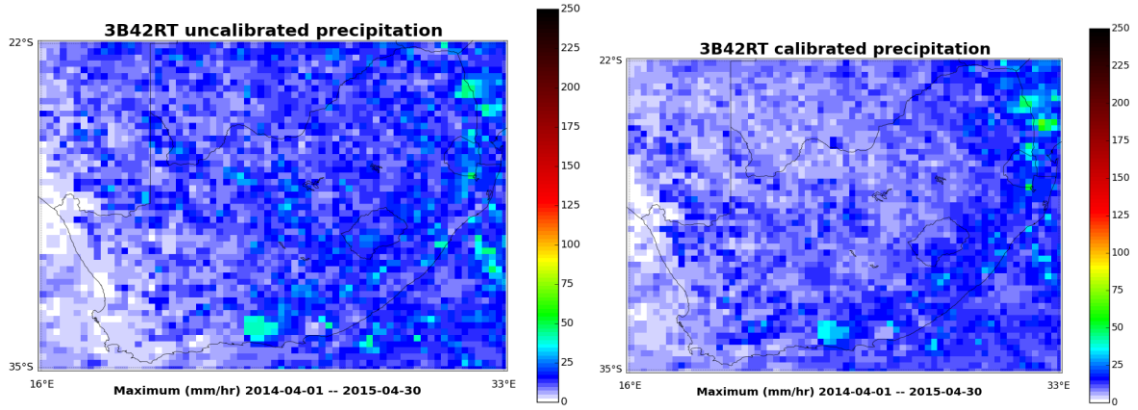


Figure 5.20: Maximum rainfall rate on each 0.125° block for the period 2014-04-01 to 2015-04-30 from the two 3B42RT combined fields. Note the generally decreased rates over much of the interior for the 'calibrated' precipitation field compared to the uncalibrated version. The overall maximum rates from 3B42RT are considerably lower than the maximum rates shown in Figures 5.21 and 5.22.

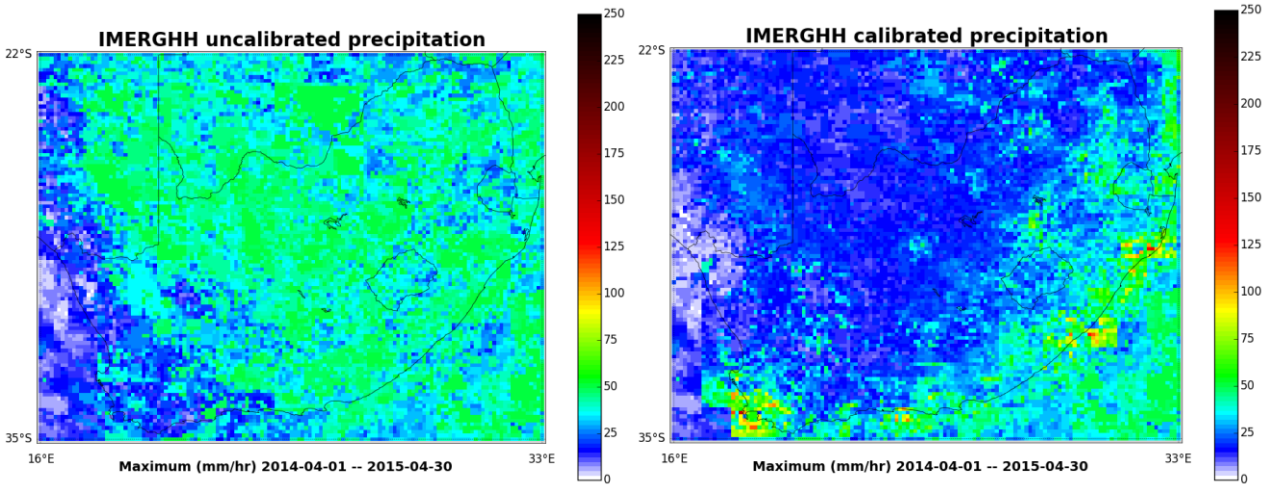


Figure 5.21: Maximum rainfall rate for the period 2014-04-01 to 2015-04-30 from the two IMERG combined fields. Note the generally decreased rates over much of the interior for the 'calibrated' precipitation field compared to the uncalibrated version, which is much more realistic. There are also a few coastal locations (near cities and especially the Cape) showing a marked increase in the expected rainfall rates in the 'calibrated' product.

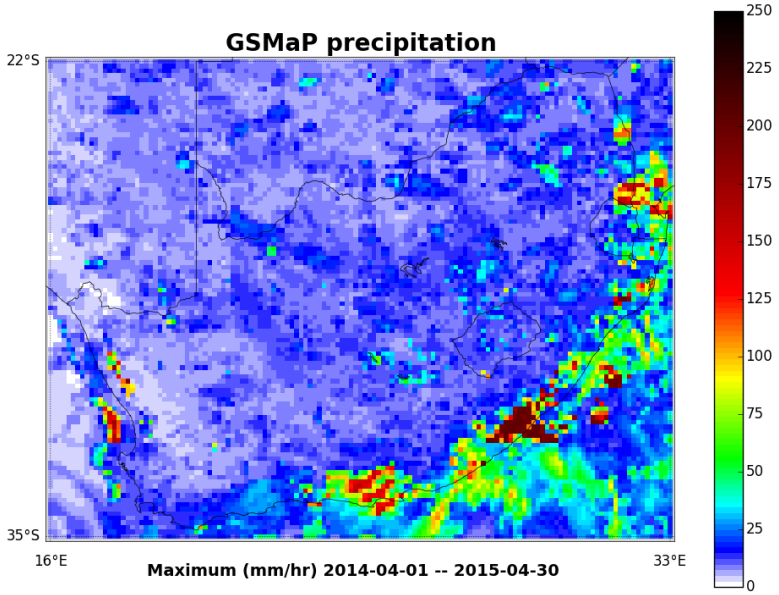


Figure 5.22: Maximum rainfall rates for the period 2014-04-01 to 2015-04-30 from GSMaP. Note the extremely high maximum rates experienced along the east coast and off the west coast.

5.2.2 Time-series and frequency distributions at selected locations

We selected several locations around South Africa with different rainfall regimes and considered the behaviour of each rainfall estimate in some detail. Figures 5.23 through 5.29 compare the time-series and cumulative distribution functions (cdfs) for each estimate at 6 different locations around the country. The series are not long, but give a feel for the comparison between the sources of information. Figure 5.28 is a comparison of cdfs of daily data over 10 years at Polokwane, and suggests that the IMERG product is more faithful to measured rainfall than either TRMM or GSMaP, however this is only one location where we have useful data. In all the other cdfs of rainfall rates, the IMERG estimates of heavy rain-rates are greater than the others' estimates, in line with our experience with raingauge estimates compared to TRMM.

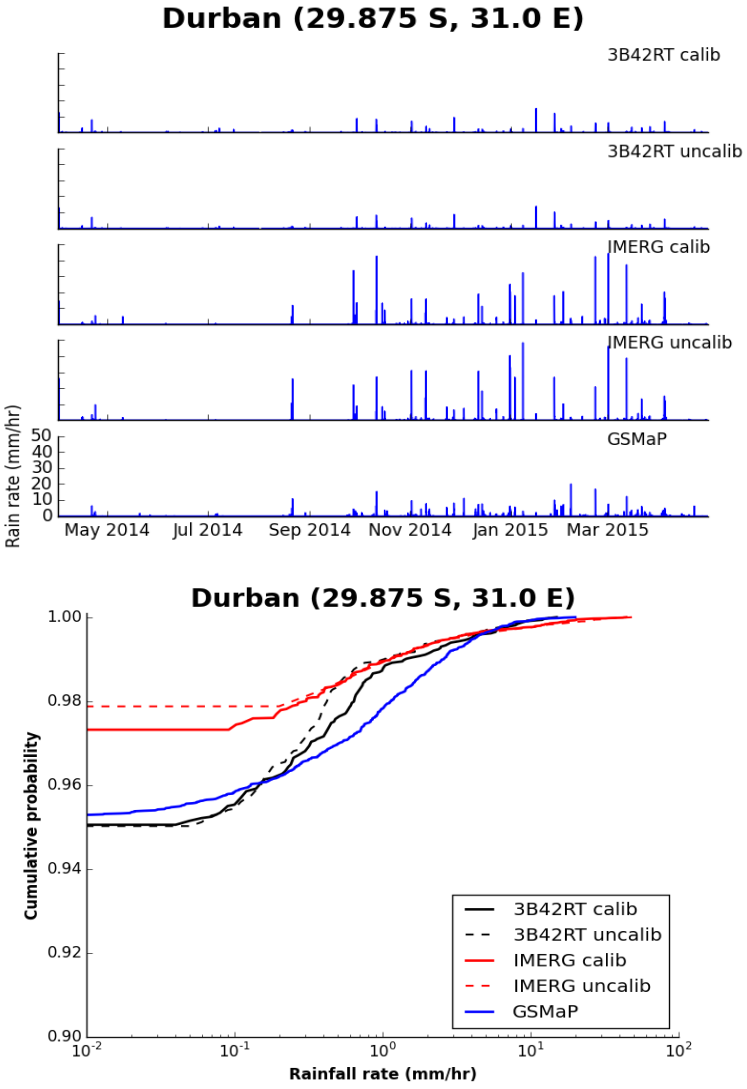


Figure 5.23: Comparative time-series and cumulative distribution functions for the estimated rainfall rates during the period 2014-04-01 to 2015-04-30 at Durban.

In most of these Comparative Figures of time series and their cdfs, we note that there is very poor timing matching between the sources of the time series. However, there is not much difference between the calibrated and uncalibrated sets of the individual products, except for Cape Town in Figure 5.24, Pretoria in Figure 5.26 and Polokwane in Figure 5.27.

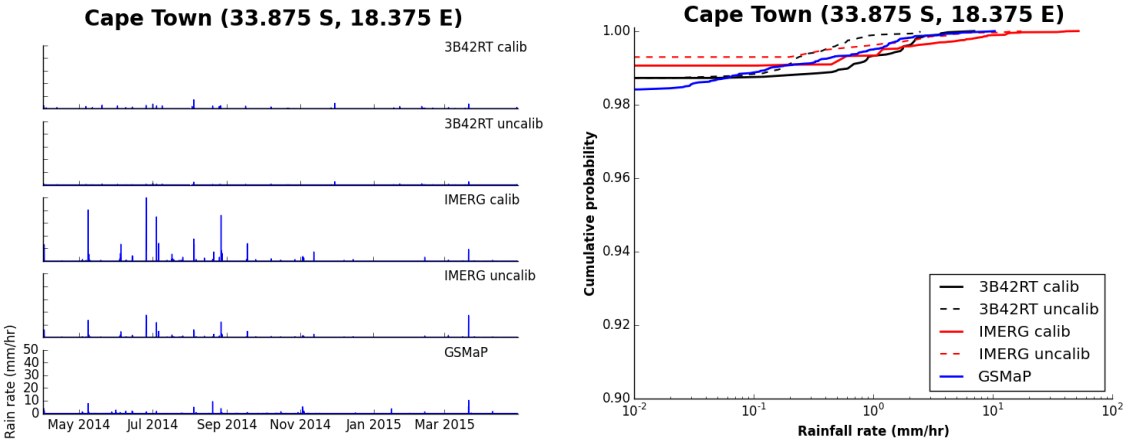


Figure 5.24: Comparative time-series and cumulative distribution functions for the estimated rainfall rates during the period 2014-04-01 to 2015-04-30 at Cape Town.

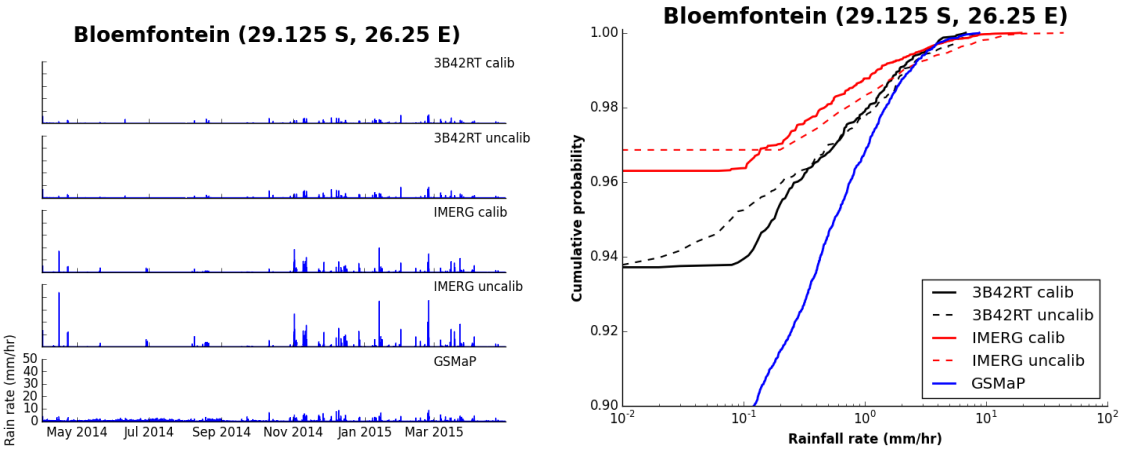


Figure 5.25: Comparative time-series and cumulative distribution functions for the estimated rainfall rates during the period 2014-04-01 to 2015-04-30 at Bloemfontein.

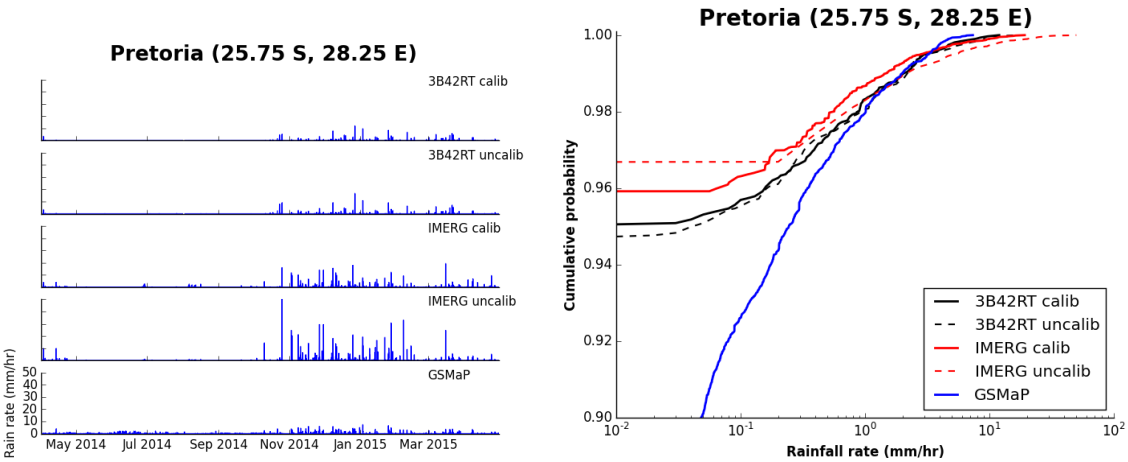


Figure 5.26: Comparative time-series and cumulative distribution functions for the estimated rainfall rates during the period 2014-04-01 to 2015-04-30 at Pretoria.

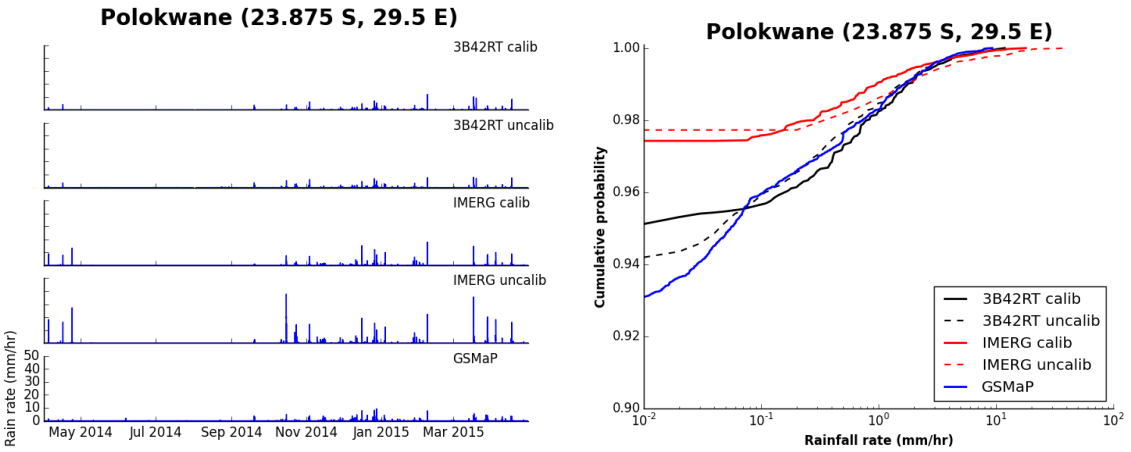


Figure 5.27: Comparative time-series and cumulative distribution functions for the estimated rainfall rates during the period 2014-04-01 to 2015-04-30 at Polokwane.

We do not have gauge rain-rate data for the above period, but we do have daily data for Polokwane over 10 years prior – see Figure 5.28 below. The character of the gauge compared to the TRMM and GSMaP cdfs in Figure 5.28 (gauges report higher maxima than TRMM) is similar to the comparison of IMERG to GSMaP in Figure 5.27, which is in the right direction, therefore lending credence to the IMERG product compared to GSMaP.

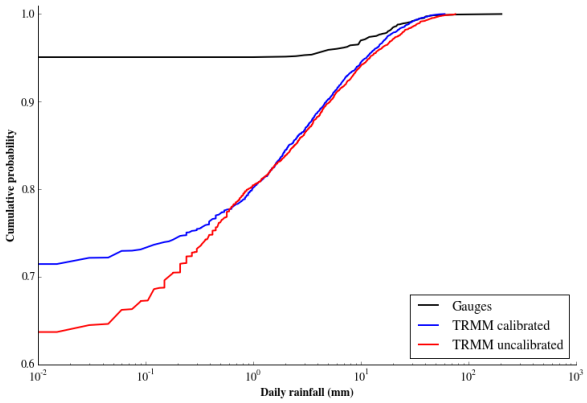


Figure 5.28: Frequency distributions of daily rainfall totals estimated by Gauge Block averages and TRMM at Polokwane, 2000 to 2010.

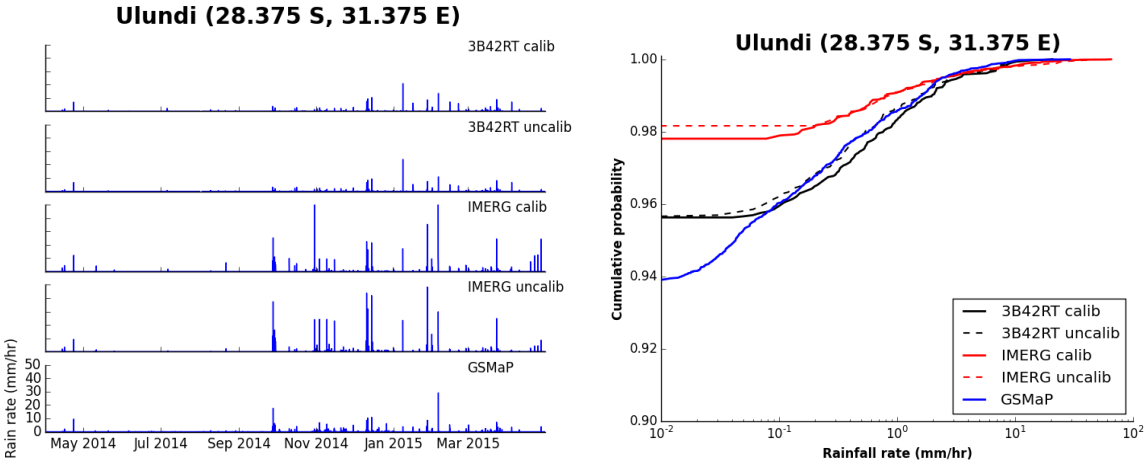


Figure 5.29: Comparative time-series and cumulative distribution functions for the estimated rainfall rates during the period 2014-04-01 to 2015-04-30 at Ulundi.

5.3 An explanation for the mismatch of gauge and TRMM rainfall

Based on Figures 5.23 to 5.29, it seems that there is poor correspondence between gauge and TRMM estimates, not only temporally, but also because the distribution functions are quite different – the gauges are 'drier', as measured by their P[0] values, in all cases. To explain this phenomenon, we include below a few paragraphs of work done on the MAP project (Sinclair and Pegram, 2013b), whence we import the relevant Subsection, indented left and right and verbatim between inverted commas, with the labelling unchanged.

'9.6.1. An experiment to determine the link between a true rainfall field and averaged gauges.

In a numerical experiment conducted by Prof Bardossy [private communication], a set of ten thousand daily images of rainfall on an area 25 km square [the size of a TRMM pixel/block] were generated. Each 1 km pixel on the square was populated with properly spatially correlated 'rainfall', generated by a Fourier transform and using a fixed Exponential distribution for the amounts, for each set. The correlation length of the spatial variogram was set at 20 km, so the generated rainfields were relatively variable, similar to fields of convective rainfall.

The 'true' block average on each image was calculated by numerically averaging all the pixel values on each 'day'. Then a set of evenly spaced points was carefully selected from each field as if they were gauge locations; the numbers of sites per image chosen were 1, 2, 4, 8 and 16 and the location of each site chosen was kept the same for the set of 10 000 estimates ('days'). Numerically averaging the individual 'daily' gauge samples and using these to compute their cfds yielded sets of gauge block average estimates to compare against the distribution of the 'true' spatial averages of the set of full fields. The numerical averaging of the gauge values was done by calculating their simple mean, but because they were equally spaced, their averages would match those of Multiquadrics. The following results are thus not only illustrative but very useful.

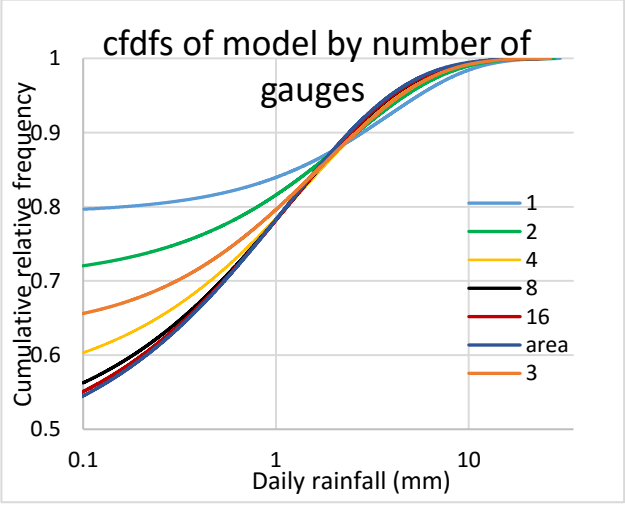


Figure 5.30: Cumulative frequency distributions of block averages of rainfall above 0.1 mm on gauges over a 25 by 25-pixel square in 10 000 days. Blue: 1 gauge; green: 2 gauges; brown: 3 gauges; yellow: 4 gauges; black: 8 gauges; magenta: 16 gauges; navy blue 625 sites (full square).

The cumulative frequency distributions (cfds) of the range of populations of gauge averages in the square shown in Figure 5.30 are somewhat surprising and a summary of these follows.

Note that the lower bound of the vertical axis in the figure has been set at a probability of 0.5 [5000 days] to help in visualising the differences. We truncate the lower estimates of 'precipitation' at 0.1 mm. The obvious reason is that the value of p_0 at a threshold of 0.1 mm depends heavily on the number of gauges when they are few, such as 1, 2 or 3, as shown by the blue, green and brown curves. There is not much difference between cumulative cfdfs when 8 (red curve) or 16 gauges (black curve) are considered and the convergence of the latter to the 'true' (navy blue) curve derived from 625 sites is quite good.

If we were to threshold the curves at 1 mm, as shown by the vertical axis (because below that measurement is technically a 'trace', which suggests that in that interval there is poor sampling of fine drizzle), then 4 gauges or more are quite adequate for a good areal estimation. A comparison of rainfall amounts in the vicinity of 0.95 (9500 days out of 10 000) shows that the 1-gauge curve reads 4.3 which overestimates the areal daily rainfall of 3.3 by about 1 mm, but that the curves of 4 gauges and above are quite faithful in their estimates of the true values. It seems we only need to be concerned when there are between 1 and 3 gauges used to get average daily rainfall on a 625 km^2 area.'

The conclusion is that there is typically a poor match between TRMM daily rainfall and that collected by singleton gauges. One is only likely to obtain a good match in wetter areas where there are 4 or more gauges in each TRMM block. More importantly, the TRMM rainfall over each block is quite good on the average over a month (as against individual days), as confirmed in a personal discussion between Geoff Pegram and George Huffman (the 'father' of TRMM) at the European Geosciences Union's Spring meeting in Vienna in 2015.

5.4 Summary of Section 5

In this Subsection we presented the results of an investigation into choosing a replacement near-realtime rainfall product to drive HYLARSMET and its extension into SADC. This is necessary because the currently used TRMM 3B42RT rainfall product will be discontinued by 2017 due to the decommissioning of the TRMM satellite in early 2015. In addition, the new GPM core satellite carries updated instruments and is central to global rainfall estimations moving into the future.

We identified two near-realtime rainfall products that use information from the GPM core observatory. These are the NASA product IMERG, and the JAXA product GSMaP. We were able to download a 13-month dataset for each of these products that overlaps in time with the TRMM 3B42RT product. These large datasets were then resampled onto the HYLARSMET 0.125° grid over South Africa and parts of its neighbours. We made various comparisons between the datasets as described in Subsection 5.2.

Although there are some differences between the products, we chose to proceed with the IMERG rainfall estimates in place of TRMM. There are several reasons for this choice:

- IMERG has fewer missing data slots

- The GSMaP product exhibits unusually high rain rates in places (especially the coastal regions) and this is of concern.

- IMERG provides both a raw satellite algorithm (uncalibrated precipitation) and a post-calibrated estimate. GSMaP provides only a post-calibrated estimate. If we intend to make our own adjustments based on gauge information, we prefer to adjust the pure satellite estimate directly.

- The GSMaP product is available in a flat binary file format, which is fast and convenient to read (if you know the file structure ahead of time). However, there is no self-describing meta-data file to assist in unpacking the information. IMERG is available in several different formats and via a number of delivery mechanisms. One of these OPeNDAP allows spatial subsets of the global file to be downloaded, which saves both bandwidth and subsequent storage space.

Based on the investigation reported here, we adapted our HYLARSMET and EXSMET soil moisture modelling procedures to ingest the new IMERG product.

-----ooOoo----

6. Determine the usefulness of SMOS remotely sensed soil moisture over RSA

6.1. Introduction and SMOS mission overview

In a previous project (Pegram et al., 2010) we had hoped to use the soil moisture estimates from SMOS (European Space Agency's Soil Moisture and Ocean Salinity satellite) in our work, but the SMOS team had not yet begun disseminating data at that time. In the intervening time, the SMOS science team has amassed a large data archive and made this available via a very user friendly online service. The purpose of Deliverable 7, ***'Determine when the Soil Moisture estimates of the European Space Agency's SMOS mission will be ready for use in model inter-comparison with EXSMET and other remote sensing estimates, and exploit it'*** was to investigate the SMOS-derived soil moisture data-sets available for Southern Africa, in order that we could carry out inter-comparisons in an effort to validate our modelling efforts. This task required a large amount of data to be downloaded and archived locally, before we were able to develop new software to extract relevant information for the Southern African region from the global SMOS data-sets.

This Section of the report is arranged into three Subsections: 6.1.1. The SMOS mission, 6.1.2. Analysis of the data obtained and 6.1.3. Summary.

6.1.1. The SMOS mission

In this Subsection we describe the characteristics of the SMOS mission (Kerr et al., 2001), describing the SMOS satellite and the relevant soil moisture products derived from SMOS observations.

The SMOS mission was designed to provide global measurements of L-band brightness temperatures, resulting in soil moisture and ocean salinity data sets from space. The mission objectives were: (1) to provide global volumetric soil moisture estimates with an accuracy of 0.04 m^3/m^3 at a spatial resolution of 35-50 km and a temporal sampling interval of 1-3 days; (2) to provide global ocean salinity estimates with an accuracy of 0.1 practical salinity scale units for a 10-30 day average for an open ocean area of 200 x 200 km^2; (3) to provide daily sea ice thickness estimates based on MIRAS observations, for the Northern hemisphere, with a spatial resolution of 100 x 100 km^2 up to maximum thickness values of 50 cm. SMOS observations additionally provide valuable information on vegetation and snow covered surfaces.

The payload of the SMOS satellite consists of the Microwave Imaging Radiometer using the Aperture Synthesis (MIRAS) instrument, a passive microwave 2-D interferometric radiometer, operating in L-band (1.413 GHz, 21 cm) within a protected wavelength/frequency band. The SMOS mission is based on a sun-synchronous orbit (dusk-dawn 6am/6pm). SMOS measurements are made over a range of incidence angles (0° to 55°) across a swath of approximately 1000 km with a spatial resolution of 35 to 50 km. MIRAS can provide measurements in dual and full polarisation, with the latter being its present operating mode.

Figure 6.1: An artist's representation of the SMOS satellite in orbit.

In a recent update study, Kerr et al. (2016) presented an overview of the performance of SMOS in global soil moisture estimation. They concluded that the data products are of high quality and meet the goals of the mission. However, their work did not include much analysis of the SMOS product performance in Southern Africa.

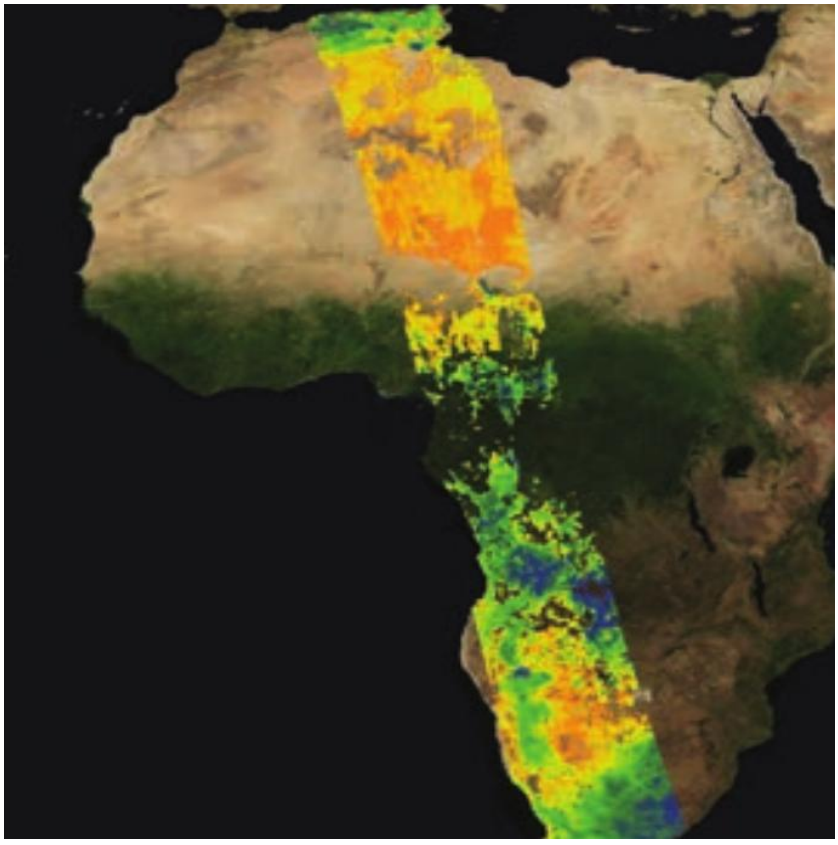


Figure 6.2: An illustration of the SMOS satellite swath extent.

6.1.2. SMOS derived soil moisture products

There are several soil moisture products available from the SMOS science team and their partners. These need to be parsed and selected from the global data set which contains the following products.

Level 2 Soil Moisture, Vegetation Optical Depth and ancillary land products:
These comprise the retrieved swath-based soil moisture, vegetation optical depth and other ancillary data derived during processing (surface temperature, roughness parameter, dielectric constant and brightness temperature retrieved at top of atmosphere and at surface) with their corresponding uncertainties. The latency of the products is 8-12 hours. Spatial resolution of grid: 15 km (ISEA 4H9 grid).

Level 3 CATDS Soil Moisture, Vegetation Optical Depth and ancillary land products:
The daily product contains filtered data processed from the ESA L1B product. The L3 processing algorithm involves the use of a multi-orbit approach, in which retrievals are done using three successive orbits within a seven-day moving window. When several retrievals are available for a given day, the best estimation of soil moisture is selected for each grid point. Besides soil moisture, the vegetation optical depth, surface roughness and dielectric constant are also included in the final global-scale product. Spatial resolution of grid: 15 km (ISEA 4H9 grid).

The 3-day, 10-day and monthly products are produced by performing a temporal aggregation of the daily global maps, after filtering for detected events. In the case of the 3-day product a moving window is used, resulting in daily availability of global maps. All products contain soil moisture, vegetation optical depth and radio frequency interference (RFI) statistics. The 3-day product additionally contains dielectric constant data. The 10-day product additionally contains surface roughness data, and minimum, maximum and median values of soil moisture. Spatial resolution of grid: 25 km (EASE grid version 2).

Level 3 SMOS-BEC Soil Moisture, Vegetation Optical Depth and ancillary land products:
Level 3 SMOS Barcelona Expert Centre (SMOS-BEC) daily maps of soil moisture, vegetation optical depth and dielectric constant are available on a 15 km (ISEA 4H9) grid; daily maps of soil moisture are also available on a 25 km (EASE) grid. The former (ISEA 4H9 grid) is generated from the ESA L2 soil moisture data set without spatial or temporal averaging. The latter (EASE grid) is generated by simple spatial averaging of the ESA L2 soil moisture data set, taking into account only the last orbit measurements in each grid cell. In both cases, ascending and descending orbits are processed separately and the maps are provided at global scale. 3-day averages, 9-day averages, monthly and annual maps of soil moisture are constructed by simple spatial averaging of ESA L2 data using all available orbits. The spatial averaging is computed in a 25 km (EASE) grid. Ascending and descending orbits are processed separately and the maps are provided at global scale.

Interim Summary: For our applications, we had to inspect, check and test the various products to decide upon a valid and applicable soil moisture estimate which closely follows our 3 hourly calculated SM variables over South Africa. This non-trivial exercise required a large amount of data handling and calculation. In any event, for our purposes we chose the level 2 swath based product because this allows for more flexibility to do our own aggregations in space and time.

6.2. Analysis and results

6.2.1. Data processing

- We downloaded the SMOS MIR SMUDP2 product for all orbits in the archive which contain data within our SA region.
- Our data archive is reasonably large at 44 Gb in size (5742 individual files) – this was all downloaded from the SMOS data archive and required the design of a usable storage system.

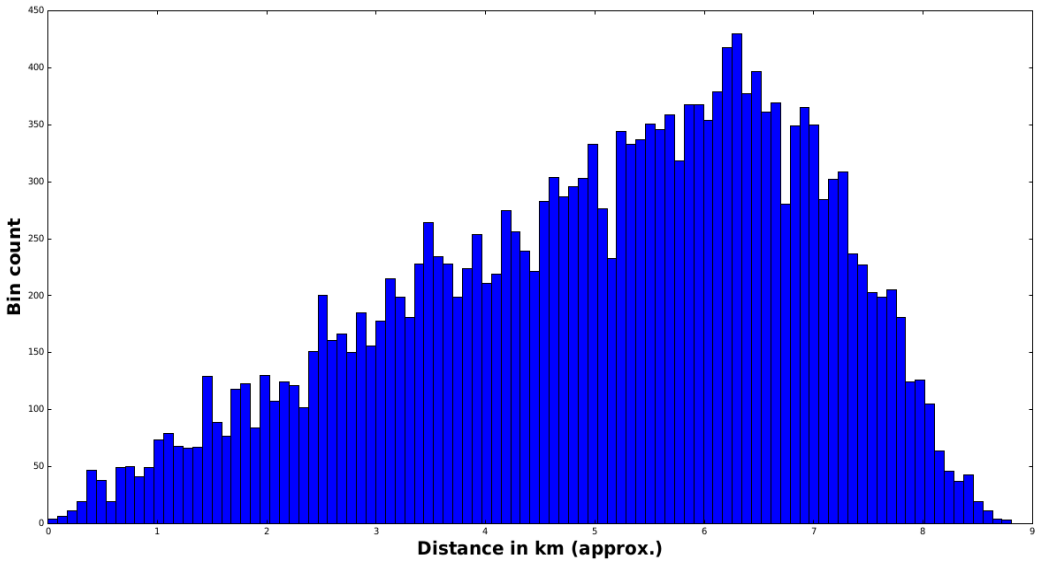


Figure 6.3: Histogram showing the approximate distances in kilometres between each HYLARSMET grid point and the nearest SMOS L2 grid point. The average is about 5 km. The SMOS grid resolution is 15 km while the HYLARSMET grid is at 12 km spacing (approximately). As a complication, the two grids are offset relative to each other, at different angles.

- Each file contains the data for a full orbit, in the ISEA 4H9 grid used by the SMOS products. These were nearest-neighbour resampled onto the HYLARSMET grid and stored in NetCDF format.
- Nearest neighbour resampling was done to obtain SMOS soil moisture estimates on our 0.125° HYLARSMET grid. Figure 6.3 shows the histogram of distances between the closest SMOS grid point and the HYLARSMET grid points. The closest point is typically within 5 km, which is reasonable given the SMOS grid resolution of 15 km and spatial resolution of approximately 40 km. This turned out to be a very laborious procedure which we had to automate.

6.2.2. Spatial comparisons

We have developed techniques to resample the SMOS data onto the HYLARSMET grid. Figures 6.4 to 6.9 give examples, the Figures are described in their individual captions.

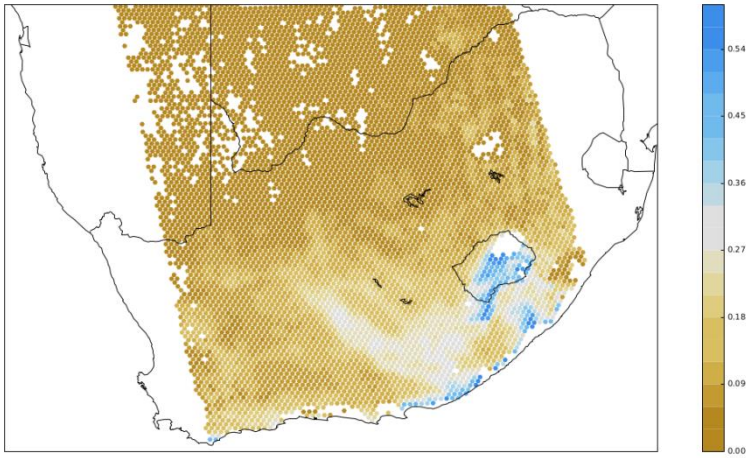


Figure 6.4: A single swath on the SMOS grid for the ascending (morning) pass on 2015-07-15. The colour scale indicates estimated volumetric soil moisture content in m^3/m^3.

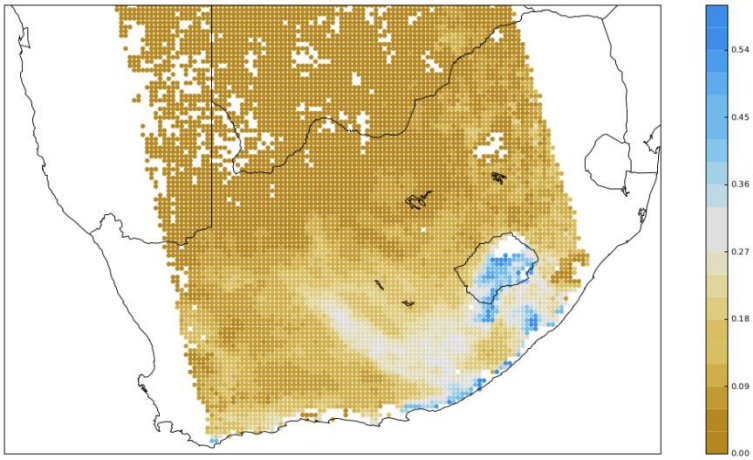


Figure 6.5: A resampled version of the single swath on the SMOS grid for the ascending (morning) pass on 2015-07-15 shown in Figure 6.4. Although this Figure appears very similar to Figure 6.4, careful inspection of the grid orientation shows that the two grids are offset and staggered relative to each other. The colour scale indicates volumetric soil moisture content in m^3/m^3.

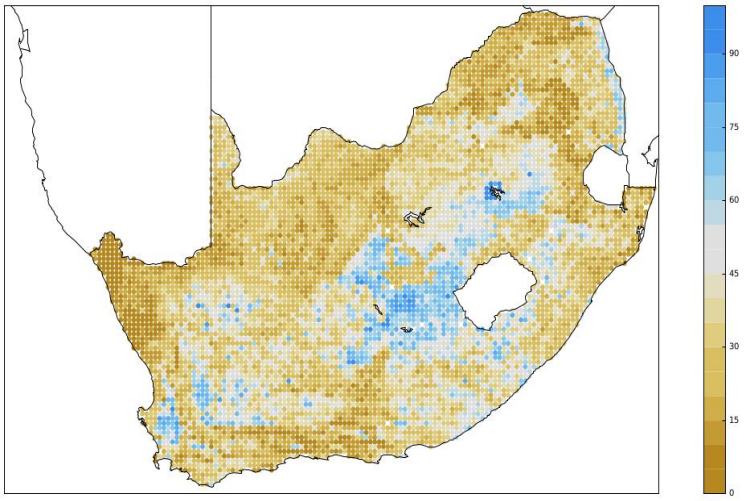


Figure 6.6: The HYLARSMET estimate matching Figure 6.4. The colour scale indicates effective saturation as a percentage – note the units and range do not match those of Figure 6.4.

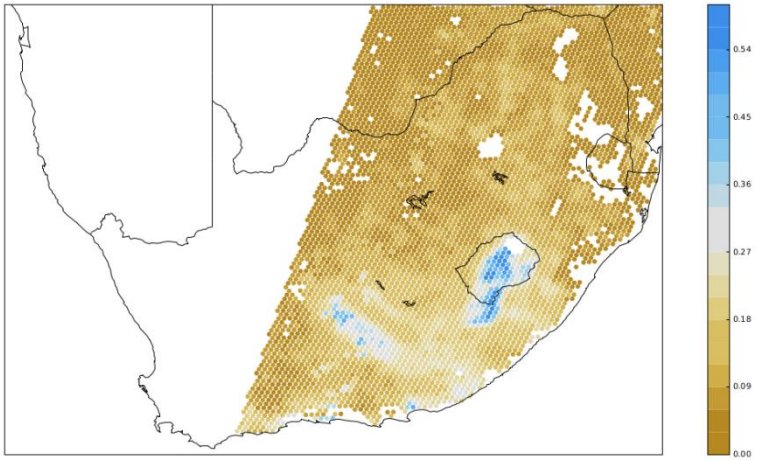


Figure 6.7: A single swath on the SMOS grid for the descending (afternoon) pass on 2015-07-15. The colour scale indicates volumetric soil moisture content in m^3/m^3.

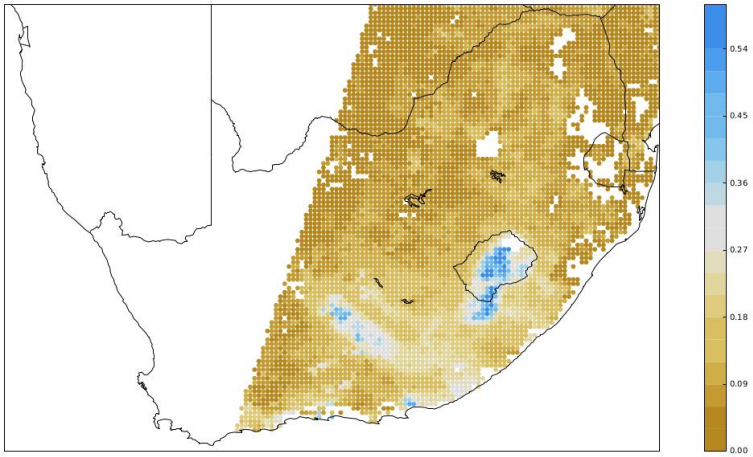


Figure 6.8: A resampled version of the single swath on the SMOS grid for the descending (afternoon) pass on 2015-07-15 shown in Figure 6.7. Although this Figure appears very similar to Figure 6.7, careful inspection of the grid orientation shows that the two grids are offset and staggered relative to each other. The colour scale indicates volumetric soil moisture content in m^3/m^3.

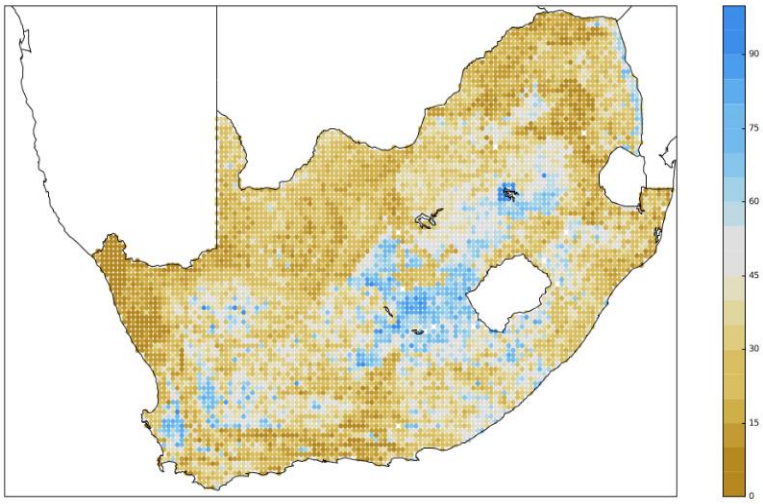


Figure 6.9: The HYLARSMET estimate matching Figure 6.7. The colour scale indicates effective saturation as a percentage – note the units and range do not match those of Figure 6.7.

6.2.3. Time-series comparison

In order to establish the correspondence in the temporal dynamics of the SMOS and HYLARSMET soil moisture estimates, we plotted the time history of each set for several locations. Figure 6.10 shows an example for a location (26.75E, 25.75S) in the eastern part of South Africa.

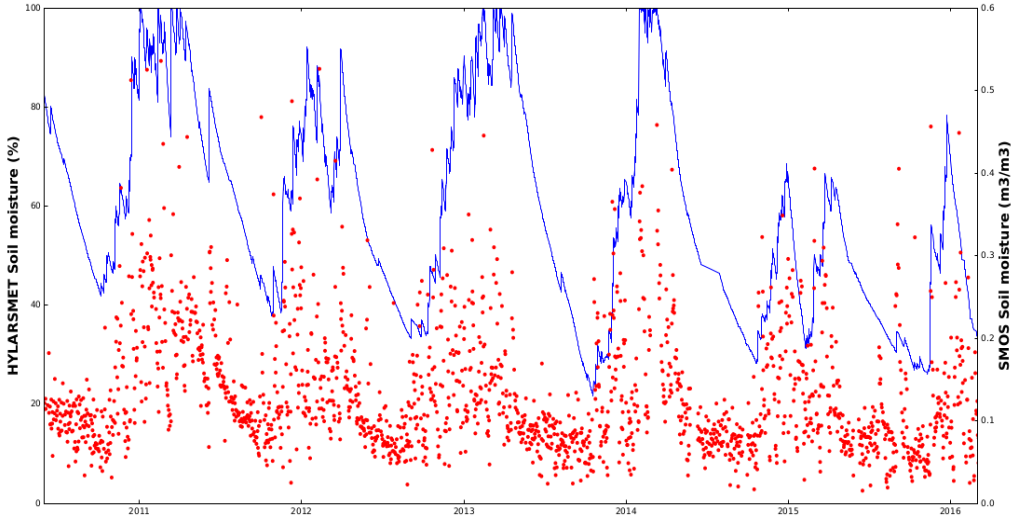


Figure 6.10: A comparison of the time-series at a single location (26.75E, 25.75S) for the full period of SMOS record in our archive. Blue line HYLARSMET, red dots SMOS L2. Note that the units are different and have separate scales, with the SMOS estimates being irregularly spaced in time.

The general seasonal cycles agree in Figure 6.10, and the drought in 2015/2016 is captured by both estimates (although it is more evident from the HYLARSMET soil moisture results). This feature is better captured in Figure 6.11, where we chose a pair of 2 year matching time series. We note at this juncture, that the rainfall product we are currently using is a blend (in time) of TRMM and GPM data from a different satellite source.

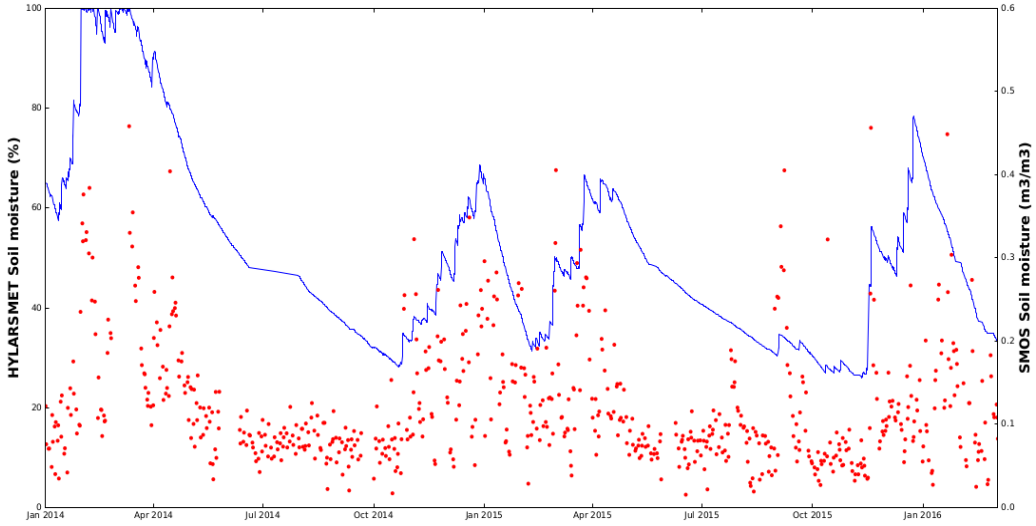


Figure 6.11: A comparison of the time-series at a single location (26.75E, 25.75S) for the 2-year period 2014-2016 of the SMOS record. Blue line HYLARSMET, red dots SMOS L2. Note that the units are different and have separate scales, with the SMOS estimates being irregularly spaced in time.

6.3. Section summary

The purpose of this deliverable was to investigate the SMOS derived soil moisture data-sets available for Southern Africa, in order that we could carry out inter-comparisons in an effort to validate our modelling efforts. We have determined that, although there are differences in the detail, the SMOS mission data are in fact ready, usable, useful and available, and have downloaded an archive of over 5 years. The data-set that we possessed at this stage was the SMOS level 2 swath data, which was manipulated and re-gridded using our analysis software, which was developed in-house. At this stage of the project, we had positively evaluated its usefulness and the research provided confidence in our modelling results.

-----ooOoo----

7. Upgrade catchment extraction tools to streamline hydrological calculations

This Deliverable 8 was defined as follows: ***'Determine how to best cope with the uncertainties associated with input parameters and forcing variables [TRMM in particular] when computing ensembles of historical and forecast data streams'.***

In previous work (Sinclair and Pegram, 2013a) we investigated the effect of uncertainty in the physical parameters that define the PyTOPKAPI model on the soil moisture state modelled in Land Surface Model (LSM) mode. There, the model cells were independent of each other, so that there was no lateral drainage between adjacent cells. Since sub-surface and overland lateral drainage (in a downhill direction) is an important hydrological process, we wanted to include its effect in this work. As a result, we chose to investigate the effect of parameter and forcing variable uncertainty on soil moisture and streamflow at the outlet of fully developed PyTOPKAPI catchments.

This Section looks forward to Deliverable 10, whose description appears in Section 9, to which it is closely related. There we will focus on the computational challenges of extending the HYLARSMET modelling system to the greater Southern African Development Countries (SADC region), where parameter and forcing uncertainty will be significant issues.

This Section 7 is arranged in four Subsections: 7.1. Overview of the selected catchments; 7.2. Description of the work carried out, including unexpected challenges; 7.3. Some first results; 7.4. Section summary.

7.1. Catchment overview

In this Subsection we provide an overview of the two catchments considered in this work: the Liebenbergsvlei (a familiar medium sized catchment of 4600 km^2) and the Crocodile catchment (a larger catchment of 10 000 km^2). Figure 7.1 illustrates the relative locations of the two catchments within the borders of South Africa, while Figures 7.2 and 7.3 give a more detailed view for each catchment, including topography, stream networks and Department of Water and Sanitation (DWS) streamflow gauging stations.

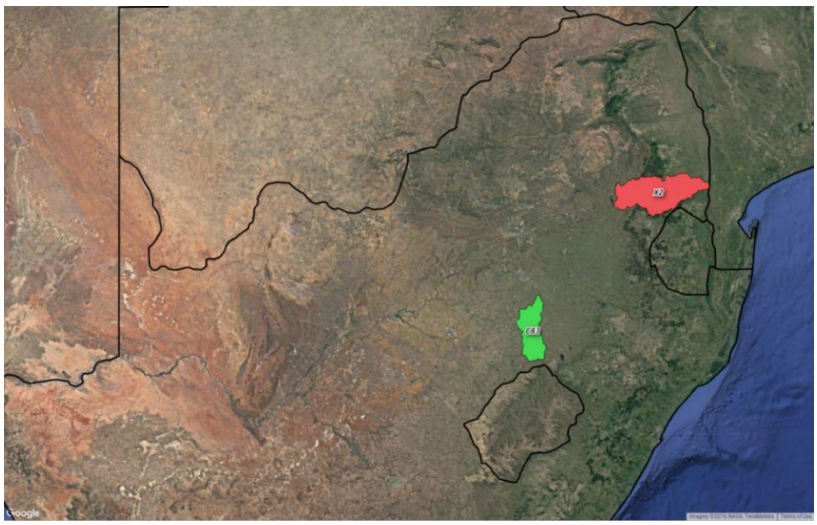


Figure 7.1: Locality map showing the relative position and sizes of the two catchments considered in this work. The Liebenbergsvlei (C83) catchment is shown in green, and the Crocodile (X2) catchment is shown in red.

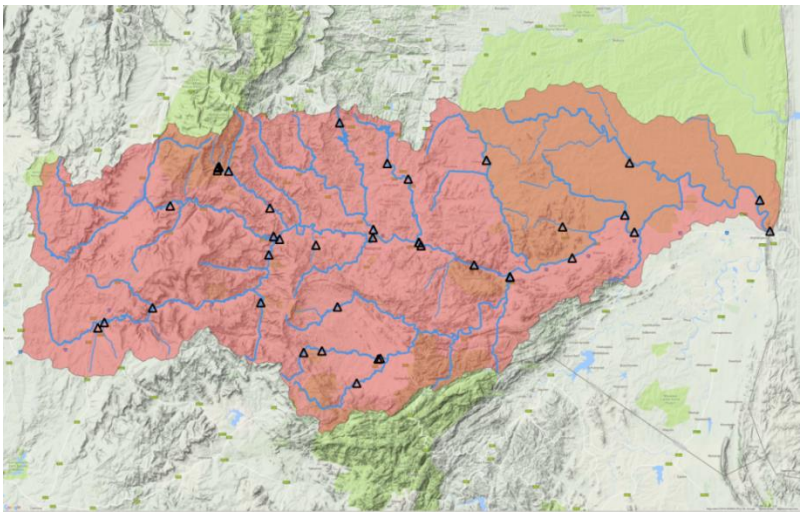


Figure 7.2: A detailed overview of the Crocodile (X2) catchment showing the topography, catchment boundary, river network and DWS streamflow gauging stations.

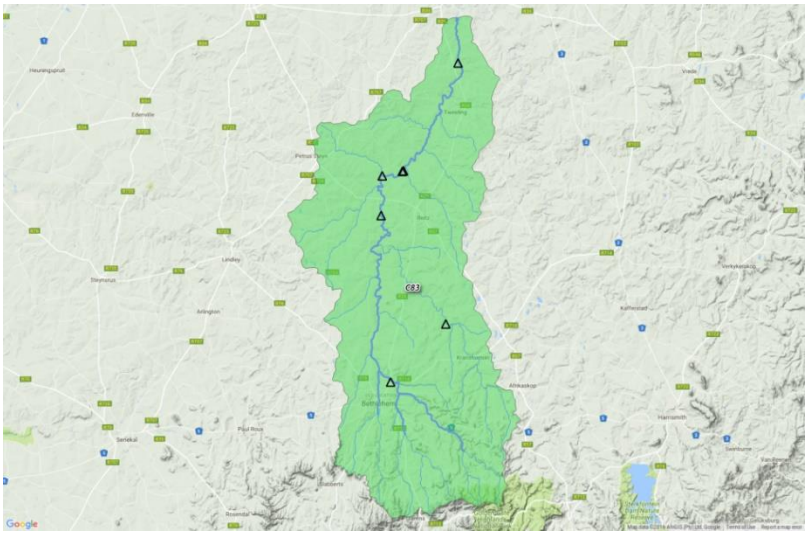


Figure 7.3: An overview of the Liebenbergsvlei (C83) catchment showing the topography, catchment boundary, river network and DWS streamflow gauging stations.

7.2. Work carried out

In this Subsection we discuss the various tasks carried out as part of this deliverable.

7.2.1. Update catchment extraction tools

We revised and updated the tools developed during WRC project K5-2024 (Sinclair and Pegram, 2013b) for extracting the information to produce a PyTOPKAPI parameter file for a given catchment of interest. This required updating the tools developed under Python 2.x to work with the newer Python versions 3.x. A more recent version of the GRASS GIS (GRASS 7.x) was also utilized to take advantage of an improved Python integration. The PyTOPKAPI tools were continually upgraded going forward to work with the most advanced versions of these GIS tools, now accessible in: PyTOPKAPI v0.4.0. Zenodo. http://doi.org/10.5281/zenodo.820640 (Sinclair, et al. 2017, June 28).

In order to model a catchment using PyTOPKAPI, it is necessary for the user to prepare a parameter file that describes the catchment's physical properties on a regular grid with the desired spatial resolution. This work needs to be carried out by model users with whatever tools and spatial information they have at their disposal typically using a GIS package to collate and manage the relevant information. We developed two toolsets to assist model users to streamline this process which *de facto* achieves the primary goal of the work reported in this Section.

First, we carefully upgraded and streamlined the existing helper scripts in the PyTOPKAPI package to cater for the sophisticated user who is able to carry out the work of delineating their catchment using GIS, or other tools with which they are familiar. One of the new tools is a Python function in the PyTOPKAPI package that produces a valid and properly formatted PyTOPKAPI parameter file when provided with the required set of georeferenced raster data files and some configuration parameters. The user needs to set up the locations of the files in the <...> brackets in the listing below, to tell PyTOPKAPI where to find the information it needs to run the model. This nontrivial innovation has considerably streamlined the setting up procedure. The listing follows:

[raster_files]

dem_fname = <path to DEM file>
mask_fname = <path to catchment mask file>
soil_depth_fname = <path to soil depth file>
conductivity_fname = <path to saturated conductivity file>
hillslope_fname = <path to hill slope file>
sat_moisture_content_fname = <path to saturated moisture content file>
resid_moisture_content_fname = <path to residual moisture content file>
bubbling_pressure_fname = <path to bubbling pressure file>
pore_size_dist_fname = <path to pore size index file>
overland_manning_fname = <path to overland Manning roughness file>
channel_network_fname = <path to channel network file>
flowdir_fname = <path to flow direction file>
flowdir_source = <source of flowdir file. Can be GRASS or ARCGIS>
[output]
param_fname = <path to output parameter file>
[numerical_values]
pVs_t0 = <initial percent saturation of soil stores>
Vo_t0 = <initial volume of overland stores>
Qc_t0 = <initial flow rate in channels>
Kc = <crop factor, currently this must be set to 1.>

The tool takes care of all the details that specify interconnection between cells and the general model setup and guarantees the user that a valid parameter file will be produced. However, producing the required input rasters for a catchment is a nontrivial task, so we developed a second tool to automate the process of delineating and analysing a catchment using standard GIS tools.

We elected to develop this second tool using the free and open source package GRASS GIS because we believe that it is important to avoid requiring potential model users to purchase expensive

proprietary GIS packages in order to take advantage of our toolset. Additionally, using open source tools makes it easier for the procedure to be critically evaluated and peer reviewed, because it is repeatable. The remainder of this Subsection outlines the procedure we have automated and then discusses two technical aspects that need special attention (infilling sinks in DEMs and computing the Strahler steam order to define the Manning streamflow roughness parameter in PyTOPKAPI).

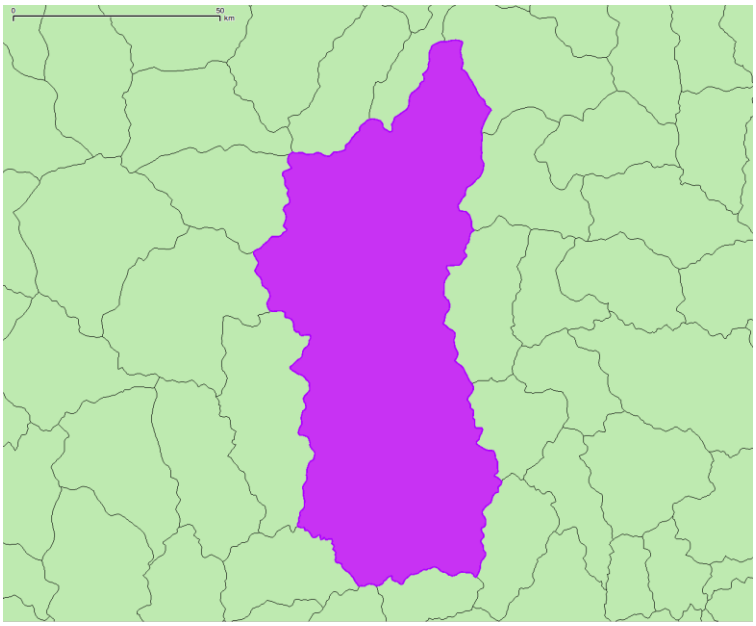


Figure 7.4: A vector polygon representing the portion of the Liebenbergsvlei catchment that we want to model using PyTOPKAPI. The green background shows WR2005 quaternary catchments, the purple overlay polygon is a concatenation of the subset of those quaternaries that make up the Liebenbergsvlei. [The bar at the top-left of the Figure is scaled to 50 km]

The procedure to produce the set of raster files needed to generate a PyTOPKAPI parameter file is outlined in Figures 7.4 to 7.9 and the supporting text; the Figures are likely to convey more information than the text. The first requirement is that the GRASS GIS location contains an elevation model and other base maps for the region where the target catchment is situated (for this work we used the 1 km countrywide base maps developed in an earlier project (Sinclair and Pegram, 2013b), but a similar set of data at an alternative spatial resolution could be used instead). The catchment modelled by PyTOPKAPI is defined by the properties of the DEM, so to ensure we have fully captured the extent of the catchment, an initial boundary is required (see Figure 7.4), for example. PyTOPKAPI now possesses code that combines the GIS tools, in an automated way, to carry out the steps to define the catchment on the basis of the DEM, as outlined in Figures 7.5 to 7.9 that follow.

The processing tool that we have introduced uses functionality in the GIS to produce a 5 km buffer around the predefined catchment boundary as shown in Figure 7.5. This is necessary because the catchment defined from the DEM will not necessarily match the boundary given in Figure 7.4 in an exact way. The buffering accounts for this possible mismatch of the boundaries by ensuring that a sufficiently large portion of the DEM is examined in the vicinity of the catchment to be modelled. The external of the buffered boundary is used to cut out a portion of the base DEM for analysis.

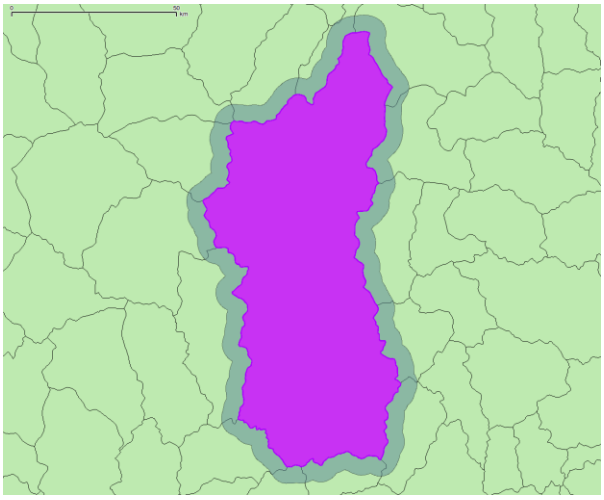


Figure 7.5: A 5 km buffer around the initial boundary is automatically provided. The buffer is indicated by the wide grey outline. [The bar at the top-left of the Figure is scaled to 50 km]

The hydrological analysis tools provided by GRASS GIS can produce sensible results even if there are inconsistencies in the DEM. The most common inconsistency is a sink, where a cell or a collection of cells is lower than all of its neighbours. PyTOPKAPI assumes that water from all cells in a catchment will eventually drain down to a single cell at the catchment outlet, so DEM sinks must be removed, despite the fact that GRASS can provide useful catchment information from flawed DEMs.

Figure 7.6 show the progress of the automated sink-filling procedure for the Liebenbergsvlei. In the first panel based on the SRTM (Shuttle Radar Tomography Mission) multiple sinks are identified (this is not uncommon) and two passes of the sink filling algorithm are typically required before all of the sinks are removed. The fourth image in Figure 7.6 is the cropped basin.

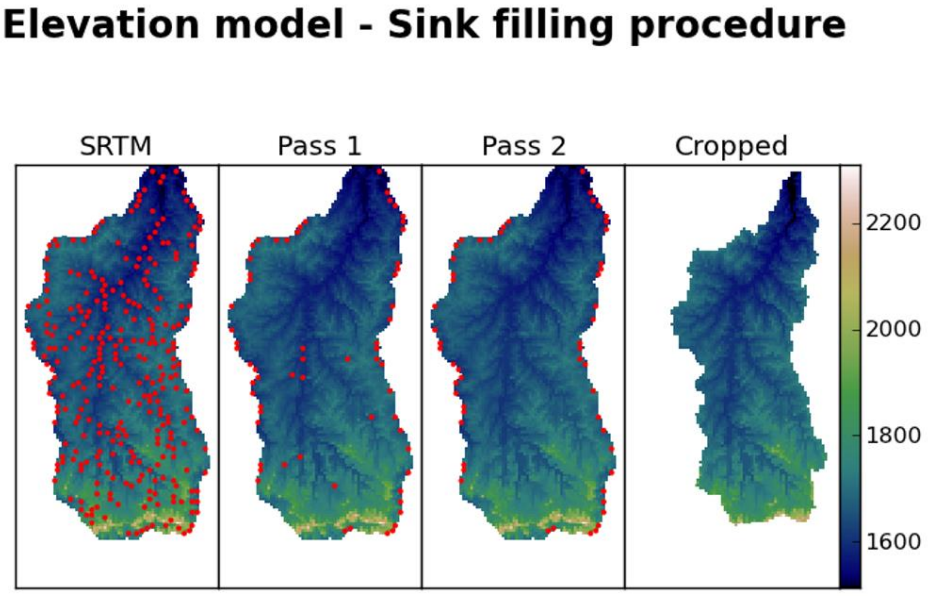


Figure 7.6: Demonstration of the automated sink filling routine for the Liebenbergsvlei DEM. The colour bar on the right describes the elevation in metres above sea level and the red dots show the locations of sinks at each stage in the process. The sinks on the border are ignored by the filling routine as they cannot be corrected without data in every cell surrounding the sink. The cropped catchment in the fourth panel is the region defined by the flow direction map, and includes all cells flowing into the outlet cell.

Once all sinks have been removed, the hydrological analysis tools in GRASS are employed to obtain (i) a flow direction raster describing the interconnectivity of the cells as well as (ii) a stream network raster. These are illustrated in Figures 7.7 and 7.8.

The catchment outlet is also identified as being the cell where the stream network intersects the original catchment boundary (red dots in Figures 7.7 & 7.8). Once the outlet cell has been identified, the GRASS tools are used to follow the flow direction information upstream until a watershed is reached, this process defines the extent of the catchment according to the DEM properties. Figure 7.7 shows the Liebenbergsvlei catchment extracted from the DEM, it matches very well (but not exactly) with the WR2005 based boundary that was initially provided to the processing tool in vector form.

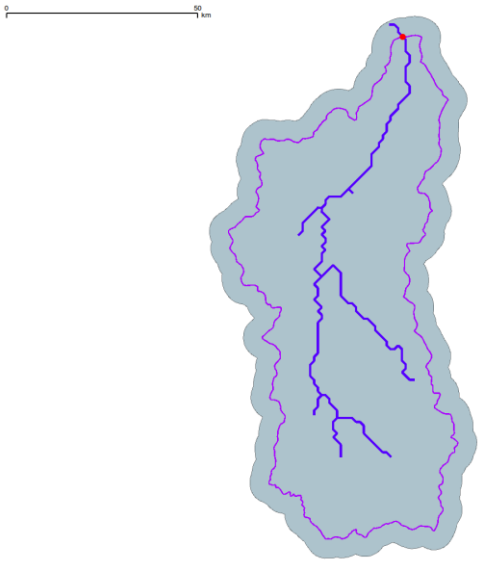


Figure 7.7: Vector version of the stream network obtained from DEM analysis (blue line). The red dot shows the catchment outlet obtained by calculating the position of the intersection of the stream network and catchment boundary. [The bar at the top of the Figure is scaled to 50 km]

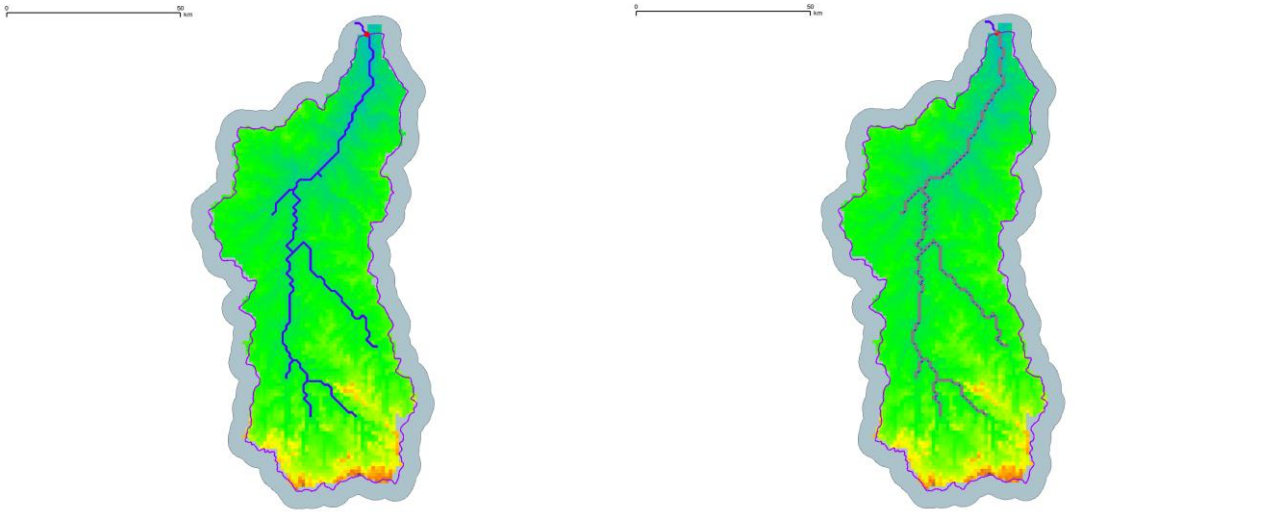


Figure 7.8: Left: The portion of the DEM extracted by computing upslope cells from the catchment outlet. Right: Raster version of the channel network, upstream of the catchment outlet.

Once the catchment has been delineated, the set of raster data files are produced to serve as input to the PyTOPKAPI parameter file generation tool. The entire process is carried out using a Python script, which controls the GRASS GIS tools, and ensures that the process is repeatable on different catchments. All that is required to process a new catchment is the vector definition of its boundary.

Figure 7.9 gives examples of model parameters obtained from two different sources, manipulated by GRASS: Surface Slope and Soil Depth.

One part of the PyTOPKAPI parameter file generation process requires determining the channel roughness coefficient based on the Strahler stream order for the given channel reach. We coded up an implementation of the algorithm described by Gleyzer et al. (2004) and this has now been made a part of the PyTOPKAPI package.

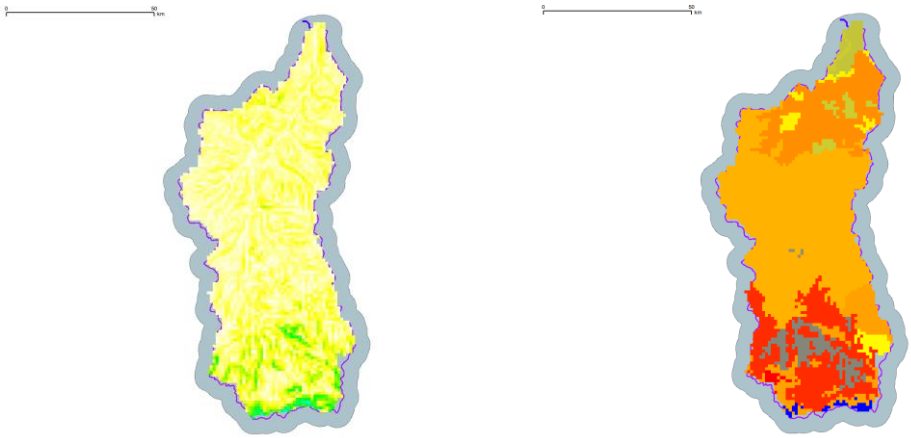


Figure 7.9: Left: Example of model parameters extracted. Surface slopes from the DEM. Right: Example of model parameters extracted. Soil depths.

Figure 7.10 following shows the stream order computed on a synthetic stream network used to test the algorithm adapted for use in the PyTOPKAPI suite.

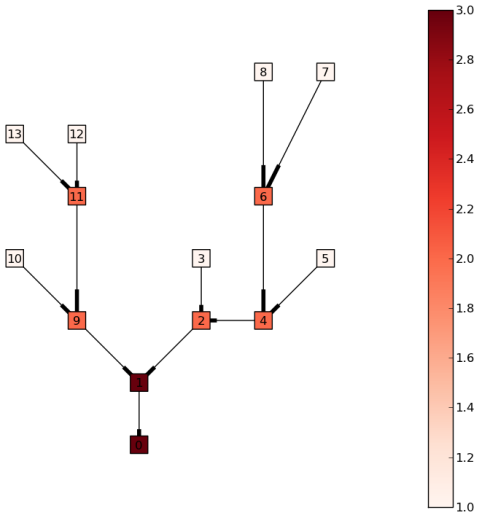


Figure 7.10: A directed graph representing a simple stream network. The numbered squares are the graph nodes and the lines joining the nodes show direction using a thickened end as an arrow head. The calculated Strahler order is shown by each node's colour.

Figures 7.11 and 7.12 show the results obtained for the stream network extracted on the Liebenbergsvlei, using the algorithm displayed in Figure 7.10.

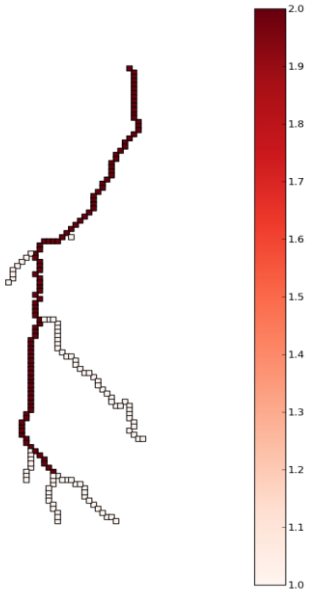


Figure 7.11: A directed graph representing the stream network extracted from the Liebenbergsvlei. The 1 km squares are the DEM graph nodes with Strahler order shown by their colour.

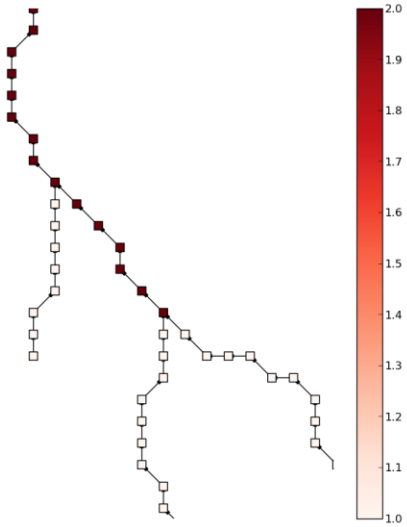


Figure 7.12: A close up showing the lower part of Figure 7.11. Two first order streams join to form a second order stream, which does not change order when joined by another first order stream.

7.2.2. Query DWS website

- We developed tools to query the DWS Hydrology website and locate gauging stations (streamflow, etc.) contained within a catchment of interest. Figure 7.13 shows an example table for the X2 catchment.

- We developed tools to download streamflow records from the DWS Hydrology website. Given a gauging station of interest, we have developed tools to download the relevant data records and store these locally.

7.2.3. PyTOPKAPI model development and fixes

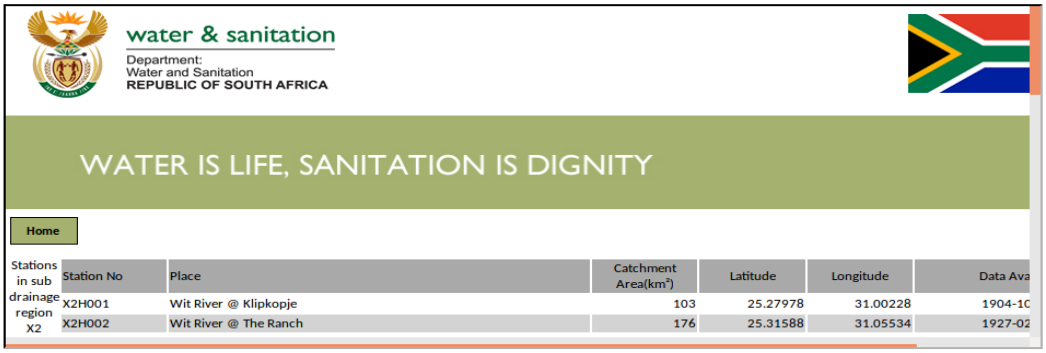
- Working with previously unvisited catchments uncovered some inconsistencies in the core PyTOPKAPI parameter datasets that had not been encountered before. We spent time developing tools to verify the integrity of the PyTOPKAPI parameter files developed using the automated tools mentioned in the previous bullet point. In addition, we coded some simple infilling algorithms to correct inconsistent parameter files. These tools were incorporated directly into the PyTOPKAPI code.
- We also discovered some troubling numerical errors in the earlier model code, related to numerical precision when the solution of the model's Ordinary Differential Equations resulted in the difference being computed between minuscule but similar numbers. We have implemented checks to ensure that these situations are correctly handled by the model.

```
In [3]: sub_region = 'X2'

        url_tmpl = 'https://www.dwa.gov.za/hydrology/HyStations.aspx?SubRegion={}&StationType=rbRiver'
        url = url_tmpl.format(sub_region)

        HTML('<iframe src={} width=950 height=350></iframe>'.format(url))
```

Out[3]:



	Station	Place	Catchment Area (sq. km)	Latitude	Longitude	Start date	End date
0	X2H001	Wit River @ Klipkopje	103	-25.2798	31.0023	1904-10-01	1952-04-01
1	X2H002	Wit River @ The Ranch	176	-25.3159	31.0553	1927-02-16	1947-03-09
2	X2H004	Krokodil River @ Nelspruit	3929	-25.4506	30.9645	1923-10-09	1928-12-31
3	X2H005	Nels River @ Boschrand	642	-25.4325	30.9658	1929-07-04	2016-08-26
4	X2H006	Krokodil River @ Karino	5097	-25.4698	31.0881	1929-10-02	2016-08-25
5	X2H007	Kaap River @ Dolton	1642	-25.5417	31.3165	1930-06-25	1947-12-01
6	X2H008	Queens River @ Sassenheim	180	-25.786	30.9243	1948-02-01	2016-08-16
7	X2H009	South Kaap River @ Daisy Kopje	280	-25.7297	30.984	1948-02-07	1966-09-30
8	X2H010	Noordkaap River @ Bellevue	126	-25.6111	30.8749	1948-02-11	2016-08-16
9	X2H011	Elands River @ Geluk	402	-25.6464	30.2775	1956-10-01	1999-12-11
10	X2H012	Dawsons Spruit @ Geluk	91	-25.6586	30.2605	1956-10-03	2016-08-17
11	X2H013	Krokodil River @ Montrose	1508	-25.4486	30.7118	1959-01-21	2016-08-22
12	X2H014	Houtbosloop @ Sudwalaskraal	250	-25.3824	30.7015	1958-12-17	2016-08-22
13	X2H015	Elands River @ Lindenau	1547	-25.4904	30.6986	1959-07-23	2016-08-22
14	X2H016	Krokodil River @ Tenbosch Kruger National Park	10365	-25.3639	31.9557	1960-08-24	2016-08-24
15	X2H017	Krokodil River @ Thankerton Van Graan se dam K...	8811	-25.4384	31.6345	1959-08-28	1998-09-01
16	X2H018	Mbyamiti River @ Kruger National Park	618	-25.2784	31.6226	1960-08-25	1997-03-04

Figure 7.13: An illustration of the automated tool developed to query the DWS website for information on the available stations in a catchment. Top – shows a query to fetch the relevant web page. Bottom – shows the reformatted results extracted from the web page in tabular form.

7.3. Section results and summary

We carried out a range of model simulation runs for the catchments introduced earlier. Figure 7.14 shows the catchment extracted for the X2 region using the updated tools described in Subsection 7.2.1.

Although we encountered a few unexpected difficulties in the process of working towards this deliverable, we were able to upgrade the catchment extraction tools to take advantage of newer versions of the Python programming language and the GRASS GIS package that PyTOPKAPI relies on to provide advanced GIS functionality. We were also able to improve the PyTOPKAPI model's ability to handle difficult corner cases that we had not encountered previously – this makes the model more robust, and was important for our coverage over the larger Southern African region.

Several things that we discovered fed well into Deliverable 10. The focus of that future deliverable was to find ways to improve the computational speed of the model and demonstrate our ability to model a significant portion of the SADC region in a convenient timeframe, which was achieved.

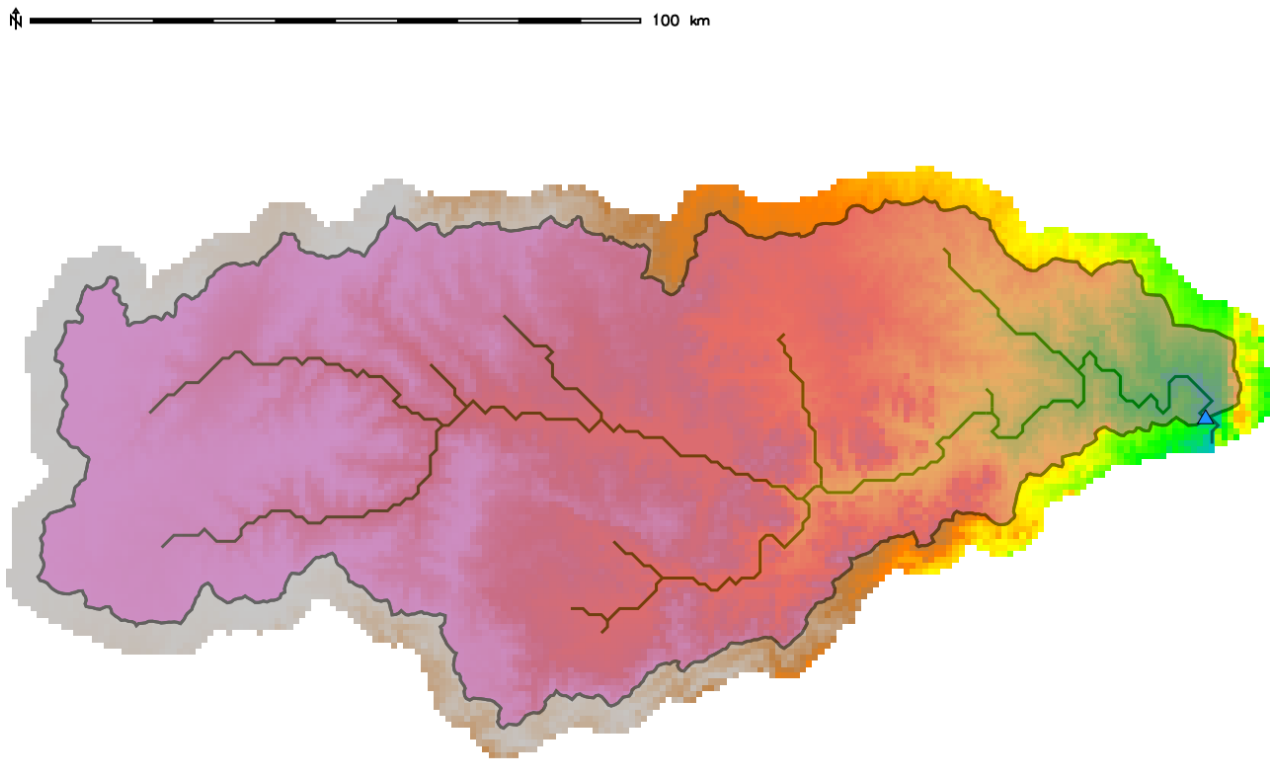


Figure 7.14: The X2 catchment extracted using the PyTOPKAPI tools. The background shows the digital elevation model (DEM) at 1 km^2 resolution, and the blue triangle marker shows the location of the catchment outlet as derived by the PyTOPKAPI catchment extraction tool. The channel network shown is that derived from the catchment DEM.

-----ooOoo----

8. Evaluating the work done part-way through the project

In this Section we describe how we took stock of the work we envisioned carrying out towards managing the computational load, as an introduction to the strategies we developed, to be described fully in Sections 9 to 11 which follow. We initially intended to model a major subset of the SADC region south of the equator. The countries targeted were: Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia and Zimbabwe. At the current modelling scale on a grid with 0.125° spacing, this equates to roughly 35 000 cells (approximately 5 times the number of cells we currently model in HYLARSMET). Figure 8.1 shows the regions covered by these countries. As it happened, we extended this area to include Kenya and Democratic Republic of Congo in Sections 9 and 10 for completeness, once we had worked with and judged the size of the problem.

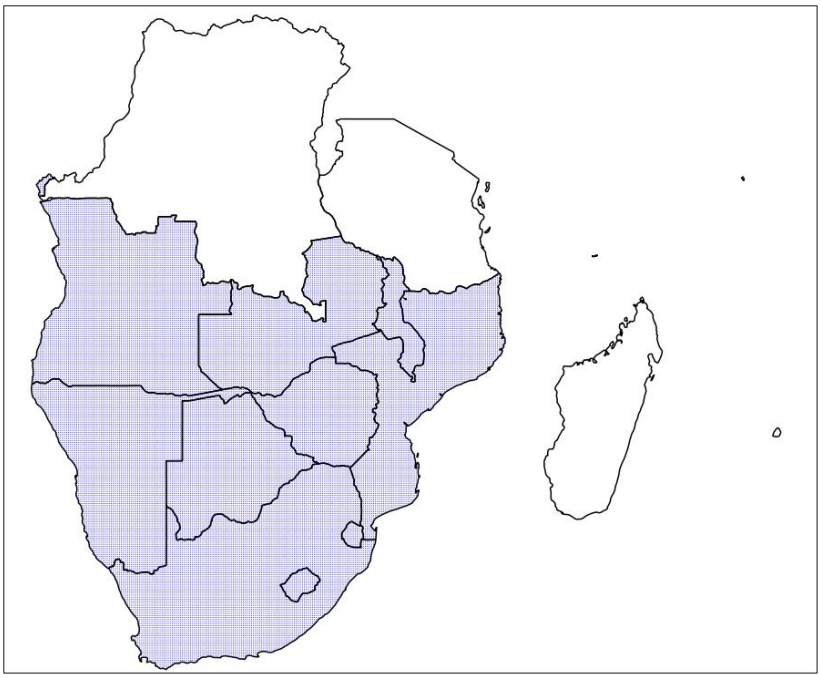


Figure 8.1: An illustration of the significant portion of the SADC region that we originally intended to model. Countries to be targeted were shaded in blue. We covered the full area of SADC in Chapters 9 and 10.

At this stage we decided to purse several avenues to increase the model's computational speed. These fall into 4 broad categories: Improve computational speed for individual cells; Exploiting additional computing power; Reducing temporal and spatial resolution; Interactions, capacity building and non-deliverable related project activities

8.1. Improve computational speed for individual cells

The core calculation component of the PyTOPKAPI model is the solution of either two or three Ordinary Differential Equations (ODEs) for each cell (a separate ODE is solved for the soil store, overland store and channel store when this exists). The original ODE solver is an implementation of the Runge-Kutta ODE algorithm coded in Python by Vischel et al. (2008). It was likely that replacing this solver with a solver coded in a compiled language by experts in ODE's would speed up the model

somewhat, in addition to the valuable property of being much more numerically robust. This is an avenue that we explored.

Since the computational time per cell is also affected by the ways that the model state and forcing information is accessed from memory and disk, we explored possible optimizations related to the model's memory and disk access patterns.

8.2. Exploiting additional computing power

In its original form the PyTOPKAPI model computes each cell sequentially on a single processor core for each time-step. By contrast, when the PyTOPKAPI model is used in Land Surface Modelling mode to compute ET and soil moisture (EXSMET – SADC region; HYLARSMET – South Africa), each model cell then is entirely independent of the others and we can take advantage of multiple computing cores, or a network of individual computers if available, to calculate the results for each cell individually. There is a large number of parallel computing options that we intend to explore here.

An interesting and related task that we tackled was the use of parallel computing techniques to speed the computations when PyTOPKAPI is set up in catchment mode as a network of interconnected cells. In catchment mode, there is a flow of water from the cells on the catchment periphery inwards to the main channels and finally to the outlet. In the catchment case, this network structure is dictated by the Digital Elevation Model of the Earth's surface, and computations on a given cell cannot begin before the computations for all of its upstream cells have been completed in order. However, at any time there may be several or many cells that do not depend on each other and can benefit from parallel computing techniques. Directed Graph structures such as those shown in Figures 7.10-7.12 are already in a suitable form to exploit well-known algorithms from Graph Theory, to determine an optimum order of computation that allows as many cells as possible to be computed concurrently.

8.3. Reducing temporal and spatial resolution

Our main aim in this project was to increase the spatial extent of our modelling efforts resting on HYLARSMET. Depending on the gains that we can achieve using the techniques described above, we thought it might be necessary to sacrifice time resolution and inter-cell spacing in order to obtain the required spatial coverage. In the event, we found that this was not necessary, so we were able to model the SADC countries to the same detail we had done in South Africa in the HYLARSMET project.

8.4 Interactions, capacity building and non-deliverable related project activities

In this Subsection we draw attention to a variety of interactions and activities that came about due to our involvement in this project and which support the WRC's aims of capacity building, knowledge dissemination and outreach.

- HYLARSMET was installed on a dedicated computer at the ARC ISCW is Pretoria, as described in Section 3. This work was carried out as a collaboration with Dr Johan Malherbe and funded by the World Meteorological Organisation (WMO). We assisted Dr Malherbe in updating the soil moisture and evaporation estimates on a monthly basis, until his departure

to the CSIR. The updates were also periodically made available on our website at http://sahg.ukzn.ac.za.

- The HYLARSMET soil moisture data produced by the system installed at ARC ISCW have been used by several parties that have made contact with us:
    - Ian Engelbrecht of SANBI investigated the use of the HYLARSMET soil moisture data in the Karoo BioGaps project (http://www.sanbi.org/biogaps), which is intended to inform environmental impact studies involved with fracking, and the HYLARSMET product will be extremely valuable for guiding the surveys in this arid part of the country. Ian's views are highlighted in this Facebook post: https://www.facebook.com/ian.engelbrecht.5/videos/10153616935491353
    - Dr Roy Williams of the ARC is using the soil moisture and evaporation estimates from HYLARSMET to help predict Rift Valley fever outbreaks
    - Nipho Msibi from the Equity Research department at Investec Bank, has expressed great interest in the HYLARSMET product to assist with their advice to listed food producers (Pioneer, Tiger Brands, RCL Astral, Clover, etc.).
- We had some good collaboration on a recently completed WRC project K5-2323 headed by Prof. Colin Everson (Everson et al., 2016), where our soil moisture and ET estimates were compared with in-situ estimates, especially soil moisture measured by a Cosmic Ray Probe. Thigesh Vather obtained his MSc from UKZN in which he showed that the HYLARSMET/EXSMET estimates compared favourably with their ground based measurements, particularly COSMOS.
- Dr Sinclair was involved in several interactions with Masters and Hydrology Honours students based in UKZN Pietermaritzburg on the topics of soil moisture and evaporation.
- Dr Sinclair was invited to present a portion of a WMO and EUMETSAT funded training workshop in Harare, Zimbabwe during 24-28 October 2016. The workshop was attended by representatives of Met offices and Departments of Agriculture for the SADC region and was a valuable platform to showcase the work being done in this project.

Looking forward, timeous and routine monitoring of the spatial distribution of Soil Moisture and Evapotranspiration over a large region in fine detail has great value for coping with two weather driven extremes: Flash Floods and Droughts. The current state of soil moisture conditions has a major impact on the runoff response of a catchment to heavy rainfall. In addition, monitoring the wetness of the soil in detail over large regions, without having to laboriously take expensive samples, is a bonus for agricultural managers who need to understand the status of crop growth potential. This is particularly relevant in the Southern and Central African countries in the SADC region which contribute importantly to the food basket of Africa.

A reminder: this project has two thrusts. The first is to extend coverage of near real-time Soil Moisture (SM) and Evapotranspiration (ET) monitoring over the SADC region using the existing methodology of operating the PyTOPKAPI hydrological model in Land Surface Modelling mode. The second thrust is to develop powerful computing techniques that will make it feasible to drive the modelling procedure for the vastly increased number of cells required to cover the SADC region. This latter thrust is the objective of the next deliverable. The two Sections following show how we exploited the ideas outlined above.

-----ooOoo----

9. Increasing computational speed and capacity

The description of this deliverable 10 in the project proposal reads '***Devise means of increasing computational speed and capacity and demonstrate ability to make estimates of SM and ET over SADC***'. The purpose was to devise techniques that will allow the modelling strategy utilized in the HYLARSMET system (Sinclair and Pegram, 2010; Sinclair and Pegram, 2013) to be extended to include a significant proportion of the SADC region for this project (EXSMET). One of the main obstacles to achieving this aim is the non-trivial increase in the number of locations to be modelled when compared to our previous work limited to South Africa, hence the associated computational workload.

We identified several avenues of investigation to increase the model's computational speed, which is the main objective of this deliverable. These fall into 4 broad categories outlined in the Subsections which follow: improve computational speed for individual cells; exploit additional computing power; reduce temporal and spatial resolution; devise parallel computing methods for implementation in PyTOPKAPI.

9.1. Improve computational speed for individual cells

The core calculation component of the PyTOPKAPI physically-based model is the solution of either two or three Ordinary Differential Equations (ODEs) for each model cell. A separate ODE is solved for (i) the soil store, (ii) overland store and (iii) channel store (when a channel exists in the cell, fed by an upstream cell). The current ODE solver is an implementation of the well-known Runge-Kutta ODE algorithm coded in Python by Vischel et al. (2008). It is likely that replacing this solver with a solver coded in a compiled language by experts in ODE's will speed up the model somewhat, in addition to possessing the valuable property of being much more numerically robust. In any event, for the purposes of this deliverable, we decided not to put energy into this modification. The reason is that profiling the model code demonstrated that the current solution speed of solving the ODEs is already extremely fast, so the optimization effort is better spent utilizing multiple CPU (Central Processing Unit) cores as described in Subsections 9.2 to 9.4.

Because the computational time per cell is affected by the ways that the model status and forcing information is accessed from memory and disk, some additional small gains are likely to be possible using optimizations related to the model's memory and disk access patterns.

9.2. Exploiting additional computing power

As mentioned above, in its original form when applied to catchment rainfall-runoff problems, the PyTOPKAPI model computes each cell sequentially on a single processor core for each time-step. In contrast, when the PyTOPKAPI model is used in Land Surface Modelling mode to compute ET and soil moisture (EXSMET – SADC region; HYLARSMET – RSA), each model cell is entirely independent of the others and we can take advantage of multiple computing cores or a network of individual computers to calculate the results for each cell individually. There is a large number of parallel computing possibilities, so we focussed on an approach which could provide the best possible benefit without making any major new demands on users of the model.

Fortuitously, there is a bonus which springs from the above, useful for catchment modelling rather than over many individual disconnected pixels. An interesting and related task that we were able to explore is the use of parallel computing techniques to speed the computations when PyTOPKAPI is set up in catchment mode as a network of interconnected cells. In catchment mode, there is a flow of water from the cells on the catchment periphery inwards to the main channels and finally to the outlet. This network structure is dictated by the DEM, and computations on a given cell cannot begin before the computations for all of its upstream neighbours have been completed in order. However, at any given time there may be several or many cells that do not depend on each other and can benefit from parallel computing techniques. Directed Acyclic Graph (DAG) structures, such as those shown in Figures 7.10 and 7.11, are already in a suitable form to employ well-known algorithms to determine an optimum order of computation that allows as many cells as possible to be computed concurrently. This bonus innovation is described in greater detail in Subsection 9.4.1.

9.3. Reducing temporal and spatial resolution

Our main aim in this project is to develop techniques to increase the spatial extent of our modelling efforts and make the applications feasible in the large scale. Initially, we were concerned that the gains that we could achieve using the techniques described above would not be sufficient, and that it might therefore be necessary to coarsen the calculation, by reducing time resolution and inter-cell spacing, in order to obtain the required spatial coverage. It turned out that this was not necessary after all.

9.4. Parallel computing implementation for PyTOPKAPI

In this Subsection 9.4, we describe the implementation of a multi-processor (parallel) algorithm for the PyTOPKAPI model. Subsection 9.4.1 gives some background, as well as our implementation strategy; Subsection 9.4.2 summarizes the results and demonstrates our ability to model the 35 000 model cells covering the portion of SADC shown in Figure 8.1, extended to about 49 000 in Section 11.

9.4.1. Strategy employed

In Subsection 9.1 we introduced the idea that the core unit of computation for the PyTOPKAPI model is the solution of three ODEs per time-step for each model cell, as listed in 9.1. In this Subsection we expand on this concept and detail the strategy employed to maximize computational speed using a DAG to represent the network topology. This construct provides us with the flexibility to take advantage of the ubiquitous multi-core processors on modern computer hardware, and opens the door to taking advantage of distributed (multi-machine) processing and possibly the use of dedicated processing units such as Graphics Processing Units (GPU).

In order to illustrate the strategy that we devised for parallel processing, consider the simple 30-cell catchment shown in the left-hand panel of Figure 9.1. The catchment outlet is coloured blue and labelled 0, while the catchment's network tree has a main trunk (given by the set of nodes {0, 1, 8, 13, 17, 22, 27, 29}) and several branches connecting to this channel. The connections between cells have a direction defined by the local topography and are shown in the figure by the bold end of the line joining connected nodes. For example, the outlet node 0 is directly connected to, and

downstream of, the nodes {1, 2, 3, 4}. Similarly, node 15 is connected to nodes {13, 18, 19}, but is downstream of nodes {18, 19} and upstream of node 13.

For the PyTOPKAPI model to produce a simulation of *m* time-steps for this catchment of *n*=30 cells, a total of *mxn* = 30*m* cell-solutions must be performed (in this report we use the term **cell-solution** to mean the block of computations required to solve all of the ODEs for a given cell at a single time-step). The total computational load of *mxn* cell-solutions is subject to some ordering constraints dictated by the structure of the model. Firstly, there is a direct serial dependence between consecutive times-steps for each cell (water flows downhill … !). This implies that the cell-solutions for a model cell must be performed sequentially in order of increasing simulation time and cannot be calculated independently. Secondly, there is a spatial dependence between the cells in the network. Due to the input transfer of water from upstream cells, each cell must wait for the cell-solutions of all its upstream predecessors to be completed before starting to calculate its cell-solution for any time-step. Thus the cells are not entirely independent of each other. The caption of Figure 9.1 describes the cell connectivity.

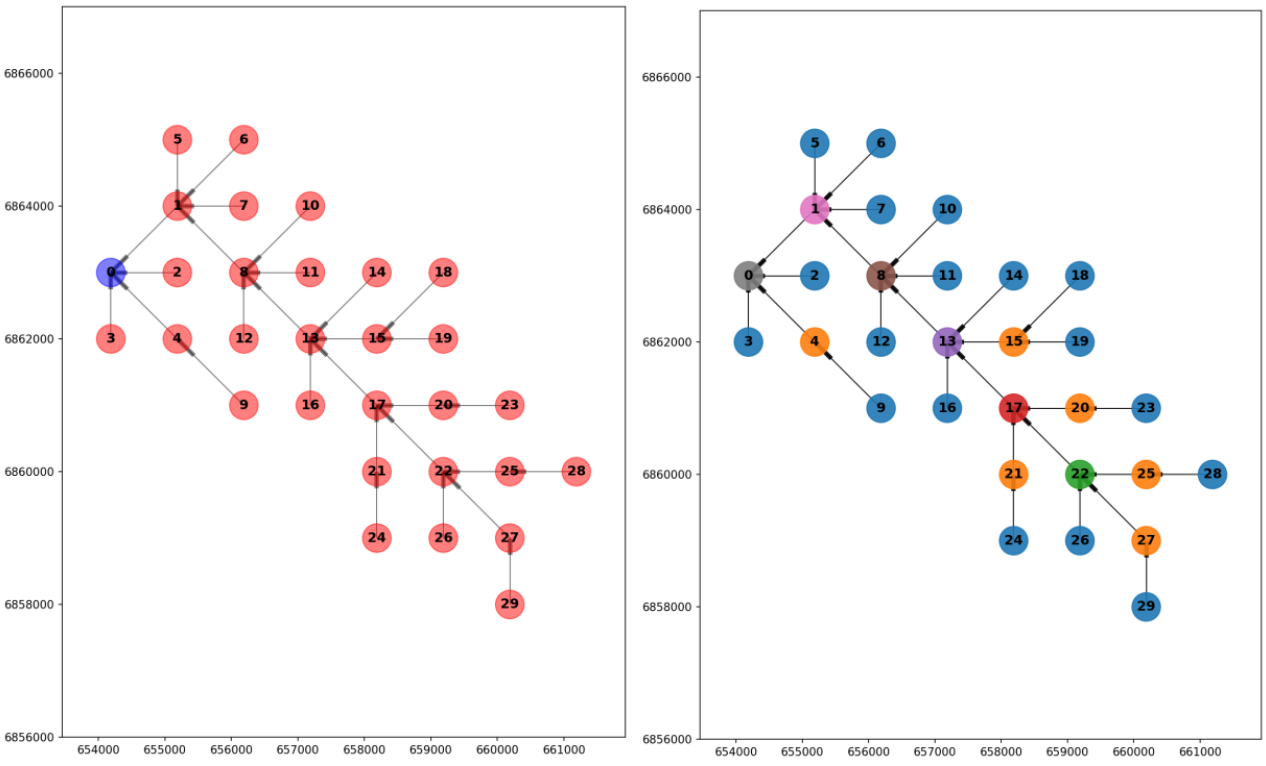


Figure 9.1: Example of a catchment network structure. The left-hand panel shows the numbered nodes of a Directed Acyclic Graph (DAG), which represents the catchment topology. The arrowed ends of the lines joining the nodes denote the downstream end of the inter-node connection, as defined by the catchment topography. The catchment outlet is the node numbered 0, in blue in the left panel. The right-hand panel shows the network decomposed into 8 levels of independent cells – the set of independent cells for each level is plotted in the same colour (refer to the text for details of the decomposition).

All of this leads quite naturally to the straightforward (and guaranteed numerically correct) computational sequencing adopted in the baseline version of the PyTOPKAPI model. In the baseline

version, the cells are first ordered by increasing distance from the outlet, where distance is defined as the number of connections which must be traversed to reach the outlet from the target cell (additional left to right sorting is performed row-wise in order to decide ties). Then, for each simulation time-step, the cell-solutions are calculated in reverse order, starting with the highest numbered cell (node 29 in Figure 9.1) and ending at the outlet node. The process is repeated serially until the simulation is complete.

The serial computational order of the baseline model makes sense and is simple to understand and implement in computer code. Unfortunately, it does not provide any context to allow for the parallel execution of independent cell-solution operations, where this might prove beneficial. The baseline computation thus runs on a single processor core with a total execution time that is a simple linear function of the total workload *m*x*n* cell solutions. The total model workload does not increase in a linear fashion with increasing numbers of model cells and time-steps, so the total execution time increases in a non-linear fashion. This is shown by the contours of equal workload plotted in Figures 10.4 and 10.5, which will be discussed in detail in Section 10.

As noted above, the solution speed of the ODEs in the individual cells is already extremely fast, and we might only envisage incremental speed-ups by changing the ODE algorithm/code-library. The primary reason to implement such changes would be to achieve improved numerical robustness, by building on the work of experts in numerically solving Differential Equations. This means that the baseline model does not provide much of an opportunity for really significant speed increases. Based on our work with the specific case of disconnected model cells in HYLARSMET, we realized that the time-series for each disconnected cell could be computed not only independently of each other, but also in parallel on multiple CPU cores (or even multiple computers). Furthermore, it is quite straightforward to generalize this idea to catchment networks of inter-connected cells represented by a DAG.

Consider the right-hand panel of the network shown in Figure 9.1 above. There we have coloured all of the outer network nodes in blue (those with no upstream predecessors), then the next level down in gold (those nodes with a single level of upstream predecessors). Finally, the set of main trunk nodes {22, 17, 13, 8, 1, 0} each depend on a sequentially larger number of upstream levels. The sets of nodes at each level (e.g. level 2 nodes {4, 15, 20, 21, 25, 27}) are entirely independent and their complete time-series of cell-solutions can be calculated in parallel. The sequence of levels and each matching set of independent nodes are easily determined by decomposing the network DAG, and the revised computation sequence becomes: *Loop through the levels and compute the full time-series for each cell in the current level*. The new sequence of computations readily lends itself to parallel processing algorithms.

For a practical catchment problem, consider the Liebenbergsvlei catchment in Figure 9.2 below.

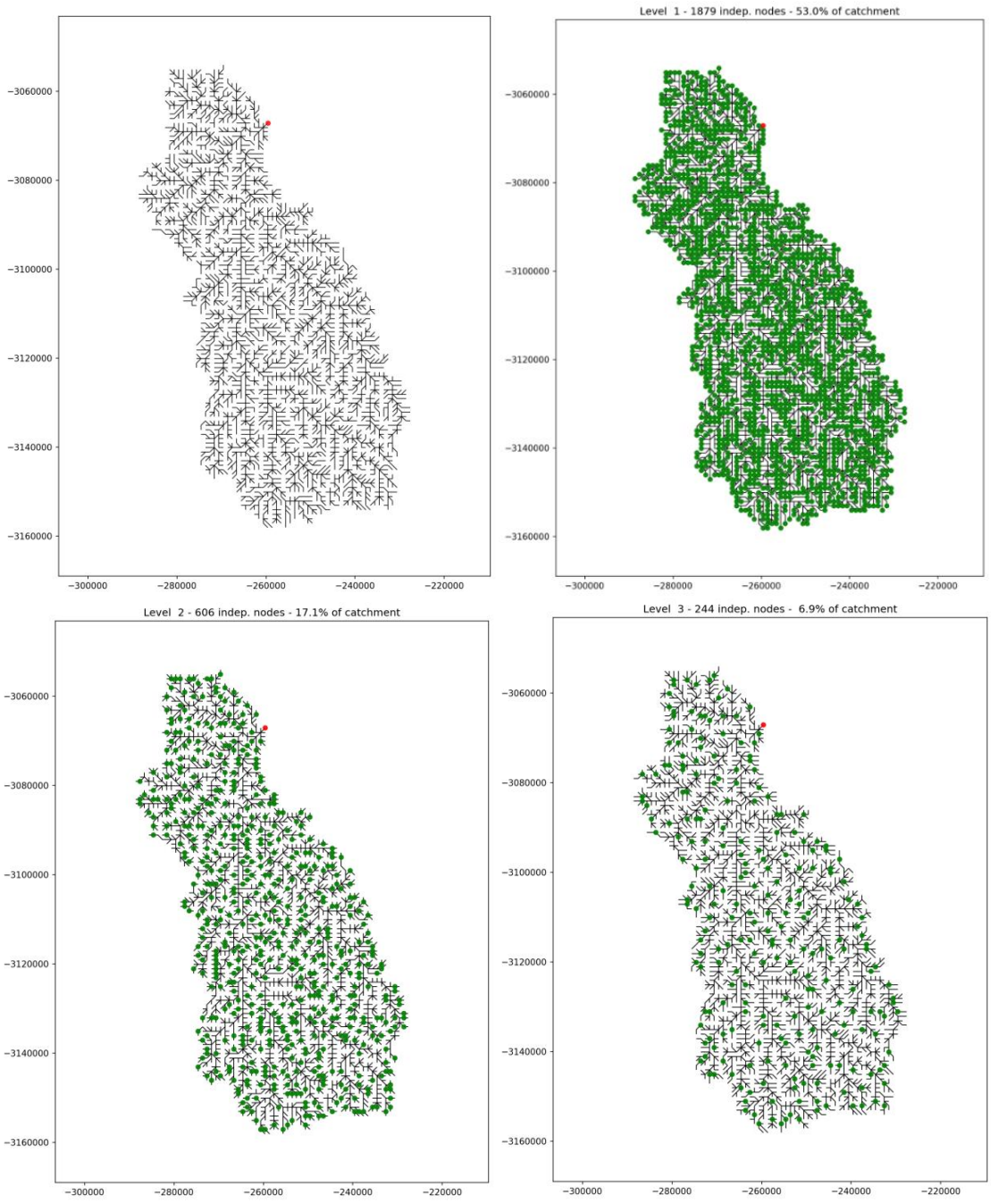


Figure 9.2: Decomposition of the Liebenbergsvlei catchment into a sequence of levels with independent cells. The top-left panel shows the inter-cell connections of the catchment network (the outlet cell is shown in red, top right of the catchment). The remaining 3 panels show the independent cells for each set of the first 3 levels of the decomposition (the lower the level, the fewer the dots). The set of cells at each level is represented as a green dot. The cells at each level have no interconnectivity and their time-series can be solved independently in parallel before moving on to solve the next level down.

For the specific case where PyTOPKAPI is run in land surface modelling mode, the 'network' DAG is decomposed to a single level where all model cells are independent (about 7 000 isolated cells for HYLARSMET in RSA – 49 000 cells for EXSMET over SADC). This is typical of a situation that computer scientists would call 'embarrassingly parallel', and it turns out that medium to large catchments are usually very likely to decompose into 'embarrassingly parallel' computational problems. Figure 9.2

illustrates this for the Liebenbergsvlei catchment. The top-left panel shows the network of connections between the 1 km^2 cells comprising the 3544 km^2 catchment, with the outlet cell shown as a red dot. The remaining three panels show the locations of independent nodes for the first 3 levels in the catchment network. It is interesting to note that these three levels account for 77% of the total cells in the catchment – which is typical. Thus it turns out that the parallel algorithms used to speed up calculations for EXSMET can be used without modification to generate valuable speed increases for catchment simulations as well.

In this Subsection we introduced the philosophy behind our strategy for calculating cell-solutions in parallel on multiple processor computers. In Subsection 9.5 we present the results of our final implementation.

9.5. Results and discussion

In Subsection 9.4 we presented our strategy for a parallel processing version of PyTOPKAPI, in this Subsection we give some results from tests of our implementation of the above strategies. The implementation required a hugely significant restructuring of the model codebase.

During the development process we used a short (170 time-step) simulation on the Liebenbergsvlei catchment network shown in Figure 9.2 (3544 cells). The total workload for this simulation was 602 480 cell-solutions (the relevance of this will be made clear in the discussion of Figures 9.4 to 9.6). Figure 9.3 shows a comparison of simulation runtimes for the baseline and parallel codes. The comparison is made in terms of the ratio of speed-up achieved, relative to the average simulation speed from five runs of the baseline model. The red horizontal bar indicates the mean speed-up of the baseline relative to itself (obviously a factor of one), while the bold black line overlaid on the bar shows the (narrow) range of speed-ups for five independent model runs. The blue bars show the relative ratios for the parallel code, again with variability of ratio.

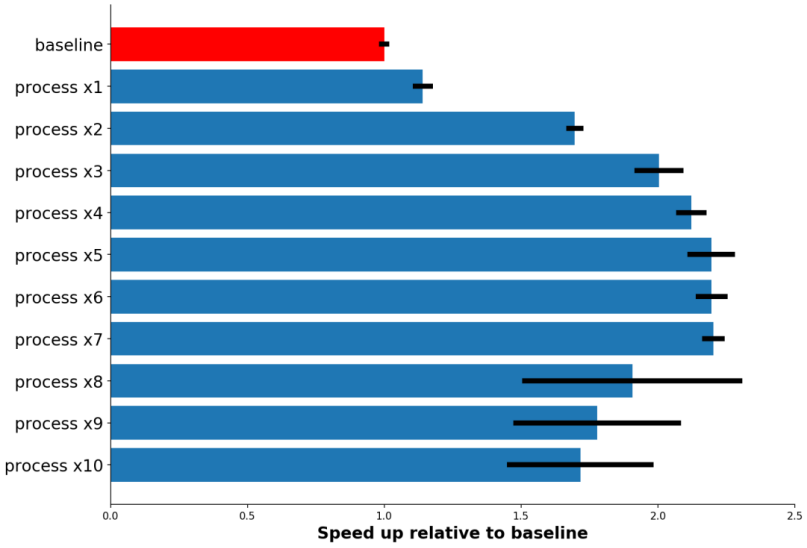


Figure 9.3: Relative speed-up for the Liebenbergsvlei catchment network with 3544 cells and 170 time-steps (see Figure 9.2 for the network layout). The horizontal bars show the speed-up relative to the baseline as an average from 5 model runs. The bold black bars show the range of speed-ups achieved from the 5 runs. The red bar represents the baseline code while the blue bars give results for the parallel code with a range of spawned worker processes from 1-10 workers.

NB. All of the technical details are captured in the **public code repository**[1] and we will not include them in this report as they are extensive. Instead we focus on the results obtained by comparing model execution speed between the baseline and parallel versions of the code. The tests described in this Subsection were carried out on a quad-core i7 laptop running the Linux operating system. This computer has 4 physical CPU cores, each of which run two virtual CPUs, so the computer operating system sees eight separate CPUs to which it can assign work. The machine has 16 GB of on-board RAM, so memory usage did not constrain the tests.

Our implementation of the parallel algorithm spawns new operating system processes as workers. Each process is passed a chunk of work (all the cell-solutions of a single cell for the entire simulation period). As the individual processes complete their work, another chunk is passed until the entire workload is complete. The operating system will typically (but not necessarily) assign each spawned worker process to a different CPU core (this depends on many factors such as the number of spawned processes, CPU workload from other programs, etc.). The blue bars in Figure 9.3 show the mean speed-ups obtained for a range of spawned worker processes ranging from one to ten.

For the test computer with 8 CPU cores, Figure 9.3 shows a steady increase in the speed-ups from 1 to 7 obtained as the number of spawned worker parallel processes is raised. The best speed-up with the smallest variability in speeds is achieved with seven worker processes. An obvious slow down and increase in variability occurs when the number of spawned processes becomes equal to or greater than the number of CPUs. This is almost surely caused by competition between the processes for CPU resources, remembering that the main process program is running in addition to the worker processes.

9.6. Technical Summary

The tests based on the Liebenbergsvlei catchment model (see Figure 9.2 and related text) were a good way of testing the procedures in a realistic and typical catchment modelling exercise. These tests incorporate a large number of the confounding factors, which introduce variability in the time taken to complete a cell-solution. For example, the solution time of an ODE can depend on the starting state of the store, the rate of upstream inputs and the parameters of the ODE equation. All of these are strongly affected by heterogeneity in the catchment characteristics as reflected in the physical properties of each cell. When we began to experiment with larger workloads, we noticed a worrying drop in the solution rate of the parallel model version for large workloads, and devised an experiment to investigate the cause.

To control for the effects of inter-cell variability, in going forward, we selected a single cell at random from the HYLARSMET grid and used its physical properties as a 'cell template'. This allowed us to create a set of duplicate cells of any number desired. We retained the inter-time-step variability using a forcing dataset of length 25 years, at a daily time-step, which we sub-sampled to create time-series of any desired length (up to the maximum). By duplicating the cells and sampling the time-series, we could construct a simulation with a chosen size of workload. We randomly selected workloads, then constructed and timed the simulations to determine the work-rate as a function of

[1] https://github.com/sahg/PyTOPKAPI

workload, number of cells and number of time-steps. The results are presented in Figures 9.4 to 9.6, with fully explanatory captions.

Figure 9.4 shows contours of equal workload in the Cartesian space of number of cells and number of time-steps. The contours are described by a sequence of rectangular hyperbolas $k = mn$ where k is the workload, m is the number of time-steps and n is the number of cells. The workload contour of constant at 8×10^6 cell-solutions is highlighted in red to match the limit of the data plotted in Figure 9.4. In that Figure, the average work-rates for each simulation (cell-solutions per second) are plotted in the Cartesian space and coloured blue according to the colour bar shown on the right-hand side. The work-rates plotted in Figure 9.4 are those computed using the baseline model version, and exhibit a stable mean with low variance (see Figure 9.6) and exhibit a relatively low work-rate compared to the other procedures.

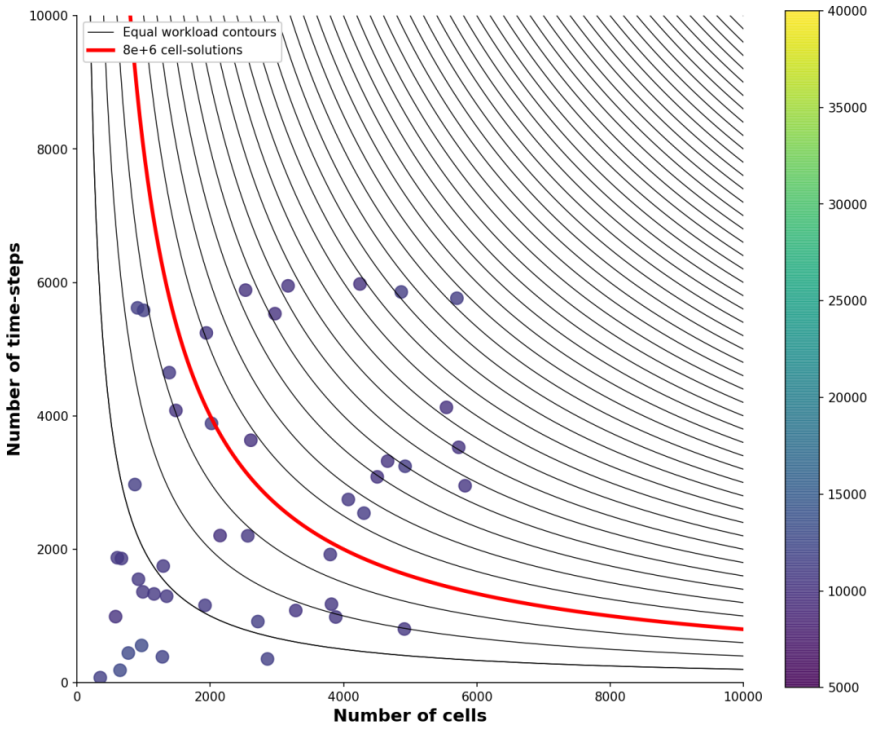


Figure 9.4: Response surface of the baseline model version; the colour bar for the dots is the number of cells solved per second. The sequence of rectangular hyperbolas shows contours of equal computational workload at intervals of 2×10^6 cell-solutions. The 8×10^6 contour is highlighted for reference (see Figure 9.6).

The number of cell-solutions per second is plotted for various combinations of number of cells and number of time-steps. The Cartesian location of each point describes the number of cells versus time-steps, while the colour indicates the rate of cell solutions per second. Note that for this version of the model, the solution rate is independent of the workload at approximately 10000 cells solved per second as all the dots have the same hue. This means that the total computation time increases rapidly with the workload.

In Figure 9.5 we have plotted the cell-solution rates for the parallel model version. Two things are immediately obvious from this Figure. First, the solution rates for small and moderate sized workloads are much higher for the parallel version (matching the results shown in Figure 9.3, as expected) and second, there is a sudden and severe drop in work-rate close to the workload contour of $4x10^6$ cell-solutions. Unfortunately, we do not yet have an explanation for this drop. Luckily it is not very difficult to break the simulation time-series into chunks that will constrain the workload to the 'good' region.

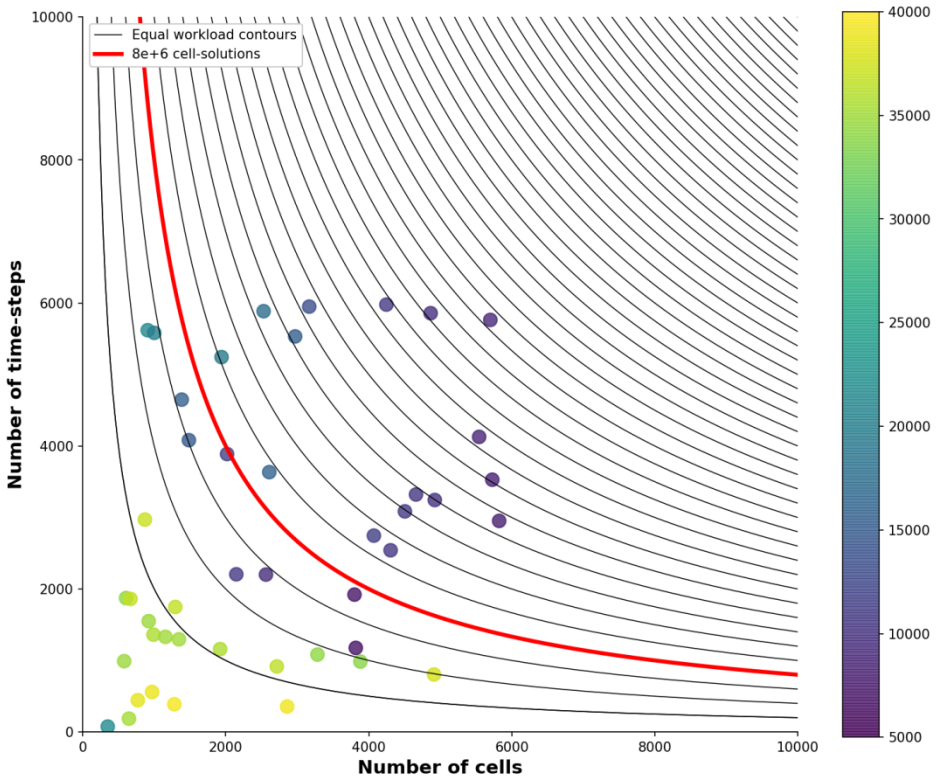


Figure 9.5: Response surface of the current parallel model version; the colour bar for the dots is the number of cells solved per second. The sequence of rectangular hyperbolas shows contours of equal computational workload at intervals of $2x10^6$ cell-solutions. The $8x10^6$ contour is highlighted for reference (see Figure 9.6).

The number of cell-solutions per second is plotted for various combinations of number of cells and number of time-steps. The Cartesian location of each point describes the number of cells/time-steps, while the colour indicates the rate of cell solutions per second. Note the higher solution rate in the lower left hand corner (more yellow than blue) compared with Figure 9.4, and also the severe drop off above the $4x10^6$ contour.

Figure 9.6 reinforces the message of Figures 9.4 and 9.5, and also demonstrates our ability to model the 49 000 cells required to produce soil moisture and ET simulations over the EXSMET region as shown in Section 11.

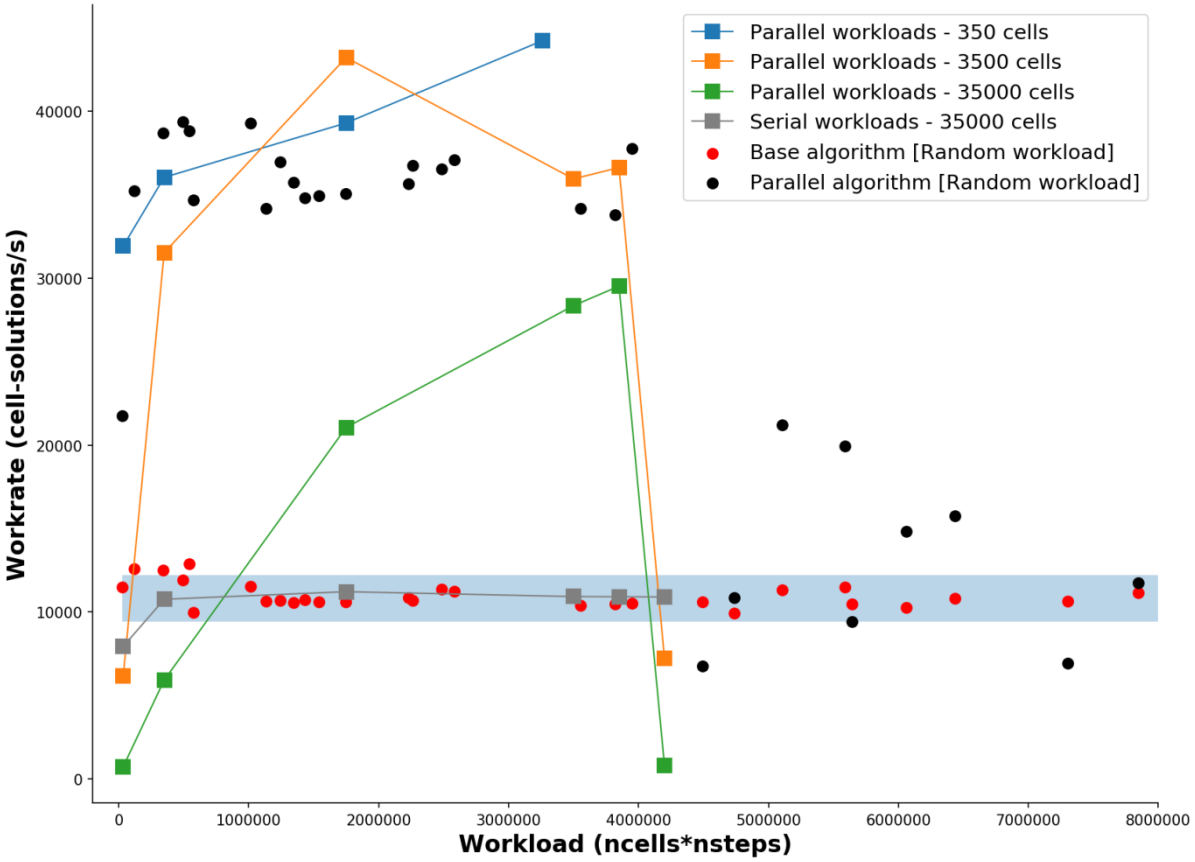


Figure 9.6: A scatter plot of the work-rates achieved for total workloads below $8x10^6$ cell-solutions (the red contours in Figures 9.4 and 9.5).

In Figure 9.6, the red dots show the solution rates achieved for the set of randomly selected workloads shown in Figure 9.4, with the blue bar showing a range of 2 standard deviations either side of their mean. The black dots show the matching solution rates achieved using the newly developed parallel version of the model code (see the yellow to green dots below the contour in Figure 9.5). The coloured squares joined by lines indicate the solution rate for a fixed number of cells and increasing number of time-steps (see the legend) – the time-steps vary by number of model cells to achieve matching workloads.

The first two square markers on each line are for points very close to the x-axis in Figures 9.4 and 9.5 (i.e. a low number of time-steps). This is because when the number of time-steps is very low, the overhead of passing data to the worker processes outweighs their benefits and high solution rates cannot be achieved. The line joining the grey squares (35 000 cells) is an average through the red dots of the serial workloads and has a considerably lower work-rate than the corresponding green line and squares. In any event, it seems that more than 400 000 cells x nsteps causes a collapse in work rate, independent of the number of cells in the parallel workloads.

9.7. Summary of Section 9

Recall that the description of this 10th deliverable in the project proposal reads '***Devise means of increasing computational speed and capacity and demonstrate ability to make estimates of SM and ET over SADC***'. We have significantly revised the PyTOPKAPI model code and implemented algorithms which allow the model to take advantage of the multiple CPU cores typically found in modern computers. This model advancement is entirely transparent to users, and does not require any special interventions on their part.

Although there are still some difficulties experienced with the parallel implementation for large workloads, work-arounds can be employed in a simple manner, and the new code is faster for 77.8% of the cases tested. The strategy is to make computations in chunks that limit the total workload to the region of maximum solution speed. In any event, the serial code path remains available through the use of a simple function parameter passed, on instruction from the user, to the main routine of the model for problem cases.

In addition, we have clearly demonstrated our ability to run the model for the 49 000 cells comprising our set-up for the full extent of SADC shown in Figure 10.1. The average solution speed improvement for moderate workloads is 2-2.5 times faster than the baseline serial version of the model code. This does not make much difference if updates are made once per day, however there are often occasions when the entire simulation must be re-run due to problems with input data, power failures and other hazards of real-time operation. This is where the speed improvements really make a significant difference to EXSMET. For example, a 25 year long daily simulation can be completed on a powerful laptop in 3 hours, compared with 8 hours for the baseline model code. This makes for a far more productive use of modelling time in real world applications of the model.

We have met the requirements of this deliverable and made important changes to the PyTOPKAPI model structure which pave the way for exciting parallel (multi-processor) and distributed (multi-machine) processing for large scale physically based hydrological modelling. As always, the latest model code is freely available at https://github.com/sahg/PyTOPKAPI under a BSD style license (Sinclair et al., 2017).

-----ooOoo----

10. Model initialization

In this 11th deliverable entitled ***'Determine the best ways of cold-starting calculations [model initialization]'*** we relied heavily on what has gone before in this project. The difference here is that we are not only on our home ground of RSA, which we have been modelling since 2007, but also North of us in the other Southern African Development Community (SADC) countries, which stretch beyond the Equator. RSA is rich in physical information that informs our PyTOPKAPI model, derived from many years of careful data collection and updating by others, to a fine spatial scale. This richness is not shared by our neighbours, so we have had to rely on sparsely sampled variables interpolated by space-borne sensors, resulting in a very much coarser data product than ours.

Nevertheless, the computational load is large, considering that the static physical parameters required for the application of the South African model are obtained from data-sets sampled onto cells of 1 km^2 equal area, in a grid covering the country. In both HYLARSMET (Sinclair and Pegram, 2013) and now EXSMET (this project), 7200 of RSA's 1.22 million cells were used in both projects and were centred on a regular latitude and longitude grid with a spacing of 0.125°. By contrast, the land area of SADC (excluding Madagascar) is 8 950 000 km^2, compared to RSA's 1 220 000 km^2, so is 6.3 times larger. This translates to modelling 48 598 locations, each 1 km^2 in size, at 3-hour intervals, which requires powerful computing capacity. In this Subsection, we have included Tanzania and the DRC for complete coverage of SADC (except for Mozambique), omitted in Subsection 8.

In this Subsection we decided to:

(i) determine the values of the PyTOPKAPI parameters at each of the 48 598 locations over SADC using the Harmonised World Soil Database (HWSD), HYDROSHEDS DEM and USGS Global Land Cover Classes (GLCC) parameter sets introduced in Section 2 of this report.
(ii) compare the GLCC parameter sets against our RSA sets, where they overlap
(iii) show proof of concept, in other words, determine the best ways of cold-starting the calculations, but not to model the whole region.

To achieve the last of the 3 tasks, we randomly selected sites over SADC and computed their 'start-up' times, in other words, the interval from initiation until nearly stable responses. We found, as expected, that there are variations depending on locations, as a result of varying soil and ground cover parameters, besides precipitation regimes. The result is that the length of the warm-up from cold start varies between sites and is between 3 and 6 months from initiation, which is much what we expected in a region which has a high variability of ambient variables (such as temperature) and precipitation.

The result is that the improved version of PyTOPKAPI, in mapping mode, is ready to be used with confidence over the region.

10.1 Introducing the Geographical Variables

In this Subsection we introduce maps of the variables used in the calculations and compare these with our original RSA ones where they overlap. Unfortunately, the colour code of each set does not match in most cases, but the tone is the key to comparison. Figure 10.1 shows the USGS Global Land Cover Characteristics Data Base, Version 2.0 (GLCC) over SADC, classed by colour, omitting Madagascar. The categories are not on a numerical scale, but are illuminated by colour. It is followed by Table 10.1, which translates these values to soil properties, as was shown in Section 4. For most of the parameters, we follow the methodology of DeLannoy et al. (2014) to convert the Harmonized World Soil Database (HWSD v1.2.1; FAO/IIASA/ISRIC/ISS-CAS/JRC (2012)) texture classes to soil hydraulic properties. However, the residual moisture content θ_r, bubbling pressure ψ_b and pore size distribution λ are calculated using the table of Rawls et al. (1982) (see Table 10.1, repeated here for convenience). The main difference from earlier work over RSA is, that for SADC in this Section, we are switching to the new texture maps from the HWSD, because our RSA maps are limited to this country.

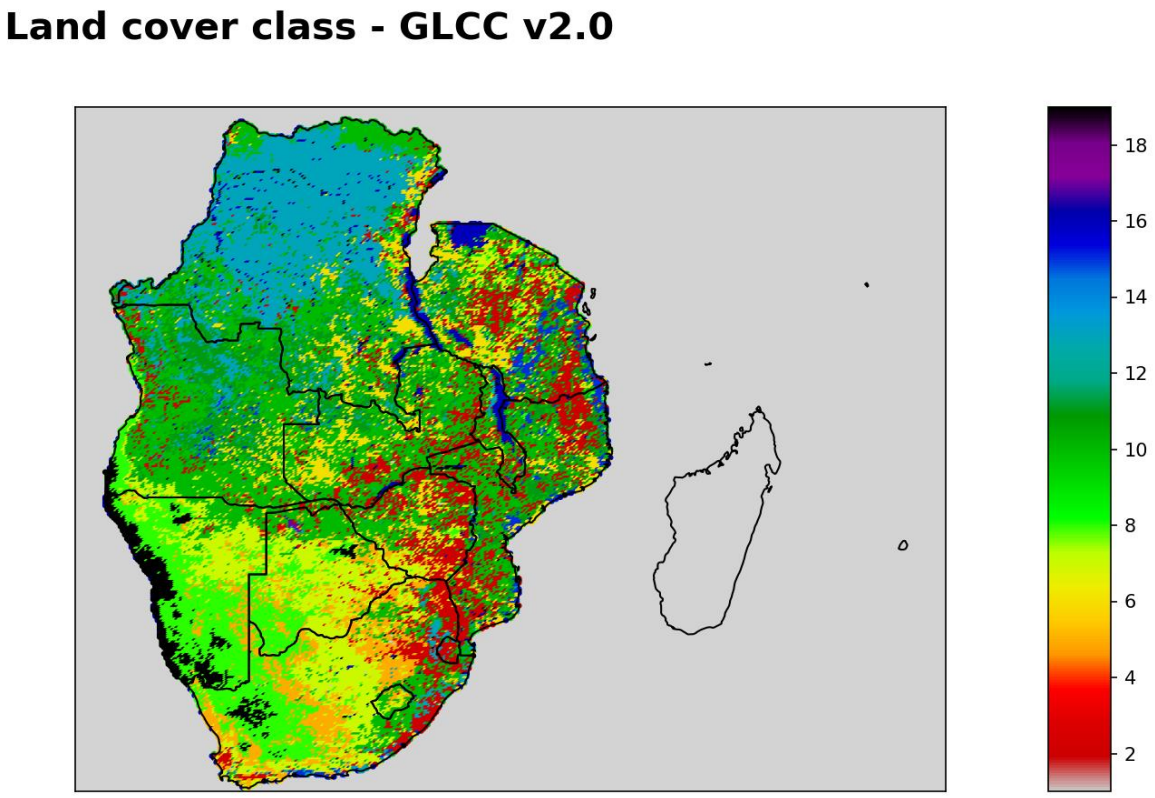


Figure 10.1: USGS Global Land Cover Characteristics Data Base, Version 2.0 (GLCC) over SADC, omitting Madagascar. Besides the countries shaded in Figure 8.1, we have included Tanzania and the DRC, to complete the set.

Table 10.1 follows.

TABLE 2. HYDROLOGIC SOIL PROPERTIES CLASSIFIED BY SOIL TEXTURE

Texture class	Sample size	Total porosity (θ), cm^3/cm^3	Residual saturation (θ_r) cm^3/cm^3	Effective porosity (θ_e), cm^3/cm^3	Bubbling pressure (ψ b) Arithmetic, cm	Bubbling pressure (ψ b) Geometric,† cm	Pore size distribution (λ) Arithmetic	Pore size distribution (λ) Geometric†	Water retained at −0.33 bar tension, cm^3/cm^3	Water retained at −15 bar tension, cm^3/cm^3	Saturated Hydraulic Conductivity‡ (K_s) cm/h
Sand	762	0.437** (0.374–0.500)	0.020 (0.001–0.039)	0.417 (0.354–0.480)	15.98 (0.24–31.72)	7.26 (1.36–38.74)	0.694 (0.298–1.090)	0.592 (0.334–1.051)	0.091 (0.018–0.164)	0.033 (0.007–0.059)	21.00
Loamy sand	338	0.437 (0.368–0.506)	0.035 (0.003–0.067)	0.401 (0.329–0.473)	20.58 (0.0–45.20)	8.69 (1.80–41.85)	0.553 (0.234–0.872)	0.474 (0.271–0.827)	0.125 (0.060–0.190)	0.055 (0.019–0.091)	6.11
Sandy loam	666	0.453 (0.351–0.555)	0.041 (0.0–0.106)	0.412 (0.283–0.541)	30.20 (0.0–64.01)	14.66 (3.45–62.24)	0.378 (0.140–0.616)	0.322 (0.186–0.558)	0.207 (0.126–0.288)	0.095 (0.031–0.159)	2.59
Loam	383	0.463 (0.375–0.551)	0.027 (0.0–0.074)	0.434 (0.334–0.534)	40.12 (0.0–100.3)	11.15 (1.63–76.40)	0.252 (0.086–0.418)	0.220 (0.137–0.355)	0.270 (0.195–0.345)	0.117 (0.069–0.165)	1.32
Silt loam	1206	0.501 (0.420–0.582)	0.015 (0.0–0.058)	0.486 (0.394–0.578)	50.87 (0.0–109.4)	20.76 (3.58–120.4)	0.234 (0.105–0.363)	0.211 (0.136–0.326)	0.330 (0.258–0.402)	0.133 (0.078–0.188)	0.68
Sandy clay loam	498	0.398 (0.332–0.464)	0.068 (0.0–0.137)	0.330 (0.235–0.425)	59.41 (0.0–123.4)	28.08 (5.57–141.5)	0.319 (0.079–0.559)	0.250 (0.125–0.502)	0.255 (0.186–0.324)	0.148 (0.085–0.211)	0.43
Clay loam	366	0.464 (0.409–0.519)	0.075 (0.0–0.174)	0.390 (0.279–0.501)	56.43 (0.0–124.3)	25.89 (5.80–115.7)	0.242 (0.070–0.414)	0.194 (0.100–0.377)	0.318 (0.250–0.386)	0.197 (0.115–0.279)	0.23
Silty clay loam	689	0.471 (0.418–0.524)	0.040 (0.0–0.118)	0.432 (0.347–0.517)	70.33 (0.0–143.9)	32.56 (6.68–158.7)	0.177 (0.039–0.315)	0.151 (0.090–0.253)	0.366 (0.304–0.428)	0.208 (0.138–0.278)	0.15
Sandy clay	45	0.430 (0.370–0.490)	0.109 (0.0–0.205)	0.321 (0.207–0.435)	79.48 (0.0–179.1)	29.17 (4.96–171.6)	0.223 (0.048–0.398)	0.168 (0.078–0.364)	0.339 (0.245–0.433)	0.239 (0.162–0.316)	0.12
Silty clay	127	0.479 (0.425–0.533)	0.056 (0.0–0.136)	0.423 (0.334–0.512)	76.54 (0.0–159.6)	34.19 (7.04–166.2)	0.150 (0.040–0.260)	0.127 (0.074–0.219)	0.387 (0.332–0.442)	0.250 (0.193–0.307)	0.09
Clay	291	0.475 (0.427–0.523)	0.090 (0.0–0.195)	0.385 (0.269–0.501)	85.60 (0.0–176.1)	37.30 (7.43–187.2)	0.165 (0.037–0.293)	0.131 (0.068–0.253)	0.396 (0.326–0.466)	0.272 (0.208–0.336)	0.06

* First line is the mean value
Second line is ± one standard deviation about the mean
† Antilog of the log mean
‡ Obtained from Fig. 2

Table 10.1: The Table 2 from Rawls et al. (1982), used to estimate soil properties from texture information.

The next set of Figures (10.2 to 10.7) displays the translation of Land Cover Characteristics, pixel by pixel from the data displayed in Figure 10.1, and HWSD soil texture information shown in Figure 2.5, to soil properties as used by PyTOPKAPI, via the values in Table 10.1.

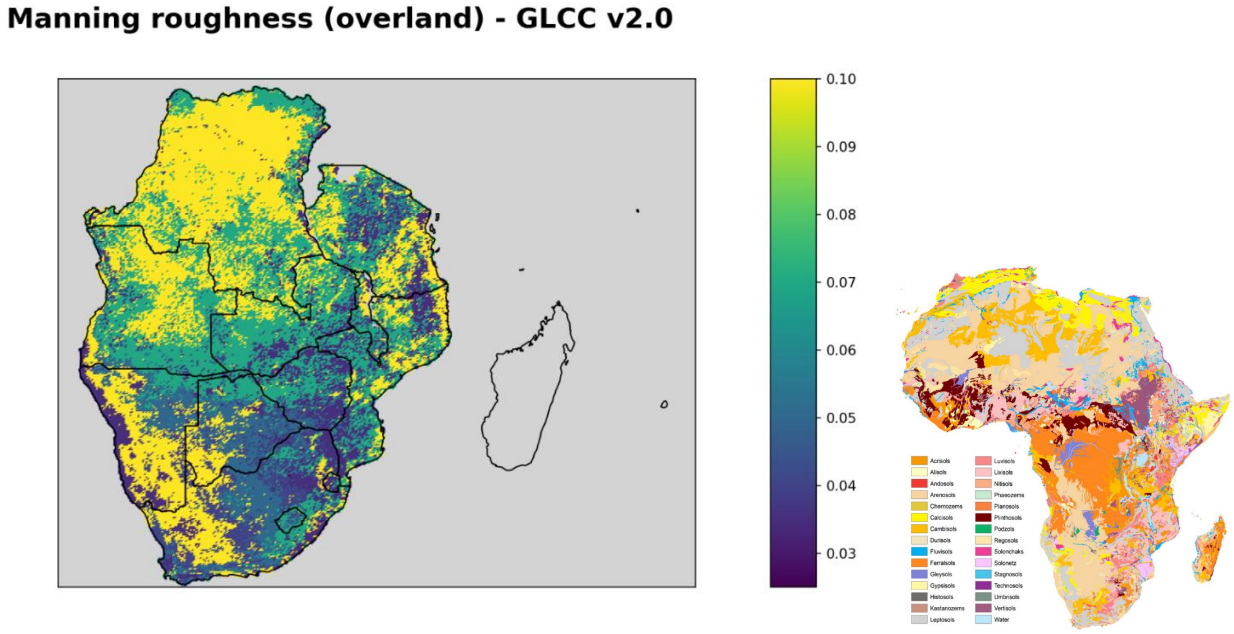


Figure 10.2: Manning's roughness coefficient [n_0] for overland flow, determined from Figure 10.1 and Table 10.1. A small copy of Figure 2.5 is on the right, for information.

Figure 10.2 displays the roughness of the ground surface for overland flow, not to be confused with channel flow resistance. We are not modelling streamflow in this project, but estimating soil saturation and consequent evapotranspiration, cell by cell, each 1 km square at a spacing of 0.125°. We now follow with maps of estimates of soil properties over SADC.

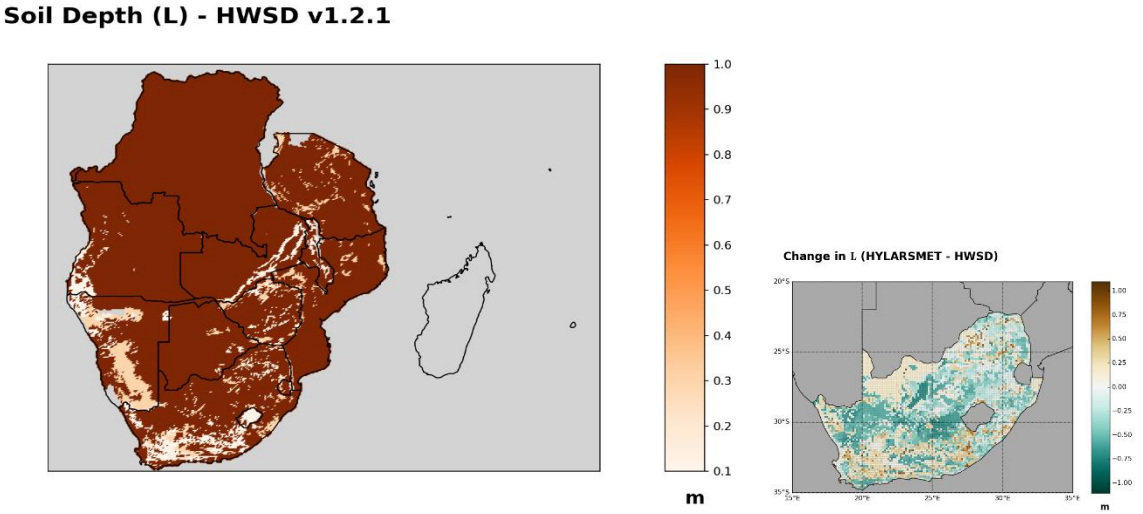


Figure 10.3: The soil depth *L* over SADC, from the Harmonized World Soil Database in the left panel. On the right is the difference in values of *L* from HYLARSMET from HWSD estimates over RSA.

Except for RSA, there is a lack of detailed information over SADC in the HSWD data-base, hence the almost universal 1 m depth over the region in the panel on the left of Figure 10.3, which is a reasonable technical compromise. Note the negative changes (green) over RSA in the right panel, which are typically dry areas with shallow soils, whose *L* is overestimated by HWSD compared to our RSA data.

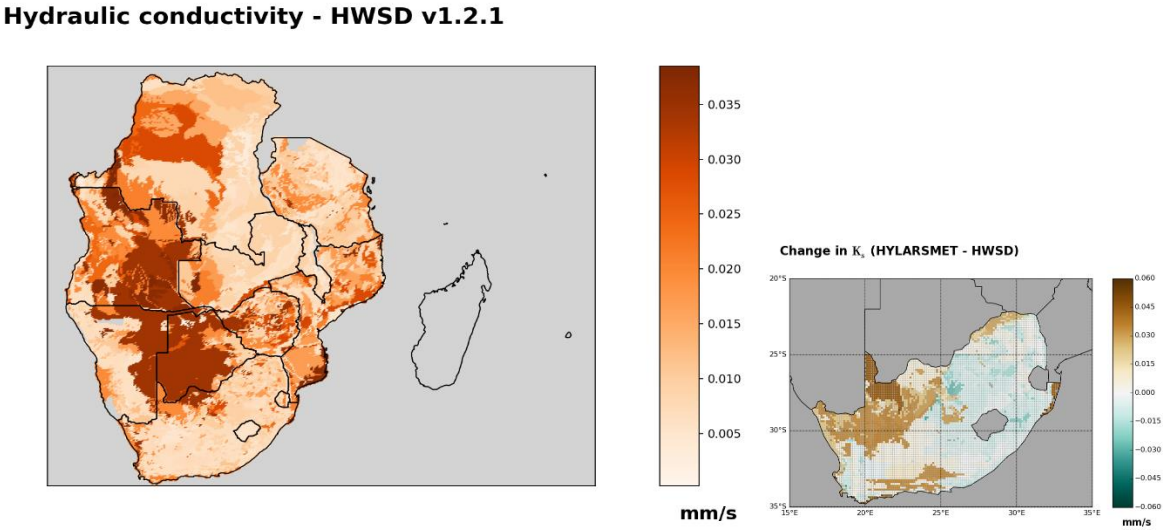


Figure 10.4: Saturated Hydraulic conductivity K_s using Figure 10.1 and Table 10.1 in the left panel, differences between HYLARSMET and HWSD in the right panel

Comparing the map on the left with the corresponding images of differences between HYLARSMET and HWSD shown on the right of Figure 10.4, it is evident that the differences over RSA are not too large – there is fair correspondence. Nevertheless, conductivity K_s estimated by HWSD is larger than

the corresponding HYLARSMET values in the drier areas and vice versa, particularly in the North-West, as shown in the panel on the right.

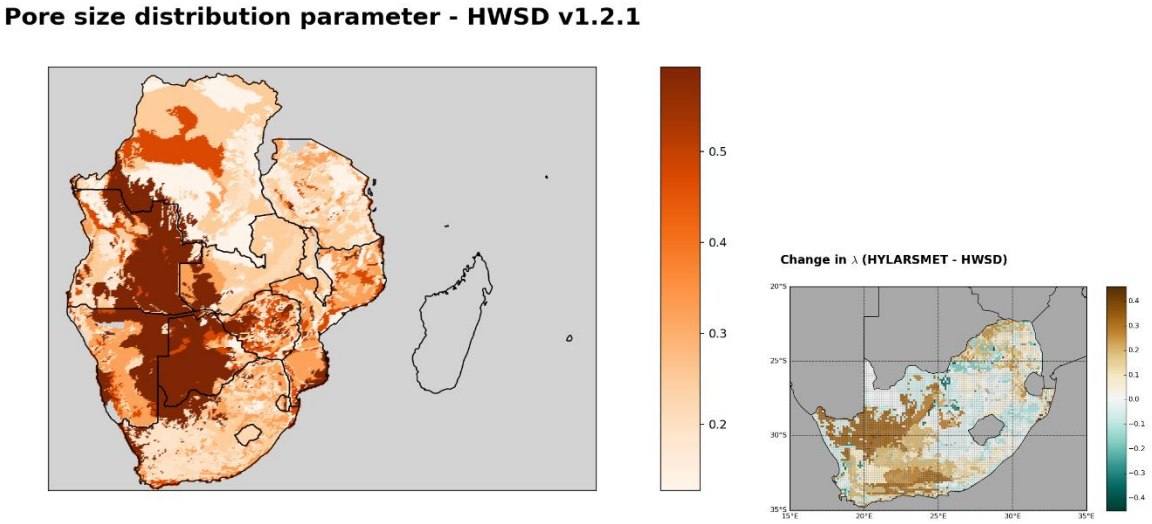


Figure 10.5: Pore size distribution λ using the information in Figure 10.1 and Table 10.1 in the left panel. Differences between HYLARSMET and HWSD in values is on the right.

Comparing the image of SADC Pore size distribution values λ on the left with the corresponding image on the right of Figure 10.5, the differences over RSA are considerable. It seems that Pore size distribution λ is larger in the drier areas and vice versa. This is particularly so over Botswana and Angola. Note that the relative differences (dark vs light shading) of HSWD in Figure 10.5 in the left panel are close to those in Figure 10.4 above, as Hydraulic conductivity K_s increases with λ, as shown in Table 10.1.

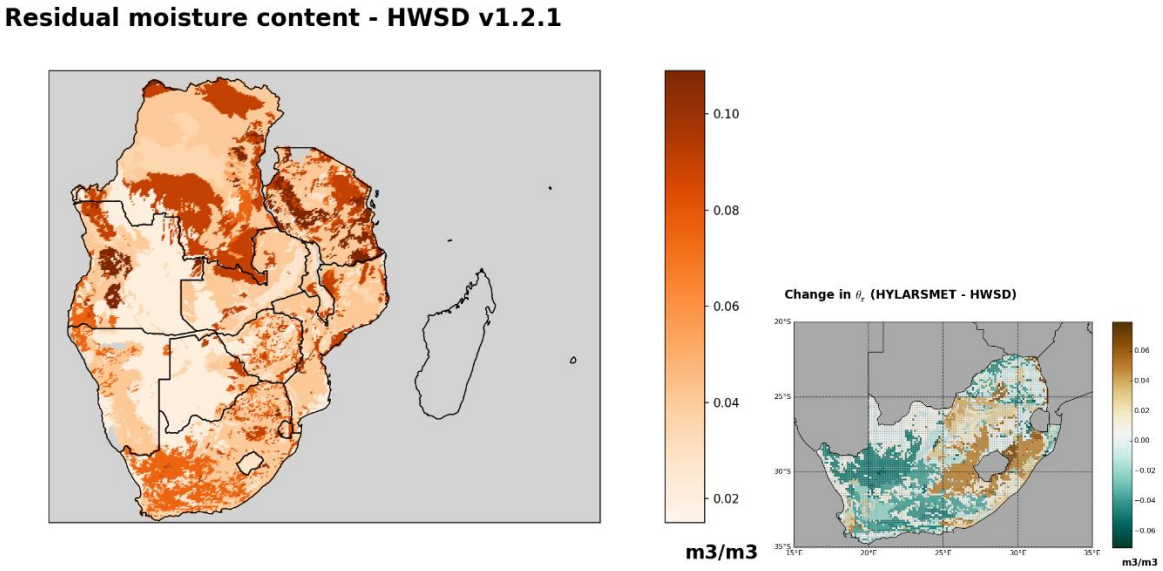


Figure 10.6: Residual soil Moisture content θ_r using Figure 10.1 and Table 10.1 in the left panel. Differences between HYLARSMET and HWSD in values is on the right.

Comparing the Figure of SADC values of Residual soil Moisture content θ_r on the left with the corresponding image of differences on the right, the differences over RSA are marked. Again, the dry areas have this parameter overestimated by HWSD, compared to HYLARSMET values, as shown by the blue-green negative changes compared to the brown positive ones in the wetter areas in the right panel.

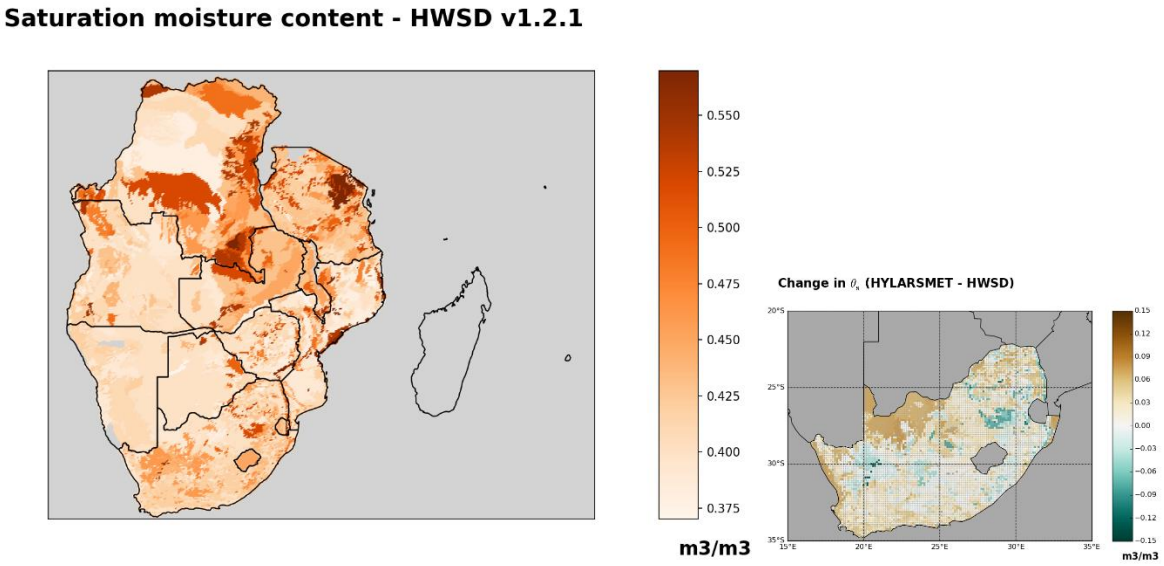


Figure 10.7: Saturation moisture content θ_s using Figure 10.1 and Table 10.1 in the left panel. Differences between HYLARSMET and HWSD values are on the right.

Comparing the Figure of HWSD values of Saturation soil Moisture content θ_s over SADC on the left with the corresponding image of differences on the right, the differences over RSA are moderate. The dry areas have this parameter slightly underestimated by HWSD, compared to HYLARSMET values, as shown by the small patch of brown in the Northwest and West coast. The remainder of RSA has been relatively well modelled, except for some green patches.

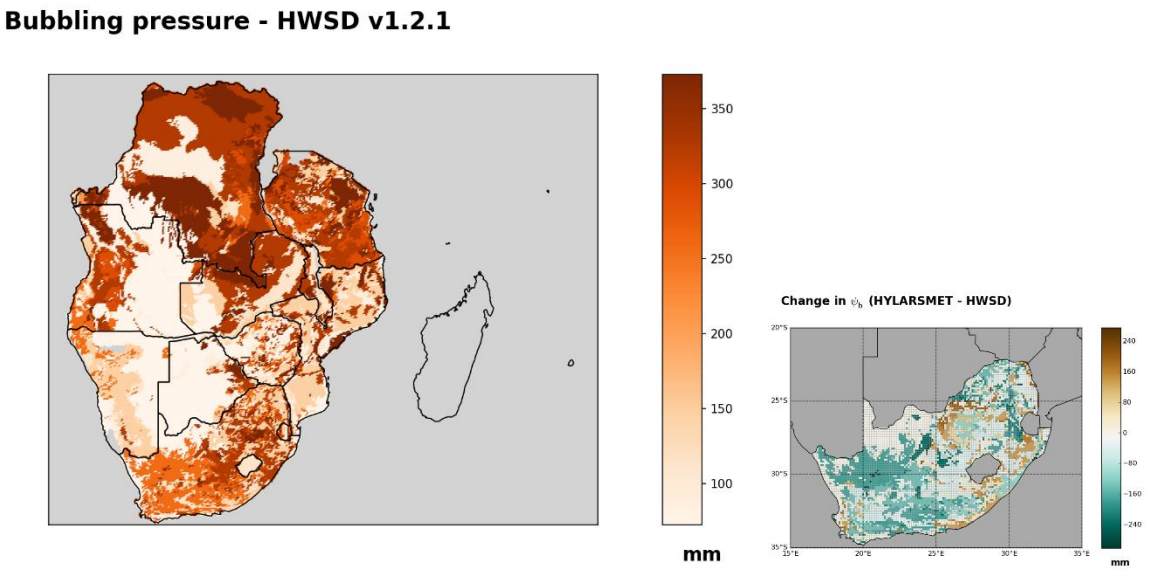


Figure 10.8: Bubbling Pressure, ψ_b, using Figure 10.1 and Table 10.1 in the left panel. Differences between HYLARSMET and HWSD values on the right.

There is considerable overestimation of Bubbling Pressure, ψ_b, as shown by the image of differences in the panel on the right, showing differences between HYLARSMET and HWSD; the differences over RSA are large. In the dry areas this parameter has again been overestimated by HWSD, compared to HYLARSMET values, as shown by the blue-green negative changes, compared to the small patches of brown positive ones in the coastal and Eastern areas of RSA.

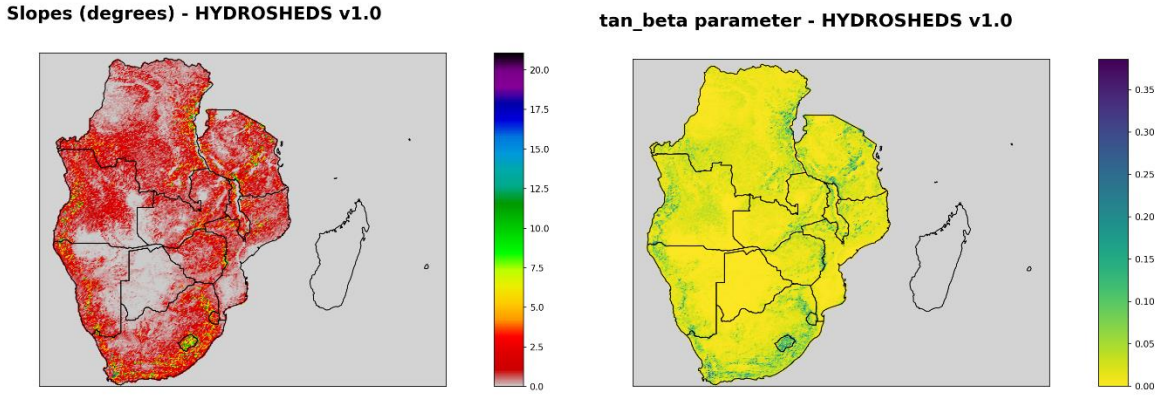


Figure 10.9: Slopes of the ground in SADC, calculated from the HYDROSHEDS product estimated by the USGS, here at 1 km resolution. The panel on the left show slopes in degrees, which have been converted to tangents [tanβ] in the right panel, as required by the Manning equation in PyTOPKAPI.

The HYDROSHEDS product of USGS is offered as a Digital Elevation Model, at a 3 arc-second resolution, over the globe. At the equator, this distance is equivalent to approximately 90 metres. As we are interested in 1 km square pixels for modelling, transformation of these data was required, requiring a considerable investment in computation skill and time. PyTOPKAPI requires the tangent of the angle of the slope [S_f = tanβ] as a parameter for the Manning's equation of flow over a surface. We did not map this variable in Deliverable 4 of the project, but decided to illustrate it here for information. It seems from the image on the left that Africa is quite flat. However, the image on the right picks out the steepness of mountainous areas of SADC in green, compared to the dominant yellow in the interior and Southern Mozambique.

10.2 Section summary

This Section was bent on data collection and comparisons for the purpose of defining the PyTOPKAPI model application – to EXport SM and ET over SADC. There is nothing more to add here, except to note that it concludes the important and arduous effort required to obtain the Geographical Variables, as Maps of the parameters, required to run PyTOPKAPI.

-----ooOoo----

11. Behaviour of soil storage volumes and ETa simulated over the SADC region

This penultimate Section does not deal with one of the individual contract deliverables as above, but is an important summary of what we have achieved in the project, by using a set of demonstration calculations. First, we use PyTOPKAPI to model 6 selected 1 km square pixels randomly chosen over SADC, to display actual Evapotranspiration (ETa) traces (as against potential ET). ETa is here used in the calculation of the Soil Storage Index (SSI) at 3-hour intervals over 90 days. These 6 cells are initialised as having a range (25%, 50% and 75%) of their soil storages, and their progress over 3 months is recorded. In addition, we determine the computation load of obtaining the SSI after 90 days of precipitation and ETa, at all cells over the whole of SADC, at each of the 1 km pixels in the 0.125° squares. Again, these are individually initialised with three different initial storage states, 25%, 50% and 75% of saturation. This information is important to obtain, because the computational load for lengthy simulation over 48 598 sites is considerable. It is therefore prudent to determine these limitations before startup of an extended calculation in space and time over this large area. The SSI responses, as time series of moisture content, are affected by three major variables:

(i) soil storage capacity, as a function of porosity and saturation moisture content
(ii) evapotranspiration rate, based on available SM, temperature and humidity and
(iii) the local precipitation regime.

The 6 sites' traces of storage follow in 6 images in Figures 11.1 a to f.

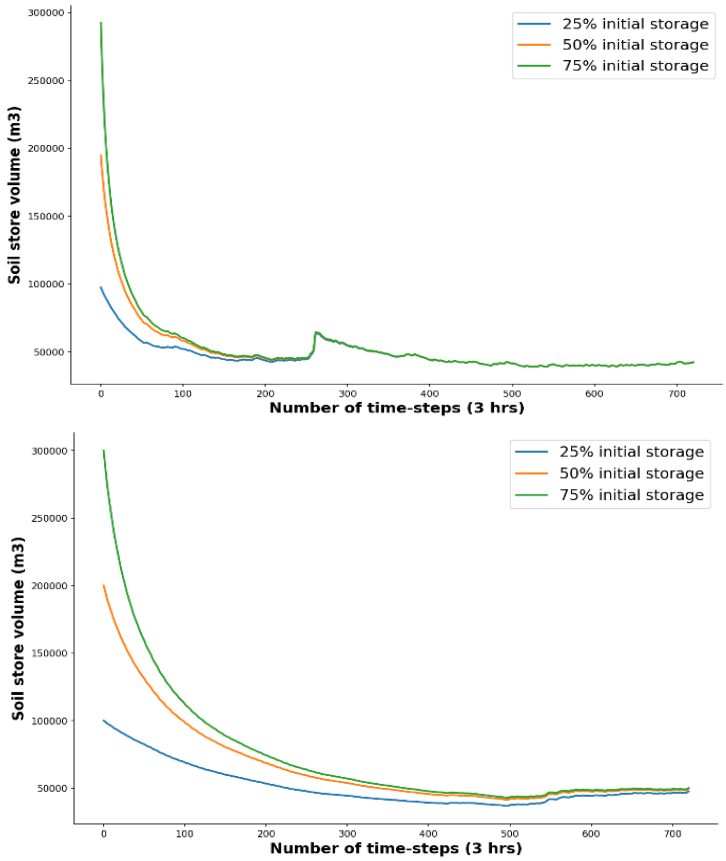


Figure 11.1 a and b: Soil water storage traces in 1 km square areas reacting to local rainfall and ETa over 3 months of computation. These are for the first 2 of 6 sites chosen at random, started at quarter, half and three-quarter full states, ranked by spread of end-states; panels c to f overleaf.

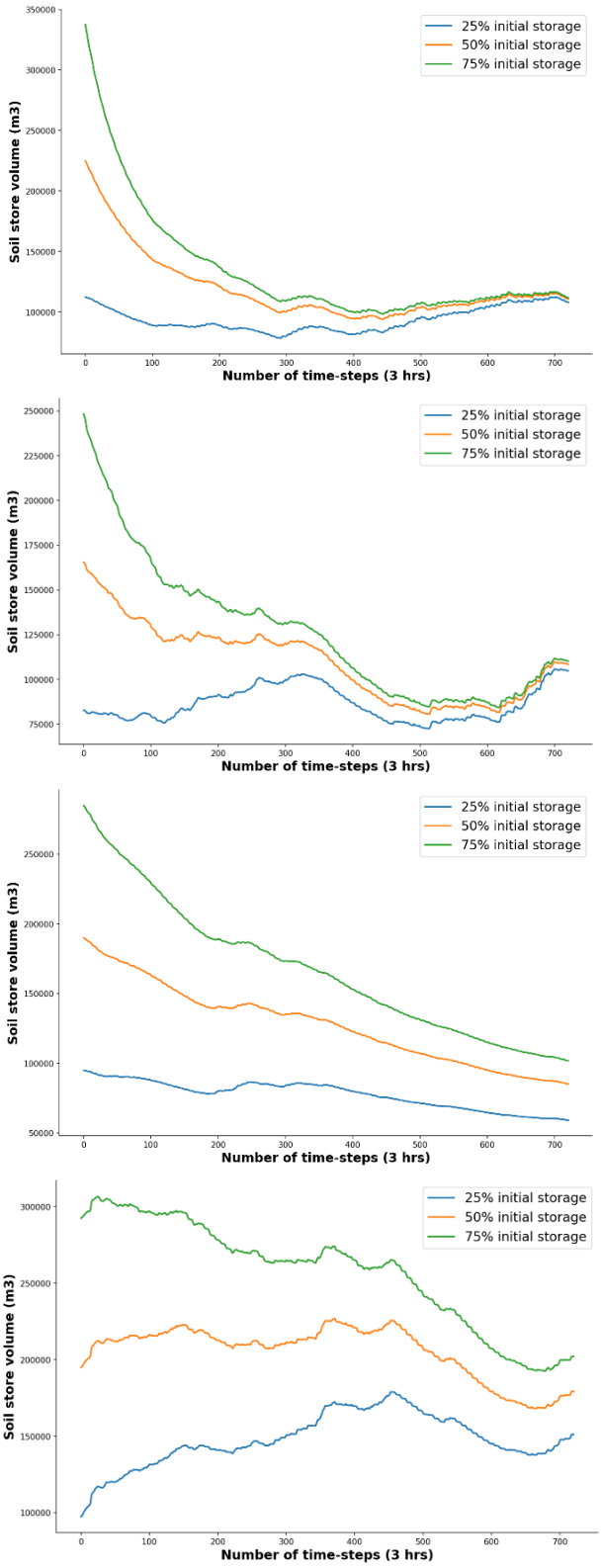


Figure 11.1 c to f: Soil water traces forced by local rainfall and ETa over 3 months of computation for another 4 sites chosen at random, started at quarter, half and three-quarter full states, ranked by spread of end-states.

The duration of the traces in Figure 11 is 90 days and the start times at each site were randomised. The 6 images in Figure 11.1 have been ranked sequentially by the tightness of the 3 individual traces at the end of the 3-month period. It is clear from the traces that the convergence depends on the soil properties, ET and the rainfall regime governing the responses, so that the wider the 3 traces

are at the end of 3 months, the longer the warm-up period required. The next set of four images in Figure 11.2 comprises the maps of initial and final Soil Saturation Index (SSI) for 3-month simulations started at 25%, 50% and 75% saturation, over the whole of SADC's 49 572 1 km-square pixels, centred in 0.125° squares, demanding a large amount of care and computation. We show the maps at start and end of three months' soil moisture estimation over the whole area of the conterminous SADC countries, each starting from 3 different initial states as in Figure 11.1.

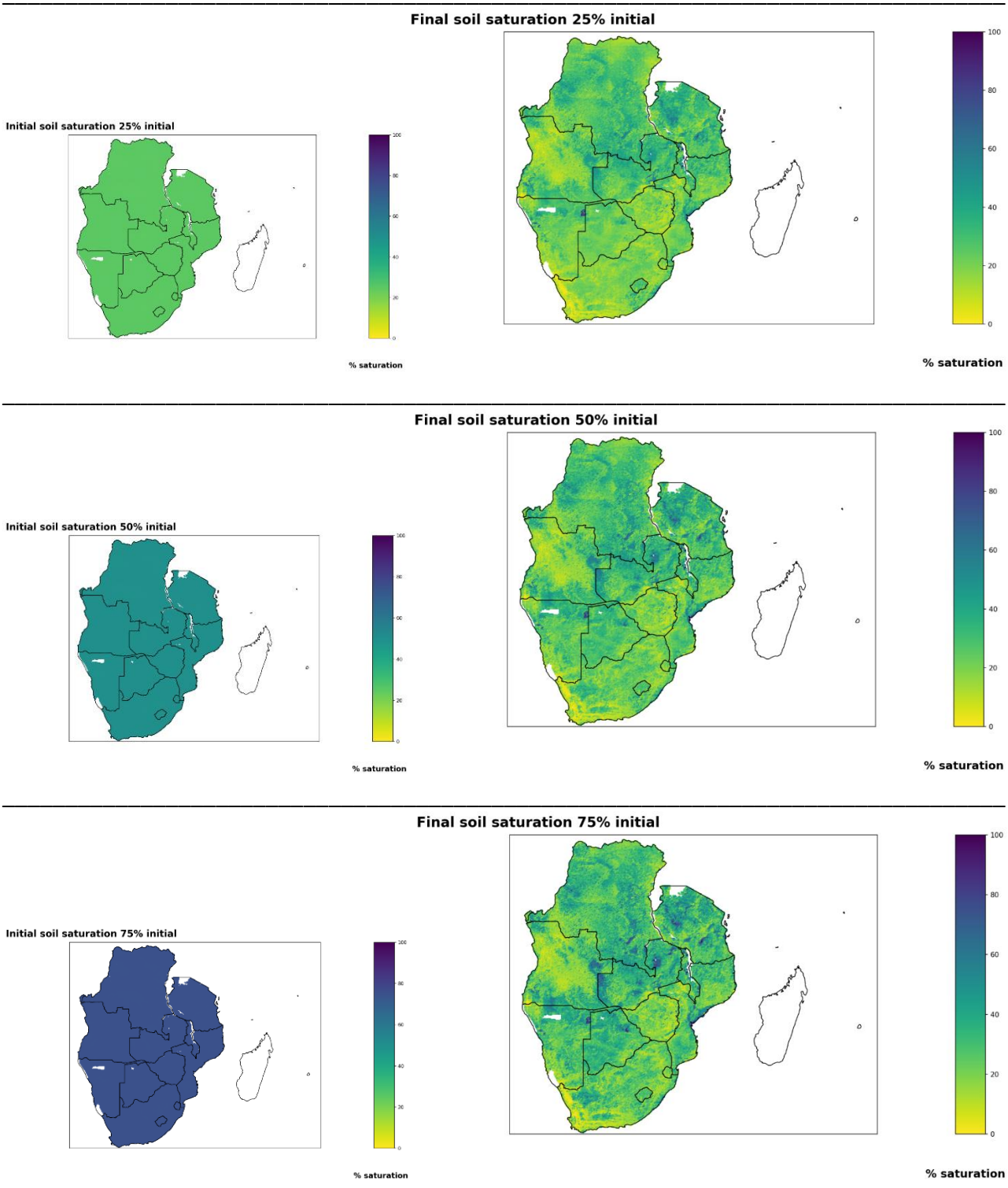


Figure 11.2: Results of simulations of soil moisture content on all of the 49 572 1 km-square pixels over the SADC region, starting from three levels of local saturation: 25%, 50% and 75% and running for 3 months.

Besides the records of SSI, also of interest is the variation, both temporal and spatial, of actual Evapotranspiration (ETa), the important driver forcing these storages. ETa is calculated over the EXSMET region in the same manner as the calculations carried out for the HYLARSMET region over South Africa. The difference for EXSMET is that we do not have access to the meteorological variables from the SAWS weather model. For the purposes of demonstrating the feasibility of running the model over SADC we used meteorological forcing from the WATCH Forcing Dataset ERA Interim (WFDEI) in Weedon et al., 2014. This is useful for historical analyses, but not for operational soil moisture modelling.

From the WFDEI paper's abstract: 'The WFDEI meteorological forcing data set has been generated using the same methodology as the widely used WATCH Forcing Data (WFD) by making use of the ERA-Interim reanalysis data. We discuss the specifics of how changes in the reanalysis and processing have led to improvement over the WFD. We attribute improvements in precipitation and wind speed to the latest reanalysis basis data and improved downward shortwave fluxes to the changes in the aerosol corrections. Covering 1979-2012, the WFDEI will allow more thorough comparisons of hydrological and Earth System model outputs with hydrologically and phenologically relevant satellite products than using the WFD.'

Shown below in Figure 11.3 is a sequence of time series of 3 months calculated ETa at a site chosen at random, and a close up of the first few days at the start of the period in Figure 11.4:

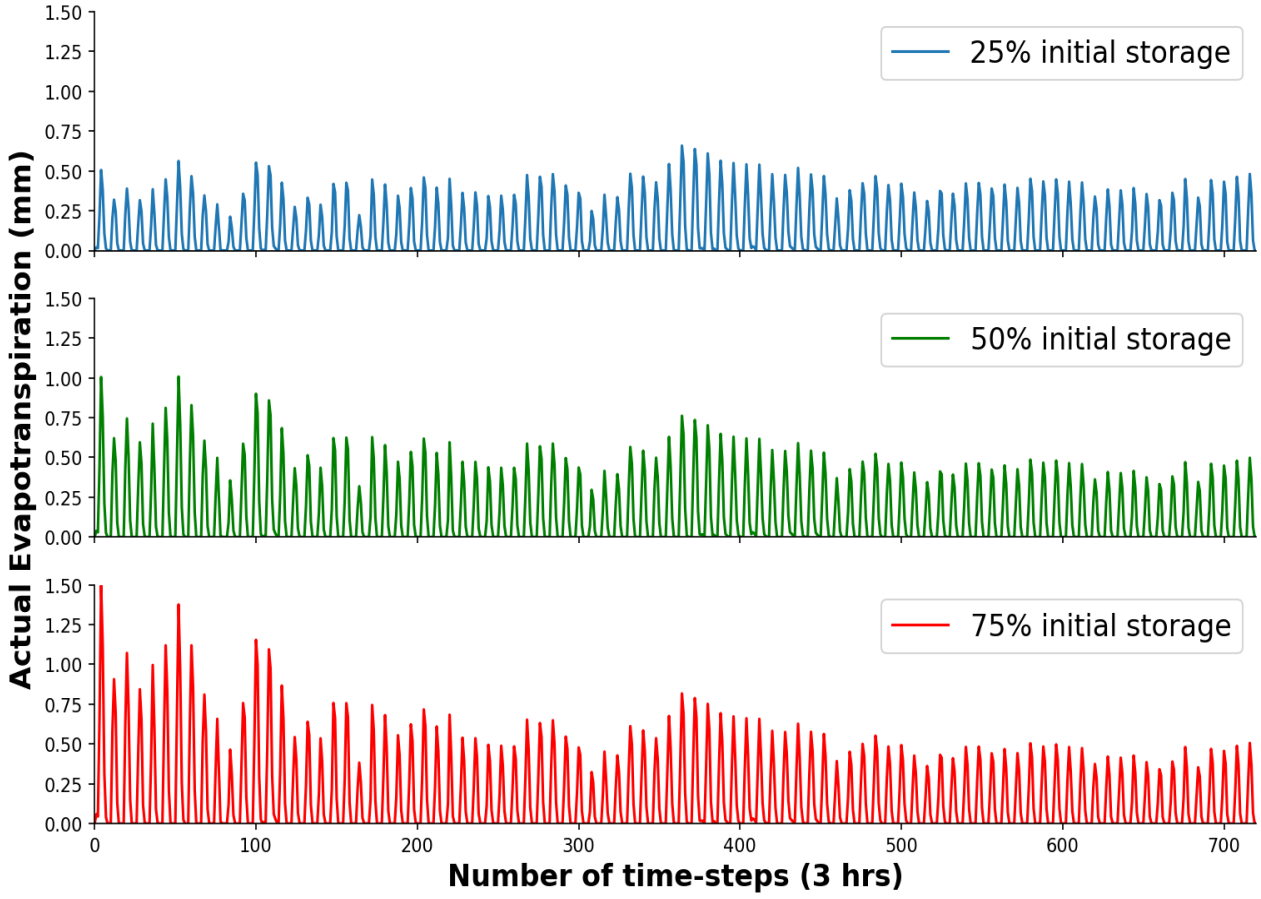


Figure 11.3: Three traces of ETa over 90 days at 3 hour intervals at a selected site. Note the gradual diminution of the red trace of ETa as the soil store dries out with time, having started 75% full, but the close correspondence between all the traces at the end of the 3-month period.

Figure 11.4 shows a combined close-up of the first 12 days of the traces of ETa in Figure 11.3.

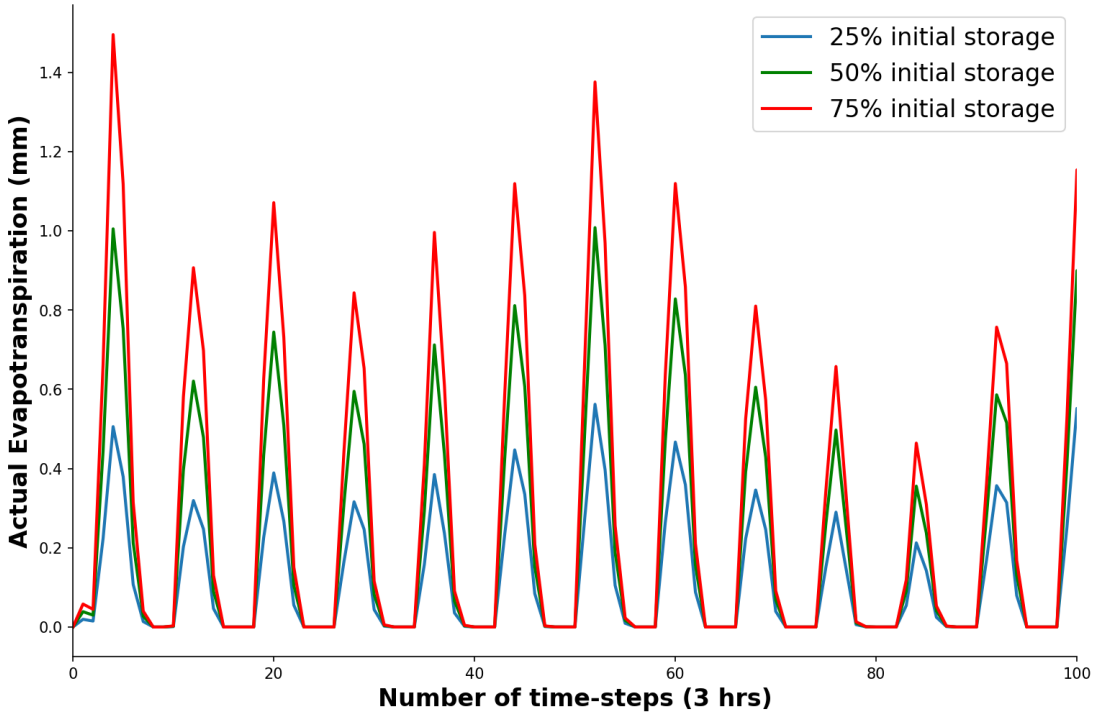
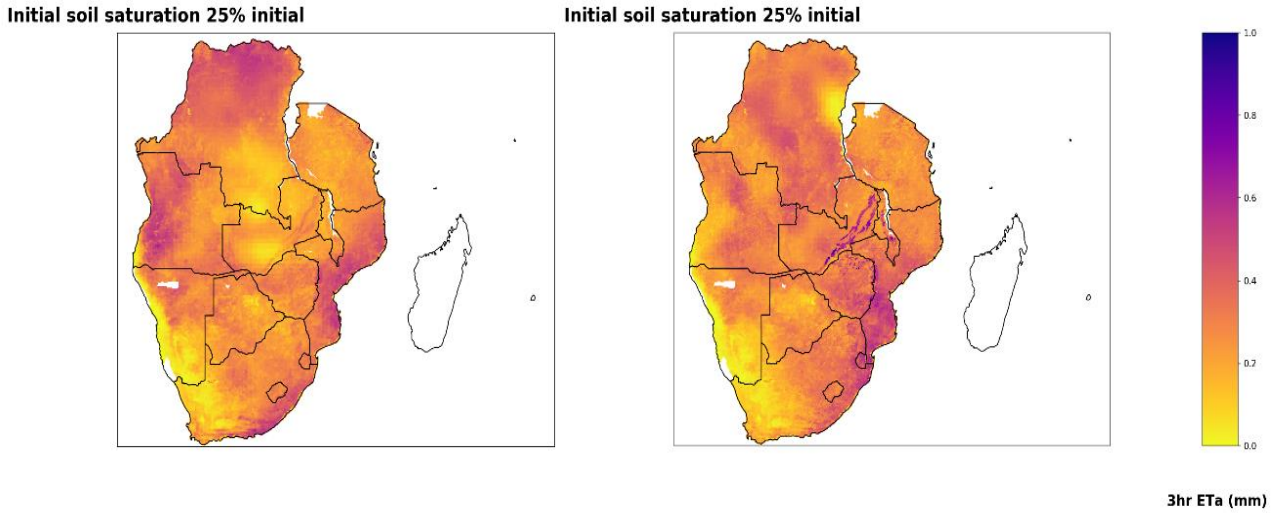


Figure 11.4: Three traces of ETa over the initial 12 days (300 hours) at 3 hour intervals at a selected site, giving detail by combining the three traces in Figure 11.3 – note the diurnal variation.

Of interest is the spatial variation of ETa over SADC. In Figures 11.5 a and b we show two maps essentially snapshots) of accumulated ETa over the SADC region. These are shown at different time intervals of 3-hours, with the initial SSI conditions set at 25%, hence the different patterns, especially in the wet interior. The scale is from 0 to 1 mm.



Figures 11.5 a and b. Two maps of 3 hour accumulated ETa over the SADC region, sampled at different times.

In Figure 11.6, we present the full period of 3-month total accumulated SSI, given the 25%, 50% and 75% SSI initial conditions at each 1 km site.

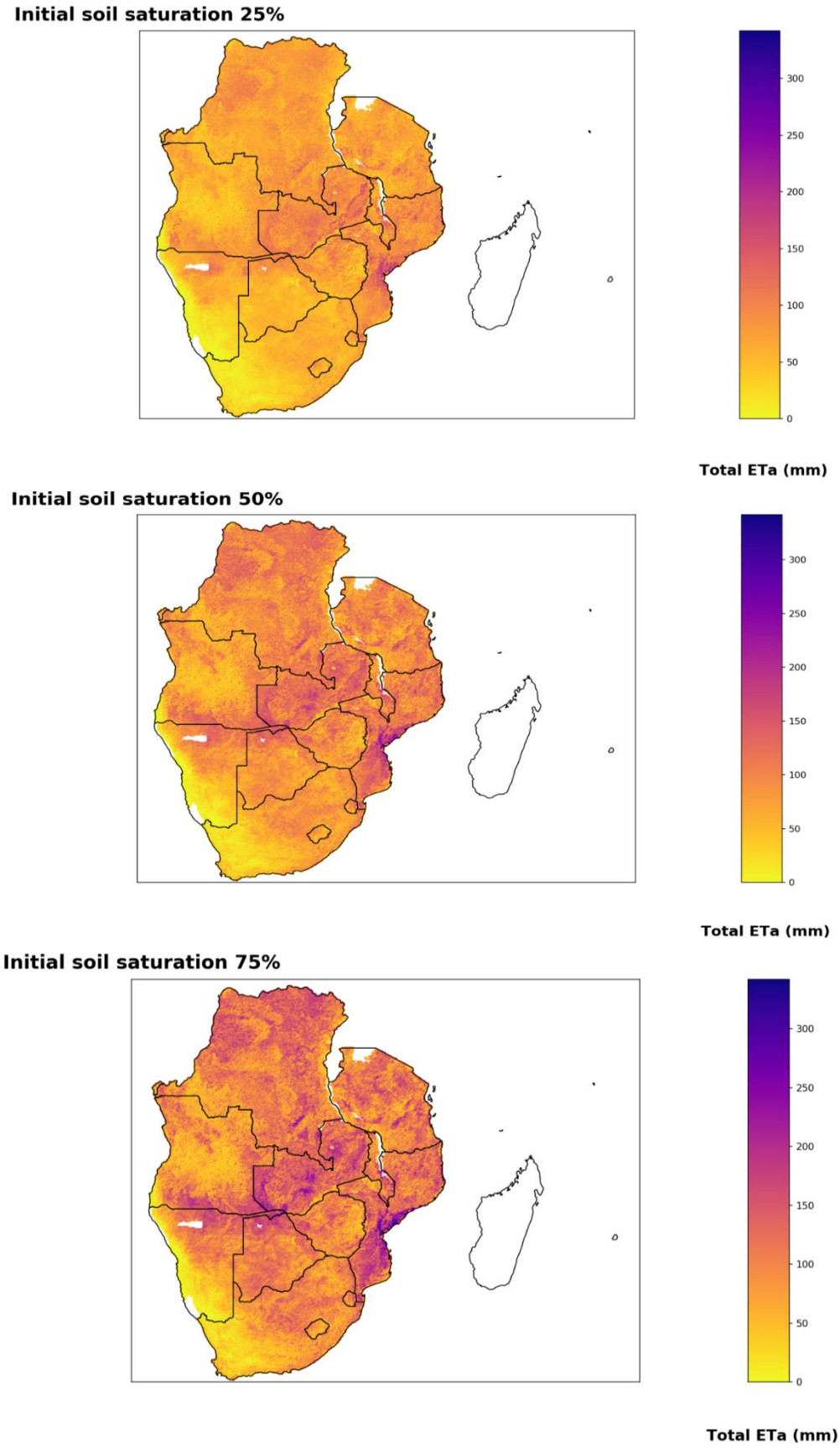


Figure 11.6: The full period of 3-month total for the 25%, 50% and 75% SSI initial condition.

In Figure 11.6, it is clear where the water-rich areas are, with significantly higher ETa losses. By contrast, the drier areas in particular are found over the South West coastline of Africa. The white patches in the maps in Figures 11.2, 11.5 and 11.6 are locations of water-bodies.

11.1 Summary of Section 11

In this chapter, based on previous work, not only have we shown that it is feasible, but have demonstrated that we have indeed managed to export PyTOPKAPI and HYLARSMET over SADC including RSA, with extended spatial and computational capacity of Soil Moisture and Evapotranspiration, for flood and drought monitoring. The tools developed herein are ready to be used by practitioners to continue the work.

-----ooOoo----

12. Conclusion

This EXSMET project has been a huge challenge, requiring significant retooling of our PyTOPKAPI software and searching through a plethora of global offerings of remotely sensed Earth surface data, in order to make a meaningful estimate of Soil Moisture and Evapotranspiration over SADC. Not only that, but we have had to make sense of which sets of global data are appropriate for the needs we had envisaged in the original contract, proposed 5 years ago. At that time, TRMM was still viable, but the decommissioning of the TRMM satellite's precipitation radar on 7 October 2014 meant that we had to consider alternative remotely sensed precipitation sources. In addition, although we had an inkling that this would happen at the time the proposal was drafted, we had to turn to the Global Precipitation Measurement program (GPM) and compare it with TRMM before its demise. Furthermore, managing the non-trivial task of understanding and exploiting the information on ground data offered by remote sensing technology required developing skills we did not have up front.

Besides the assessment of the Earth's surface properties (land-cover, soil types, etc.) we expected that we needed a means of estimating Evapotranspiration (ET) in detail over SADC. Where we had access to Meteorological data from SAWS in our HYLARTSMET project to so do, we now had no similar meteorological data-sets available over Africa North of RSA for the estimation of ET. We were able to find sources of soil properties and parameters from the Food and Agricultural Organisation's (FAO) global maps. These space-born products were therefore sufficient to drive our PyTOPKAPI software in land surface, rather than hydrological mode, as was similarly done in HYLARSMET. As it happened, it turned out in this EXSMET project, as in HYLARSMET, that the Soil Saturation Index was most sensitive to the rainfall input, then to the soil parameterization and much less to actual evapotranspiration (ETa).

In more focused detail, here follows the list of aims that we listed in our proposal/contract, all of which have been achieved satisfactorily:

1. To ascertain what rainfall data are available in the SADC countries outside our borders and perform checks on their suitability for modelling
2. To obtain suitable ground cover and soil maps over the whole SADC region (e.g. FAO and others), for comparison with those already available in RSA
3. To exploit the sensitivity calculations performed on the RSA data-set under HYLARSMET
4. To determine the best way to compare FAO data-sets with our RSA sets of ground-based data
5. To determine if there are better alternative rainfall inputs to TRMM, for near real-time precipitation data input. If not, exploit bias adjustment of the TRMM product
6. To exploit the Soil Moisture estimates of the European Space Agency's SMOS mission when they are ready for use in model inter-comparisons
7. To determine how to best cope with the uncertainties associated with input parameters and forcing variables [TRMM in particular] when computing ensembles of historical and forecast data streams
8. To devise means of increasing computing capacity and the speed of calculations by improving key parts of the code and employing parallel [or High Performance] computing power
9. To determine the best ways of cold-starting calculations [model initialization] for both gauged and ungauged catchments

Our focus in EXSMET was to develop and utilise calculations at isolated pixels in 0.125° (about 12 km) squares, driven by hydrological modelling through PyTOPKAPI, leading to Soil Moisture assessments. It took advantage of gains made in an improved understanding of Soil Moisture and surface water hydrological processes through the completed project HYLARSMET, which was blessed by the availability of better hydrological data, especially the various forms of more accurate remotely-sensed data with better coverage, like TRMM, to be replaced by GPM. Throughout this project, strategic partnerships with international expertise in climate, soil water and surface water hydrological research flourished. Hydrological tools that were developed in the past under HYLARSMET were upgraded, re-developed and accelerated by alternative tools that are suited to current data availability, improved knowledge and the recent technological advances in hydrological modelling using the PyTOPKAPI model. In this project, the continued deterioration of hydrological gauging processes and other installed earth measurement devices were managed through the intensive use of new data sources from remote sensing, coupled with the limited earth-based measurements. Finally, the major outcomes we achieved here are that (i) we were able to model Soil Moisture and Evapotranspiration over SADC in detail and (ii) model hydrological runoff on 'ungauged' catchments over RSA, both with estimated levels of confidence, using accelerated parallel computing designs which were especially devised for this project.

_____ooooo///O\\\ooooo_____

References

Adler, R., Huffman, G., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. and Nelkin, E. (2003). The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). J. Hydrometeor., 4:1147-1167.

Aonashi, K. and Liu, G. (2000). Passive microwave precipitation retrievals using TMI during the Baiu period of 1998. Part I: Algorithm description and validation, J. Appl. Meteor., 39(12), 2024-2037.

Bartholomé, E. and Belward, A. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. International Journal of Remote Sensing, 26:1959-1977.

Bódis, K. (2009). Development of a data set for continental hydrologic modelling – Input layers related to topography, channel geometry, land cover and soil characteristics of European and African river basins. Technical report, European Commission, Joint Research Centre, Institute for Environment and Sustainability.

Burek, P., van der Knijff, J. and de Roo, A. (2013). LISFLOOD – Distributed water balance and flood simulation model. Revised user manual. Technical report, European Commission Joint Research Centre, Ispra, Italy.

Chen, M., Xie, P., Janowiak, J. and Arkin, P. (2002). Global Land Precipitation: A 50-yr Monthly Analysis Based on Gauge Observations. J. of Hydrometeorology, 3:249-266

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., Van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Holm, E., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J., Park, B., Peubey, C., deRosnay, Tavolato, C., Thepaut, J. and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137:553-597

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L. and Leroy, M. (2006). GLOBCOVER: A 300 m global land cover product for 2005 using ENVISAT MERIS time series. In ISPRS Commission VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes, Enschede (NL).

DeLannoy, G.J.M., Koster, R.D., Reichle, R.H., Mahanama, S.P. and Liu, Q. (2014). An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. Journal of Advances in Modeling Earth Systems, 6:957-979.

Dewitte, O., Jones, A., Spaargaren, O., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Gallali, T., Hallett, S., Jones, R., Kilasara, M., Roux, P.L., Michli, E., Montanarella, L., Thiombiano, L., Ranst, E.V., Yemefack, M. and Zougmore, R. (2013). Harmonisation of the soil map of Africa at the continental scale. Geoderma, 211-212:138-153.

Durre, I., Menne, M.J., Gleason, B.E., Houston, T.G. and Vose, R.S. (2010). Comprehensive automated quality assurance of daily surface observations. Journal of Applied Meteorology and Climatology, 49:1615-1633

Everson, C., Mengistu, M.G. and Vather, T. (2016). The Validation of the Variables (Evaporation and Soil Moisture) In Hydrometeorological Models: Phase Ii, Application of Cosmic Ray Probes for Soil Water Measurement (WRC Project K5/2323/1).

FAO/IIASA/ISRIC/ISS-CAS/JRC. (2012). Harmonized world soil database (version 1.2). Technical report, FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., Woodcock, C., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C. (2002). Global land cover mapping from MODIS: algorithms and early results. Remote Sensing of Environment, 83(12):287-302. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring. 12

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Rowland, J., Romero, B., Husak, G., Michaelsen, J. and Verdin, A. (2014). A quasi-global precipitation time series for drought monitoring. Technical report, U.S. Geological Survey Data Series 832

GLCC : see USGS (Global Land Cover Classes)
https://edcftp.cr.usgs.gov/project/glcc/globdoc2_0.html

Gleyzer, A., Denisyuk, M., Rimmer, A. and Salingar, Y. (2004). A Fast Recursive GIS Algorithm for Computing Strahler Stream Order in Braided and Nonbraided Networks. Journal of the American Water Resources Association (JAWRA), 40(4):937-946.

Green, W. and Ampt, G. (1911). Studies on soil physics, part 1, the flow of air and water through soils. *J. Agric. Sci.*, 4(1):1-24.

Hansen, M., Defries, R., Townshend, J. and Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing, 21:1331-1364.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G. and Gonzalez, M.R. (2014). SoilGrids1km – Global Soil Information Based on Automated Mapping. PLOS ONE, 9(8):e105992.

Huffman, G., Adler, R., Morrissey, M., Bolvin, D., Curtis, S., Joyce, R., McGavock, B. and Susskind, J. (2001). Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. J. Hydrometeor., 2:36-50.

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Hong, Y., Stocker, E. and Wolff, D. (2007). The TRMM Multi-satellite Precipitation Analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. J. Hydrometeor., 8(1):38-55.

Huffman, G., Adler, R., Bolvin, D. and Gu, G. (2009). Improving the Global Precipitation Record: GPCP Version 2.1. Geophys. Res. Lett., 36:L17808.

Huffman, G.J. (2015). The Transition in Multi-Satellite Products from TRMM to GPM (TMPA to IMERG). Technical report, NASA.

HWSD: see FAO/IIASA/ISRIC/ISS-CAS/JRC (2012). Harmonized world soil database (version 1.2). Technical report, FAO, Rome, Italy and IIASA, Luxemburg, Austria.

HYDROSHEDS: see USGS https://hydrosheds.cr.usgs.gov/data.php

Jarvis, A., Reuter, H., Nelson, A. and Guevara, E. (2008). Hole-filled seamless SRTM data V4.

Jones, A., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Dewitte, O., Gallali, T., Hallett, S., Jones, R., Kilasara, M., Le Roux, P., Micheli, E., Montanarella, L., Spaargaren, O., Thiombiano, L., Van Ranst, E., Yemefack, M. and R, Z. (2013). Soil Atlas of Africa. Technical report, European Commission, Publications Office of the European Union, Luxembourg.

Joyce, R.J., Janowiak, J.E., Arkin, P.A. and Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. Journal of Hydrometeorology, 5:487-503.

Kerr, Y., Al-Yaari, A., Rodriguez-Fernandez, N., Parrens, M., Molero, B., Leroux, D., Bircher, S., Mahmoodi, A., Mialon, A., Richaume, P., Delwart, S., Bitar, A. A., Pellarin, T., Bindlish, R., Jackson, T., Rdiger, C., Waldteufel, P., Mecklenburg, S. and Wigneron, J.-P. (2016). Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation. Remote Sensing of Environment, 180:40-63.

Kerr, Y., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J.-M., Font, J. and Berger, M. (2001). Soil moisture retrieval from space: The soil moisture and ocean salinity mission (SMOS). IEEE Trans. Geosci. Remote Sens., 39:1729-1735.

Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., Hirose, M., Takayabu, Y.N., Ushio, T., Nakagawa, K., Iwanami, K., Kachi, M. and Okamoto, K. (2007). Global Precipitation Map Using Satellite-Borne Microwave Radiometers by the GSMaP Project: Production and Validation. IEEE Trans. Geosci. Remote Sens., 45(7):2259-2275.

Liu, Z. and Todini, E. (2002). Towards a comprehensive physically-based rainfall-runoff model, Hydrol. Earth Syst. Sci., 6, 859-881

Loveland, T., Reed, B., Brown, J., Ohlen, D., Zhu, Z., Yang, L. and Merchant, J. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. International Journal of Remote Sensing, 21:1303-1330.

Maidment, D. (editor). (1993). Handbook of Hydrology. McGraw-Hill, New York.

Masson, V., Champeaux, J.-L., Chauvin, F., Meriguet, C. and Lacaze, R. (2003). A Global Database of Land Surface Parameters at 1-km Resolution in Meteorological and Climate Models. Journal of Climate, 16(9):1261-1282.

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E. and Houston, T.G. (2012). An Overview of the Global Historical Climatology Network-Daily Database. Journal of Atmospheric and Oceanic Technology, 29:897-910

Middleton, B. and Bailey, A. (2009). Water resources of South Africa, 2005 study (WR2005). Technical report, WRC report number TT 380/08, Pretoria, South Africa.

Pegram, G.G.S., Sinclair, Scott and Bárdossy, András. (2016). New methods of infilling Southern African raingauge records enhanced by Annual, Monthly and Daily Precipitation estimates tagged with uncertainty. Water Research Commission, WRC Research Report No. 2241/1/15, ISBN 978-1-4312-0758-9, March

Pegram, G., Sinclair, S., Vischel, T. and Nxumalo, N. (2010). Soil moisture from satellites: Daily maps over RSA for flash flood forecasting, drought monitoring, catchment management & agriculture. Technical report, Water Research Commission Report No. K5-1683, Pretoria, South Africa.

Pegram, G.C. and Pegram, G.G.S. (1993). Estimating Areal Rainfall by Integration of Multi-Quadric Surfaces over Polygons. Journal of Hydraulic Engineering, American Soc. Civ. Eng. Vol.119, pp.151-163

Rawls, W., Brakensiek, D. and Saxton, K. (1982). Estimation of soil water properties. Trans. of the ASAE, 25(5):1316-1320 and 1328.

Schulze, R., Maharaj, M., Warburton, M., Gers, C., Horan, M., Kunz, R. and Clark, D. (2008). South African atlas of climatology and agrohydrology. Technical report, Water Research Commission Report No. 1489/1/08, Water Research Commission, Pretoria, South Africa.

Shangguan, W., Dai, Y., Duan, Q., Liu, B. and Yuan, H. (2014). A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems, 6:249-263.

Sinclair, S. and Pegram, G.G.S. (2010). – A comparison of ASCAT and modelled soil moisture over South Africa, using TOPKAPI in land surface mode, Hydrol. Earth Syst. Sci., 14, 613-626

Sinclair, S. and Pegram, G.G.S. (2013a). A sensitivity assessment of the TOPKAPI model with an added infiltration module. Journal of Hydrology, 479:100-112.

Sinclair, S. and Pegram, G.G.S. (2013b). HYLARSMET: A Hydrologically Consistent Land Surface Model for Soil Moisture and Evapotranspiration Modelling over Southern Africa using Remote Sensing and Meteorological Data, Water Research Commission, WRC Research Report No. 2024/1/12, ISBN 978-1-4312-0392-5, March.

Sinclair, Scott, Vischel, Theo & Pegram, Geoff. (2017, June 28). PyTOPKAPI v0.4.0. Zenodo. http://doi.org/10.5281/zenodo.820640

Tchuenté, A.T.K., Roujean, J.-L. and Jong, S.M.D. (2011). Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. International Journal of Applied Earth Observation and Geoinformation, 13:207-219. 13

Tian, Y., Peters-Lidard, C.D., Adler, R.F., Kubota, T. and Ushio, T. (2010). Evaluation of GSMaP precipitation estimates over the contiguous United States. Journal of Hydrometeorology, 11:566-574.

Vischel, T., Pegram, G.G.S., Sinclair, S. and Parak, M. (2008). Implementation of the TOPKAPI model in South Africa: Initial results from the Liebenbergsvlei catchment, Water SA, 34(3), 1-12.

Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J. & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resources Research, 50, 7505-7514

-----ooOoo----