SOIL MOISTURE FROM SATELLITES: DAILY MAPS OVER RSA FOR FLASH FLOOD FORECASTING, DROUGHT MONITORING, CATCHMENT MANAGEMENT & AGRICULTURE

Report to the
Water Research Commission

by

Geoff Pegram, Scott Sinclair, Theo Vischel and Ntokozo Nxumalo
Pegram & Associates

WRC Report No. 1683/1/10
ISBN 978-1-4312-0048-1

December 2010

DISCLAIMER

This report has been reviewed by the Water Research Commission (WRC) and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the WRC, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

EXECUTIVE SUMMARY

The benefits of regularly updated Soil Moisture information

Of all the hydrometeorological variables we monitor, one of the most difficult to collect and interpret is Soil Moisture. Soil Moisture is more spatially variable than rainfall (with which it is strongly coupled) because of the spatial variability of soil type and depth, local slope of the ground and of course vegetation. One can collect Soil Moisture data at a "point" using one of many designs of soil water probes and this information is valuable in the vicinity of the probe. However to interpolate this information between widely spaced probes is difficult. It was the purpose of this project to find a way to estimate the soil moisture over the country in real time at fine spatial and temporal detail and share the results freely.

In order to plan ahead for agriculture, and monitor soil water availability to assist in crop and drought management, it helps to have up-to-date widespread estimates of soil moisture status. At a short time-scale, to anticipate the hazard of flash flooding due to forecast heavy rainfall, it is useful to know how wet a vulnerable catchment is. At longer time-scales, information about the spatial behaviour of soil moisture over the nation will assist with the monitoring of droughts as well as the assessment of the degree of change of soil water availability due to anthropogenic and climate change effects. It is the purpose of the research in this project to provide soil moisture information

- freely (at no cost),
- frequently (daily or shorter),
- currently (the status in near real time within a few hours),
- in detail (with as small an area of estimation as possible),
- extensively (over the whole nation),
- reliably (with minimal interruptions to the flow of information)
- and easily (readily accessible to non-expert inquiries in a friendly format)

This report details what we have achieved in four years and how we have designed a system to share the information. The remainder of this executive summary outlines what we planned to achieve, what avenues we investigated, what we actually achieved.

What we planned

As early as 2004, we were aware that there were remote sensing products which had been developed by various research teams in collaboration with space agencies (such as NASA, ESA and EUMETSAT), which were aimed at estimating soil moisture automatically over large regions. We intended to exploit that information to assist us in Southern Africa. Our intention was to devise a multi-pronged approach to determine how good the remote sensing products are on the ground. We were advised by Dr Tom Jackson of US Department of Agriculture (who heads up the International Soil Moisture Working Group focussing on Ground Validation of remote sensing – GSSM-GV) that a network of soil moisture probes would give us point measurements at selected sites and that these would be the first call for validation. To get an idea of what the soil wetness status was in areas between the probes, we planned to interpolate the measurements using meteorological information available from the SA Weather Service (SAWS).

These measurements were to be supplemented by two other techniques. The first idea was to exploit the information readily available to us from EUMETSAT (through UKZN and SA Weather Service we have a contract with that organisation to receive all METEOSAT data streams in real time) and select the Infra-Red channels to estimate early morning temperatures in detail over the country at 15 minute intervals. From these observations we could compute the rate of warming of the ground at each site between 8 and 9 am local time to assess how wet it was. The second idea was to use a suitable physically based hydrological model to calculate the soil moisture status of every square kilometre of a chosen catchment after calibration with historical rainfall and runoff records. We would then combine the information from gauges, using sophisticated statistical models to provide the "best soil moisture estimate for the country, available on the web".

What we investigated

The hydrological modelling approach was a resounding success. With the assistance of a visiting French Post-doctoral fellow (Theo Vischel) we implemented our own version of a physically based hydrological rainfall-runoff computer model called TOPKAPI, to model the Liebenbergsvlei catchment in the Free State. TOPKAPI is the brain-child of Professor Ezio Todini of the University of Bologna who hosted the PI for a 6-month sabbatical in 1996; since then a strong research link has been maintained. The rainfall data available was from a superb set of pluviometer data (40 tipping bucket gauges at a spacing of 10 km on the 'Vlei) installed by SAWS to monitor the MRL5 radar sponsored by the WRC for radar rainfall research in the 1990s. The TOPKAPI model was calibrated to the rainfall and runoff data and the history of the modelled water content of each 1 km cell was calculated and recorded. This was compared to the remotely sensed estimates of soil moisture (SM) from a radar scatterometer onboard the ERS series of satellites, available to us through an ESA contract called TIGER-SHARE. In this contract we were partners with Professor Wolfgang Wagner and his team at the Technical University of Vienna, who was the author of the algorithm used to estimate soil moisture from the ERS scatterometer data. The correspondence between the estimates was so good it resulted in publications in Water SA and HESS.

Due to unforeseen delays, the soil moisture probe network that SAWS offered to install during 2006 never materialised during the term of this project, although it might for the next one. This was quite outside our control and the lack of information was unfortunate; as a result we had to rely on other methods. To compensate, we spent substantial energy on the Temperature Gradient work but were disappointed by the very noisy signals we obtained which required temporal and spatial filtering to extract the information. A suggestion from the GSSM-GV group was that the interference was most likely due to variability of wind-speed and humidity, which were not accounted for in the algorithm. We also found that the cloud cover occluded far more information than we expected, especially during the summer months when much of the country experiences its greatest rainfall, and consequently its highest variations in soil moisture.

By far the most successful approach was to take the cell elements of the TOPKAPI model (about 7000 of them) which had worked so well on the 'Vlei and centre them on a 1/8th degree grid over the entire country. We operated these mini hydrological models in a mode that meteorologists call Land Surface Models (LSMs), but with a difference. The standard LSM allows for only vertical fluxes – rain falls downwards, evapotranspiration goes up and drainage is accounted for using an empirical formulation, usually based on average topography in the region. In contrast, in our application of TOPKAPI cells in LSM mode, the drainage occurs downhill using the local slope and soil properties, as it does in the hydrological model. For each cell we applied TRMM spatial estimates of rainfall provided in near real time (within 12 hours) by NASA and computed the forcing function of potential Evapotranspiration (ET_0) at matching time intervals. The ingredients for computing ET_0 were obtained from two sources. The elusive one was the radiation reaching the Earth's surface at each pixel – we obtained that via EUMETCast from the LSA SAF laboratory in Lisbon at half-hour intervals in near real time (within 1 hour of observation). The remainder (temperature, humidity, wind-speed and air pressure) were obtained from SAWS Forecasting division as the products of their 24 hour Numerical Weather Predictions (NWP). We checked and found that these were very close in most cases to the measurements of the same variables at the 169 SAWS automatic weather stations.

Having obtained the SM estimates from our 7000 LSMs at 3-hour intervals for about a year, we were ready to compare them with a remotely sensed SM estimate. Fortuitously EUMETSAT launched a new satellite during 2008 which had on board an Advanced Scatterometer instrument – ASCAT, which was undergoing trials. The instrument flies about 700 km above the Earth's surface beaming down a C-band radar (which is not much affected by attenuation due to cloud in the vertical) and collecting the reflected energy, whose magnitude depends on the wetness of the soil surface. It is calibrated against ground validation data in Europe and USA and takes account of each area's response over some time. We were alerted to this news by our collaborator in Vienna, Prof Wagner (whose team was responsible for developing the algorithm). We immediately compared the TOPKAPI/LSM estimates with the ASCAT estimates and found remarkably good correspondence over the wetter part of the country (the Summer rainfall region). The results were so encouraging (as well as being confirmation of the 'Vlei results) that we have published a new paper in HESSD in Dec 2009 with our findings – it is currently under review. At this stage we were ready to compare the TOPKAPI/LSM with the ASCAT and the Temperature Gradient estimates. It was then that we found that the latter (sadly) did not add value to the estimation procedure.

In summary, the ASCAT estimates are useful for checking at a longer time scale, as their temporal coverage of a particular piece of ground is at a frequency of about three days. By contrast, the TOPKAPI/LSM estimates are available at an interval of 3 hours, every day, over the whole country on a 12 km grid in all weathers.

How to share this bounty?

It needs to be emphasised that this endeavour requires a sophisticated support-base of national and international data providers, to whom we are grateful for their collaboration and support. What may be easy to ignore, when looking at the static images presented in this report, is the sheer magnitude of the data streams that go into making the product. The amount of design and programming involved with their sorting, quality control and archiving before turning them into information is considerable and tends to be hidden in a "back room". The successful calculation on-line, using PCs and the internet successfully is a mark of the care, skill and knowledge developed and possessed by the team as a result of this project.

Once the calculations have been done, every 3 hours, every day, every year (we had to install Uninterrupted Power Supply facilities in 2008!) we have an estimate of the Soil Saturation Index in detail over the country. As a spin-off, we compute other useful things like potential and actual evapotranspiration. These have all been available on our web-site (as browse-able images) since October 2007, while the Soil Saturation Index has been available since August 2008. Our final offering, in addition to the images, is the set of data and estimates that make the figures possible. This “best estimate of the Soil Moisture over RSA” and the background data are now available, on demand, by remote access, by any enquirer who has access to the internet. There is a catch – we currently rely on the University of KwaZulu-Natal to host this product on our server; at this time the arrangement works, but we will need to secure a dedicated host in the future. The system aims to be both self-sufficient and reliable and should be a boon to those involved with the anticipation of extreme weather events and monitoring the effects of climate change, by building a stable data-base.

Acknowledgements

The project researchers would like to express their thanks to the WRC for providing funding and in particular, the project steering committee for their valuable support, comments and suggestions throughout the life of this project.

The members of the reference group committee were:

Dr R Dube	:	Water Research Commission – Chairman
Mr Chris Moseki	:	Water Research Commission – Chairman
Prof ATP Bennie	:	University of the Free State
Mr DB Du Plessis	:	Department of Water Affairs and Forestry
Prof C Everson	:	Council for Scientific and Industrial Research
Dr GC Green	:	Independent Advisor
Prof B Hewitson	:	University of Cape Town
Dr WA Landman	:	South African Weather Service
Dr J Malherbe	:	Agricultural Research Council (ISCW)
Dr M Rouault	:	University of Cape Town
Prof M Savage	:	University of KwaZulu-Natal
Mr Eric Becker	:	SAWS (alternate to Dr Landman)
Mr A Clulow	:	CSIR (alternate to Prof Everson)

We developed some valuable national links which enriched this work.

- South African Weather Service for providing meteorological data, in particular Louis van Hemert for the Unified Model outputs, as well as Tracey Gill and Daan Malan for providing the Automatic Weather station data.
- The solar radiation data from CSIR site was generously provided by Prof Colin Everson and Alistair Clulow.
- Dr Mathieu Rouault of the Department of Oceanography of UCT was not only a reference group member but also PI of the WRC project MAHYVA which facilitated our collaboration. He was also a co-researcher in SAFeWater.

Valuable international links were engendered by the WRC with the following organisations, either directly or indirectly:

- European Space Agency via the TIGER-SHARE project – the ESA project manager Diego Fernandes announced at the report back meeting in Frascati “this is an excellent project”;
- Prof Wolfgang Wagner and his group at TUWien, as a direct offshoot from TIGER;
- Through him, Dr Tom Jackson of USDA who organises the International Soil Moisture Working Group Workshops:
- Prof Jeff Walker of University of Melbourne who gave considerable early encouragement
- Prof Sandrine Anquetain of Josef Fourier University of Grenoble, through the SAFeWater project, funded in RSA by DST and administered by the WRC. This enabled the valuable contribution of Dr Theo Vischel and later Ms Julia Pfeffer.

List of Acronyms

AWSs	Automatic Weather Stations
CEO	Chief Executive Officer
CNRS	Centre National de la Recherche Scientifique (French Research center)
CHARM	Collaborative Historical African Rainfall Model
CRU TS	Climate Research Unit Timeseries
CSIR	Council for Scientific and Industrial Research
DEM	Digital Elevation Model
DLSI	Digital Elevation Model for South Africa
DST	Department of Science and Technology
DWAF	Department of Water Affairs and Forestry
ECMWF	European Centre for Medium-range Weather Forecasting
EGU	European Geosciences Union
ERA 40	ECMWF re-analysis
ERS	European Remote Sensing satellite
ESA	European Space Agency
EUMETCast	Satellite broadcasting system used by EUMETSAT
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GLCC	Global Land Cover Characteristics
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project dataset
HESS	Hydrology and Earth System Sciences journal
HYDRO1K	Global Hydrologically consistent DEM 1 km resolution
HYDROS	HYDROsphere State mission
IPWG	International Precipitation Working Group
IR	Infra Red
ISMWG	International Soil Moisture Working Group
LSA SAF	EUMETSAT Land Surface Analysis Satellite Applications Facility
METEOSAT	Geostationary meteorological satellite operated by EUMETSAT
METOP	Polar orbiting meteorological satellite operated by EUMETSAT
MOU	Memorandum of Understanding
NASA	National Aeronautics Space Administration
NDMC	South African National Disaster Management Centre
NDVI	Normalized Difference Vegetation Index
PI	Project Investigator
PCs	Personal Computers
RSA	Republic of South Africa
SAHG-UKZN	Satellite Application and Hydrology Group of UKZN
SIRI	Soil and Irrigation Research Institute
SMOS	Soil Moisture and Ocean Salinity instrument onboard METOP
TIGER	Initiative for supporting African countries with Earth Observation Information
TOPKAPI	Topographic Kinematic Approximation and Integration hydrological model
TRMM	Tropical Rainfall Measuring Mission
UCT	University of Cape Town
UKZN	University of KwaZulu-Natal
UM	Unified Model
USA	United States of America
USDA	United States Department of Agriculture
WRC	Water Research Commission

Table of Contents

1. INTRODUCTION

The background to the study

The amount of water in the soil acts as a vital switch between atmosphere and ground, governing many Earth-bound water processes: infiltration, evapotranspiration, interflow and ground-water recharge. Soil Moisture (SM) even affects Numerical Weather Prediction modelling because it controls the vapour returning to the atmosphere. SM is painfully difficult to measure and is therefore seldom used explicitly in hydrological modelling, even though modern distributed models implicitly demand it. Nonetheless, SM has traditionally been measured, at point locations, by probes, primarily for agricultural purposes. If accurate spatial estimates of SM over large areas were available, they would be useful in many applications in Hydrology and Meteorology, besides Agriculture. Because of its importance in these areas, huge international effort is continuing to be made to use remote sensing to estimate SM at increasingly finer resolution. The Space Agencies are serious about SM measurement: EUMETSAT launched METEOSAT-8 (MSG) in 2003 followed by METEOSAT-9 in 2005 and launched the operational MetOp series in 2006, the European Space Agency (ESA) was to launch SMOS in late 2006, but this was delayed until 2009 – it is operational but still under test. NASA was to have launched HYDROS in 2010, but the program was cancelled due to lack of funds. A replacement satellite is on the cards. We in RSA can be beneficiaries of this energy and technological bounty, if we are prepared for it.

It was therefore important to put RSA in the position where it has the capability, capacity and competence to make use of these data, which, because trust and links have already been established, have been freely available from EUMETSAT, ESA & NASA. The purpose of this proposal was to develop, and put in place, the scientific capacity to exploit the hardware, software and skill that exist as a result of agreements between various key organizations. To set the stage, it is relevant to highlight the work done prior to the project being granted.

A summary of what had been achieved prior to the project follows.

The precursor to this study was the WRC Consultancy K8/598 “In Connection with Soil Moisture Mapping from Satellites”: via this vehicle contracted in 2005, the team had half a year to make a preliminary investigation into this exciting and potentially valuable area of research and it laid the ground for much of our thinking before the current contract was let.

RUMSA (Research in the Utilization of MSG in South Africa): in 2004, SAWS made a contract with EUMETSAT including ARC, CSIR & UKZN as partners. This resulted in funding enabling each organization to purchase a 2.5m satellite dish, computers and software (with some HR support) to receive MSG data directly. This facility, installed in May 2005, enabled UKZN to collect data routinely to enable estimates of soil moisture to be made by inference using early morning temperature gradients at fine scale over the region. Many additional products disseminated via the EUMETCast system, such as radiation and NDVI have directly benefited this project. Selected data have been archived since May 2005.

TIGER: UKZN (with SAWS, ARC, DWAF & NDMC as partners) entered into an agreement with ESA to use their satellite data, to be provided free for SM estimation; the agreement was signed in November 2005.

TIGER-Innovator: Through contacts established at international meetings by the Proposer (IPWG in Monterey, 2004 and EGU in Vienna, 2005) a joint proposal came out of collaboration in TIGER. This TIGER-SHARE project was between UKZN and Prof Wolfgang Wagner of Vienna University of Technology (TUWien) and was awarded by ESA in November 2005; this relationship was further strengthened at ISSM-WG in Noordwijk, Netherlands in March 2006 and at the TIGER workshop in Cape Town in November 2006. The purpose of the project, from our point of view, was to harness TUWien's expertise to exploit the microwave information coming from the polar orbiter MetOp ASCAT in order to improve the SM estimates from the geostationary MSG.

The MOU between SAWS & UKZN: This memorandum of understanding, cementing 15 years of scientific collaboration between SAWS' Observation research Division and UKZN's Civil Engineering Programme, was signed in June 2005.

Bethlehem, May 10, 2005: The PI was convener and chair of a meeting which comprised senior SAWS management and the Deputy CEO of WRC. Inter alia it was agreed that SAWS has the responsibility of Flash Flood Forecasting (FFF) in RSA, noting that DWAF's mandate is to monitor catchments with response times greater than 12 hours. To enable SAWS to perform FFF, it needs SM probes.

As a result of the link formed between UKZN and ARC via TIGER, ARC intended to install SM probes for their own use.

The SAWS & ARC probes were to provide a valuable ground-truthing function for remote sensing of SM and the deployment of the network(s) had the enthusiastic support of ESA and USDA, whose emails were on the agenda of the May 10 meeting in Bethlehem.

SAFeWater: This French/South African collaborative research agreement was initiated in 2005 and signed in March 2007. It is funded by our Department of Science and Technology and by the French CNRS and is administered in South Africa by the WRC. Its funding enabled two visitors from France to make contributions to our research: Dr Theo Vischel who was instrumental in applying TOPKAPI in South Africa and Ms Julia Pfeiffer who assessed the available remote sensing products in 2008. The agreement forged strong links between our Satellite Application and Hydrology Group at UKZN and the Department of Oceanography – particularly Dr Mathieu Roualt – of the University of Cape Town.

Technical Background on the need for Soil Moisture Estimates

The content of water in the first active metres of soil plays a central role in the regulation of the hydraulic and energy transfers between the soil, the surface and the atmosphere. Soil moisture is thus widely recognized as a key variable in numerous environmental disciplines especially in meteorology, hydrology and agriculture. For hydrological and agricultural purposes, the estimation of soil moisture is crucial since it controls (i) the quantity of water available for the growth of vegetation (Rodriguez-Iturbe, 2000), as well as the recharge of deep aquifers (Hodnett and Bell, 1986); (ii) the saturation of soils which controls the partitioning of rainfall between runoff and infiltration (Merz and Plate, 1997). In meteorology, the soil moisture content has a great impact on the transfer of energy from the surface into the atmosphere since it controls the evapotranspiration fluxes (Entekhabi et al., 1996).

An accurate estimation of soil moisture is difficult to obtain since it is highly variable in both space and time (Western and Blöschl, 1999). The two main sources of soil moisture information come from ground-based and remote sensing estimations. In the field, data can be obtained from gravimetric sampling; this gives the most accurate measurement of the soil water content but is obviously not suitable for automation. Probes (Neutron or Time Domain Reflectometry) can be calibrated to also provide an accurate and possibly automated estimation of soil moisture. Ground observations have helped to document soil moisture patterns at plot to hillslope scales (less than 1 km^2) in different regions of the world (e.g. Grayson et al., 1997; McNamara et al., 2005; De Lannoy et al., 2006; Hébrard et al., 2006). However, when catchment scales are of interest, one is rapidly confronted with scaling issues (Western and Blöschl., 1999) since ground measurements provide soil moisture estimation limited (i) to small spatial support (from few centimetres for probes, to 1 metre for gravimetric sampling) and (ii) to relatively small areas (extension in the order of a few hectares) since the implementation of a probe network of large extent is subject to obvious logistical and economic constraints.

Remote sensing of soil moisture from satellites is a promising alternative to ground measurements. Microwave frequencies are most often used, both in active (scatterometer or SAR) and passive (radiometer) instruments, to estimate soil moisture (see Wagner et al., 2007 for a detailed review). The advantage of microwave remote sensing is that it provides extended soil moisture estimations, gridded on averaged surface (footprint) from tens of metres to 50 km resolution, scales more suitable for catchment hydrology. However microwave estimations are only representative of the top few centimetres of soil, provided that the vegetation is not too dense, and the data availability is often dependent on a low frequency repeat cycle at a point (from 1 day to several weeks depending on the satellite).

Due to the uncertainties associated with the estimation of soil moisture, Kostov and Jackson (1993) suggest that the ideal approach for estimating soil moisture is to combine soil moisture measurements with hydrological models by using assimilation techniques. In fact, remotely sensed soil moisture is often directly assimilated into hydrological models (Ottlé and Vidal-Madjar,1994; Pauwels et al., 2002 Parajka et al., 2006) or into land surface schemes (Bruckler and Witono, 1989; Houser et al., 1998; Reichle et al., 2001; Walker et al., 2001) in order to initialize, drive, update and/or re-calibrate models, with the main objective of improving the simulations of river discharges or atmospheric fluxes respectively.

However, very few studies in the literature detail the comparison between the estimations of soil moisture from remote sensing with the estimations from hydrological models (Biftu and Gang, 2001; Parajka et al., 2006). One must however be able to know *a priori* the compatibility between the model and remotely sensed soil moisture estimations to better evaluate the effective potential of (i) hydrological models to provide back-calculated estimations of soil moisture for evaluating remotely sensed soil moisture, followed by the use of physical disaggregation tools to improve the low resolution typical of remotely sensed soil moisture fields, (ii) remotely sensed soil moisture estimates to be assimilated into hydrological models. Wagner et al. (2003) point out the necessity of comparing remotely sensed soil moisture with independent data derived from ground observations, models and/or other remote sensing techniques. Blyth (2002) mentions the necessity of modelling the soil moisture in detail and making comparisons between models and data. Pellenq et al. (2003) argue that it is essential to accurately understand all the processes involved in the soil moisture variability and their scale interactions. For that purpose, Western et al. (2002) point out the potential of process-based hydrological models that explicitly represent the dynamic and the spatial scales of the processes that control the soil moisture.

Outline of the report

The topics covered in Chapters 2 to 7 of the report, in brief summary are:

Chapter 2: we compare two independent approaches of soil moisture estimation on a regional size catchment in South Africa (Liebenbergsvlei, 4625 km^2). The first estimates are derived from the physically-based distributed hydrological model TOPKAPI (Liu and Todini, 2002). The second set of estimates are derived from the scatterometer on board the European Remote Sensing satellite ERS (Wagner et al., 1999).

Chapter 3: reference crop Evapotranspiration (ET_0) estimates are computed routinely at a spatial resolution of roughly 12 km over Southern Africa. The required variables are obtained from the South African Weather Service's Unified Model runs and an estimate of ET_0 computed for each model grid cell. The result is that ET_0 estimates are produced on a 0.11° square spatial grid over South Africa at ½ hourly intervals.

Chapter 4: TOPKAPI model cells in Land Surface Model mode are used to estimate SM & Actual ET using methodologies similar to those used in the assimilation of evapotranspiration and energy fluxes into Meteorological models of the atmosphere. The difference here is that the drainage from the cells is dependent on the local slope of the topography. The cells are driven by the 3-hourly spaced estimates of rainfall from NASA's TRMM satellite and the ET_0 product developed in Chapter 3.

Chapter 5: we attempted to determine Soil Wetness from ground surface early morning temperature gradients. Although this was an interesting idea (which incidentally forced the team to handle and archive vast quantities of data from EUMETSAT) it turned out not to be a fruitful methodology because of its intermittency (due to cloud) and noisy scatter (possibly due to wind and humidity effects).

Chapter 6: the hydrologically based TOPKAPI/LSM estimates of SM described in Chapter 4 were compared against two remote sensing based estimates: the temperature gradient ones described in Chapter 5 and a recently available remote sensing estimate from the ASCAT instrument aboard the MetOp satellite. It was here established that the temperature gradient estimates were inferior to the ASCAT ones; regression lines are fitted to the ASCAT~ TOPKAPI/LSM estimates showing high correlations over the summer rainfall region of RSA.

Chapter 7: finally, the procedure for sharing the 'Best Soil Moisture Estimate on the Web' is described. This product is the culmination of all the work that went into this project. It is designed to make these valuable data readily available to practitioners in Agriculture, Catchment Management, Drought Monitoring and Flood Forecasting. The information will also find use in initialising hydrological models of catchments. In addition, we suggest that the archive will be valuable for monitoring the reactions in the country's soil water and evaporation due to changes in weather and climate.

CAVEAT IN ADDENDUM

At the final reference group meeting of the project in February 2010, the research team was asked to specifically address the following issues in the introduction to the report as they were absent in the body:

- a comment that mentions the exclusion of snow in the model
- an indication of how to go about estimating errors in and uncertainty of the final products which are the result of uncertainty in the forcing fields (rainfall, ET and products used to derive these) and other relevant factors. [There was not to be any extra work done in this project.]

It was also noted that the Unified Model has a fixed albedo, which is not varied by season. How would this affect the relationships between UM forecasts and observations as seasons vary?

In response to these requests, we first note that the snow module was not included in the TOPKAPI model at this juncture. This point is mentioned in Chapter 2 in the section 'The TOPKAPI model Background', but it was not made clear that it was not included in the country-wide application of the subsequent TOPKAPI LSM calculations.

The following addresses the error structures and uncertainty of the products of the research.

The differences between alternative estimates of Energy flux at the ground and Reference Crop Evapotranspiration (ET_0) were addressed individually in Chapter 4, but their propagation through the computational procedure to evaluate the errors of the SSI products was not done at this stage. The only work done on error structures of the TOPKAPI SSI estimates is that done in comparison with the ERS and ASCAT SWI estimates, where regression lines were fitted to the pairwise relationships. These are summarised in Chapters 2 and 6.

We note that the differences between the observations of the meteorological variables at the weather stations and those forecast by the Unified Model seem to be unbiased, but it is difficult to determine which are 'correct' at the 12 km pixel scale. Nevertheless, to calculate the possible error associated with the determination of SSI, a formal computation using ensembles of variables with errors added will be conducted in the follow-on project.

There may be a problem with the fact that the Unified Model has a fixed albedo (does not vary with season) and the suggestion was that this might affect the UM forecasts. The effect of the constant albedo was not observed in the comparisons done on ET_0 calculations using the SAWS AWS data and the UM forecasts reported in Chapter 3, however this will be revisited in the follow-on project.

We are grateful to the reference group for drawing these issues to our attention. We are also deeply indebted to the group for their support and valuable guidance throughout the project.

2. APPLYING THE TOPKAPI HYDROLOGICAL MODEL IN SOUTH AFRICA

Planning and management of water resources are of major concern in Southern Africa. Droughts and floods are extreme hydrological hazards that regularly face the population. Food security and public health are directly dependent on these hydrological phenomena (Kleinschmidt et al., 2001; Jury et al., 2002). In this context, environmental scientists and engineers are involved in two important challenges:

- Providing short-term forecasts of the extreme hydrological events
- Assessing how the frequency and nature of these events could possibly evolve in the context of ongoing global climate change.

For this purpose, the numerical modelling of the water cycle is essential in order to provide tools for operational hydrology in order to assist decision makers to develop suitable environmental strategies, and for improved understanding of the hydrological processes involved in the variability of the water cycle.

The TOPKAPI model was adapted and coded using algorithms obtained from the literature. The early development performed by Parak (2007) was reported in Pegram et al. (2007). One of the objectives within the current project is to use the distributed TOPKAPI catchment model to estimate the soil moisture from hydrological data and then to compare this estimate with remote sensing estimates using two different types of satellite data. Thus the research-based model had to be further developed to run with practical hydrological applications in an operational manner. The initial results of the application of the model to compare simulated and remotely sensed soil moisture estimates have been reported on in Vischel et al., (2008a). The development and application of the TOPKAPI model to a South African catchment, including the details of the model formulation was published in Vischel et al., (2008b).

In this chapter the TOPKAPI model, initially developed by Liu and Todini (2002), is presented and applied to the Liebenbergsvlei catchment (4 625 km^2). TOPKAPI meets most of the requirements expected by both operational and research hydrology:

- It is a fully distributed model; the spatial range of the grid cell discretization within which the model is valid is up to 1 km (Martina, 2004)
- It is physically-based, meaning that it explicitly represents the hydrological processes on the basis of the fluid mechanics and soil physics, while the input parameters can be directly obtained from existing spatial datasets
- There are relatively few (15) input parameters for a distributed model of which only three or four typically require calibration (Liu and Todini, 2002; Liu et al., 2005).

The TOPKAPI model has already been successfully implemented as a research and operational hydrological model in several catchments in the world (Italy, Spain, France, Ukraine, China) (e.g. see Liu and Todini, 2002; Bartholomes and Todini, 2005; Liu et al., 2005; Martina et al., 2006). The existing literature describing the TOPKAPI model was the point of departure of the work presented here in order to develop the code and devise an efficient hydrological modelling tool for South Africa. In this way, this work is also the first attempt to implement TOPKAPI in Africa. This section presents the main aspects of the model:

- Its principle and physical concepts

- An example of an application of TOPKAPI to the modelling of a regional-size catchment in South Africa.

The TOPKAPI model Background

TOPKAPI is an acronym which stands for TOPographic Kinematic APproximation and Integration and is a physically-based distributed rainfall-runoff model. In the original version proposed by Liu and Todini (2002), TOPKAPI consists of 5 main modules comprising soil, overland, channel, evapotranspiration and snow modules. The first 3 modules take the form of non-linear reservoirs controlling the horizontal flows. The evapotranspiration module has been slightly modified in this application, compared to the original module presented in Liu and Todini (2002). The snow module component is ignored in the present study as the influence of snow-falls on runoff is negligible in the Liebenbergsvlei catchment.

Model assumptions

- The TOPKAPI model is based on 6 fundamental assumptions (Liu and Todini, 2002): Precipitation is constant in space and time over the integration domain (namely the single grid cell or pixel and the basic time interval, usually one to a few hours). This assumption simply means that the model is lumped at the grid scale.
- All precipitation falling on the soil infiltrates, unless the soil is already saturated (Dunne, 1978)
- The slope of the groundwater table coincides with the slope of the ground
- Local transmissivity, like horizontal subsurface flow in a cell, depends on the integral of the total water content of the soil in the vertical plane
- In the soil layer, the saturated hydraulic conductivity is constant with depth and, due to macro-porosity, is much larger than in the deeper underlying layers, which do not form part of the model structure
- During the transition phase, the variation of water content in time is constant in space over each cell.

The TOPKAPI model has the added benefit that the soil moisture store in each pixel is a physically realistic representation of the physical soil moisture quantities. Therefore it is the model of choice in this study to complement the remote sensing estimation of the elusive variable, Soil Moisture. It is important to have worked on this in advance of the remote sensing comparison, because:

- The model structure and development help to define the parametric need for atmospheric data, putting pressure on the accurate determination of spatial rainfall and evapotranspiration potential.
- It turns out that besides being a useful hydrological model, its elements are admirably suitable for estimating soil moisture (SM) on a grid of disconnected cells acting as land surface models, so that SM can be relatively easily computed in near real time. It turns out that this forms the core model of the project, so this section will carefully set the stage for its implementation.

The TOPKAPI Model: Adaptation of Parak's work to a small virtual catchment

The development follows the preliminary MSc Eng thesis work of Parak (2006), partly reported in Pegram et al. (2007). From the existing literature, Parak (2006) carried out a detailed description and a simple application of the TOPKAPI hydrological model (Liu and Todini, 2002). The TOPKAPI model is a physically-based distributed hydrological model that applies the kinematic wave approximation on three non-linear reservoirs governing respectively the flows into the soil (subsurface), on the overland hillslope and in the channel network.

As a first case study, Parak (2006) created a simple idealized catchment composed of 4 fundamental aligned cells (1 hillslope cell without channel network and 3 cells with channel network), on which the model was applied. The implementation of the model was achieved through the use of a standard spreadsheet in Microsoft Excel. The purpose of this was to ensure that all calculation procedures were fully understood before extending the model to natural catchments.

Several technical issues arose from this first application, mainly concerning the capacity of the model to numerically close the mass balance. Before any application of the model to a realistic case, this problem of continuity has to be resolved. It is addressed in detail in Appendix A mainly by analyzing the numerical aspects of the model through the simple 4-cell case study.

There, we give a synthetic description of the model. The question of mass continuity as simulated by the model is also addressed by testing several methods to solve the ordinary differential equations that control the flux and storages in the model. To test the adaptability of the model for further applications on real catchments, a generic catchment of 32 cells is then created and used as a more complex intermediary case study. The purpose of including the detail in the Appendix is to have a record of the changes and improvements made to the model in its local application; these are too detailed to occupy the main body of this report as their presence here would break the flow.

Due to the obvious limitation of the Microsoft Excel spreadsheets in realistic catchment modelling, the TOPKAPI model was implemented using a programming language. The Python language was chosen, primarily since it is suited for the management of large data arrays and because it can be interfaced with many existing languages (FORTRAN, C, C++, etc.). The model code has been structured in the form of a library and supporting tools for data management and running the model. These are free under a BSD style Open Source license. A follow on project will fund improved dissemination of the code and documentation to model users and collaborators in its further development.

Ordinary Differential Equations controlling the reservoir flows

The continuity equations of each of the three reservoirs (soil, overland and channel) that compose a cell can be written as a classical balance differential equation:

$$\frac{dV_i}{dt} = Q_i^{in} - Q_i^{out} \tag{2.1}$$

where all the variables are observed at time t: V_i is the total volume stored in the reservoir, $\frac{dV_i}{dt}$ is the rate of change of water storage, Q_i^{in} is the total inflow rate to the reservoir, Q_i^{out} is the total outflow rate from the reservoir. The energy equation in the soil reservoir is the Darcy equation; the Manning equation applies to the overland and channel flows.

The kinematic wave approach used to resolve the continuity and mass balance in TOPKAPI (by neglecting the dynamic acceleration terms in the energy equation) leads to a nonlinear relationship between Q_i^{out} and V_i, turning equation (2.1) into to an ordinary differential equation (ODE) of the form:

$$\frac{dV_i}{dt} = Q_i^{in} - b_i V_i^{\alpha} \qquad (2.2)$$

where b_i is constant in time (it is frequently spatially variable) and is a function of the geometrical and physical characteristics of the reservoir. The coefficient b_i also depends on the exponent coefficient α which originates from infiltration equations for soil reservoir behaviour and from Manning’s equations for overland and channel reservoir (see Pegram et al., 2007 or Lui and Todini, 2002 for more details). For the three reservoirs (soil, overland and channel), the expressions for b_i and α are reported in Table 2.1. Depending on the type of reservoir, Q_i^{in} is a combination of the forcing variables (rainfall and evapotranspiration in the case of the soil reservoir and interconnecting flows between the elemental storage reservoirs within the cell and from upper connected cells; Figure 2.1).

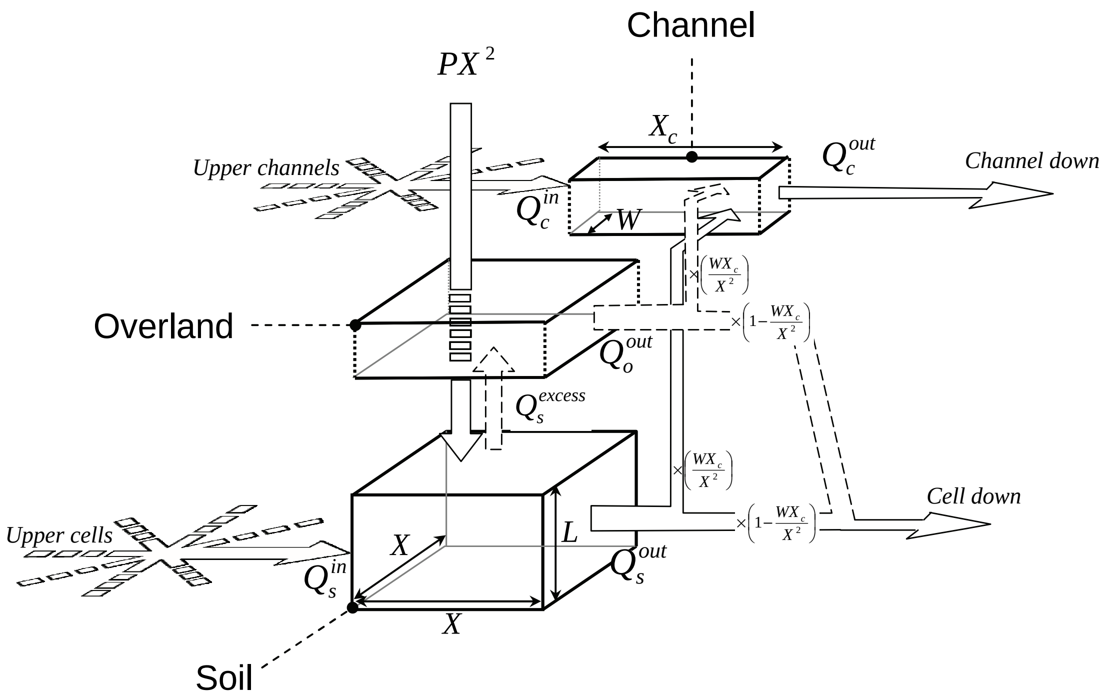


Figure 2.1 The water balance of a cell in the TOPKAPI model

At each simulation time step, the inflow rate Q_i^{in} is computed, assumed to be a constant over the interval, then the ODE equation is solved analytically (or by numerical integration, when necessary) over the time interval Δt, dependent on the initial volume stored in the reservoir at time t, to obtain the volume $V_i(t+\Delta t)$ stored at $t+\Delta t$. In Table 2.2 all the variables that are computed for each reservoir from the ODE finite difference solution showing the reservoir and cell connectivity are reported. Table 2.2 is associated with Figure 2.1 which illustrates the fluxes and connections for a typical modelled cell.

Table 2.1 Expressions and/or typical values of the coefficient b_i and α of equation 2.2 for each component store in a cell

	b_i	α
Soil	$b_i = \frac{C_{s_i} X}{X^{2\alpha}}$ with $C_{s_i} = \frac{L_i k_{s_i} \tan(\beta_i)}{(\theta_{s_i} - \theta_{r_i})^{\alpha} L_i^{\alpha}}$ where: - L_i is the soil depth - k_{s_i} is the saturated hydraulic conductivity - $\tan(\beta_i)$ is the tangent of the ground slope β_i - θ_s is the saturated soil moisture content - θ_r is the residual soil moisture content	$\alpha = \alpha_s$ with $2 \leq \alpha_s \leq 4$ Where α_s is a pore-size distribution parameter (Brooks and Corey, 1964)
Overland	$b_i = \frac{C_{o_i} X}{X^{2\alpha}}$ with $C_{o_i} = \frac{1}{n_{o_i}} \sqrt{\tan(\beta_i)}$ - n_{o_i} is Manning's roughness coefficient - $\tan(\beta_i)$ is the tangent of the ground slope β_i	$\alpha = \alpha_o = 5/3$
Channel	$b_i = \frac{C_{c_i} W_i}{(XW_i)^{\alpha}}$ with $C_{c_i} = \frac{1}{n_{c_i}} \sqrt{\tan(\beta_i)}$ - W_i is the width of the channel - n_{c_i} is Manning's roughness coefficient - $\tan(\beta_i)$ is the tangent of the channel slope β_i	$\alpha = \alpha_c = 5/3$

Table 2.2 Variables computed at each cell i between time t and Δt: see Figure 2.1 for flow paths

	Initial value: **Volume at *t***	**Inflow rates during [*t, t+Δt*]**	***ODE solution:*** **Volume at *t+Δt***	**Outflow rates during [*t, t+Δt*]**	***Flow partitioning:*** **Flow rate to next cell during [*t, t+Δt*]**
Soil	$V_{s_i}(t)$	$Q_{s_i}^{in} = P_i X^2 + Q_{s_i}^{up} + Q_{o_i}^{up}$	$V_{s_i}(t+\Delta t)$	$Q_{s_i}^{out} = Q_{s_i}^{in} - \frac{V_{s_i}(t+\Delta t) - V_{s_i}(t)}{\Delta t}$	*To next soil reservoir* $Q_{s_i}^{out} - Q_{s_i}^{excess} - \frac{W_i}{X} Q_{s_i}^{out}$
Overland	$V_{o_i}(t)$	$Q_{o_i}^{in} = Q_{s_i}^{excess} = \max\left(0, Q_{s_i}^{out} - Q_{s\max i}\right)$ with $Q_{s\max i} = XK_{s_i} L_i \tan(\beta)$	$V_{o_i}(t+\Delta t)$	$Q_{o_i}^{out} = Q_{o_i}^{in} - \frac{V_{oi}(t+\Delta t) - V_{o_i}(t)}{\Delta t}$	*To next soil reservoir* $Q_{o_i}^{out} - \frac{W_i}{X} Q_{o_i}^{out}$
Channel	$V_{c_i}(t)$	$Q_{c_i}^{in} = Q_{c_i}^{up} + \frac{W_i}{X} Q_{s_i}^{out} + \frac{W_i}{X} Q_{o_i}^{out}$	$V_{c_i}(t+\Delta t)$	$Q_{c_i}^{out} = Q_{c_i}^{in} - \frac{V_{ci}(t+\Delta t) - V_{c_i}(t)}{\Delta t}$	*To next channel* $Q_{c_i}^{out}$

Adaptation of TOKAPI to the Liebenbergsvlei Catchment

Two main issues are presented in this subsection: (1) the way the initial internal parameter values affect the TOPKAPI model, (2) the choice of the simulation periods depending on the simultaneous availability of rain gauge and flow data. According to the results of these two explorations, a proposal is made for a modelling experiment. The calibration and validation depend crucially on the quality of the data. As a result some detailed analysis of this aspect is included in the following.

Summary of initial values of model parameters

Different data and tables from literature were sourced to determine appropriate values of the internal parameters of the TOPKAPI model. Table 2.6 describes the model parameters and the sources used to determine their values. The spatially varying parameters represented in the model are mapped in Figure 2.2, which also illustrates the source data used to infer the final parameter values. The tables that are used to link the environmental properties (land use, soil texture, etc.) to the physical model parameters are reported in Table 2.3, Table 2.4 and Table 2.5.

texture class	Total porosity θ_s (mean)	Total porosity θ_s (std)	Residual soil moisture θ_r (mean)	Residual soil moisture θ_r (std)	Pore-size distribution λ (mean)	Pore-size distribution λ (std)	Brooks and Corey coefficient α_s=3+2/λ	Ks (mm/s)
sand	0.437	0.063	0.020	0.019	0.694	0.396	5.882	6.54E-02
loamy sand	0.437	0.069	0.035	0.032	0.553	0.319	6.617	1.66E-02
sandy loam	0.453	0.102	0.041	0.065	0.378	0.238	8.291	6.06E-03
loam	0.463	0.088	0.027	0.047	0.252	0.166	10.937	3.67E-03
silt loam	0.501	0.081	0.015	0.043	0.234	0.129	11.547	1.89E-03
sandy clay loam	0.398	0.066	0.068	0.069	0.319	0.240	9.270	8.33E-04
clay loam	0.464	0.055	0.075	0.099	0.242	0.172	11.264	5.56E-04
silty clay loam	0.471	0.053	0.040	0.078	0.177	0.138	14.299	5.56E-04
sandy clay	0.430	0.060	0.109	0.096	0.223	0.175	11.969	3.33E-04
silty clay	0.479	0.054	0.056	0.080	0.150	0.110	16.333	2.78E-04
clay	0.475	0.048	0.090	0.105	0.165	0.128	15.121	1.67E-04

Table 2.3 Water retention properties and hydraulic conductivity classified by soil texture (Maidment, 1993). The grey columns are the variables that were effectively extracted from the table for the modelling experiment. The others are presented since they could be useful for further experiments.

Table 2.4 The related soil properties of the landtype map (SIRI, 1987) of the Liebenbergsvlei catchment. The grey columns are the variables that were effectively extracted from the Table for the modelling experiment. The others are presented since they could be useful for further experiments.

Land-type code	Depth of top soil (Layer A) (m)	Depth of sub soil (Layer B) (m)	Wilting point (A) (m/m)	Wilting point (B) (m/m)	Field capacity (A) (m/m)	Field capacity (B) (m/m)	Porosity (A) (m/m)	Porosity (B) (m/m)	Saturated drainage rate (A to B) (fraction)	Saturated drainage rate (B to C) (fraction)	Impervious areas next to stream (fraction of landtype)	Other impervious areas (fraction of landtype)	USLE erodibility factor (A) (fraction)	USLE erodibility factor (B) (fraction)	Plant available water (cm)
1156	0.23	0.29	0.23	0.25	0.28	0.30	0.44	0.44	0.21	0.21	0.002	0.040	0.42	0.42	27.9
1198	0.26	0.41	0.22	0.24	0.28	0.30	0.44	0.43	0.21	0.21	0.002	0.050	0.41	0.41	37.7
1312	0.23	0.37	0.18	0.26	0.25	0.31	0.41	0.42	0.24	0.24	0.002	0.003	0.37	0.37	34.8
1354	0.22	0.32	0.16	0.22	0.23	0.29	0.43	0.43	0.20	0.20	0.000	0.008	0.58	0.58	36.3
1481	0.24	0.41	0.14	0.20	0.22	0.27	0.44	0.42	0.24	0.24	0.002	0.006	0.48	0.48	47.9
1518	0.25	0.43	0.15	0.21	0.23	0.28	0.42	0.41	0.31	0.31	0.002	0.040	0.39	0.39	48.7
1522	0.21	0.23	0.15	0.16	0.23	0.24	0.44	0.41	0.22	0.22	0.000	0.140	0.41	0.41	34.8
1687	0.21	0.21	0.18	0.26	0.25	0.31	0.41	0.41	0.24	0.24	0.000	0.005	0.46	0.46	25.0
1745	0.28	0.51	0.13	0.19	0.22	0.27	0.44	0.42	0.32	0.32	0.002	0.010	0.45	0.45	59.9
1789	0.23	0.34	0.18	0.25	0.25	0.30	0.41	0.41	0.29	0.29	0.002	0.010	0.39	0.39	34.7
1823	0.28	0.53	0.14	0.20	0.22	0.27	0.43	0.40	0.44	0.44	0.000	0.040	0.37	0.37	57.9
2093	0.23	0.34	0.13	0.15	0.21	0.23	0.45	0.43	0.39	0.39	0.003	0.240	0.45	0.45	48.0
2125	0.22	0.43	0.18	0.19	0.25	0.26	0.43	0.41	0.44	0.44	0.002	0.080	0.35	0.35	46.4
2243	0.20	0.13	0.23	0.23	0.28	0.29	0.41	0.42	0.25	0.25	0.000	0.610	0.27	0.27	18.3

Table 2.5 The related soil properties of the land use map (GLCC, 1997) of the Liebenbergsvlei catchment. The grey columns are the values that were effectively extracted from the table for the modelling experiment. The others are presented since they could be useful for further experiments.

	Manning coefficient ($m^{-1/3}.s^{-1}$)		
Description \ Source	Chow (1959)	Maidment (1993)	Music (2000)
Urban and Built-Up Land	?	?	0.025
Cropland/Grassland Mosaic	0.04	0.02-0.04	0.04
Grassland	0.05	0.03-0.05	0.1
Savanna	0.07	0.05-0.2	?
Evergreen Broadleaf Forest	0.1	0.05-0.2	0.2
Barren or Sparsely Vegetated	0.035	0.03-0.05	0.1

It is worth noting that the soil types as defined by SIRI (1987) are separated into two soil horizons A and B. In its original version, the TOPKAPI model only accounts for a single soil layer. The properties of the A and B horizons appear to be generally very similar in terms of porosity, which we use as a proxy for the saturated soil moisture content, on this basis we chose to sum the depth of the two horizons and average their porosity.

A point that is not highlighted in Table 2.6 or Figure 2.2 is the choice of a value to be given to the initial saturation of the model reservoirs (soil, overland and channel) at the beginning of the simulation. Typically, regardless of which hydrological model is used, the first simulated values are highly influenced by the initial values of reservoirs. Then, after a "spin-up" period, simulations started with different initial values tend to converge (for the same model parameters and input data). This means that the simulations become reliable only after an initialization period. A sensitivity analysis is thus required to ascertain the real impact of the initial saturation level and determine the length of the initialization period that is necessary for different applications. An analysis of the flow data discussed later in this chapter (e.g. see Figure 2.4), shows that a convenient characteristic of the Liebenbergsvlei catchment is the strongly seasonal cycle, with the river often drying after the rainy season ends. The simulations can thus be initiated at the beginning of a hydrological year when the river is still dry (overland and channel reservoirs level equal to zero). If the time between the previous rainfalls is sufficiently long, the initial value of the reservoirs at the end of the dry period can be reliably taken equal to zero. This intuition is of course to be verified by the simulations.

Table 2.6 TOPKAPI model parameters.

		Realistic experimentation	Comments
Channel cell definition	-	From DEM (DLSI,1996)	Obtained after GIS treatment
Flow directions	-	From DEM (DLSI,1996)	Obtained after GIS treatment
Surface Slope	$\tan\beta$	From DEM (DLSI,1996)	Spatially variable over the catchment (cf. Figure 2.2)
Depth of surface soil layer (m)	L	From landtype map SIRI (1987)	Spatially variable over the catchment (cf. Figure 2.2)
Saturated hydraulic conductivity ($m.s^{-1}$)	k_s	Soil texture map WRC90+ Tables Maidment (1993)*	Spatially variable over the catchment (cf. Figure 2.2)
Residual soil moisture content	θ_r (0.01–0.1)	Soil texture map WRC90+ Tables Maidment (1993)**	Spatially variable over the catchment (cf. Figure 2.2)
Saturated soil moisture content	θ_s (0.25–0.7)	Soil texture map WRC90+ Tables Maidment (1993)**	Spatially variable over the catchment (cf. Figure 2.2)
Manning's surface roughness coeff.	n_o (0.05–0.4)	From landuse map GLCC (1997)+Tables in Chow (1988)***	Spatially variable over the catchment (cf. Figure 2.2)
Manning's channel roughness coeff.	n_c (0.02–0.08)	Strahler order method (Liu and Todini 2002)	Spatially variable over the catchment (cf. Figure 2.2)
Horizontal dimension of cell (m)	X	1000	-
Time step (s)	θt	3600	-
Non-linear soil exponent	α_s (2–4)	Soil texture map WRC90+ Tables Maidment (1993)**	Constant over the catchment, as a first experimentation
Non-linear overland exponent	α_o	$^5/_3$	-
Non-linear channel exponent	α_c	$^5/_3$	-
Max. channel width at outlet (m)	W_{max}	35	Coarsely estimated from Google Earth
Min. channel width for $A_{threshold}$ (m)	W_{min}	3	-
Area required to initiate channel (m^2)	$A_{threshold}$	2500000	-

* Reference to the initial works of Rawls and Brakensiek (1983)
** Reference to the initial works of Rawls and Brakensiek (1985)
*** Reference to the initial works of Chow (1959)

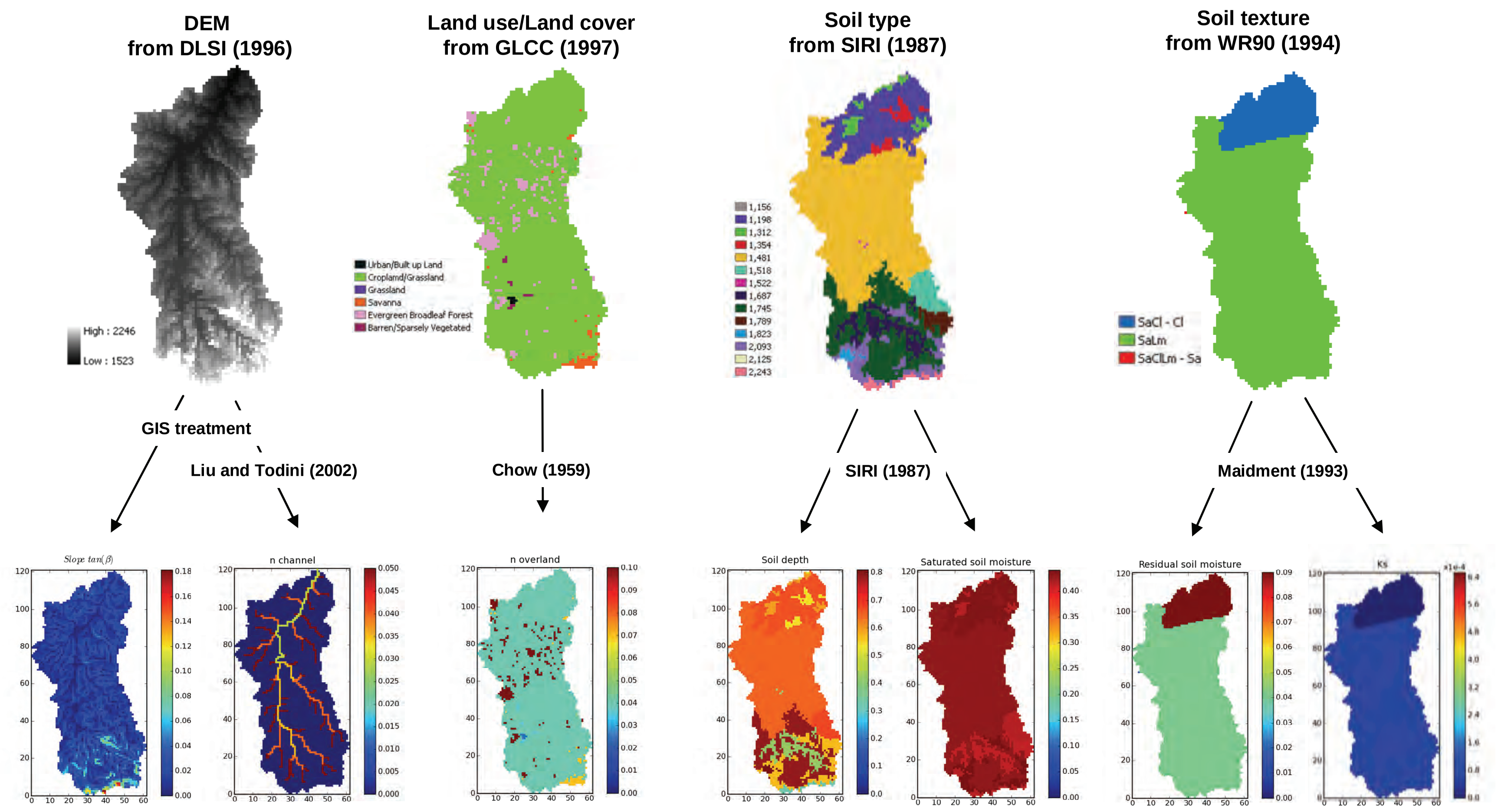


Figure 2.2 Maps of initial data and associated estimated TOPKAPI parameters over the Liebenbergsvlei catchment.

Choice of flow and rainfall series used for the TOPKAPI simulations

In the process of selecting a study period, it was important to consider the requirements associated with the calibration and the validation of the model. We will not discuss the different methods and theoretical frameworks used to calibrate and validate a hydrological model in detail here. However, the minimum requirement to evaluate a model is to have two independent simulation periods that can be used either for calibration or for validation.

This means that two periods of suitable length must be found where both rainfall and flow data are of sufficiently good quality to be considered as a reference for the model evaluation.

The rainfall and flow data are first presented separately without considering the modelling aspects. The question of defining periods for hydrological simulation is then addressed.

Rainfall and flow data availability on the Liebenbergsvlei catchment

As illustrated in Figure 2.3 the hydrological network available on the Liebenbergsvlei catchment is very dense.

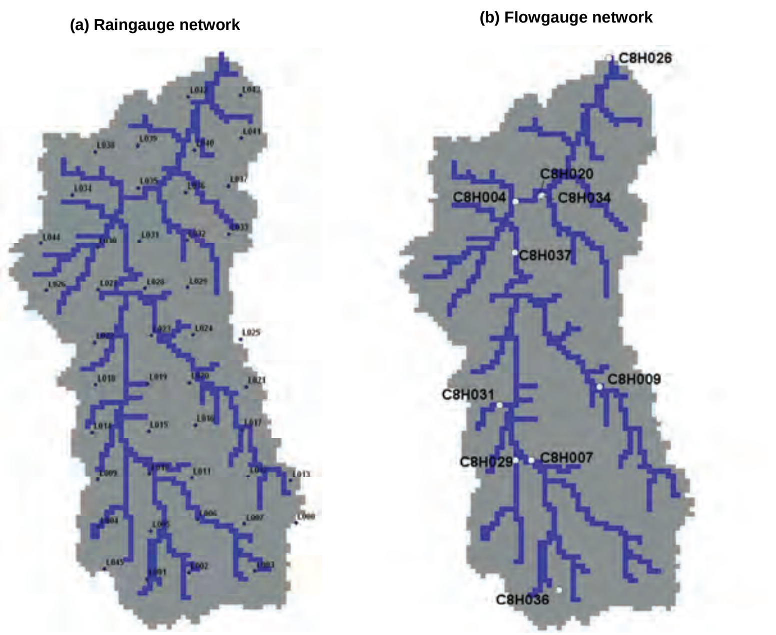


Figure 2.3 Complete rain gauge and flow gauge network on the Liebenbergsvlei catchment.

Although the density of rain and flow gages is very impressive and certainly unique in Africa, the choice of a simulation period is not straightforward since (i) a lot of flow data are missing, and (ii) the period of flow and rainfall measurements does not always coincide.

Analysis of flow data:

The flow data availability is presented in Figure 2.4, where the hydrograph of each flow gauge is plotted for the whole period of available measurements at daily time-steps. The data base has the advantage of having been carefully documented and associates a quality code with each measurement. This quality code is also reported in Figure 2.4, using a different colour for each quality measure to construct the curves.

As a first comment it can be seen that stations C8H031, C8H029 and C8H007 have very little, or very old, data. These stations can thus immediately be ignored for the study.

The second comment concerns stations C8H037, C8H034 and C8H009. Even if the period of measurement covered by these stations is sometimes attractive (especially for the station C8H009), the quality code shows that the capacity of the gauges to measure high flows is strongly limited. These stations will thus not be helpful for our purposes.

Thus four remaining flow stations that are potentially of good quality are C8H026 (catchment outlet), C8H020, C8H004 and C8H036.

The station C8H036 was installed in order to measure the external flows that come from Lesotho via a tunnel into the Ash River since 26th September 1997. This aspect was a source of worry since the precise knowledge of these external flows is a necessary condition to simulate the catchment flows after 1997, so it is quite comforting to find that they were properly measured. It is worth noting that the measurements begin on the 29th of October 1997, i.e. one month after the tunnel flows started.

The station C8H026 at the outlet of the catchment is of relatively good quality but presents a long period of missing data between the middle of 1996 to the middle of 1999, while the peak of a major event that occurred in 1996 was apparently missed. From the end of 1999 to the present, the data seem to be of very high quality.

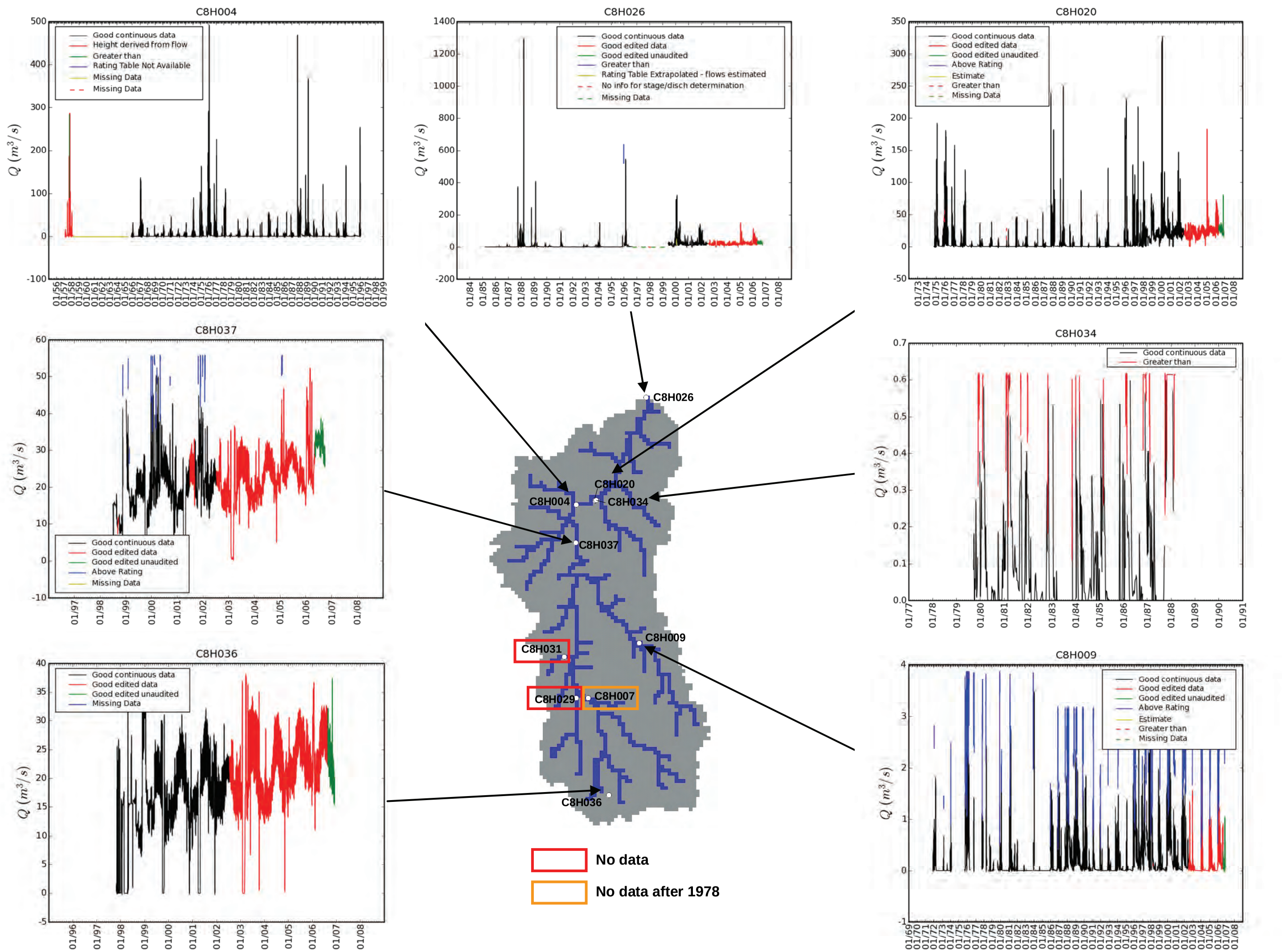


Figure 2.4 Flow data available on the Liebenbergsvlei catchment. On each station hydrograph, the colours correspond to different qualities of data that are explained in the legend. Note that the colours used do not have the same meaning for the different stations.

The station C8H020 is the most complete station since it displays long term continuous data from the end of the 1970s to present, with very few missing data.

The station C8H004 is at first sight an interesting station for the period before 1996. But some measurements are puzzling when compared to the stations C8H026 and C8H020. As shown in Figure 2.5 for a sub-period between 1988 and 1995, the flows measured by the station C8H004 are often higher than flows measured by the following station C8H020. This is not sufficient to affirm that the measurements of station C8H004 are not reliable since the station C8H020 could underestimate the flows. But the flows measured at the station C8H004 are also sometimes higher than the flows at the catchment outlet (station C8H026) that give sufficient justification to question the quality of the measurement of station C8H004. It is also noted that there is no photograph of the station C8H004, although photographs are available for all the other stations in the DWAF database (see for example the photos of the stations C8H026, C8H020 and C8H036 in Figure 2.6). There is thus no visual clue that could help us to interpret the puzzling station behaviour.

Consequently, it is certainly more cautious to dismiss station C8H004 and to keep only the three stations C8H020, C8H026 and C8H036 to be used as reference data for the hydrological simulations.

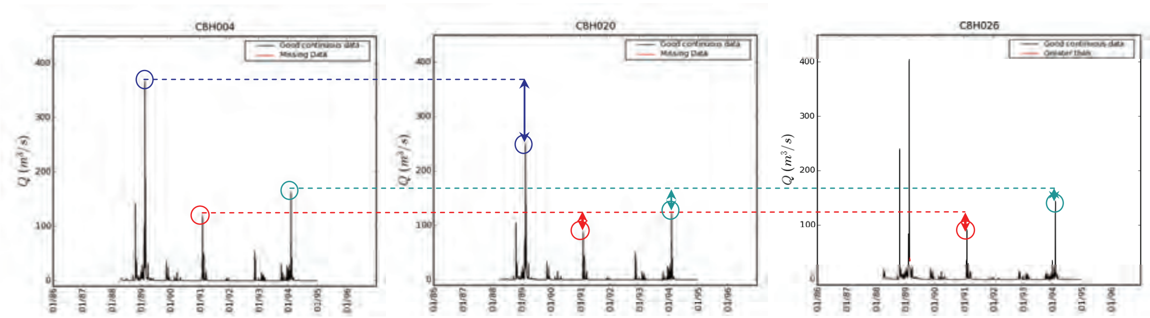


Figure 2.5 Comparison between flows measured by the station C8H004 with the flows measured by the two downstream stations C8H020 and C8H026. The flows measured by the station C8H004 are not only higher than those measured by the station C8H020 (as in the case of the peak circled in blue) but they are sometimes higher than the flows at the catchment outlet (case of peaks circled in red and green).

C8H026 (outlet) **C8H020** **C8H036 (tunnel from Lesotho)**

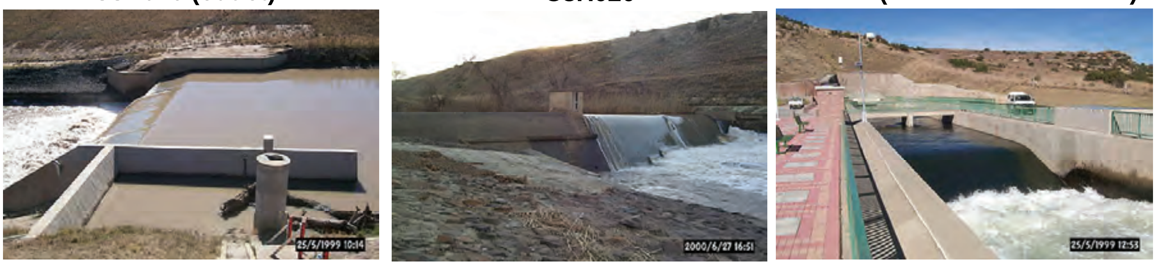


Figure 2.6 Photos of the flow stations in the Liebenbergsvlei catchments. These photos were obtained from the DWAF website (http://www.dwaf.gov.za).

<u>Analysis of rainfall data:</u>

The rain gauge network is composed of 45 tipping bucket rain gauges allowing the estimation of rainfall at small time steps.

The graph of Figure 2.7 shows the period of availability for each station. The first measurements are available from the end of 1993 to the beginning of 2002. But some sub-periods have to be commented on since some data are missing.

First it is noted that almost the entire network delivered data for the three years 1994, 1995 and 1996.

During the year 1997 the stations identified from L001 to L011 were stopped from the middle of February to the end of July. Even if for some years this period corresponds to the end of the rainy season, the rainfall was still active between February and July 1997 according, for example, to the hyetograph of the station L012 (Figure 2.8). This is quite important from a hydrological modelling point of view since the stations L001 to L011 cover the part of the basin which is the steepest and most elevated and certainly contributes significantly to the hydrological response of the catchment.

The data availability of the network is still very good during the year 1998; but the density of data decreases for few months in the second half of 1999 where about the half of the network is available.

The density of the network remains good for the year 2000, while it begins to decrease strongly at the middle of 2001.

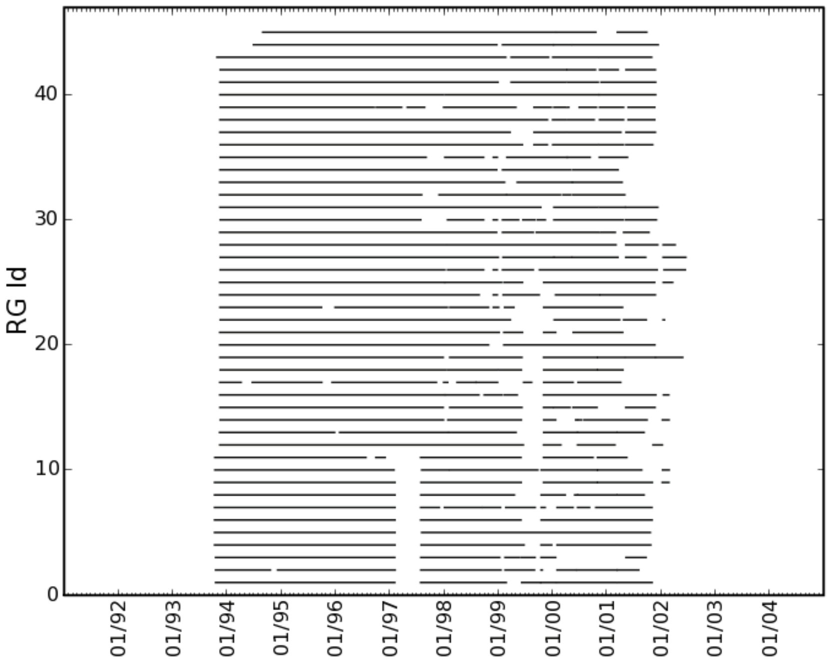


Figure 2.7 Availability of rainfall data over the total period covered by the network measurements.

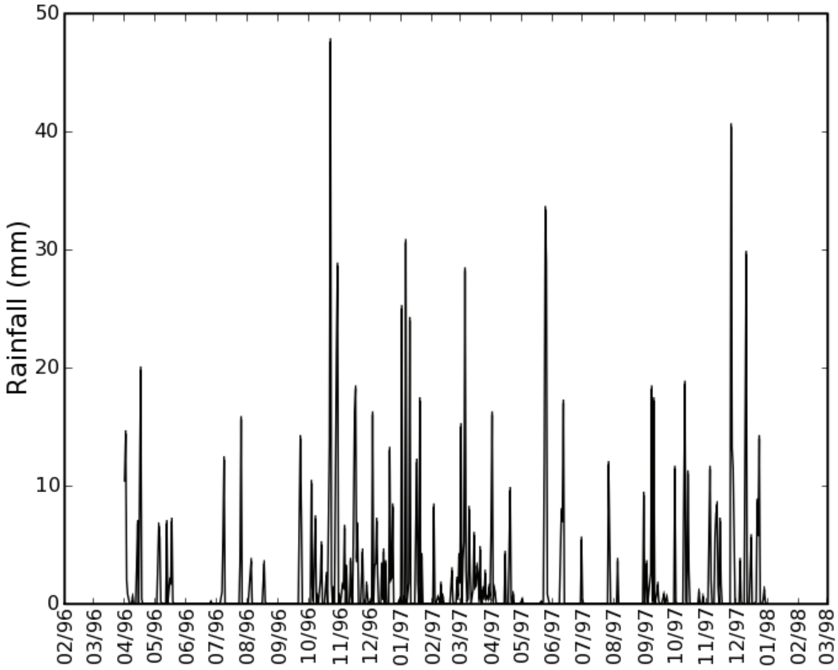


Figure 2.8 Example of hyetograph of station L012 from April 1996 to December 1997, showing that the period between February and July 1997 was still experiencing rainfall.

NB: The flow data were analyzed at a daily time step since they were directly available at this time step in the database. But the raw data have also been downloaded in order to reconstitute one hour time step flow sequences. This computation was subsequently done for the calibration and validation.

A program was designed to compute the cumulative rainfall at any time step from the tipping bucket records. Thus the rainfall series at one hour time steps were computed for the whole period of measurement.

Selection of a simulation period

Since the flow stations were chosen, it was now possible to more carefully sort out the definition of periods that contain both good quality rainfall and flow data. Note that (i) the period is now focused on 1994-2002 that corresponds to the period of availability of rain gauges, and (ii) the two flow stations C8H020 and C8H026 are chosen while the station C8H036 has not been chosen for the definition of a simulation period and is only used to determine the catchment inflows due to the Lesotho Highland water transfers.

To get a first glimpse of the common availability of data, Figure 2.9 shows (in the same picture) the rainfall availability per station (similar to Figure 2.7) and the hydrographs of stations C8H020 and C8H026. This figure was used as the basis for selecting a minimum of two independent periods for reference simulation periods.

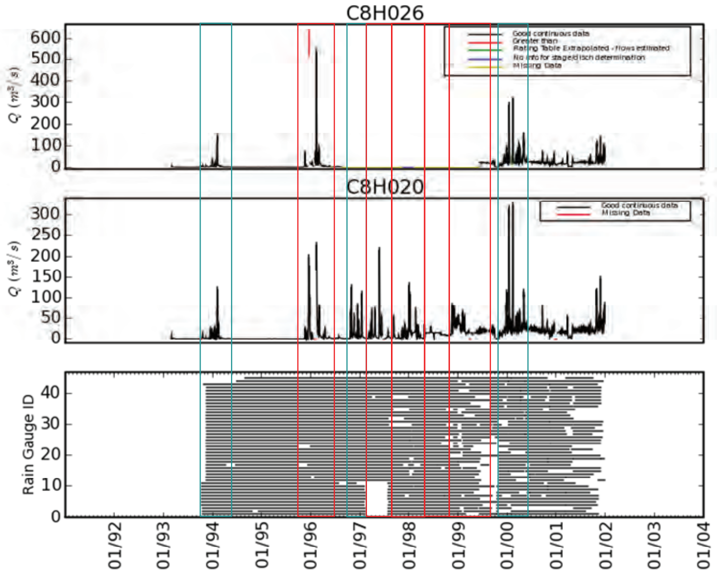


Figure 2.9 Rainfall data availability and hydrographs on the two selected streamflow stations inside the Liebenbergsvlei catchment. The vertical rectangles show the periods with major events that were selected. The green rectangles correspond to the periods selected for hydrological simulations and are shown in greater detail in Figure 2.10.

Sub-periods displaying major events were selected and are plotted in Figure 2.10. From the figures, three of these sub-periods were considered to match the characteristics required to be selected as reference periods for hydrologic simulations.

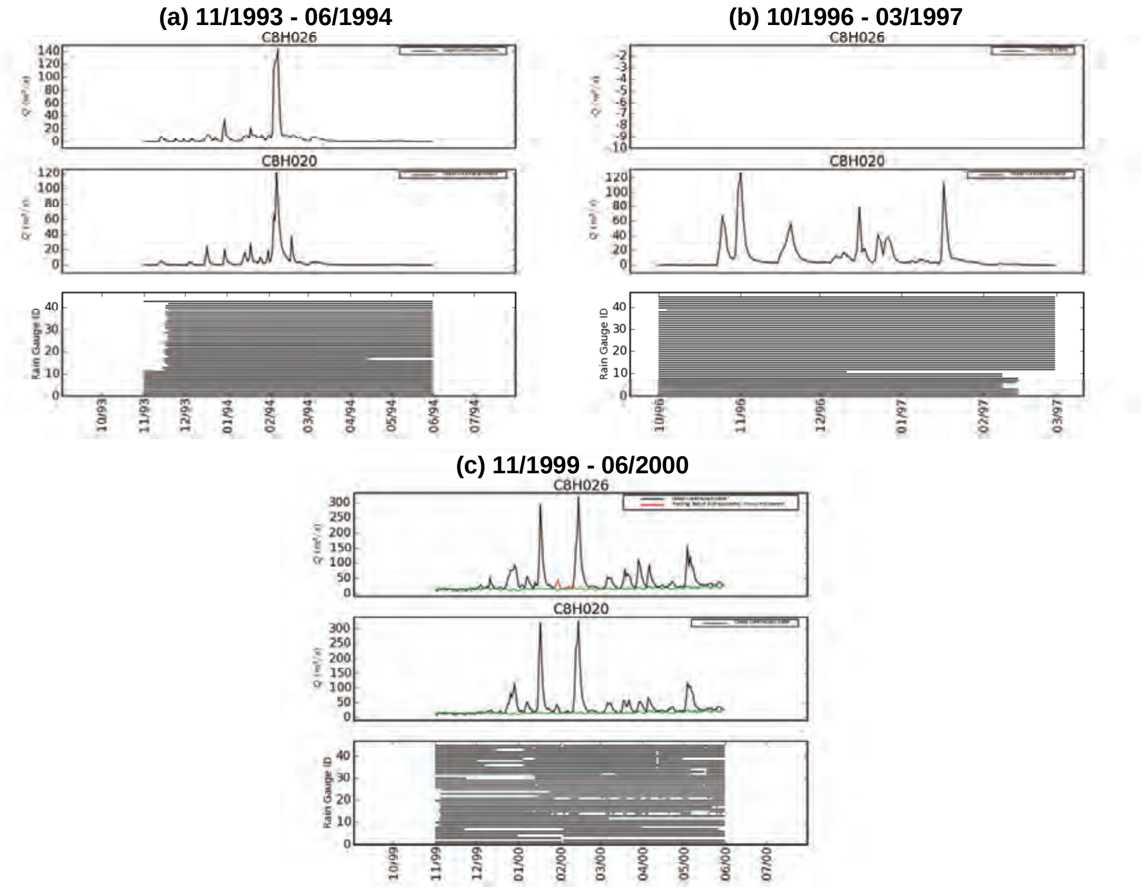


Figure 2.10 Sub-periods selected to be simulated with TOPKAPI. The green curve on Figure (c) is the flows at the output of the tunnel coming from the Lesotho as measured by the station C8H036. CH026 has missing data in (b); this accounts for the (spurious) negative flows.

The first selected period is between November 1993 and June 1994. It is the only period that has concurrently: (i) flow data at the outlet station (C8H026), (ii) flow data at the internal station C8H020, (iii) the complete rainfall network without any missing data during the most intense flows.

The second selected period is from October 1996 to March 1997. The flow data are missing at the catchment outlet but the flows are measured continuously at the internal station. The rain gauge network density is excellent with only missing data at the end of the event, but this should not influence the simulations according to the recession shape of the hydrograph.

The third selected period is after 26 September 1997, meaning that external flows arrive from the Lesotho (these flows are represented by the green curve on the hydrographs of Figure 2.10). It is noted that there is a good correspondence (with a routing delay of the order one day) between flow measured at the station C8H036 and the flows measured at the station C8H020 and C8H026 (when data are available). This confirms the coherence of the flow measurements along the catchment. For this period flow data are available for the both stations. Only some flow values have been estimated for a few days at the outlet but they seem to be reliable according to the measurements of the station C8H020. The rainfall network is not complete but remains very good according to the initial density of the network.

The adopted modelling strategy

A classical test scheme

Klemes (1986) proposes a hierarchical scheme for the systematic testing of hydrological simulation models. The scheme is in four steps:

- the *split-sample test (SS):* it consists of taking two periods, calibrating the model in the first period, validating (or verifying) the model on the second period. The procedure can eventually be done again by switching the role of the two periods.
- the *differential split-sample test(DSS)*: the test *SS* is applied under non-stationary climatic forcing conditions.
- the *proxy-basin test (PB)*: the model is calibrated on the basin A and applied on the basin B. Then the model is calibrated on the basin B and applied on the basin A. If the test is satisfactory, the model can be reliably applied to ungauged catchments.
- the *proxy-basin differential split-sample test(PB-DSS)*: the test *PB* is applied under non-stationary climatic forcing conditions.

The split-sample test is the most widely used procedure to calibrate the models in hydrological modelling studies.

Application to this study

Based on the availability of the selected data, we propose to use the classical split-sample test in the following manner:

Calibration: the best period to be used as calibration period seems to be the period from 11/1993 to 06/1994. It is first proposed to calibrate the model by only comparing the simulated and measured discharge at the outlet.

Validation:

(1) the model simulations can be validated for the same period (from 11/1993 to 06/1994) in the intermediate station (C8H020).

(2) If the simulations are acceptable, the model can be tested over the second period 10/1996 to 03/1997 according to the flows measured at the station C8H020.

(3) As a final validation test, the model can be applied for the third period (from 11/1999 to 06/2000) by injecting into the model channel cell corresponding to the tunnel output, the flow measured by the station C8H036.

In summary, the concurrence of the various data sets available for calibration and validation are presented in Figure 2.11.

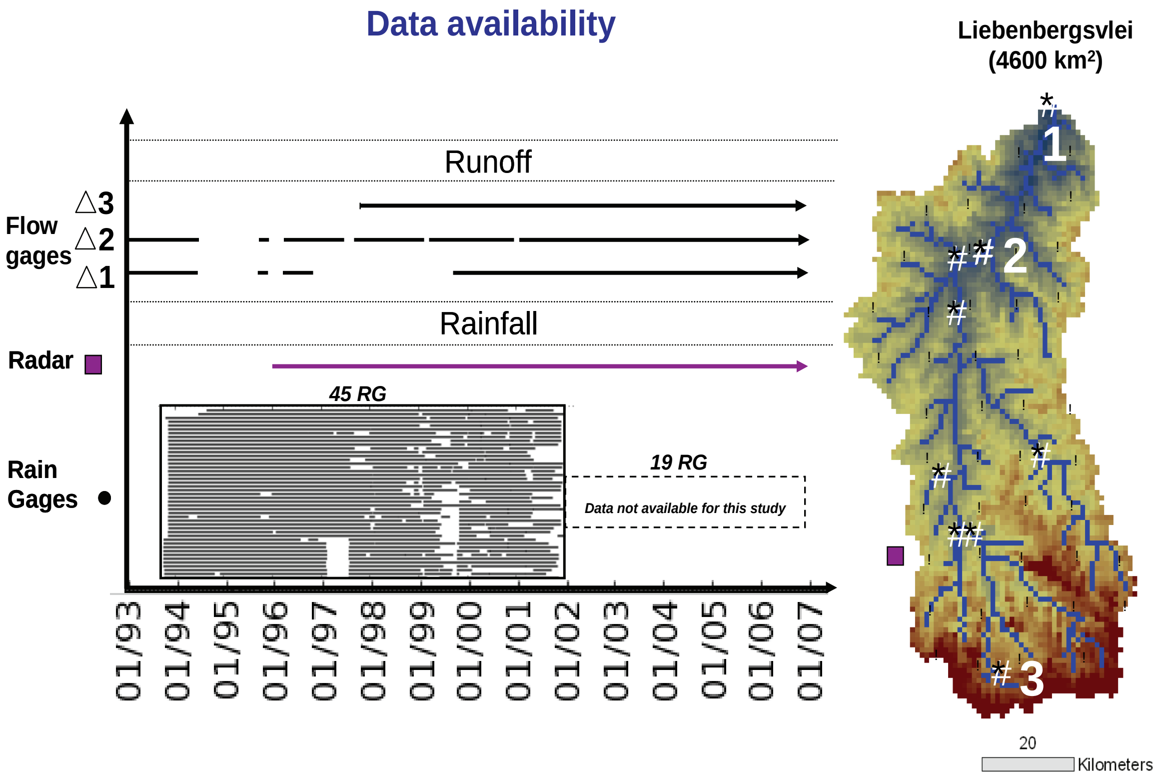


Figure 2.11 Summary of the Hydrometeorological data availability on the Liebenbergsvlei catchment.

From catchment DEM to cell connection

The automation of the TOPKAPI model requires the definition of a numerical grid that divides the catchment space into squared cells that must be connected in order to model the transfer of flow (surface and subsurface) within the catchment. The Digital Elevation Model is used as the base map to:

- Define the grid and thus the spatial resolution of the model
- Delineate the stream network.

These two steps can be achieved by using GIS software (here the ESRI ArcGIS software package has been used) and are thoroughly detailed in Parak (2007) and Pegram et al. (2007).

The spatial resolution of the model (parameter X in the model equations, see Table 2.2) is imposed by the grid characteristic of the DEM. In this application it was desirable to use the standard 1 km resolution usually provided by the freely available USGS DEM (USGS DEM, 2006). Thus, the initially fine resolution of 200m provided by the DEM available over the Liebenbergsvlei catchment (DLSI, 1996), was intentionally re-sampled to a resolution of 1 km (Parak, 2007).

The 1 km DEM was then treated in order to delineate the streamflow directions. A common problem in DEMs is the occurrence of sinks. A sink, also referred to as a depression or pit, is a cell or area that is surrounded on all sides by higher elevation values. As such, a sink is an area of internal drainage and prevents the down-slope routing of water and, unless it is an actual case such as a lake or swamp, is an error in the data-set. These errors often arise due to the sampling techniques used in processing a DEM or due to the rounding off of elevation values to integers (Mark, 1988). In order to create an accurate representation of the flow direction, it is best to use a DEM that is free of sinks, unless they are naturally occurring. In order to simplify the DEM treatment, Liu and Todini (2002) suggest the consideration of only 4 drainage directions (D4). However, limiting the drainage to 4 directions can lead to an unrealistic representation of the relief variability. Indeed, the filling of the sinks using D4 in the Digital Elevation Model treatment results in a strong smoothing of the relief variability. For this reason, the TOPKAPI model was adapted to be compatible with 8 direction drainage (D8), which includes the 4 extra pixels beyond the diagonals. To fill a sink in D8, the methodology proposed by Mark (1988) was used, exploiting the hydrology toolbox available in the ArcGIS software.

From the depressionless DEM so created, the outflow direction of each cell is determined, i.e. the direction of the steepest descent from an active cell to the neighbouring downstream cells. A drainage direction code is assigned to each cell depending on the direction of maximum slope, which is calculated as the maximum difference in elevation divided by the horizontal distance from the centre of the active cell to the centres of the four surrounding cells. If the maximum slopes to several cells are the same, then the neighbourhood around the active cell is enlarged until the direction of steepest slope is found. The D8 flow direction raster map for the Liebenbergsvlei catchment is shown in Figure 2.12a.

The next step in the delineation of stream networks is to determine the number of upslope cells that contribute flow into each cell, i.e. the flow (in terms of contributing cells) accumulated in each cell. This is also achieved in ArcGIS using the "Flow Accumulation" tool from the hydrology toolbox. Figure 2.12b shows the flow accumulation raster for the Liebenbergsvlei catchment. The colour palette indicates, for each cell, the number of upslope cells that feed it. The main trunk of the stream network is easily identified from this image, since it has the largest number of upstream cells.

The final step in delineating the stream network from a DEM is to assign a threshold value to the flow accumulation raster, for the minimum number of upslope cells that are required to initiate a channel in an active cell. The determination of this threshold value depends, according to Tarboton et al. (1991), on climate, slope and soil characteristics. They present procedures in order to 'rationally select the scale at which to extract channel networks' which correspond to networks obtained through more traditional methods, such as from topographic maps or fieldwork. Figure 2.12c shows a comparison between the stream network delineated from the DEM in the manner described above and a stream network digitised from a topographic map of the catchment (digitized by Midgley et al., 1994) at a spatial scale of 1:250 000. The comparison shows good correspondence between the two sources of networks, which is an important check. The threshold value chosen for the area over which a cell is considered to initiate a river channel (a parameter referred to as $A_{threshold}$ hereafter) was eventually fixed at 25 km^2. In reality, the river network extension evolves within the seasons and with flow intensities but this value was considered to be representative of the average river network. Setting $A_{threshold}$ equal to 25 km^2 is in accordance with Todini's (1996) recommendation that the ratio between the number of cells containing channels and the total number of cells takes on a value ranging between 5% and 15% of the total catchment area; this threshold defined a total channel length of 522 km by 1 km pixel width, which is 11.3% of the total Liebenbergsvlei area of 4 625 km^2.

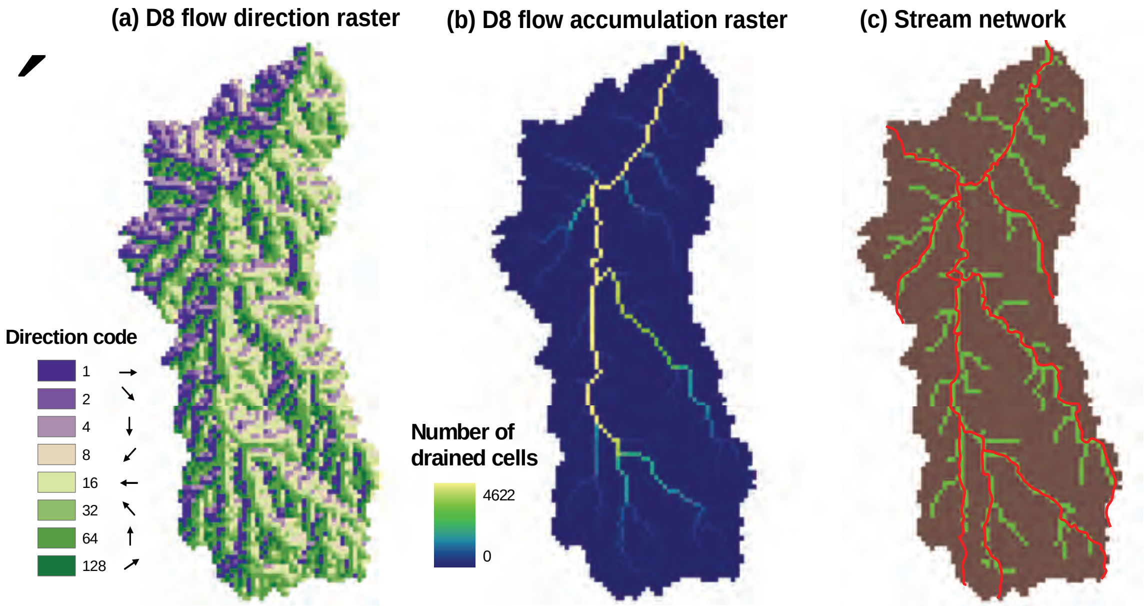


Figure 2.12 An illustration of the steps in moving from the DEM to stream channel delineation. The red line superimposed on (c), is the river network digitalized from a topographic map at a spatial scale of 1:250 000 (Midgley et al., 1994).

From catchment data to physical model parameters

One of the most noticeable advantages of the TOPKAPI model, underlined by all the studies dealing with the model, is its physical basis that allows the user to link the model parameters with the catchment characteristics. The model parameters can be estimated *a priori* from elevation data, soil and surface properties such as those available on the Liebenbergsvlei catchment but also available on the major part of South Africa.

A total of 15 parameters have to be assigned in the TOPKAPI model, as listed in Table 2.7. Among them, 11 are cell specific, meaning that they are potentially spatially variable (depending on the detail of the information available), and they mainly refer to physical characteristics ($tan(\beta)$, $tan(\beta_c)$, L, K_s, θ_r, θ_s, n_o, n_c, α_s, k_c, W see Table 2.7). The 4 others are constant and refer to geometric characteristics of the channel or grid cell (X, $A_{threshold}$, the minimum width of channel W_{min}, the maximum width of channel W_{max}). All the parameters and their associated values or range of values are reported in Table 2.7, as well as the references used to infer the parameter values.

As already noted in earlier, the constant parameters X and $A_{threshold}$,, have been assigned according to the DEM treatment and their respective values are 1 km and 25 km^2. W_{min} and W_{max} were fixed at 5 m and 35 m (estimations made from photographs taken at the flow stations after checking Google Earth).

For the cell-specific parameters, different degrees of treatment have been used to assign the values from the available data. Some of the parameters were directly provided by available datasets, some of them required references to the existing literature. Figure 2.2 illustrates how the catchment data maps were used to derive the maps of spatially variable model parameters.

Maps of soil depths L and saturated soil moisture content θ_s were already available over the catchment in the data set of soil properties (SIRI, 1987). The slopes of the ground $tan(\beta)$ (used for the soil and overland reservoirs) were computed from the DEM according to a neighbourhood function more representative of the mean slope within the cell and thus more representative of the transfers inside the cell (in and over the soil). This was achieved by using the "slope" tool in the ArcGIS toolbox. The slopes used to transfer the flows in the channel drainage network $tan(\beta_c)$ were computed from cell to cell in a down-stream direction using differences in altitude. Following Liu and Todini (2002), the pore-size distribution parameter α_s was uniformly set to the value 2.5.

A sensitivity analysis (not presented herein) showed that varying the value of α_s in the realistic range of its values (between 2 and 4 according to Liu and Todini, 2002) had only a small influence on the simulations. As a first approximation, and because of the relatively homogeneous cropland/grassland landcover, the crop factor k_c was assumed to be spatially uniform over the catchment and equal to 1 (the consequence of this choice is that the evapotranspiration forcing applied in the simulations is assumed to be the actual ET). The width of channel in each channel cell was assigned by applying a linear relationship between the drained area and the channel width as proposed by Liu and Todini (2002):

$$W_i = W_{\max} + \frac{W_{\max} - W_{\min}}{\sqrt{A_{total}} - \sqrt{A_{threshold}}}\left(\sqrt{A_{drained_i}} - \sqrt{A_{total}}\right) \qquad (2.3)$$

where:

A_{total} is the area drained in the catchment

$A_{drained_i}$ is the area drained by a given cell i

Table 2.7 Values of the TOPKAPI model parameters estimated *a priori* from data and literature, and values of multiplying factors obtained from the calibration procedure

Parameter		Values *a priori*	Origin and references	Post-calibration multiplying factor value
Cell specific				
Ground Slope	$tan\ \beta$	1.7E-4 – 1.81E-1	DEM (DLSI,1996)	
Channel Slope	$tan\ \beta_c$	1.7E-4 – 1.81E-1	DEM (DLSI,1996)	
Depth of surface soil layer (m)	L	0.33 – 0.81	Soil type map (SIRI,1987)	fac_L 1.
Saturated hydraulic conductivity (m s^{-1})	K_s	1.67E-4 – 5.18E-2	Soil texture map (Midgley et al., 1994 ; Maidment,1993)	fac_{Ks} 0.6
Residual soil moisture content (cm^3/cm^3)	θ_r	0.02 – 0.09	Soil texture map (Midgley et al., 1994; Maidment,1993)	
Saturated soil moisture content (cm^3/cm^3)	θ_s	0.41 – 0.44	Soil type map (SIRI,1987)	
Manning's surface roughness coeff. (m$^{-1/3}$s^{-1})	n_o	0.025 – 0.1	Landuse map (GLCC, 1997; Chow et al.,1988)	fac_{no} 1.
Manning's channel roughness coeff. (m$^{-1/3}$s^{-1})	n_c	0.035 – 0.045	Strahler order method (Liu and Todini, 2002)	fac_{nc} 1.7
Non-linear soil exponent (-)	α_s	2.5	Liu and Todini (2002)	
Crop factor (-)	k_c	1.0	Land-use map (GLCC, 1997)	
Channel width (m)	W	See Eq. (7)	Liu and Todini (2002)	
Constant				
Horizontal dimension of cell (m)	X	1 000	DEM (DLSI,1996)	
Max. channel width at outlet (m)	W_{max}	35	Field photographs	
Min. channel width (m)	W_{min}	5	-	
Area required to initiate channel (m^2)	$A_{threshold}$	25 000 000	Todini (1996)	

* Comment on fac_{Ks} : In Vischel et al. (2008a), the calibrated value of fac_{Ks} was reported as 60, we subsequently discovered that the *a priori* soil conductivity values used in that study were too small by a factor of 100. This has been corrected in subsequent work and in this report.

Particular attention was given to the 4 remaining parameters (K_s, θ_s, n_o, n_c) that are not easily measurable *in situ* and have thus to be indirectly inferred by using literature references and tables. The ordering method of Strahler (1957) was used to infer the values of the channel roughness Manning's coefficients n_c. In Liu and Todini (2002), channel orders of 1, 2, 3 and 4 were assigned values of 0.045, 0.04, 0.035 and 0.035 for the Upper Reno catchment in Italy. These values were assumed to be suitable as starting values for the Liebenbergsvlei catchment for which no more precise information was available. The values of the overland roughness Manning's coefficient n_o were derived from the land use/cover map (GLCC, 1997), using the widely used tables proposed by Chow et al. (1988). The residual soil moisture θ_r and the hydraulic conductivity at saturation K_s were derived from the soil texture map (Midgley et al., 1994) according to parameter tables for the Green-Ampt infiltration model (Maidment, 1993).

Model configuration

Once the model parameters have been estimated *a priori* from the catchment data, the application of the model requires some specific features defined in four main steps:

- The definition of a modelling period
- The choice of a simulation time-step
- Pre-treatment (aggregation/disaggregation) of the forcing data to conform with the spatial scale and time-step of the simulation
- If necessary, an adjustment of some parameters in order to improve the model performance and account for the uncertainties in the *a priori* estimates.

Selected period

From the data set presented above, 2 seasons of 8 months each were selected during which the rainfall and flow data were both continuous and of good quality. The first season (Season 1) between November 1993 and June 1994 was used to adjust the *a priori* parameters of the TOPKAPI model. The second season (Season 2) between November 1999 and June 2000 is used further as a model validation period. As shown in Figure 2.13, Season 1 is extracted from a relatively dry year (580 mm) and Season 2 from a relatively wet year (780 mm), compared to the mean annual rainfall value of 636 mm computed from the 1st of November to the 31st of October over the period 1993-2002.

Model resolution

The spatial resolution of the model is imposed by the resolution of the DEM selected to delineate the catchment and extract the channel network. However, the simulation time is chosen by the user. Here, a 6h time-step was chosen, which is small enough to model the main discharge variations, since the catchment response time is estimated to be between 1 and 2 days.

It is worth noting that, at this time-step, only 20 min on a current PC are needed to run an 8 month season over the entire catchment (4 625 cells of 1 km).

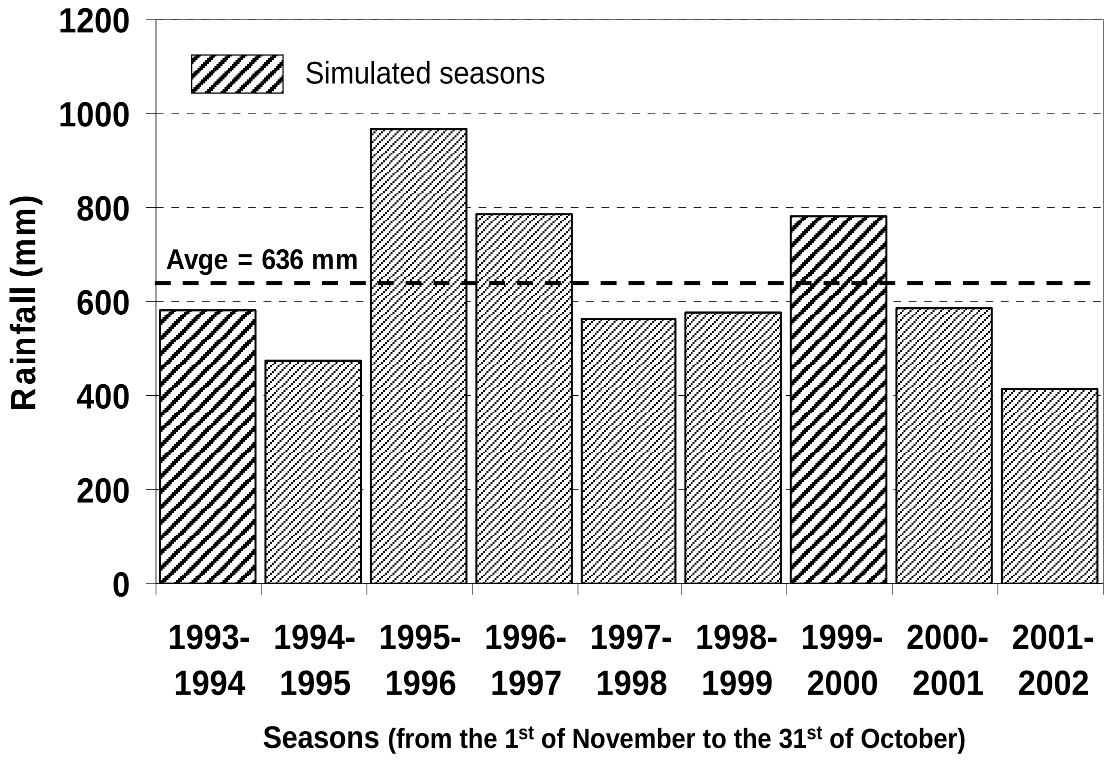


Figure 2.13 Mean (wet) seasonal rainfall computed between 1st November and 31st October from 1993 to 2002 over the Liebenbergsvlei catchment. These seasons match the duration of the calibration and validation periods in the study as indicated in Figure 2.14.

Forcing variables

For the 2 seasons considered in this study, 6 h time-step rain-fields were computed by Kriging the rain-gauge measurements at 1 km resolution using a climatological variogram with range of 30 km and a zero nugget (Wesson and Pegram, 2006). Because of the very good density of the network (10 km spacing) for the chosen period, the rainfall estimates from the weather radar were not used to create the rain-fields.

As no evapotranspiration data are available for the simulated periods on the catchment, the mean annual evapotranspiration over the region was used and disaggregated at a 6h time-step, according to a mean seasonal signal determined from McKenzie and Craig (1999).

Calibration of the parameters

The TOPKAPI model is considered to be a physically based model, with all its parameters having physical meaning and which can be measured directly through fieldwork. But, as in every physically-distributed model, the TOPKAPI model is subject to several uncertainties associated with the data on:

- The information on topography, soil characteristics and land cover
- The approximate methods and tables used to infer physical parameters from the data
- The approximations introduced by the scale of the parameter representations.

For these reasons, Liu and Todini (2002) suggest that the calibration of parameters is still necessary but all studies dealing with the TOPKAPI model maintain that the calibration of the model is 'more of an adjustment' that can be achieved through simple trial- and- error methods. In the present application of the model on the Liebenbergsvlei catchment, the parameters estimated *a priori* from the catchment data were did not result in satisfactory flow simulations and as a result, a calibration procedure was required.

In order to be convinced that only a small adjustment of the parameters is sufficient for efficiently modelling the Liebenbergsvlei catchment, an automated method of calibration was implemented. Inspired by the so-called Ordered Physics-based Parameter Adjustment Method (OPPA) proposed by Vieux et al. (2004), the method consists of a dissociated calibration of the parameter responsible for the production of the runoff, from those responsible for the routing of runoff. According to a sensitivity analysis (not shown here, see also Liu et al., 2005) the most sensitive parameters controlling the runoff production are the soil depth L and the soil conductivity K_s, while the Manning roughness of channel n_c and overland n_o are the primary routing parameters. In the absence of any quantitative information, the initial soil moisture $V_{s\_initial}$, which was shown to have a strong influence on the simulations, was also calibrated. Ten values of mean catchment saturation between 1% and 90% were tested.

Note that the *a priori* parameters are not individually calibrated for each cell. This would lead to an extreme over-parameterization of the model and to multiple and inconsistent combinations of parameter values.

The spatial pattern of the parameter maps is relevant information that was chosen to be conserved in the calibration procedure by using a multiplicative factor applied uniformly in space to the maps of the *a priori* parameters. For our application the 4 multiplicative factors to be applied were fac_L (for the soil depth), fac_{Ks} (for the hydraulic conductivity), fac_{no} (for the overland roughness) and fac_{nc} (for the channel roughness).

The triplet (fac_L, fac_{Ks}, $V_{s\_initial}$) was then adjusted in order to minimize the root mean square error (RMSE) objective function comparing modelled and observed discharge volumes aggregated at a monthly time-step. Once the optimum triplet was found, the pair (fac_{no}, fac_{nc}) was adjusted using the R^2 coefficient (as the objective measure) to match the timing of observed and modelled discharges at a 6 h time-step.

The calibration was first carried out using the flows estimated at station 2 (see Figure 2.11). At this station, the upstream drainage area is 3 563 km^2, which effectively preserves the predominant patterns of soil heterogeneity for the entire catchment.

Results Calibrated period

Calibration was done on the flows of Station 2 in Season 1, shown in Figure 2.14a. All other simulated flows in (b), (c) & (d) were verifications based on this single calibration. There is a good visual correspondence between observed and modelled hydrographs that is supported by a high value of the coefficient of efficiency (Nash and Suttcliff, 1970) of 0.78 and a good correlation illustrated in Figure 2.15.

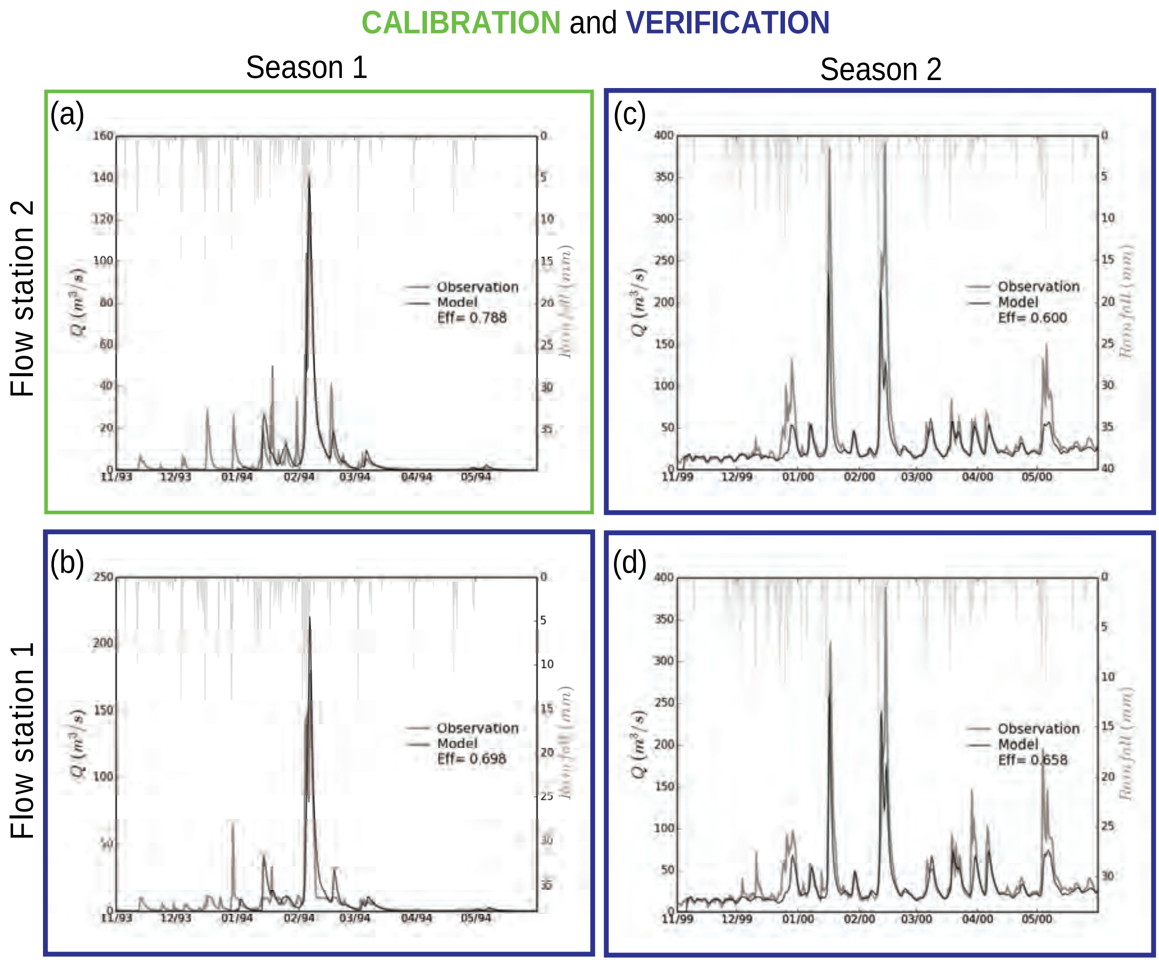


Figure 2.14 Comparison between modelled and observed discharges on the Liebenbergsvlei catchment for calibration and validation procedures. Calibration was done on the flows of Station 2 in Season 1, shown in (a). All other simulated flows in (b), (c) & (d) were verifications based on this single calibration.

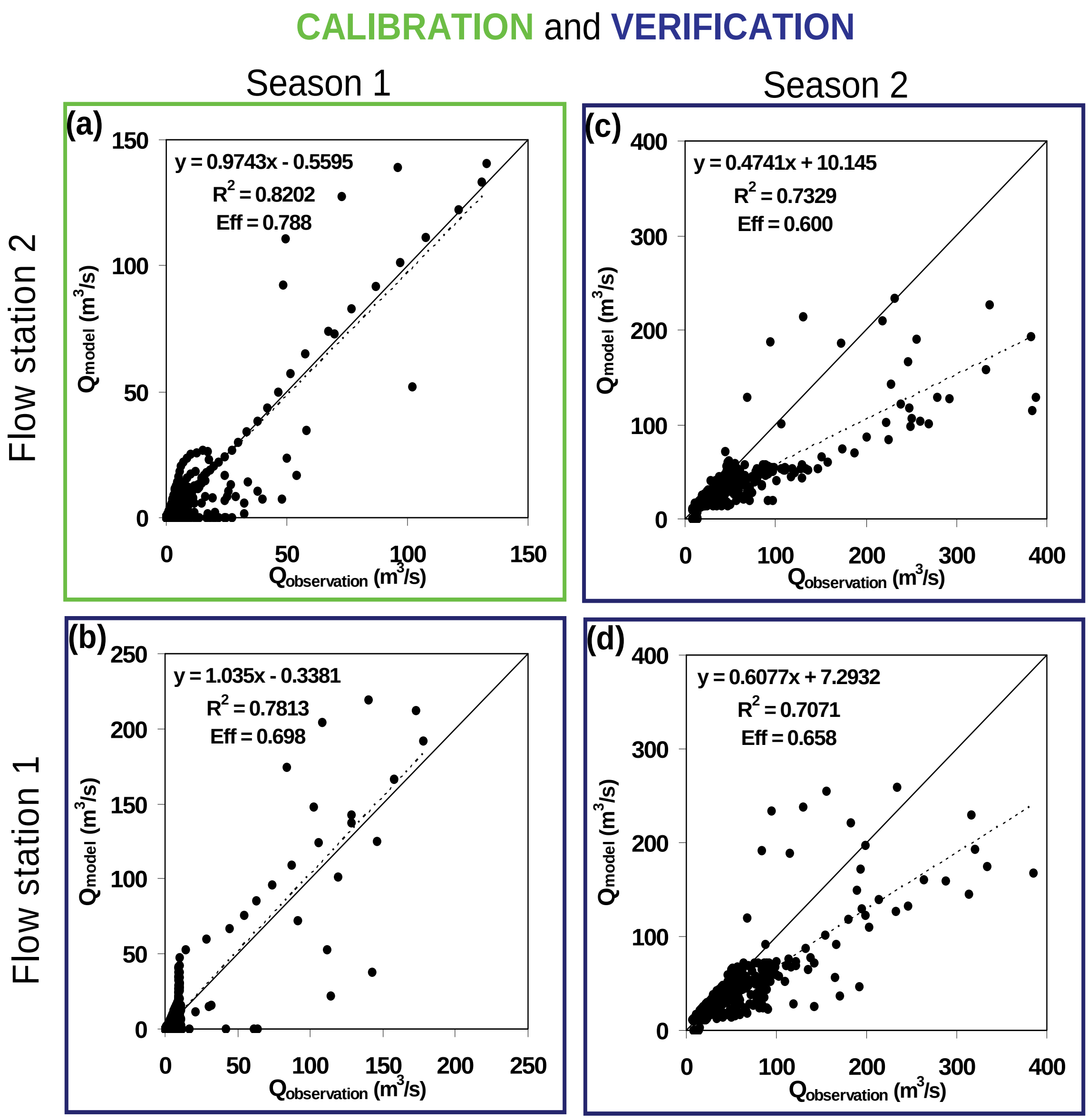


Figure 2.15 Scatter plot of observed versus simulated discharges on the Liebenbergsvlei catchment for calibration and validation procedures. Calibration was done on the flows of Station 2 in Season 1, shown in (a). All other simulated flows in (b), (c) & (d) were verifications based on this single calibration.

In Table 2.7 the values of the 4 calibrated multiplying factors are reported. It is worth noting that all the values of the parameters estimated *a priori* were quite appropriate except for the soil conductivity which has been multiplied by a factor of 0.6. The initial soil moisture was adjusted by calibration to a mean value of 40% over the catchment.

Evaluation of the calibration relevance on the entire catchment

As a verification of the relevance of the calibration procedure and its effect on other discharge time series, the calibrated model was applied to the entire catchment. For the same season (Season 1) the observed and modelled discharges at the outlet of the catchment (Station 1) are plotted in Figure 2.14b. Globally, there is once again a good correspondence between observed and modelled flows, however at some points, the observed data seem to be unreliable since some peaks recorded at Station 2 do not appear as they should at the outlet and the recession shape of the main peak discharge seems somewhat unrealistic. This is clearly indicated on Figure 2.15b where the shape of the scatter plot of simulated versus observed discharge values presents an unexpected accumulation of observed discharge around the value 10 $m^3.s^{-1}$.

Validation of the calibration procedure using data from another season

In order to validate the model calibration on more reliable data, the model was then applied to a different season (Season 2) some years later, using the same parameter values as for Season 1. During this season, the discharges are influenced by the inter-basin transfer flows now arriving from Lesotho. In order to reliably compare the modelled and observed discharges, the external flows observed at Station 3 had to be taken into account in the simulations. The modular structure of the TOPKAPI model allows the user to easily integrate additional fluxes that could come from new modelled processes, external flows or reservoirs. This is facilitated by the fact that a cell can be considered as an independent entity controlled by its own characteristics and its input. Here the external tunnel flows were simply added to the 'natural' input term Q_i^{in} of the cell positioned at the location the closest to Station 3. Again in the absence of any information about the initial soil moisture, the value of 40% calibrated for Season 1 was assumed to be applicable for Season 2. Hydrographs are plotted in Figure 2.14c and d, and scatter plots of observed versus simulated discharge are plotted in Figure 2.15c and d. Again, good simulations of the hydrographs were obtained even if the main peak discharges are underestimated. One can also note that the timing of the flows is remarkably good, especially at the beginning of the season, when in the absence of rainfall, the flows are mainly explained by the external flows that are routed from the upper part of the catchment; these appear pulsed due to hydropower generation.

Summary

TOPKAPI is a physically-based distributed hydrological model that couples the continuity equation with the soil and surface dynamic equations, resulting in a set of three kinematic wave equations applied to soil, overland and channel non-linear reservoirs within each cell, to which an evapotranspiration module is added. Its implementation is mainly based on elevation data (provided by a Digital Elevation Model) and also requires information about catchment surface properties and land use.

The present study is the first application of the TOPKAPI model on an African catchment, while the performance of the model has already been demonstrated in several catchments throughout the world. Applied on the Liebenbergsvlei catchment (4 625 km^2, South Africa), the model showed good ability in modelling the river discharges at a small (6 h) time-step with a limited adjustment of the parameters, and low computation times.

The exercise of producing a software implementation of the TOPKAPI model from the existing literature, and applying this on the Liebenbergsvlei, has resulted in significant expertise and understanding of the model being gained. Several improvements to the original formulation have been made, for example an improved evapotranspiration scheme and a competitive algorithm for solving the differential equations, as outlined in Appendix A. From this experience, the following points indicate that TOPKAPI is a very promising hydrological model in the South African context:

- ***Physical realism***. A comprehensive and robust physical framework has been developed by Liu and Todini (2002) to give an explicit representation of the most significant hydrologic processes explaining the hydrological response of catchments. This response is the runoff production due to the subsurface flows and contributing saturated areas, surface transfers on hillslopes and river channels and evapotranspiration losses.

- ***Spatial representation of the catchment properties***. The spatial discretization of the catchment at fine resolution (1 km or finer) displayed by the DEM gives a distributed knowledge of the water fluxes that occur throughout the catchment, both surface and subsurface. This fine resolution is also coherent with the scales at which the processes effectively occur.
- ***Sensible parameterization based on observed values from the field.*** The parsimonious parameterization of the TOPKAPI model limits the number of parameters to 15, many of which are preset by the data available. Because of its physical basis, the parameters can be assigned according to field data obtained from in-situ measurements or remote sensed data, the parameters have a physical meaning attached to them, only a limited number of the parameters have to be adjusted due to *a priori* uncertainty, the parameters are relatively scale independent up to the limit of 1 km resolution as shown by Martina (2004).
- ***Accuracy and computation speed***. In this application of TOPKAPI a combination of a quasi–analytical solution (Liu and Todini, 2002) with a numerical integration procedure based on the Runge-Kutta-Fehlberg method was used. This hybrid scheme, which differs from the scheme initially suggested by Liu and Todini (2002), was chosen after a comparison with other methods, by carefully analysing the performances in terms of computation speed, numerical stability and accuracy of the simulations (detail in Appendix A).
- ***Modularity***. The structure of the TOPKAPI model is based on the connection of independent cell entities. This allows the easy addition of reservoirs or dams, external flows (as was done for the Liebenbergsvlei case study) or supplementary hydrological processes that might modulate the hydrological response of the catchment.
- ***Ease of use***. Because of its simple and highly comprehensive reservoir structure, its parsimonious parameterization and the direct physical link between the catchment data and the model parameters, the TOPKAPI model is very easy to use.

Combining the representation of the rapid flows associated with flash flood events and the fast computation time, TOPKAPI is a suitable tool for operational hydrology. It is also a useful tool for scientific and research hydrology. The TOPKAPI model has already been used as a research modelling tool in the project SHARE (a TIGER Innovator contract with the European Space Agency) to evaluate the soil moisture products derived from the scatterometers onboard the satellites ERS 1-2. Beside its relevance of modelling the flow discharges of the Liebenbergsvlei catchment, the model also shows very good performances in modelling spatial soil moisture within the catchment as shown by the strong correspondence found between remotely sensed and modelled soil moisture (see Vischel et al., 2007), to be discussed in the next section.

Comparison of TOPKAPI SSI and Remote Sensing estimates of SM

In this section, we compare two independent approaches of soil moisture estimation on the Liebenbergsvlei catchment. The first estimates are derived from the physically-based distributed hydrological model TOPKAPI. The second set of estimates is derived from the scatterometer on board the European Remote Sensing satellite ERS.

The remotely sensed soil moisture estimates used in this study are derived from scatterometers on-board of the satellites ERS-1 and ERS-2 (Wagner et al., 2003). The ERS scatterometer is a C-band radar (5.3 GHz), operational since 1991 up to at least 2007, which has acquired data with a spatial resolution of 50 km with a vertical polarization at a 25 km grid spacing. Global coverage is achieved by the satellite every 3 or 4 days on average, but since the ERS scatterometer is in operational conflict with the ERS Synthetic Aperture Radar, only a part of the coverage is effectively available for scatterometer measurements. The repeat cycle at one point is thus 7 days on average, varying irregularly from 3 to 10 days. This global product is used here in its finalized form. Over Southern Africa, ERS data are available until 2000.

Scatterometer microwave radiation data are very sensitive to the moisture content of the surface soil layer due to the strong variation of the dielectric constant of the soil with water content. However other factors influence the scatterometer backscatter signal. Soil moisture retrieval methods must mainly take into account the effects of vegetation, surface roughness and heterogeneous land cover. The retrieval method technique adopted for the data used here is based on the change detection method proposed by Wagner et al. (1999a). To account for effects of roughness and heterogeneous land cover, the lowest and highest values of backscatter coefficients are determined (respectively σ^0_{dry} and σ^0_{wet}) over the nine-year measurement period 1992-2000. The two limiting reference values are assumed to be representative of the vegetated land surface under respectively dry and saturated soil conditions. The measured backscatter coefficients are then compared to σ^0_{dry} and σ^0_{wet}, resulting in the definition of topsoil moisture contents m_s (<5 cm) interpreted as a surface soil moisture index (i.e. a relative quantity) ranging between 0 and 1 (respectively, 0–100%), scaled between zero soil moisture and saturation. At any time t, m_s is then defined as:

$$m_s(t) = \frac{\sigma^0(t) - \sigma^0_{dry}}{\sigma^0_{wet} - \sigma^0_{dry}} \tag{2.4}$$

The effects of plant growth and decay are taken into account through the application of varying seasonally σ^0_{dry} and σ^0_{wet} values as proposed by Wagner et al. (1999b). This method exploits the multi-incidence capabilities of the ERS scatterometer to describe the effect of enhanced volume scattering in the vegetation layer and the corresponding decrease of the ground scattering contribution.

Definition of a remotely sensed Soil Water Index (SWI) and modelled Soil Saturation Index (SSI)

As already noted, the remotely sensed soil moisture estimation is representative of the relative water content of the first 5 centimetres of topsoil effectively 'seen' by the scatterometer. However, for the purpose of the present study, which is to compare the soil moisture as modelled by TOPKAPI and the remotely sensed estimate of soil moisture, the variable of concern is the soil moisture in the entire hydrologically active soil layer.

In order to provide a reliable comparison, the soil moisture in the whole soil layer must thus be obtained from the surface soil moisture estimated by the satellite. In addition to the surface soil moisture index available in the global ERS soil moisture product, a Soil Water Index (*SWI*) is provided that aims to estimate the soil moisture profile in the soil horizon from the ERS product, initially estimated as relative surface soil moisture (SSM). The method used here to estimate *SWI* was proposed by Wagner et al. (1999c). It is a simple conceptual infiltration model based on an exponential filter, temporally smoothing the signal of the (instantaneously estimated) relative surface soil moisture (SSM) to give the Soil Water Index, *SWI*:

$$SWI(t) = \frac{\sum_i m_s(t) e^{-(t-t_i)/T}}{\sum_i e^{-(t-t_i)/T}} \quad \text{for} \quad t_i \le t \tag{2.5}$$

where m_s is the surface soil moisture estimate SSM from the ERS scatterometer defined in equation 2.4. *T* represents a characteristic time length depending to the soil properties (mainly soil depth, diffusivity and moisture state). To maintain the crucial independence of the physically based approach of TOPKAPI and the remotely sensed soil moisture estimates, it was decided not to refine the estimation of the parameter *T* for the particular study area by using the soil data. Thus the value of *T*=20 days, suggested by Wagner et al. (1999c) as an average value, was retained.

A surrogate for *SWI* can easily be defined for TOPKAPI by computing the relative soil saturation at each catchment cell, for each time step of the simulation. We choose to call it a Soil Saturation Index, defined as SSI = $(\theta - \theta_r)/(\theta_s - \theta_r)$; here θ_s = porosity (direct from the SIRI data base) and θ_r is obtained from table 5.3.2 of Maidment (1993) based on WR90 (Midgley et al., 1994) texture classes. The reason for distinguishing between these estimates is that the physical quantities primarily measured/estimated are quite different, even though they are attempting to infer the same quantity – the local amount of water in the soil.

Two different scales are considered to make the comparison between the modelled and remotely sensed soil moisture. The first is the catchment scale, at this scale: (i) the mean catchment SSI is computed from the hydrological model by averaging over the catchment the SSI computed at each cell, (ii) the mean catchment SWI is computed from the scatterometer data, by averaging over the catchment the SWI computed for the scatterometer grid points in and surrounding the catchment (the average being weighted according to Thiessen polygons). The second scale is the scatterometer footprint scale, which is smaller than the catchment scale. This corresponds to the original scatterometer resolution defined by a circle of diameter 50 km. The corresponding footprint SSI is computed from the hydrological model by averaging the SSI computed at each cell within the footprint. In order to make a robust comparison, only the three footprints showing the largest areal coverage of the catchment were considered.

Results

At Catchment scale

The modelled SSI and remotely sensed mean catchment SWI are compared in Figure 2.16 for the two modelled seasons, at the time step of ten days imposed by the ERS sampling interval. There is a very good correspondence between the two estimates, as illustrated by the regression coefficients (R^2) of 0.759 for the first (calibration) season and 0.923 for the second (validation) season. According to the regression equation, a relative bias is observed (which seems to be independent of the season) that is likely to be due to the uncertainties associated with each of the two. Despite this, the order of magnitude of the remotely sensed SWI and the modelled SSI is still very similar. As an interesting example, the value of the initial soil moisture, which was calibrated at 40% for the catchment model, could have been estimated by using the remotely sensed value. This result is very encouraging since the initialization of hydrological models after a dormant period remains a constant problem in hydrology.

At catchment scale

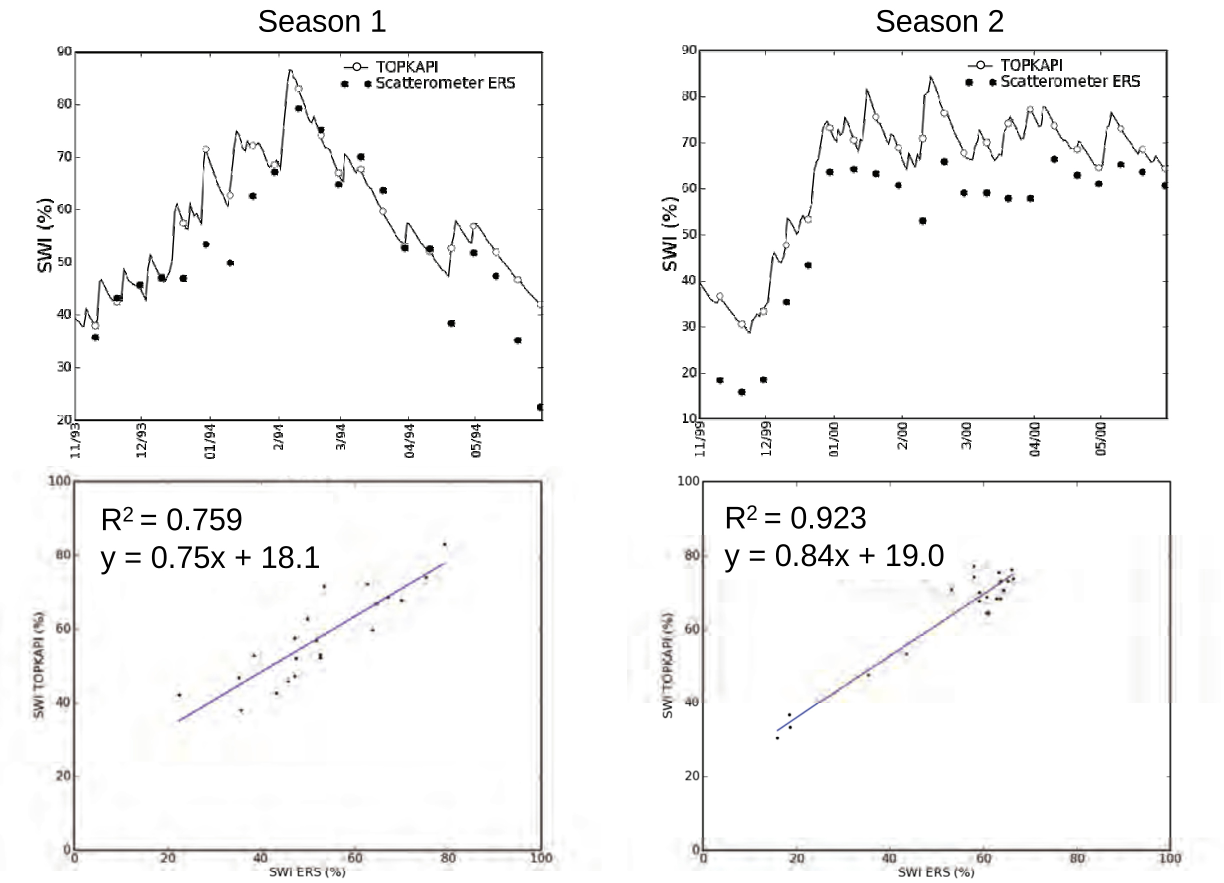


Figure 2.16 Comparison between the modelled SSI and the remotely sensed SWI computed at catchment scale. The open circles are the TOPKAPI SSI estimates at time corresponding to the scatterometer SWI estimates (filled circles).

At footprint scale

Figure 2.17 and Figure 2.18 show respectively the remotely sensed SWI and the modelled SSI at footprint scale and the associated scatter plots. The results show that the good correspondence already found at catchment scale is still evident at the smaller scale of the footprint. The correlations are still fair (greater than 0.678), while according to the regression equations, the bias between the two independent estimates is relatively stable and appears to be independent of season and location.

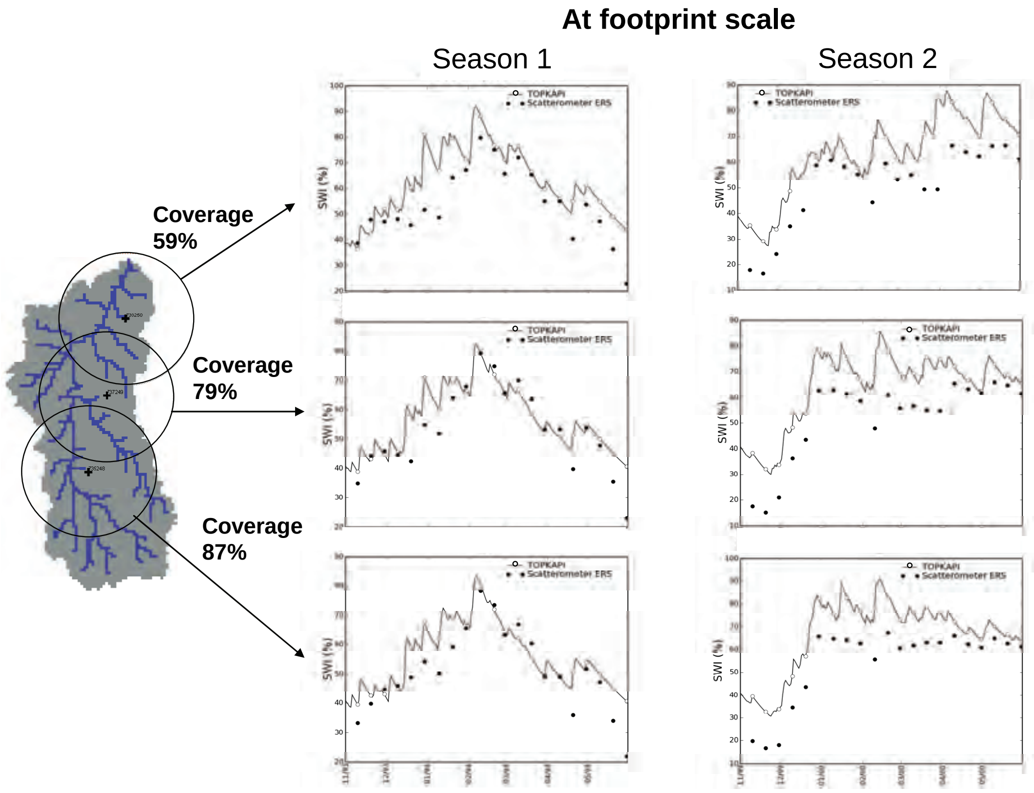


Figure 2.17 Comparison of the modelled Soil Saturation Index (SSI) and remotely sensed Soil Water Index (SWI) at the scatterometer footprint scale.

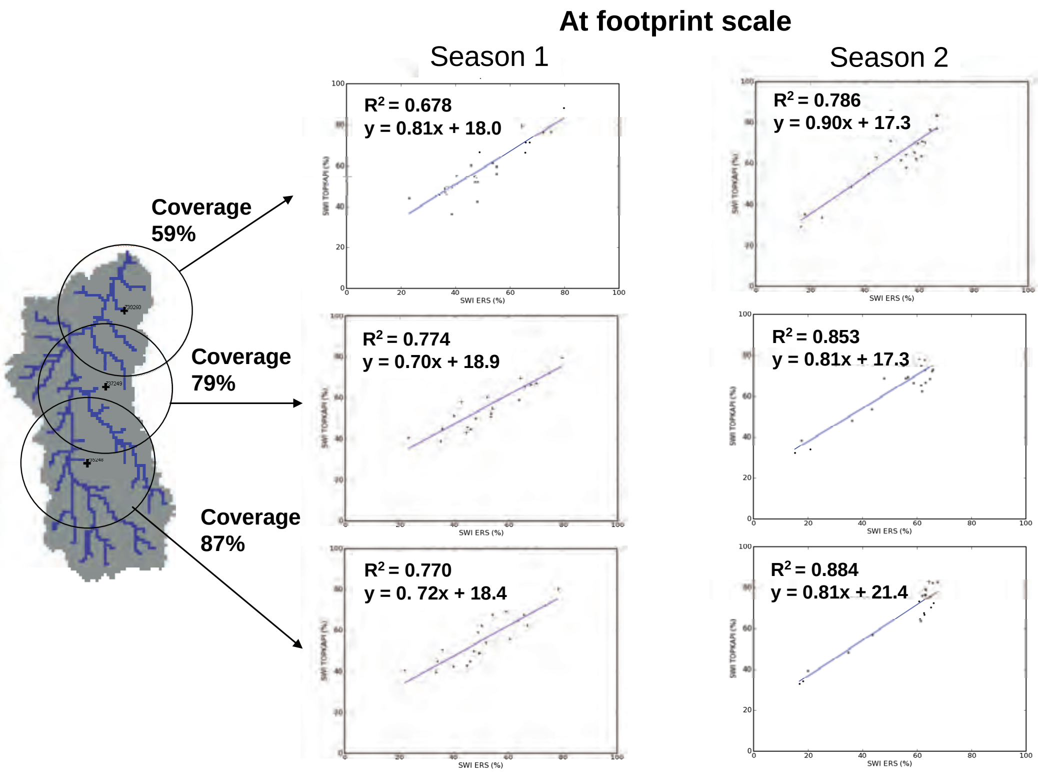


Figure 2.18 Comparison of the modelled SSI and remotely sensed Soil Water Index (SWI) at the scatterometer footprint scale (Scatter plots and regression equations).

Summary

The purpose of this Chapter was to compare, for the purpose of corroboration, not validation, two independent approaches used to estimate soil moisture at the scale of a region-sized catchment (Liebenbergsvlei, 4625 km^2, South Africa). The first estimation was derived from physically based hydrological modelling of the catchment using the TOPKAPI model (Liu and Todini, 2002) and the second was derived from the remotely sensed observations of the scatterometer on board the ERS satellite. A calibration procedure of the TOPKAPI model has been carried out consisting of the adjustment of the four most sensitive parameters of the model according to runoff production and routing. A good agreement was found between observed and modelled hydrographs of the Liebenbergsvlei catchment for both the calibration and the verification period. The comparison between the modelled SSI and the remotely sensed SWI soil moisture estimates was made over catchment and footprint scales. As the satellite only provides soil moisture for the topsoil layer (first 5 centimetres), a conceptual infiltration model developed by Wagner et al. (1999c) was applied to the remotely sensed surface soil moisture estimates in order to estimate an SWI. The comparison between the modelled SSI and remotely sensed SWI was shown to be good with regression coefficients between 0.678 and 0.923. Even if a constant bias of around 19% is identified, the dynamic of the soil moisture behaviour is very coherent between the two approaches.

Discussion

Comments on the bias between the modelled and remotely sensed soil moisture

As there is no possibility of obtaining the “true” value of soil moisture at catchment scale, it is difficult to precisely assess the reasons for the bias identified between the modelled and remotely sensed soil moisture. It is clear however that the two approaches have their own uncertainties that potentially explain this bias.

Uncertainties associated with hydrological modelling

The results of the calibration showed that the values of the parameters of the TOPKAPI can be estimated *a priori* with a good reliability from information about the topography and the soil properties associated to parameter tables from the literature. The two exceptions were the channel roughness, multiplied by a factor 1.7 to get the right flow timing, and the hydraulic conductivity at saturation, which had to be multiplied by a factor of 0.6, instead by a factor of 10 as suggested by Liu and Todini (2002).

The increase of the channel roughness value n_c is clearly due to the uncertainty of the DEM that does not reflect precisely the slopes of the drainage network that are particularly flat in the lower part of the catchment. When considering the multiplying factor of K_s, one has to be aware that the values of K_s estimated *a priori* were derived from the Green-Ampt infiltration model tables that are associated with the local scale of a column of soil and for vertical infiltration fluxes. Another consideration when evaluating the calibration, might be that the production of runoff can also be due to infiltration excess mechanisms (or Hortonian processes). Such processes are indeed likely to occur especially in semi-arid areas, as in the Liebenbergsvlei catchment. The difficulty of the model to respond to observed precipitation in the beginning of the wet season is probably linked to the production of Hortonian runoff, when the soils are dried and potentially crusted and the vegetation is not fully developed. However, the assumption of the predominance of subsurface flows and the associated saturation excess runoff production seems to be realistic in the area for the major part the season. Some field experiments have been conducted at the hillslope scale in the region which confirm that saturation excess production of runoff is predominant (Colin Everson, 2007; personal communication). These experiments suggest that the basic TOPKAPI hypothesis is quite realistic on the Liebenbergsvlei catchment. It is also worth noting that a part of the reason the *a* *priori* values require some calibration could be explained by the precision of the DEM, the reader is referred to Wechsler et al., 2007 for an interesting review of the hydrological model uncertainties associated with a DEM.

Generally speaking, it is accepted that there is uncertainty in the definition of a unique set of optimal parameters. The optimal parameter set calibrated on Season 1 gives relatively poor results when used in the verification process to simulate discharges of season 2. The low performances of the model applied on season 2 can be attributed to the arbitrary choice of the initial soil moisture, the Hortonian processes that may have a longer influence for season 2 than season 1. But a more probable reason is that the calibration conducted on station 1 in season 1 only has few chances to give a robust representation of the mean behaviour of the catchment over a long period, especially on the Liebenbergsvlei catchment where rainfall and runoff are subject to a strong inter-annual variability.

In order to determine how the choice of the parameter values can influence the results of the comparison of the estimates of soil water, Figure 2.19 shows for season 1 the impact on the simulated discharge and SSI of a change in the values of the two main parameters influencing the runoff and the soil moisture production in the model. In association with Figure 2.19, Table 2.8 reports the values of the criteria characterizing (i) the model performance according to the discharge and (ii) the comparison of the simulated SSI with the remotely sensed SWI. This sensitivity analysis shows, by examining the responses, that of the parameters influencing the discharge simulation: *Ks* mainly influences the volume of runoff, while *L* mainly influences the values of the main peaks of discharge. The modelled soil moisture is also naturally influenced by the parameter values. The soil depth *L* mainly controls the variability of the soil moisture and both *L* and *Ks* influence the bias value. But, whatever the values of the parameters, the correlation between the modelled and remotely sensed soil moisture remains fair (R^2 always higher than 0.6). It means that, even if there is an uncertainty in the parameter values, the main result of the study remains unchanged.

Table 2.8 Sensitivity analysis on the effect of the parameters *Ks* and *L* on the simulated discharge and SSI. The Nash efficiency is computed on the calibration period over the subcatchment. The coefficient of determination R^2 and the regression equation are computed by comparing the simulated and the remotely sensed SWI. This table is associated with Figure 2.19 that shows the corresponding curves of simulated discharge and soil moisture. We note that the values of fac_{Ks} reported in Vischel et al., (2008a) are presented in this table (see discussion below Table 2.7).

		Nash Q period1 subcatchment	**R² SSI~SWI period 1 catchment**	**Regression line SSI~SWI y=ax+b**	
				a	**b**
$\boldsymbol{fac_{Ks}}$	20	0.600	0.780	0.81	16.4
	40	0.759	0.771	0.78	17.3
	60 (optimal)	**0.788**	0.759	0.75	18.1
	80	0.719	0.748	0.72	18.7
	100	0.605	0.737	0.69	19.2
$\boldsymbol{fac_L}$	0.6	0.227	0.629	0.88	10.3
	0.8	0.669	0.72	0.82	14.2
	1 (optimal)	**0.788**	0.759	0.75	18.1
	1.2	0.755	0.765	0.68	21.4
	1.4	0.744	0.749	0.62	24.1

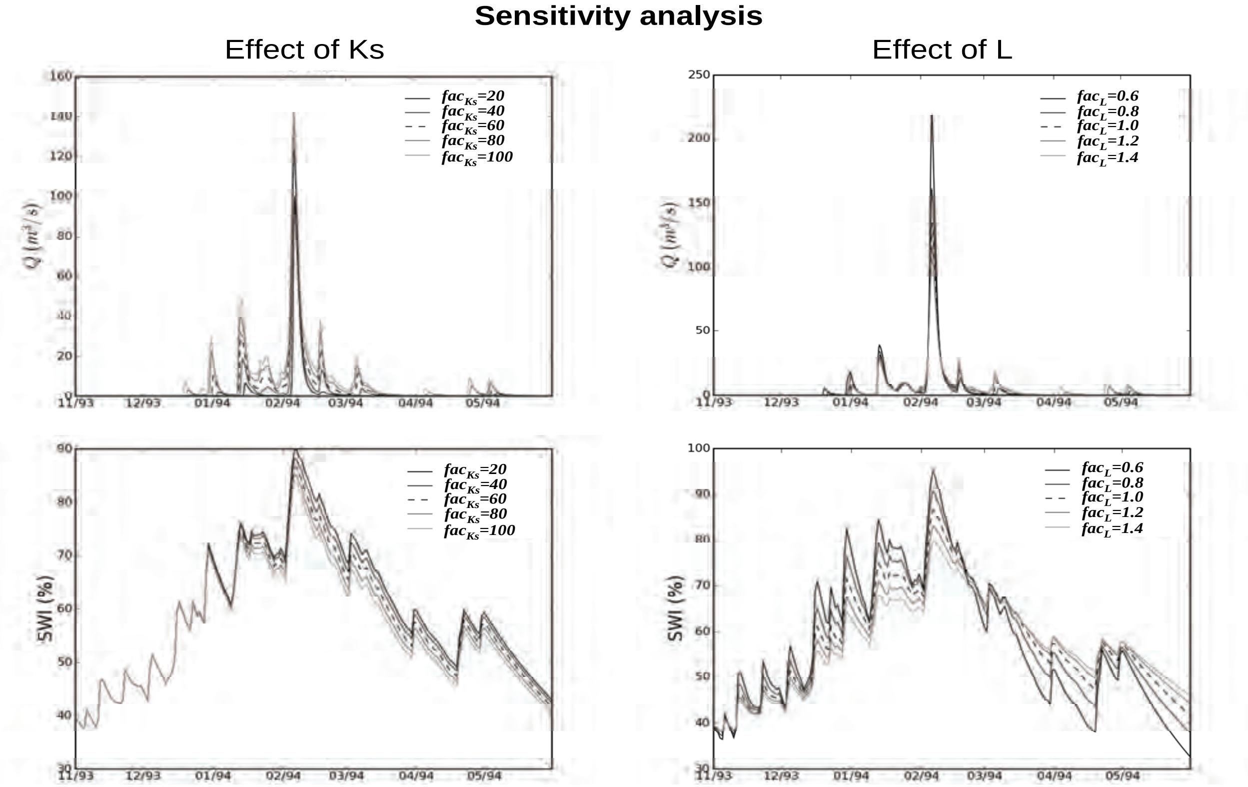


Figure 2.19 Sensitivity analysis on the effect of the parameters *Ks* and *L* on the simulated discharge and SSI. This figure is associated with Table 2.8 in which are reported the criteria characterizing (i) the model performance according to the discharge and (ii) the comparison of the simulated SSI with the remotely sensed SWI.

Comments on the good correlation between the modelled and remotely sensed soil moisture

Despite the uncertainties of each one of the approaches, that explain the bias identified in the comparison between the modelled and remotely sensed soil moisture, the study shows that there is a good correspondence in the dynamic behaviour of the soil moisture between the two independent soil moisture estimates. One might therefore question the reason for such a good correspondence. This is likely due to three main reasons:

i. The Soil Saturation Index (SSI) is considered in the present study, meaning the relative water content along the soil depth. Many studies focus on vertical transfers (Soil Vegetation Atmosphere Transfer model) but ignore the lateral transfers (horizontal subsurface flows) that occur in the soil layer and partly control the soil moisture. In the present study, the lateral transfers are explicitly modelled by the TOPKAPI model to represent the subsurface flow processes. The study shows the benefit of using a distributed hydrological model that is able to explicitly represent the horizontal transfers in the soil in order to spatially redistribute modelled soil moisture.

ii. The scatterometer estimations are very sensitive to the vegetation and are better in less vegetated regions (Wagner et al., 1999b). In the Liebenbergsvlei, the grassland and cropland surfaces are likely to result in reliable estimates of soil moisture from this source.

iii. The remotely sensed SWI comes from a conceptual infiltration model (Wagner et al., 1999c) whose parameters produce estimates consistent with what were found using a more physically based (distributed catchment modelling) approach.

iv. The raingauge network is characterized by a very high spatial density of well calibrated pluviometers that gives a reliable estimation of the precipitation amount at the catchment scale.

Perspective

The results obtained at this stage are very encouraging for (i) hydrological modelling and the possibility of using remotely sensed soil moisture to validate the models and also to initialize them; assimilation of the remotely sensed soil moisture data into hydrological models during simulations is also an exciting possibility, (ii) remote sensing and the possibility of using physically based hydrological models to validate and disaggregate the soil moisture estimations down to fine spatial scales.

Further research will aim to improve the modelling of the vertical fluxes that explicitly represent the vertical water transfers in the soil and will allow direct comparison between the remotely sensed soil moisture at the surface (first 5 centimetres of soil) without being dependent on the conceptual infiltration model used in the present study to infer the soil moisture profile from the surface remotely sensed soil moisture. Such a complete physically based model should help to better understand the processes that control the soil moisture patterns at regional scale and will be applied as a physically based soil moisture back-calculation and disaggregation tool.

3. SPATIAL ESTIMATES OF REFERENCE CROP EVAPOTRANSPIRATION (ET_0)

Introduction

Evapotranspiration is a significant component in the water balance at a range of different space and time scales but is difficult to measure directly. This is particularly important in Southern Africa, where a large proportion of the rainfall is lost through evaporative processes. Since evapotranspiration is driven by the surface energy balance (equation 3.1), its spatial distribution is determined by the spatial behaviour of the components of this energy balance and can therefore be quite complex (particularly at small space-time scales). The surface energy balance can be written as:

$$R_n = \lambda ET_a + H + G \tag{3.1}$$

Where R_n is the net radiation, H is the sensible heat flux, G is the soil heat flux, ET_a is the actual evapotranspiration and λ is the latent heat of vaporization for water. As a core forcing variable for nationwide hydrological modelling of soil moisture, a spatial grid of reference crop evapotranspiration (ET_0) estimates is produced routinely using the methodology described in Allen et al. (1998). ET_0 can be related to ET_a through the application of location and seasonally dependent land cover and water stress coefficients.

In order to test the methodology, point estimates were produced using a Penman-Monteith approach with measurements of the meteorological variables required obtained from a network of automatic weather stations. In parallel, the required variables were obtained from the South African Weather Service's Unified Model runs and an estimate of ET_0 computed for each model grid cell.

The result is automatically produced ET_0 estimates at an hourly time-step on a 0.11° square spatial grid over South Africa and a portion of the neighbouring states. The methodology and initial results are presented in this chapter, with the primary focus of this product being input to a distributed hydrological model used to compute estimates of soil moisture. However, there are many other potential uses of the spatially distributed ET_0 product resulting from this study in Hydrology, Agriculture and Disaster Management.

First we summarize the methodology for computing reference crop evapotranspiration from observed or forecast meteorological variables. Next, a description of the data sources used is given and then we present some examples of the ET_0 product computed routinely.

Methodology for Computing ET_0 from Meteorological Variables

Since it is difficult to solve the Surface Energy Balance directly for actual evapotranspiration (ET_a) without measuring all of the radiative fluxes (see for example Savage et al., 2005), the estimation at each station (or alternatively, each Unified Model grid point) uses the Penman-Monteith equations recommended in FAO-56 (Allen et al., 1998). The result is an estimate of ET_0 for a well watered (sufficient soil water to meet maximum evapotranspiration demand) reference crop defined as

> *"A hypothetical reference crop with an assumed crop height of 0.12m, a fixed surface resistance of 70sm^{-1} and an albedo of 0.23"* – Allen et al. (1998).

An implementation of the algorithm described in Allen et al. (1998) has been developed using the Python programming language. This code has been applied to process SAWS' AWS data and produce an estimate of ET_0 at each Unified Model grid point (or weather station) in hourly increments. The hourly estimates of ET_0 can easily be summed to produce a daily total, when this is required.

The 'reduced form' Penman-Monteith equation is given in equation 3.2 below.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{C_n}{T + 273} u_2 [e_s - e_a]}{\Delta + \gamma(1 + C_d u_2)} \tag{3.2}$$

Where Δ is the slope of the saturation vapor pressure versus temperature curve, R_n is the net radiation flux, G is the soil heat flux, γ is the pyschrometric constant, T is the temperature, u_2 is the wind speed at 2 m height, e_s is the saturation vapor pressure, e_a is the actual vapor pressure and the coefficients C_n and C_d vary with the aerodynamic and bulk surface resistance and are therefore specified according to the calculation time step, reference crop type (grass in this case) and, as suggested by Allen et al. (2006), the time of day.

The meteorological data required to derive the variables to solve equation 3.2 are: Air temperature, Relative Humidity, Wind Speed and Radiation, the detailed formulation is not repeated here as it's well documented in Allen et al. (1998). In this case an estimate of the solar radiation based on Meteosat data is used (LSA SAF (2006), described later in the chapter) because this variable is not measured at the weather stations. Since the SAWS weather stations and the Unified model provide wind speed at 10m height, the required 2m wind speed is estimated from equation 3.3 (Allen et al., 1998; Cai et al., 2007).

$$u_2 = u_z \frac{4.87}{\ln(67.8z - 5.42)} \tag{3.3}$$

Where u_2 is the required 2 m wind speed and u_z is the wind speed at height z.

Output from the SAWS Unified Model runs is used to compute estimates of ET_0 over the region at 12 km grid spacing. To test this approach, independent estimates were computed at each of the 164 automatic weather stations operated by SAWS.

Description of Data Sources

This section describes the sources of data used to produce the spatially distributed ET_0 estimates.

Automatic Weather Station Network

The SAWS Automatic Weather Station network provides real-time surface meteorological information via SMS at 15 minute intervals to a central data-collection facility. The network is shown in Figure 3.1, indicating the relatively sparse coverage over the country. Due to this sparse coverage the weather stations can not be used as the sole source of information for producing spatial ET_0 estimates as they are unable to efficiently sample the spatial detail.

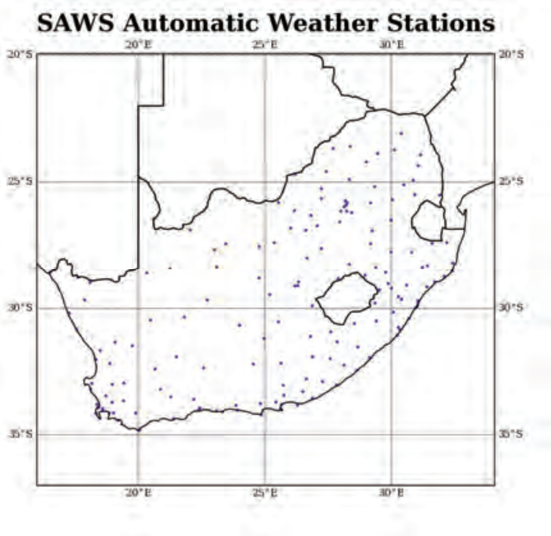


Figure 3.1 A plot showing the locations of the stations in SAWS automatic weather station network. Data from this network is used to compute ET_0 estimates at the station locations.

This network forms the core of SAWS surface data for forecasting and climate services. The SAWS recapitalization plan is designed to significantly increase the number of stations between 2008 and 2012 (pers. comm. N. Kroese, 2007). The meteorological variables measured at each station, which are relevant to the computation of ET_0, are temperature, relative humidity and wind speed. Unfortunately no radiation measurements are made at the AWSs.

Unified Model

SAWS installed the UK Met office's Unified Model (UM) in 2006, which is run at a grid resolution of 0.11° with 401 rows and 601 columns, covering Africa and the surrounding oceans south of the Equator. The bounding co-ordinates of the grid are defined by 10°W to 56°E and 0°S to 44°S, this region is shown on a square Latitude-Longitude grid in Figure 3.2. The model is run twice daily in a number of different configurations. Assimilation of observed data and the subsequent model runs occur at 02:00 and 14:00 SAST (this corresponds to 00:00 and 12:00 UTC), and *hourly* forecasts are produced out to 48 hours from each model run time. The model fields used in this study are the 02:00 analysis fields and the corresponding forecast fields out to 23 hours ahead (i.e. 01:00 the following day).

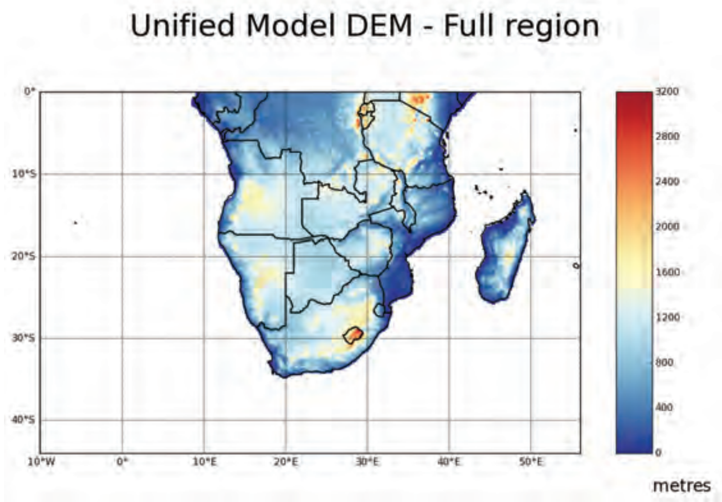


Figure 3.2 A plot of the digital elevation model used in the SAWS configuration of the Unified model, with a pixel scale of 0.11°. The figure is included to show the spatial extent of the modelled region, the bounding box of the figure is the full Unified model region.

Solar Radiation

Since it is difficult to find automated surface radiation observations at an hourly (or even daily) frequency, solar radiation estimates based on Meteosat data are used. The data products are produced by the Land Surface Analysis Satellite Application Facility (LSA SAF http://landsaf.meteo.pt) in Lisbon and are disseminated in real time at 30 minute intervals via the EUMETCast system available at SAHG-UKZN (Pegram et al., 2006). The advantage of this product over sparse surface observations is that a detailed spatial coverage is available over large areas at frequent intervals. Figure 3.3 shows a typical map of the estimated solar radiation flux for Africa, south of the equator. Figure 3.4 shows a spatially resampled version of the data in Figure 3.3. The information has been resampled onto a subset of the SAWS Unified model grid. The resampling method used is a nearest neighbour technique, because it is not yet clear how the smoothing effects typically introduced by more complex sampling methods will affect the result, especially near the boundaries of clouds, whose positions are coloured yellow through to blue, indicating various degrees of occlusion.

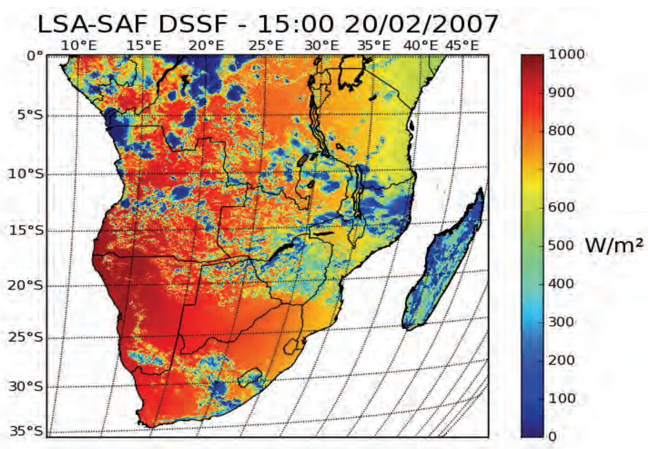


Figure 3.3 An estimate of the down-welling surface shortwave radiation flux (W/m^2) from the LSA-SAF. The data products are produced at 30 minute intervals and disseminated via the EUMETCast system. The data are stored in HDF5 format and produced in the Geostationary Satellite co-ordinate system at the nominal Meteosat resolution.

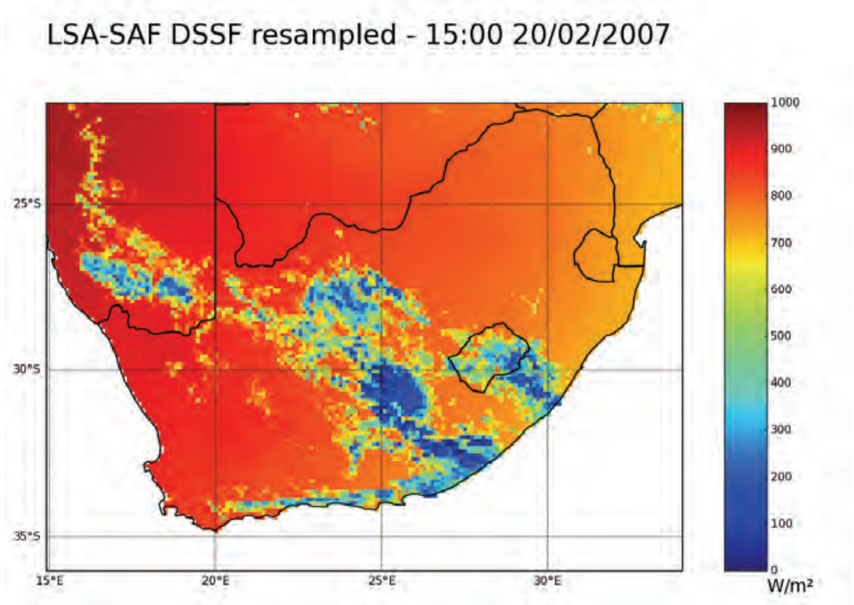


Figure 3.4 A spatially resampled version of the LSA-SAF DSSF product shown in Figure 3.3. The estimates of solar radiation flux have been resampled onto a regular Latitude-Longitude grid matching that of the SAWS configuration of the Unified Model (approximately 12 km pixels). A nearest neighbour sampling technique has been used as it is not clear how smoothing by more complex resampling techniques will affect the result, especially in the vicinity of cloud boundaries.

To the best of the authors' knowledge, the LSA-SAF product has only been validated under European conditions, (LSA SAF, 2006) where it was shown to be unbiased, and it's applicability in Southern Africa is not known. As an exercise to increase confidence in the estimates, a basic comparison with some observed data was carried out.

A time series of solar radiation data collected from the CSIR study catchments (30.67°S, 29.19°E), situated at the Mistley-Canema Estate (Mondi Forests) in the Seven Oaks district, approximately 70 km from Pietermaritzburg, was obtained. These observed data were compared with the LSA SAF estimates at the same location and times. Figure 3.5 and Figure 3.6, show comparisons between the measured and estimated solar radiation over a 5 day period between the 20th and 24th of February 2007. This initial test of applicability is very encouraging, with the coefficient of determination (R^2) for the best fit linear regression line (Figure 3.6) being equal to 0.918. The scatter from the line are points relating to the intermittent cloud cover in the last 3 days of the sequence; timing is a possible contributing cause.

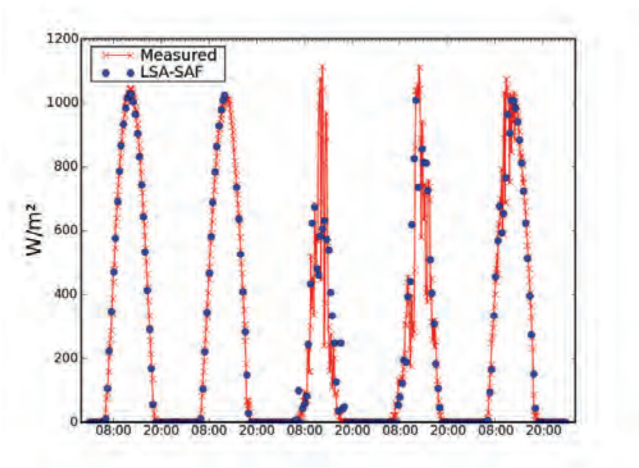


Figure 3.5 A 5-day comparison between the DSSF product (blue dots at 30 minute intervals) and measured incoming solar radiation (red lines and crosses at 10 minute intervals) at the CSIR site. There is evidence of a slight timing mismatch that is due to Meteosat’s scan strategy, a closer match is expected if this mismatch is accounted for.

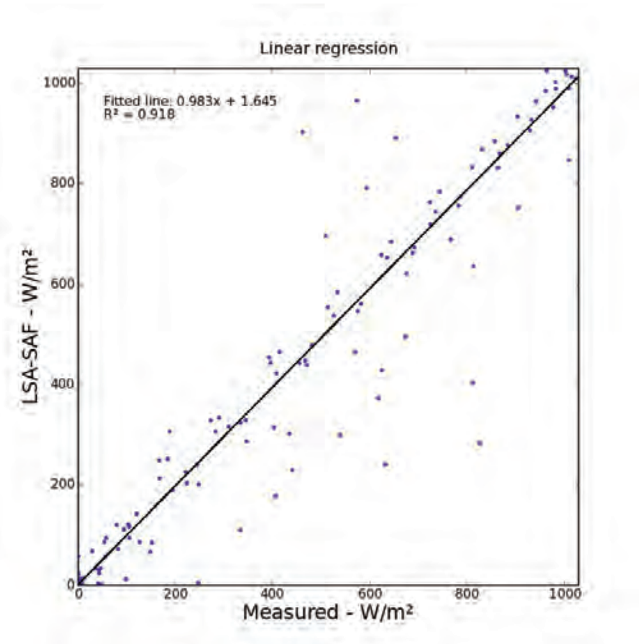


Figure 3.6 A linear regression between the DSSF product and measured incoming solar radiation at the CSIR site for 5 days of data. The coefficient of determination for the regression (R^2 = 0.918) indicates strong agreement between the estimated and measured radiation flux. Although this is a short sample of data at a single location, the results are encouraging.

Spatially Distributed Estimates of Evapotranspiration

In Figure 3.7 a typical map of ET_0 calculated at an hourly time step is shown. The estimate of ET_0 is computed at each grid point for a subset of the Unified model field. In this case (12:00 SAST) a 10 hour ahead forecast of the meteorological variables has been used along with the corresponding LSA SAF radiation estimate. A linear regression through the origin is shown in Figure 3.8. The regression compares the ET_0 computed from station observations at each station in the SAWS network (Figure 3.1) to the ET_0 computed at the nearest Unified model grid point in the map shown in Figure 3.7. Since it is well known that the R^2 value is reduced for regressions through the origin (Gordon, 1981) and is in fact disputed as a measure of the goodness of fit by some (Eisenhauer, 2003), the Pearson correlation coefficient is also given for comparison.

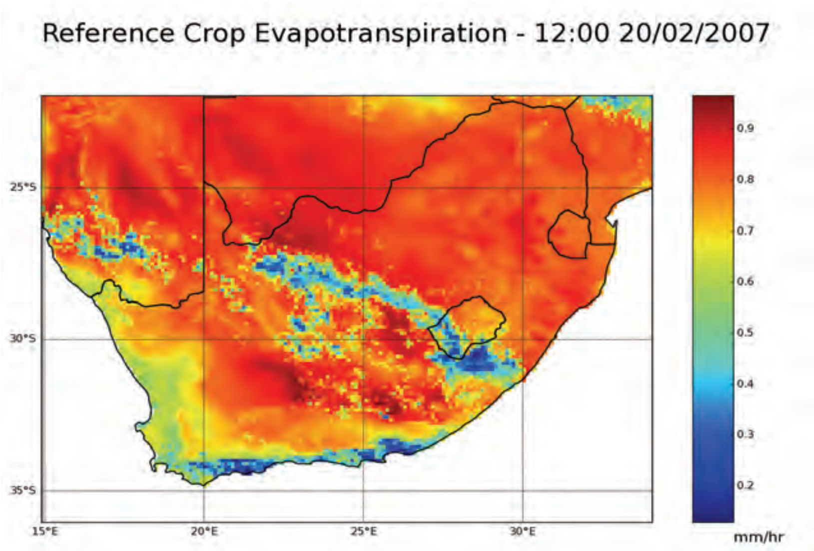


Figure 3.7 A typical map of hourly reference crop evapotranspiration computed from Unified Model forecast data and resampled LSA SAF solar radiation estimates. The pixel resolution is 0.11° and the ET_0 estimates are plotted on a regular Latitude-Longitude grid.

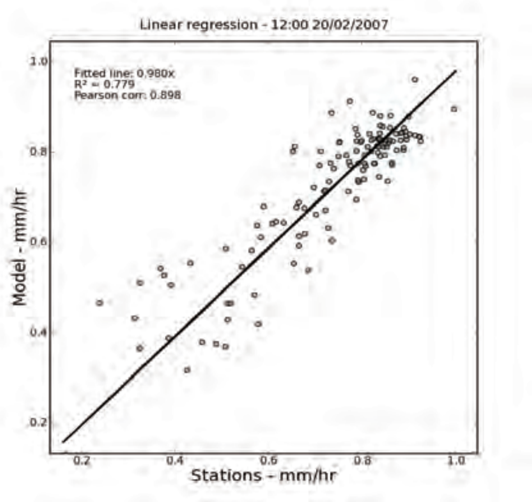


Figure 3.8 A typical linear regression, through the origin, between the hourly reference crop evapotranspiration computed from Unified Model forecast and LSA SAF radiation data (Figure 3.7) and ET_0 computed from the closest SAWS automatic weather station. The two calculations use the same estimate of incoming solar radiation R_S but all other inputs to the FAO56-PM equation are independent. $R^2 = 0.779$.

The map of daily ET_0 shown in Figure 3.9 is produced by summing the hourly estimates (e.g. Figure 3.7) over a 24 hour period. In this case the map has been produced for the 22nd of February 2007. The hourly ET_0 computed from the SAWS station observations have also been accumulated to daily totals and Figure 3.10 shows these compared to the spatial estimates in the same way as was done to produce Figure 3.8. The strong correlations between the station and spatially distributed ET_0 estimates at the station locations indicates that the spatial estimates based on Unified model forecasts reproduce the ET_0 well at many locations throughout South Africa. There are some outliers, for which an explanation is still sought.

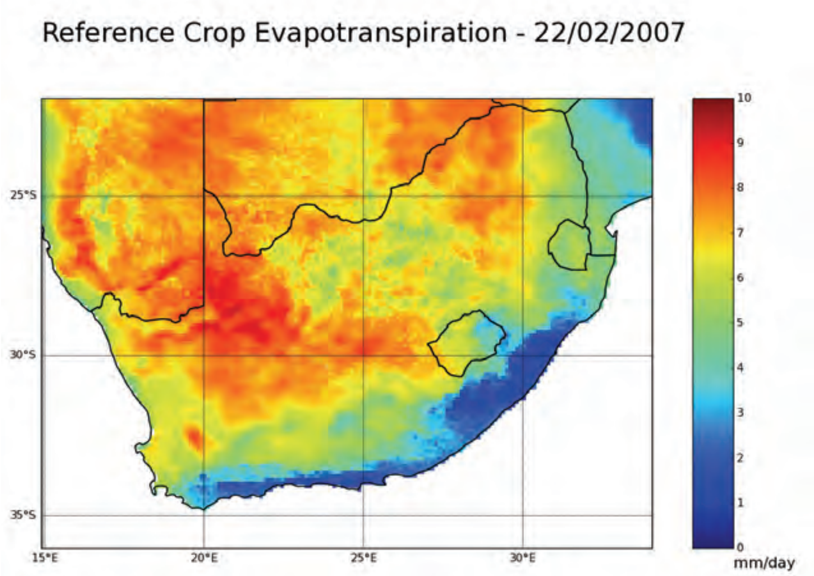


Figure 3.9 A typical map of the daily total reference crop evapotranspiration computed from Unified Model forecast data and resampled LSA-SAF solar radiation estimates. The pixel resolution is 0.11° and the ET_0 estimates are plotted on a regular Latitude-Longitude grid. This map is produced by accumulating the hourly ET_0 estimates for a 24 hour period.

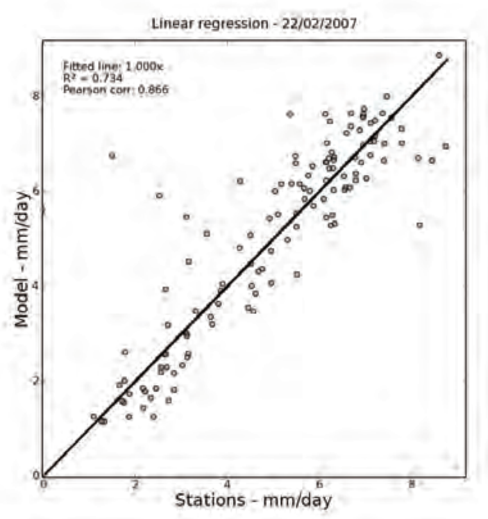


Figure 3.10 A typical linear regression, constrained through the origin, between the daily total reference crop evapotranspiration computed from Unified Model forecast data (Figure 3.9) and ET_0 computed from the closest SAWS automatic weather station. The two calculations use the same estimate of incoming solar radiation R_S but all other inputs to the FAO56-PM equation are independent. $R^2 = 0.734$.

Summary

This chapter has presented a brief explanation of the methodology employed for producing spatially distributed estimates of reference crop evapotranspiration using the guidelines set out by Allen et al. (1998). Novel use has been made of remotely sensed radiation estimates and forecast fields from a numerical weather prediction model. Initial results show that the correspondence in ET_0 estimated from the model forecast fields are strongly correlated with ET_0 calculated from weather station observations at both the hourly and daily timescales. This builds significant confidence in the Unified Model forecasts produced by SAWS, as well as suggesting that spatially distributed ET_0 estimation is viable, using this model data.

4. USING TOPKAPI IN LAND SURFACE MODEL MODE TO ESTIMATE SM AND ACTUAL ET

Scientific Background and Methodology

The soil moisture content in a 1 km square area surrounding each SM gauge-site is modelled by a TOPKAPI cell, which is a single layer store, with surface and soil properties defined by maps, stored in a GIS model. In this limited application, the TOPKAPI cell acts as a Land Surface Model (LSM) similar to those used in the assimilation of evapotranspiration and energy fluxes into Meteorological models of the atmosphere (e.g. Rodell et al., 2004). The following development in this sub-section is adopted and modified from Teuling et al. (2008) and sets the stage very well for the work presented here.

For a single-layer LSM, with effective root-zone depth L between 600 and 1000 mm, the rate of change of the water content of the soil θ (mm/mm) or the Soil Saturation Index (SSI) s, which is the degree of saturation as a ratio of θ - θ_r to the porosity θ_s - θ_r (where θ_r and θ_s are the residual and saturated soil moisture contents), can be expressed as:

$$\frac{ds}{dt} = \frac{1}{L\theta_s}[P - I - R - ET - Q] \tag{4.1}$$

where P is precipitation, I the rate of interception by vegetation, R the saturation excess runoff, ET the total actual evapotranspiration (through root water uptake and soil evaporation), Q the drainage from the base of the soil profile, typically all in mm/day.

In their investigation, for both arid and moderate climates (much of RSA), I and Q were found to be relatively negligible and R only occurs when the soil store is saturated – these are the same assumptions as are made in the TOPKAPI specification. The sensitivity of the SM to the depth of the root zone was found in their study to be relatively insignificant, the most important parameters being:

the soil parameters:

- θ_w the permanent wilting point
- θ_c the critical moisture content below which stomatal opening is reduced due to soil moisture stress
- θ_s the saturated water content (in more humid environments)

the evaporation drivers:

- LAI (leaf area index which is vegetation and season dependent)
- $r_{s,min}$ (unstressed stomatal resistance)
- Δq the specific humidity gradient between the surface and the lowest atmospheric level

and the mean frequency and depth of precipitation

Teuling et al. (2008) chose to model evapotranspiration directly as a function of the drivers mentioned above. The approach followed in TOPKAPI is to treat the evapotranspiration as a model forcing variable and therefore assume that it has been estimated separately. This allows for some simplification of the model and provides the flexibility to use appropriate and well verified ET products.

The soil parameters are defined in Figure 4.1 (after Teuling et al., 2008).

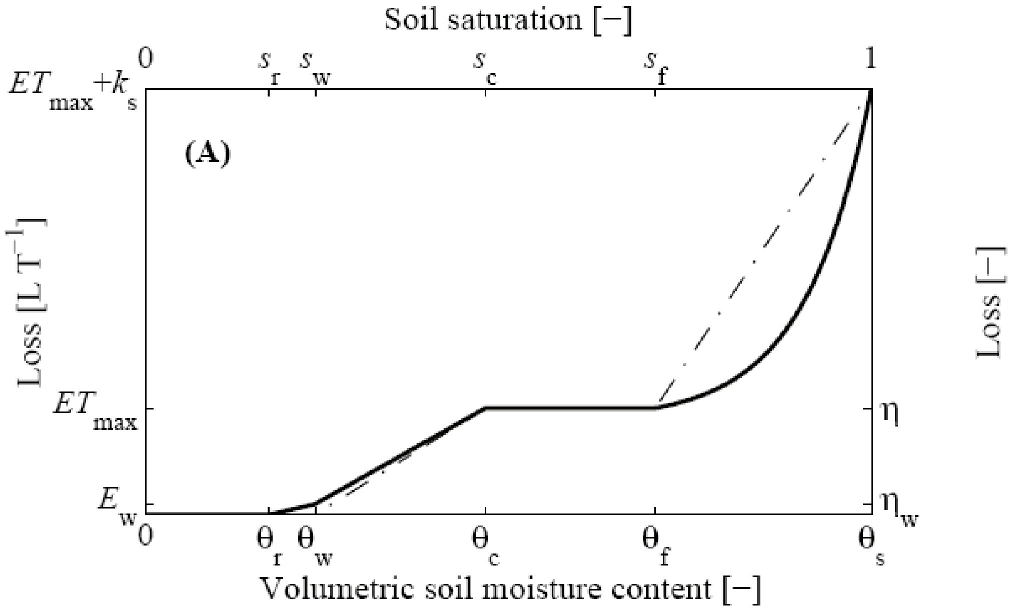


Figure 4.1 Loss functions and associated variables for the regular (bottom/left) and dimensionless notation (top/right) and for linear (dash-dotted line) and nonlinear (solid line) drainage. k_s is the saturated conductivity, which may be orders of magnitude greater than ET_{max}; θ_r and θ_f are residual soil moisture and field capacity respectively; other symbols are defined above. [Teuling et al. (2008)]

In our application of TOPKAPI cells as LSMs in this study, we use the linear assumption, supported by the observation that it makes very little difference to the results when compared to the non-linear model (Teuling et al., 2008).

In summary, our strategy here was to obtain as good a real-time rainfall input as is available, exploit the rich source of mapped soil parameters and use the best estimates of ET that are available both operationally and countrywide.

It is important to note that the infiltration behaviour of the TOPKAPI model is unaffected by the intensity of the rainfall (due to the absence of an infiltration layer). It is only the depth of rainfall during an interval that matters as overland runoff is initiated at saturation, making TOPKAPI highly suited to the TRMM rainfall product which has a pixel scale of approximately 25 km at our latitudes (and therefore cannot capture the spatial detail of rainfall intensities at a fine spatial scale). In addition, because the ground-water recharge rate is very slow in arid and moderate climates, we have neglected the vertical drainage Q and have ignored the interception I as a second-order effect. Using Q to represent lateral drainage due to the ground slope in each cell, the continuity equation thus reduces to:

$$\frac{ds}{dt} = \frac{1}{L\theta_s}[P - R - ET - Q] \tag{4.2}$$

Model forcing data

Rainfall

There are numerous freely available rainfall datasets for use by the scientific community, but their quality and the ease of obtaining these datasets is highly variable. We went in search of suitable real-time, online, high frequency rainfall products with a valid coverage at fine spatial scale over Southern Africa. We were naturally led to out research partners in the University of Cape Town's Oceanography Department who are working on hydrological response to climate variability – the MAHYVA project: WRC K5/1747. In the context of examining historical time series to detect links between ENSO and Southern African rainfall, Dr Mathieu Rouault and his team evaluated the following freely available datasets: C INTERP, SAWS, CRU TS 2.1, CHARM , TRMM 3B42, GPCP version 2, ERA 40, & GPCC version 4

However, the goals of the MAHYVA study (a historically focused desk-top study) are somewhat different to ***the goals of this project*** where the focus is on ***producing an operational prototype*** of a system that will be capable of producing ***automatically updated soil moisture maps over the regio****n* using a combination of information from many disparate sources. It is important therefore to ensure that the rainfall forcing data is available at sufficiently short time steps to capture the relevant dynamics and that it is reliably available for operational use.

All of the products evaluated in the MAHYVA study are daily time-step products except for TRMM 3B42, which is 3 hourly and ERA 40 which is available at 6-hour time-steps but only available historically up until 2001.

The search for a suitable rainfall product

A wider search was therefore made and a selection of the ***currently available*** (near real-time with sub-daily timesteps) rainfall products considered ***in this project*** are:

NRL Monterey – http://www.nrlmry.navy.mil/sat-bin/rain.cgi?GEO=g25

> Blended satellite IR and passive Microwave rainfall estimates, providing 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 who is well published and respected in the field of satellite rainfall estimation).
> ***Conclusion – Does not appear useful, beyond qualitative comparisons with other products. Not suitable.***

PERSIANN – http://chrs.web.uci.edu/research/satellite_precipitation/activities00.html

> Geostationary satellite IR based rainfall rates, updated with rainfall estimates from low earth orbiting satellites (TRMM etc.) using a neural network model. Spatial resolution at 0.25° and a temporal resolution of 30-mins provided as 6 hourly accumulations.
> ***Conclusion – Data access appears to require manual downloads. Not suitable.***

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

> Combines precipitation estimates from low earth orbiting microwave satellite platforms using geostationary IR data to provide motion fields describing the movement of

rainfall 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. Also available in 8 km and 30 minute resolution.
Conclusion – Rejected due to the longer time-lag relative to TRMM 3B42RT. Not suitable.

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.
Conclusion – Rainfall forecasts are typically poor quality with large uncertainties. Not suitable.

Global Rainfall Map – http://sharaku.eorc.jaxa.jp/GSMaP/
Combination microwave and IR rainfall estimates in near real-time, only available as maps (images). Historical data up to 2006 available for download. ***Conclusion – Not suitable.***

TRMM 3B42 V6 –
http://daac.gsfc.nasa.gov/precipitation/TRMM_README/TRMM_3B42_readme.shtml
TRMM Multi-Satellite Precipitation Analysis, 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. A subset of this data is available in NetCDF3 format on the SAHG data server for the period January 2002 to April 2007 and for a limited region including South Africa, due to our collaboration with the MAHYVA project.
Conclusion – Useful product for historical analyses/case studies as suggested by the work under MAHYVA. Not suitable.

TRMM 3B42-RT – ftp://trmmopen.gsfc.nasa.gov/pub/merged
TRMM Real-Time Multi-Satellite Precipitation Analysis, combined microwave and microwave calibrated IR estimates of precipitation rate. Spatial resolution is 0.25° and 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.
Conclusion – Most useful product for routine rainfall estimates. Best available.

After evaluation, TRMM 3B42-RT was the product we decided to use.

The chosen near real-time precipitation data stream

We adopted the TRMM 3B42RT product for the purposes of SM estimation using TOPKAPI in LSM mode. Figure 4.2 shows an example of the instantaneous rainfall rate provided by 3B42RT at 18:00 on the 11th of January 2009.

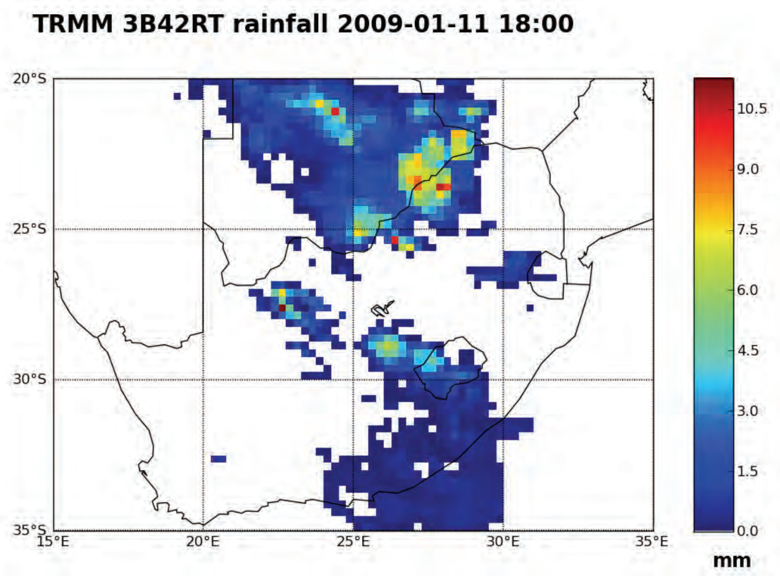


Figure 4.2 The TRMM 3B42RT rainfall estimate at 18:00 on the 11th of January 2009. Note the relatively large 0.25° pixels.

Evapotranspiration

In Chapter 3, we developed a reference crop evapotranspiration (ET_0) product following the methodology of Allen et al., (1998), for the purpose of driving the ET budget in TOPKAPI. ET_0 provides a measure of the evapotranspiration expected from a well watered grass crop as a result of the prevailing environmental conditions (temperature, radiation, wind, humidity). The conventional way of using this product is to reduce the potential using a Crop Factor, which is a seasonal ratio, depending on the local vegetation type and its transpiration potential, in turn dependent on its Leaf Area Index (LAI). It would be very helpful if there was a spatially distributed ET_a product, suitable for operational use, available for free at the time and space scales appropriate for this study; however, such an online product is not available at this time.

Nevertheless, as part of the information gathering done for this study, we exploited conversations held with Francoise Gellens-Meulenberghs, a Belgian scientist at EGU in Vienna in 2007 and again in 2008, who expressed a desire to collaborate with us in evaluating a potentially useful product. As a result of these conversations, Dr Sinclair attended the 2nd EUMETSAT LSA SAF workshop in Portugal for a week during June 2008; the primary goal being to make contact with the team from the Belgian Meteorological institute (KMI) which is one of the LSA SAF consortium members responsible for the development of an operational ET_a product. We have been offered the Beta version of the Africa product, and look forward to working with this in the future. In the interim, we have chosen to force the TOPKAPI model with the ET_0 product we discussed in Chapter 3 (using suitable crop and water stress factors to account for the differences between ET_0 and ET_a).

Reference Crop Evapotranspiration - 08/09/2008

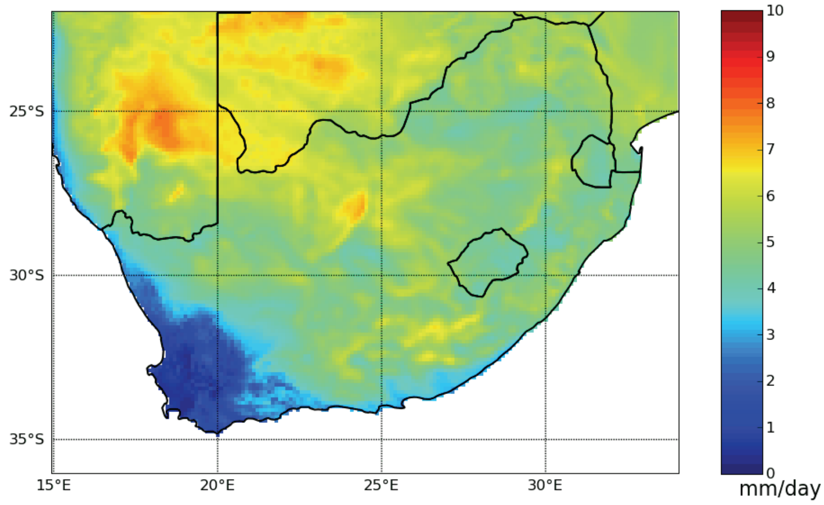


Figure 4.3 An example of a daily ET_0 map posted on the SAHG website. These daily totals of ET_0 are calculated by summing the hourly estimates that are automatically produced and archived; the product was developed as described in Chapter 3.

Estimating the required TOPKAPI parameters

Each cell in the TOPKAPI model is 1 km square in plan and is as deep as the active soil layer (approximately 1 m 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:

<u>for the soil store:</u> local slope of cell, depth, saturated soil moisture, residual soil moisture and soil conductivity

<u>for the overland flow store:</u> local slope of cell, surface roughness.

These were available for the Liebenbergsvlei study and are shown in Figure 4.4. However, there was a need to set up a procedure for determining the parameters at “randomly” scattered sites by creating a country-wide data base for TOPKAPI pixels.

For these purposes, we use the SIRI mapped data to obtain the local soil depth and saturated soil moisture and the WR90 soil texture data set to obtain residual soil moisture and saturated conductivity, as before. However, these are not ideal for our purposes, so we will be looking out for better in future. The reasons are that the SIRI mapped data misses former homeland states and the WR90 soil texture data set is very coarse, but these sets are all we have at this time. We note however that the ARC apparently has a newer soil map which we will investigate later. Figure 4.4 itemises the data-sets. The cross in the figure indicates that we do not need channel roughness in the LSM version of TOPKAPI, but only in catchment mode.

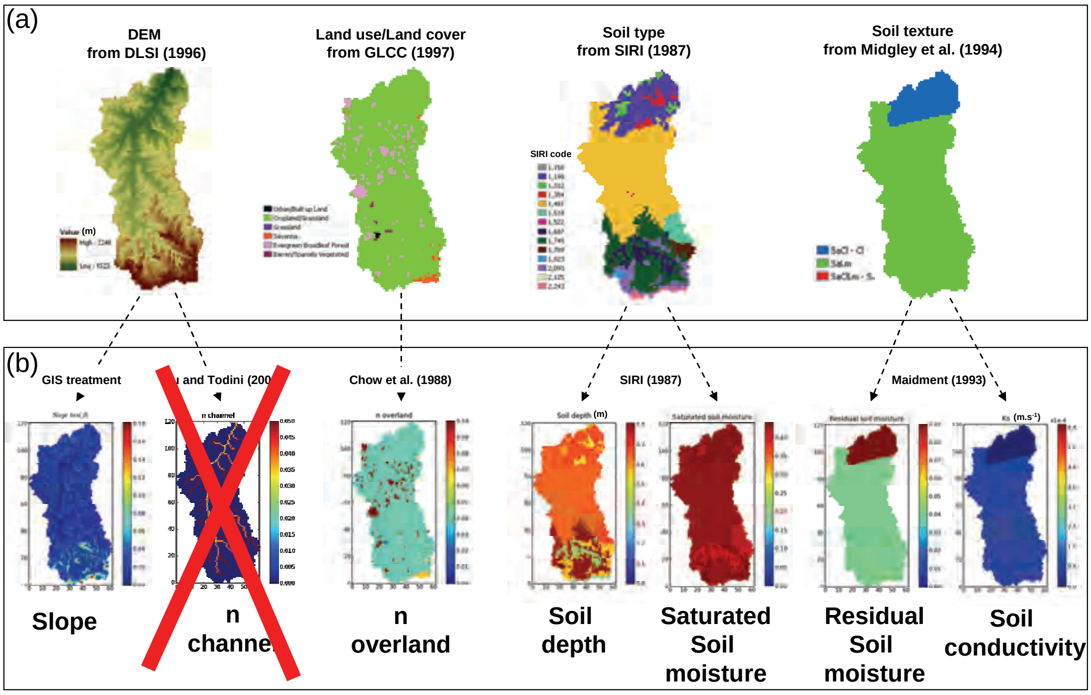


Figure 4.4 A diagrammatic overview of the methods used to compute the *a priori* physical parameters for TOPKAPI. This figure is based on the work carried out for Liebenbergsvlei (see Vischel et al., 2008).

Because the originally adopted USGS digital elevation model (DEM) indicated in Figure 4.4 was found to be flawed and needed repair every time a new area was chosen for study, a new DEM was sought and found. This is a 90 metre DEM covering all the land surface of Earth; it was used to create a new 1 km DEM for our purposes, as described in the next section.

These new data-sets put a fresh demand on the code written specifically for hydrological modelling of the Liebenbergsvlei. To obtain the cell properties on every square kilometre in the country required a complete overhaul of the data library files. This impacted the specification of the ***TOPKAPI model,*** which was ***extended and partially re-written*** to allow it to model individual cells as LSMs and also to model the hydrological response of catchments.

The following subsections explain the data-collection methodology for hillslopes, overland roughness and soil properties in more detail.

Determination of hillslopes

A high quality digital elevation model (DEM) forms the basis on which the TOPKAPI model is built. The DEM provides a definition of the location, size and shape of the model cells as well as the hill slopes and delineation of the drainage network. The original intention was to use a 1 km resolution global DEM (HYDRO1K) to specify the TOPKAPI model cells. However, in earlier work on the Liebenbergsvlei by Parak (2007) and Vischel et al., (2008) it became clear that HYDRO1K was not sufficiently detailed to reliably reconstruct the drainage network and a more detailed DEM was adopted to develop a suitable 1 km DEM for the model. Since the chosen DEM will form the basis for future work (where stream delineation is important) and because the 200m DEM adopted in these earlier studies is not available over the entire region, a suitable freely available alternative to HYDRO1K was sought. The DEMs considered are described below:

GTOPO30 – http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html

This is the USGS global DEM at a grid spacing of 30" of arc. Projection is Cylindrical (lat-lon) referenced to the WGS84 datum.

Conclusion – Not Hydrologically correct. Not suitable.

HYDRO1K – http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html

A suite of Hydrological useful derivative products based on GTOPO30. The basis of the derivative products is a Hydrologically correct DEM in a Lambert Azimuthal Equal Area projection with a cell size of 1000m. The entire HYDRO1K data-set is available on the SAHG data server.

Conclusion – Difficulties delineating drainage network in previous studies. Not suitable.

GLOBE – http://www.ngdc.noaa.gov/mgg/topo/gltiles.html

1 km spatial resolution, however internet sources indicate some issues with data quality.

Conclusion – Not suitable.

SRTM 90m – http://srtm.csi.cgiar.org/

This is a post-processed version of the Shuttle Radar Topography Mission (SRTM) DEM data-set (http://srtm.usgs.gov/) to provide a complete hydrologically sound DEM at a grid spacing of 3" of arc. Projection is Cylindrical (lat-lon) referenced to the WGS84 datum horizontally and the EGM96 vertical datum. Vertical accuracy is to within 16m. Data for a set of 5°square tiles covering South Africa is available in GeoTIFF format on the SAHG data server.

Conclusion – Hydrologically correct high spatial resolution DEM. Suitable for upscaling and model development.

The adopted 90m DEM with global cover (Jarvis, 2008), based on the output of the Shuttle Radar Topography Mission (SRTM), was chosen as the primary data required for TOPKAPI in the Southern African region. Figure 4.5 illustrates the base 90 m DEM, while Figure 4.6 shows typical slopes calculated from a (1 minute of arc pixel) up-scaled version of the base DEM.

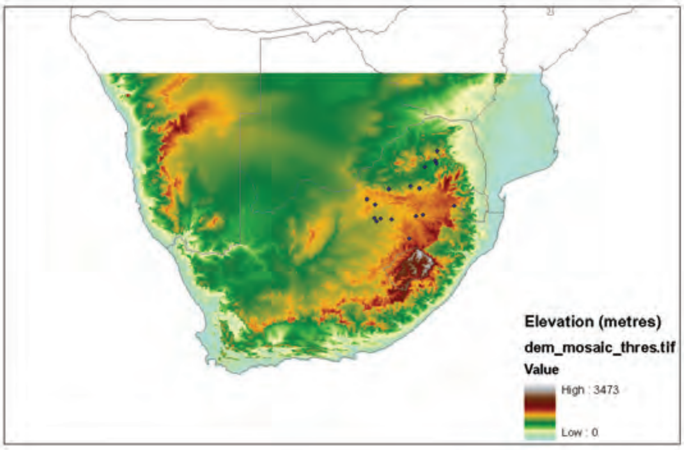


Figure 4.5 A 90m resolution digital elevation model (Jarvis, 2008). The elevation model is based on SRTM data and has been filled to remove sinks and produce a hydrologically correct DEM. The elevations indicated in the colour scale are in metres; the dark blue points are the locations of 17 soil moisture probes operated by the ARC that were discontinued.

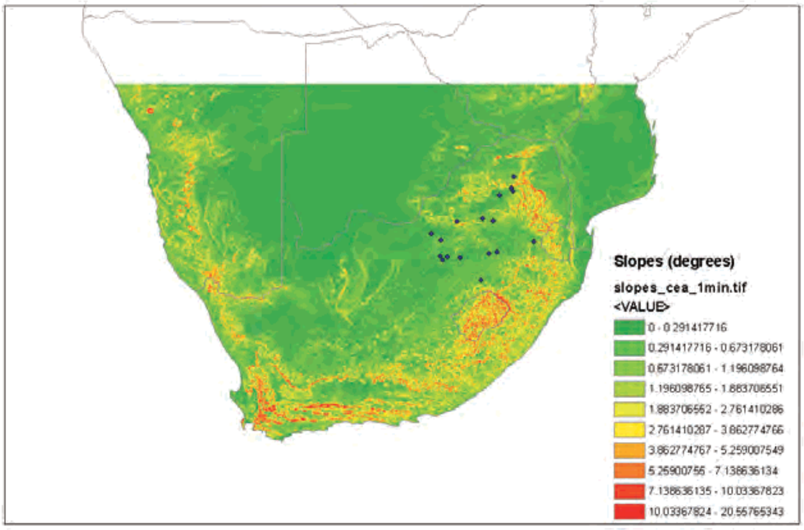


Figure 4.6 A regional map of slopes computed from an up-sampled version of the high resolution 90 m DEM shown in Figure 4.5. The area of each pixel in this image is roughly 3.0 km^2, chosen to approximate a pixel size of 1 minute of arc on an Equal Area projection. The dark blue points are the locations of 17 soil moisture probes operated by the ARC that were discontinued.

Overland roughness coefficients

In order to estimate the Manning roughness coefficients for overland flow (n_o) in each cell, a global dataset of landcover classifications has been used (GLCC, 1997). The dataset is provided separately for each continent and a subset of the African continent data has been processed. Tables from Chow (1988) relating landcover to n_o were digitized to allow a direct translation of the landcover classes for use in the TOPKAPI model. Figure 4.7 illustrates the extent of the region for which this work was carried out, while Table 4.1 shows the description of the landcover classes in Figure 4.7.

Table 4.1 GLCC landcover classifications (see Figure 4.7 for a map of Southern Africa).

Class ID	Description
1	Urban and Built-Up Land
2	Dryland Cropland and Pasture
3	Irrigated Cropland and Pasture
4	Mixed Dryland/Irrigated Cropland and Pasture
5	Cropland/Grassland Mosaic
6	Cropland/Woodland Mosaic
7	Grassland
8	Shrubland
9	Mixed Shrubland/Grassland
10	Savanna
11	Deciduous Broadleaf Forest
12	Deciduous Needleleaf Forest
13	Evergreen Broadleaf Forest
14	Evergreen Needleleaf Forest
15	Mixed Forest
16	Water Bodies
17	Herbaceous Wetland
18	Wooded Wetland
19	Barren or Sparsely Vegetated
20	Herbaceous Tundra
21	Wooded Tundra
22	Mixed Tundra
23	Bare Ground Tundra
24	Snow or Ice
99	Interrupted Areas (Goodes Homolosine Projection)
100	Missing Data

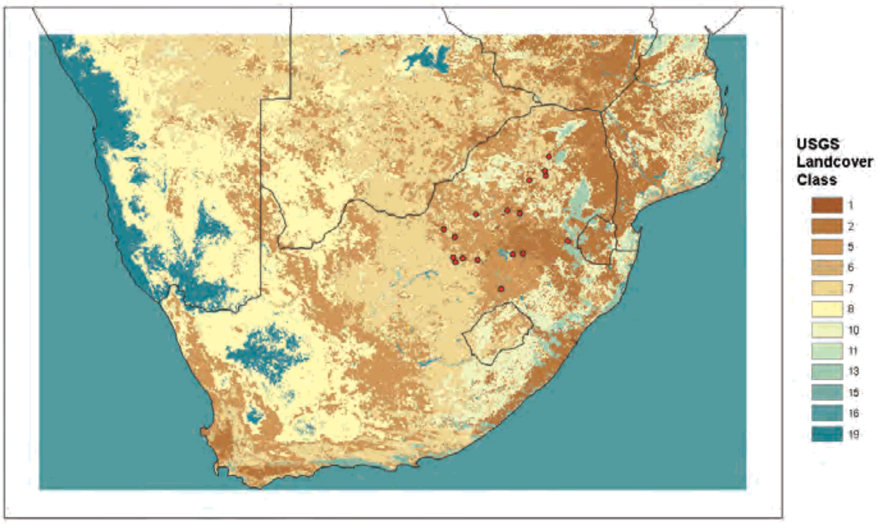


Figure 4.7 A regional map of land cover classes derived from the GLCC Global Landcover dataset. A description of the classes is given in Table 4.1.

Soil properties

The four soil properties required by the TOPKAPI model have been estimated from SIRI (1987) and Midgley et al. (1994), following the work of Vischel et al. (2008a, 2008b). The dataset from SIRI (1987) provides soil depths and porosity (interpreted here as saturated soil moisture content θ_s) for the A and B soil horizons. The soil depth used in the TOPKAPI model has been taken as the sum of the depths for the A and B horizons while θ_s is taken to be the average value of the quoted values for the A and B horizons. Figure 4.8 shows the SIRI (1987) classes for South Africa. Of concern are the gaps in the data for the former Transkei and Ciskei. In follow up work it is anticipated that a more recent (and complete) soil dataset from the ARC might be used, after it has been evaluated. Figure 4.9 shows the soil depths calculated from the SIRI (1987) data. We note that the mid-range of the depths of all soils in RSA is 575 mm and that the ARC probes (shown as dots in Figure 4.8 through Figure 4.10, following) are apparently sited in slightly deeper soils on average. Figure 4.10 is a map of the WR90 soil classes (Midgley et al., 1994). These soil classifications are used to determine soil texture information and hence to estimate conductivity and residual soil moisture (θ_r) from tables presented in Maidment (1993).

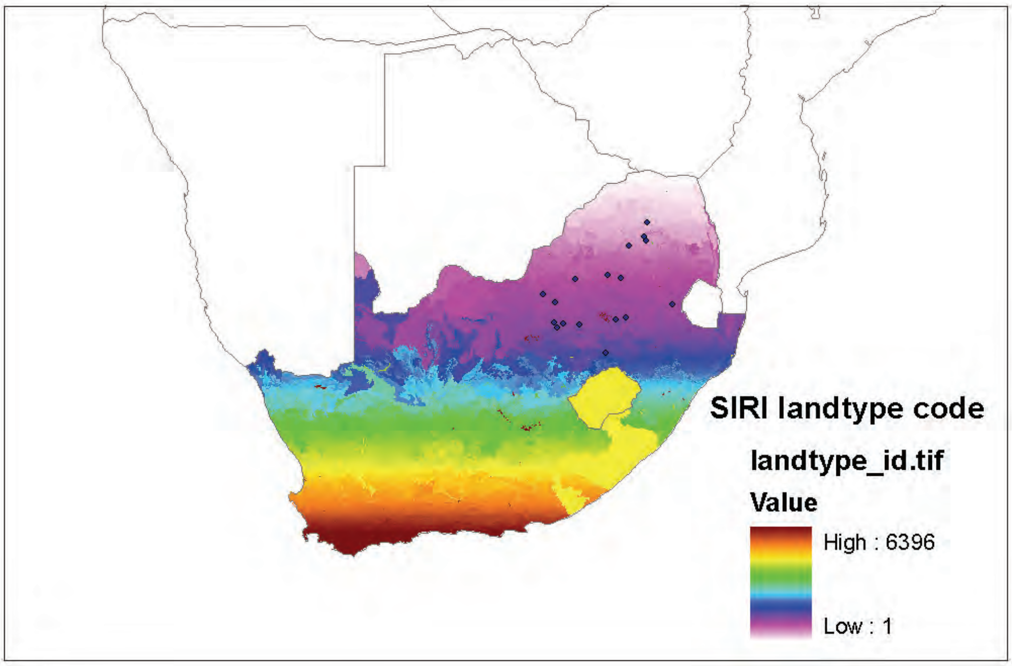


Figure 4.8 A map showing the distribution of the SIRI (1987) landtype classifications. The strong North-South stratification is due to the large number of factors used for classification rather than a stratification of the properties at each location (see Figure 4.9 for example).

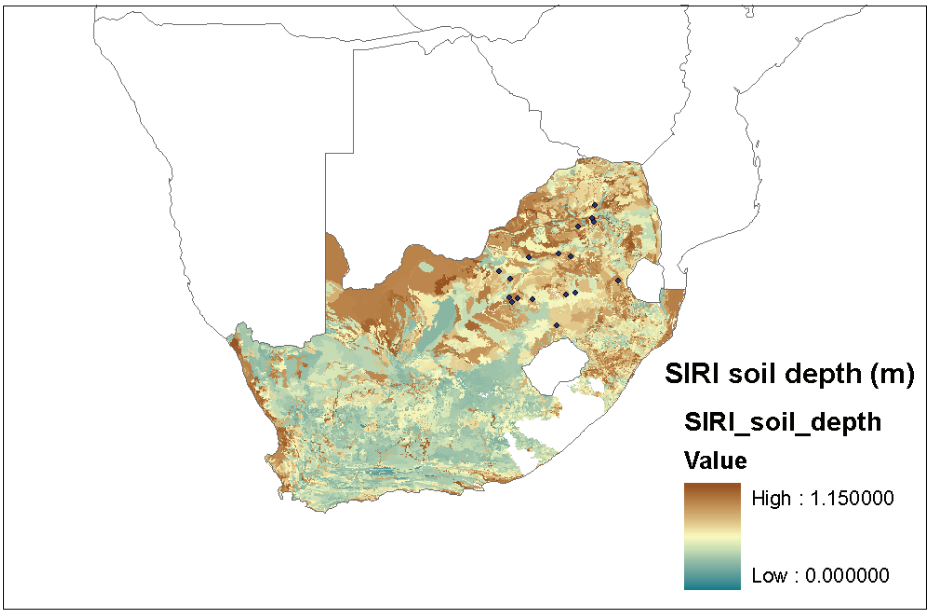


Figure 4.9 A map of combined A and B horizon soil depths calculated from the information contained in the SIRI (1987) soil types database. The locations of the 17 ARC soil moisture probes indicated by dots lie in soils with median depths near 700 mm, near the middle of the range of all soils.

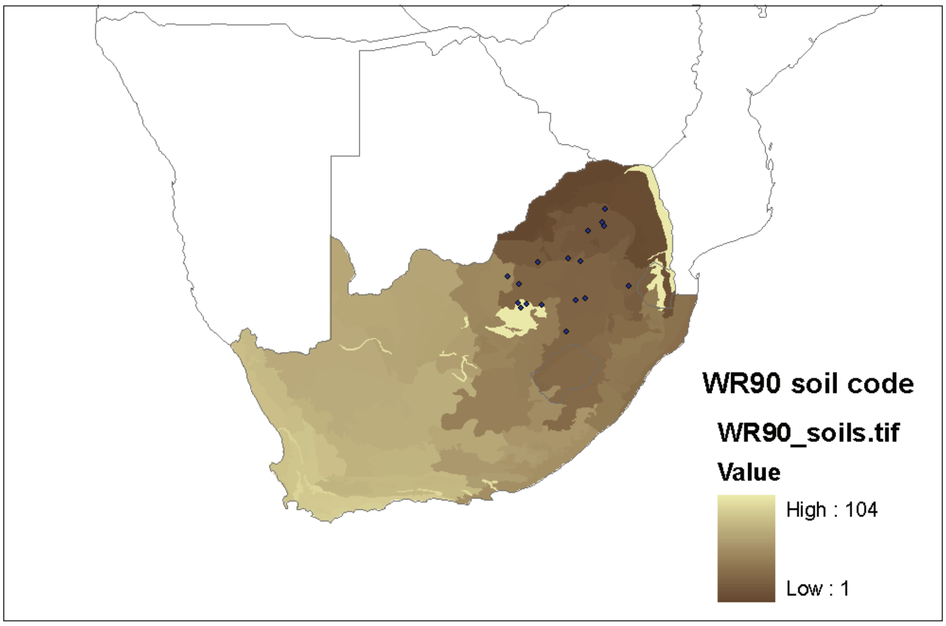


Figure 4.10 A map of the WR90 soil classes (Midgley et al., 1994). These soil classifications are used to determine soil texture information and hence to estimate conductivity and residual soil moisture (θ_r) from tables presented in Maidment (1993).

Putting it all together

The work described above entailed the following:

Developing the ET_0 spatial product on the SAHG website

- by combining the Unified Model data obtained from SAWS and the LSA SAF radiation product obtained from EUMETSAT, all in real time
- obtaining the TRMM rainfall data in near real time from the NASA website

Modelling the individual cells at desired locations using a specially designed extension of the TOPKAPI model in LSM mode, exploiting the static data from:

- Jarvis's SRTM based 90 m DEM
- SIRI for soil depth and saturated soil moisture
- WR90 for residual soil moisture and conductivity
- GLCC Land cover for overland/surface roughness Manning's *n*

These are then combined automatically to produce a Soil Water product at 3-hourly intervals, which can be mapped or compared with other estimates. The flow chart showing these operations follows in Figure 4.11.

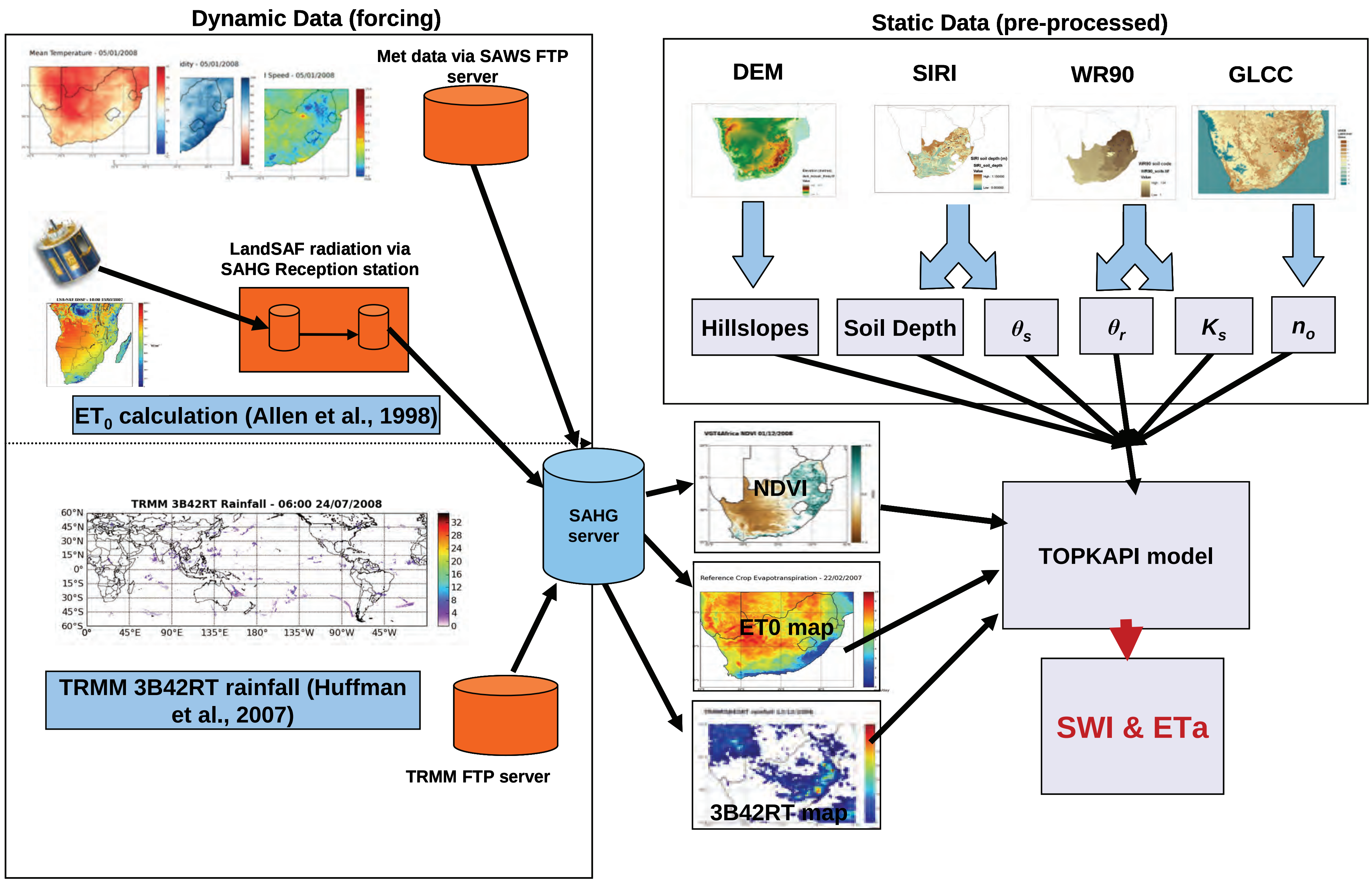


Figure 4.11 Flowchart of the data streams required to model SSI at TOPKAPI-LSM locations

TOPKAPI-LSM Simulations

Once all the "ingredients" were available, then it was time to ingest them into the "new-look" TOPKAPI-LSM model as shown in Figure 4.11, which is a flow chart describing the data streams required for simulation of Soil Saturation Index (SSI).

The preliminary simulations of Soil Saturation Index (SSI – the proportion of the soil store voids occupied by water) to "test" the code, were done at the 17 sites where ARC had been collecting data for some time; this choice of locations was suggested by the knowledge that ARC had been recording data at probe locations since mid July 2008. This meant that we would be able to compare our work against these data when they became available. The simulations at the sites were started at a nominal value of SSI = 50%. It turned out after we obtained the data, that this was a reasonable estimate compared with the measured values at the beginning of the records, which ranged from 30% to 60% with most near 45%. After performing the calculations, we were informed at the 2009 Reference Group meeting of the project, that the probes had been found by ARC to be faulty. As a result, the calculated trajectories of the recordings have not been included here. Nevertheless, the description of the procedure used to estimate SM and ET at the sites is translateable to every TOPKAPI-LSM site, so will follow.

Difficulties encountered during operational simulations

Figure 4.12 illustrates the one of the problems. The horizontal time scale is the number of intervals of 3 hours (the TRMM data sampling/arrival interval). All the SSI records at the 17 ARC station locations were initialised at 50%, as mentioned above, and seemed to be decaying as expected until time-step 30 (90 hours after initiation), when there is a sudden jump, some SSI values reaching 100%. After that, the SSI traces decay again at different rates as expected. It was discovered that this unrealistic behaviour was caused by one frame of anomalous TRMM rainfall data; it is difficult to automate detection algorithms to eliminate this oddity, but this has been accomplished.

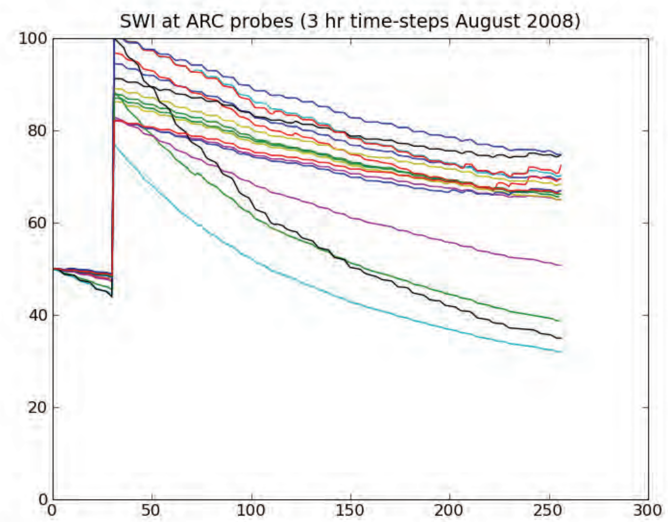


Figure 4.12 SSI calculated at the 17 ARC probe locations using physical parameters based on location. Rainfall forcing is TRMM 3B42RT (Huffman et al., 2007) and evapotranspiration forcing is the SAHG ET_0 product. Note the extreme (unrealistic) increase in SSI at time-step 30. The horizontal axis unit is: numbers of time steps of 3 hours.

Figure 4.13 shows the “maverick” data from TRMM, followed in Figure 4.14 by the frame downloaded 3 hours later in time step 31.

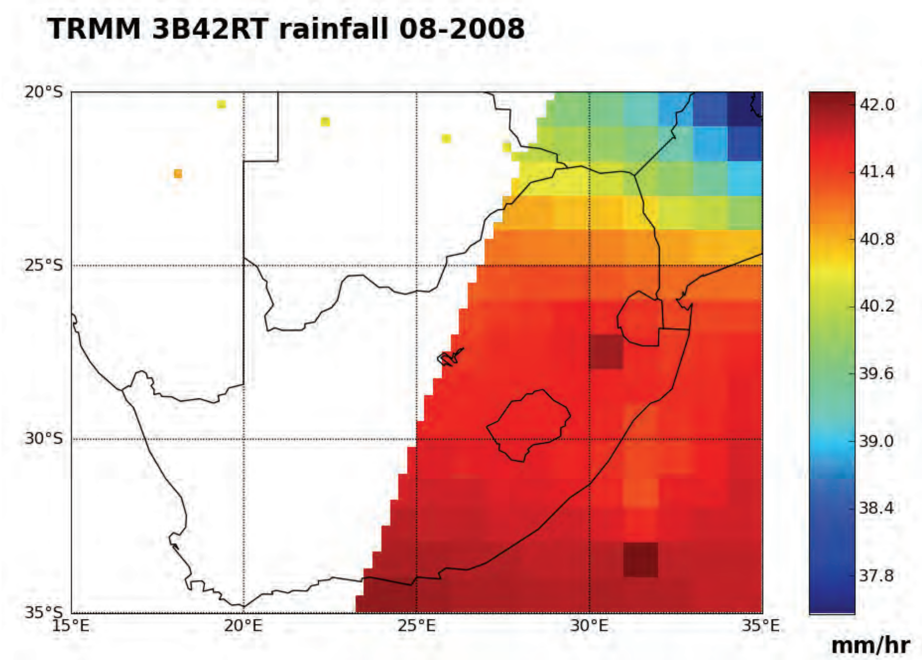


Figure 4.13 TRMM 3B42RT rainfall at time step 30, used in the modelling process shown in Figure 4.12. There is an obvious problem with this data-set that is difficult to detect automatically. In our experience this is a relatively infrequent occurrence, but has now been rectified.

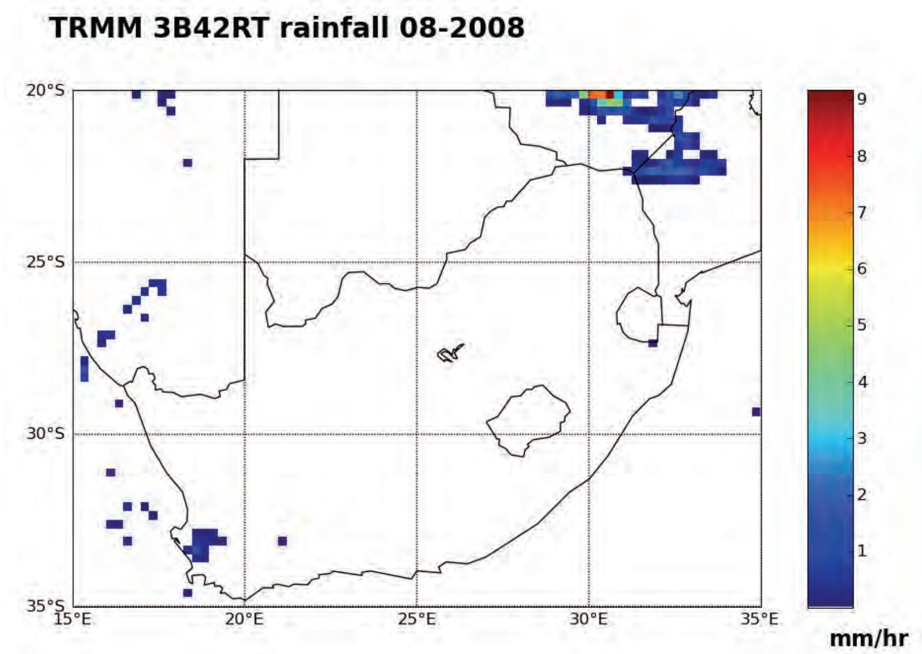


Figure 4.14 TRMM 3B42RT rainfall 3 hours after that shown in Figure 4.13 (for comparison).

Extension of SSI Estimation to SAWS AWS sites

Here we show that it is straightforward, using TOPKAPI-LSM, to extend the estimation of physical parameters to many isolated sites anywhere in the country and hence compute SSI routinely. Figure 4.15 includes all the SAWS automatic weather station (AWS) sites and is the state of the SM at many sites around the country after a month's run of the TOPKAPI-LSM in multiple site mode. We note that there are problems with coastal locations that are affected by the land-sea mask applied to ET data; we subsequently devised a means of rectifying that.

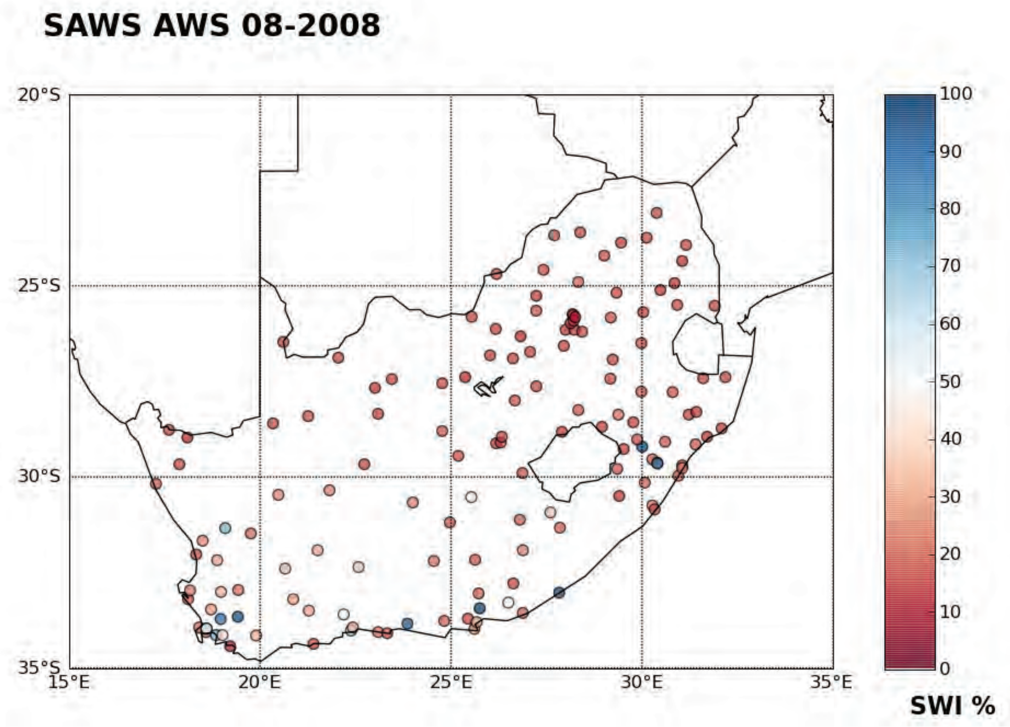


Figure 4.15 An example of SSI (or percentage saturation) calculated at 164 SAWS automatic weather station locations. The figure is a snapshot of one time-step of the simulation. It should be noted that the size of the circles is not indicative of the areal extent of each cell modelled, and neighbouring circles may in fact be a significant distance apart.

EXTENDING THE TOPKAPI CELLS FROM A FEW SITES TO MANY OVER THE COUNTRY

We now summarise the work done to enable the production of the daily updated 3 hourly spatial SSI product. All our work is computer-based, so is not tangible except as code which is designed to capture our ideas and the needs of the techniques. This code has to be validated each time there is a change. The major effort in this part of the project was to move from a desktop study on a few points to a near real time system covering the country in fine spatial and temporal detail. Problems and complications which had to be resolved included:

- finding a sensible compromise due to the mismatch of resolutions of the various dynamic data streams and static data-bases;
- resolving the limitations imposed by the spatial and temporal resolutions within the computing capacity of the system;
- managing the erroneous data which occasionally contaminate a data-stream;
- adapting potentially useful algorithms to real-time applications;
- designing a computational vehicle to convey the value and information in the final product, which is robust and can perform without supervision.

Conversion of analysis scripts used in the desktop study of Deliverable 6 into a robust data processing system. In the work described earlier, all of the required dynamic forcing data (TRMM 3B42RT rainfall and reference crop evapotranspiration estimates) were available to the project team for the entire simulation period considered as fixed data-sets. This greatly simplified the modelling procedure and analysis of the results, because it was done in desktop mode, working with historical (static) sets of data. For the purpose of near real time automatic, unsupervised processing, significant reorganisation of the existing software was required in order to ensure a robust system for retrieving and processing forcing data, prior to carrying out model runs. The final status of each of the 6989 cells modelled by the TOPKAPI model (Liu and Todini, 2002) needs to be stored at the end of each day's model run while waiting for new data, which required restructuring of the software. Each new day's data are now processed independently of the previous day except for the carry-forward of the SM state of the TOPKAPI cells (this is different and more complex than the way it is done in desktop studies where the state is easily retained in memory and all of data is processed in a single run).

Resolving problems related to the software configuration required to run the modelling system on the SAHG server. All of the input data streams and static data sets are collected on the server via various mechanisms. The new tools and software required to fetch, read and process data in a multitude of formats required significant levels of problem solving and careful software design to build and maintain. Typical tools took the form of compiled extension modules, additional to the core Python tools originally used to develop the system offered by EUMETSAT. Building these extension modules required ingenuity, research and development knowledge, skill in trial and adjustment procedures and advanced computer programming expertise to ensure that the correct dependencies are available on the system. In fact, in many cases, it required building the shared libraries from scratch, as the versions provided by the operating system's packaging tools were not sufficiently up to date to cope with the new data formats.

Quality control of dynamic data streams. Since the quality and availability of the dynamically changing forcing data (rainfall and actual evapotranspiration ET_a) cannot be guaranteed, the modelling procedure must be capable of handling missing data and automatically identifying poor quality data and rectifying the situation in an unsupervised way. These algorithms and procedures were new problems to be overcome in the near real time computational procedures; they were not required in previous work, because in a desktop study a coherent period of data could be chosen with which to develop ideas.

Producing a better estimate of evapotranspiration demand. The reference crop evapotranspiration (ET_0) (Allen et al. ,1998) is an estimate of the evaporative forcing which should be applied to a model cell if it were covered in a well watered grass crop, subject to the atmospheric forcing experienced at a particular time (radiation, wind, temperature etc., which are all variable in time). In reality, water stress and vegetation type and condition have a moderating effect on ET_0 and the actual evapotranspiration (ET_a) is typically lower than ET_0. We discovered that the relatively high evaporative demand predicted by ET_0 resulted in a much faster drainage rate than was observed by in-situ soil moisture probes. We therefore sought a way to remedy this shortcoming. The approach taken was to combine a crop factor approach using NDVI (Tasumi et al., 2005) with additional processing to account for soil evaporation under sparsely vegetated conditions (this is described in detail elsewhere in this report). Since our method, based on the ideas of Tasumi et al (2005), fundamentally requires NDVI, we selected a decadal NDVI product disseminated by the VGT4Africa project (http://www.vgt4africa.org) via the EUMETCast system. Software tools to read and manage this new dataset were developed; consequently the data processing performed by our satellite receiving station also required changes to ensure that the NDVI products are properly archived.

Assessing the level of modelling detail that is feasible without requiring more powerful computing facilities. The spatial resolution of the modelling grid directly affects the level of computational effort required to run the TOPKAPI model, at all cells in the grid, at each time-step. Initial experiments in the development of this deliverable were carried out at a grid spacing of 0.25° with grid centres matching those of the TRMM 3B42RT rainfall product (Huffman et al., 2007). At this spacing, the number of model cells completely covering the country (except for those places where soil properties are not available on the SIRI data-base, was about 1700. We were able to increase the grid resolution to 0.125° (6989 cells and closely matching the spatial scale of the Unified Model meteorological fields) without incurring a significantly increased computational load. It appears that the increase in time with increase in grid resolution is reasonably linear and would be well suited to parallel computation, since the cells are independent of each other when operated in Land Surface Model mode. Investigation of these techniques is an interesting and potentially fruitful study, but beyond the scope of this project.

Initialisation of the TOPKAPI soil stores. In order to produce simulations of SSI over the region, it was necessary to have sensible initial values. Building on earlier work in this project, the remote sensing soil moisture product from the National Snow and Ice Data Centre (NSIDC) University of Colorado at Boulder, was used for SSI initialisation. This product provides a gravimetric moisture content (grams per millilitre) at 0.25° spatial resolution and was converted by calculation to SSI (relative volume of water in a TOPKAPI cell's soil store) using the soil properties at each cell location. The available ascending and descending passes of the AMSR-E based estimates were combined, by averaging at grid locations where both an ascending and descending pass was available on 01 August 2008, choosing whichever was available at other grid points, and using a constant regional average at the few remaining locations.

A representative SSI image

The following Figure 4.16 is an example of the end product – a single frame of estimated Soil Saturation Index (SSI) on 17 December 2008 at midnight. It is the last of the 1113-long sequence of 3 hour time-steps from 03:00 on 31/07/2008 to 00:00 on 17/12/2008, whose images at midnight of each day appear later in the report.

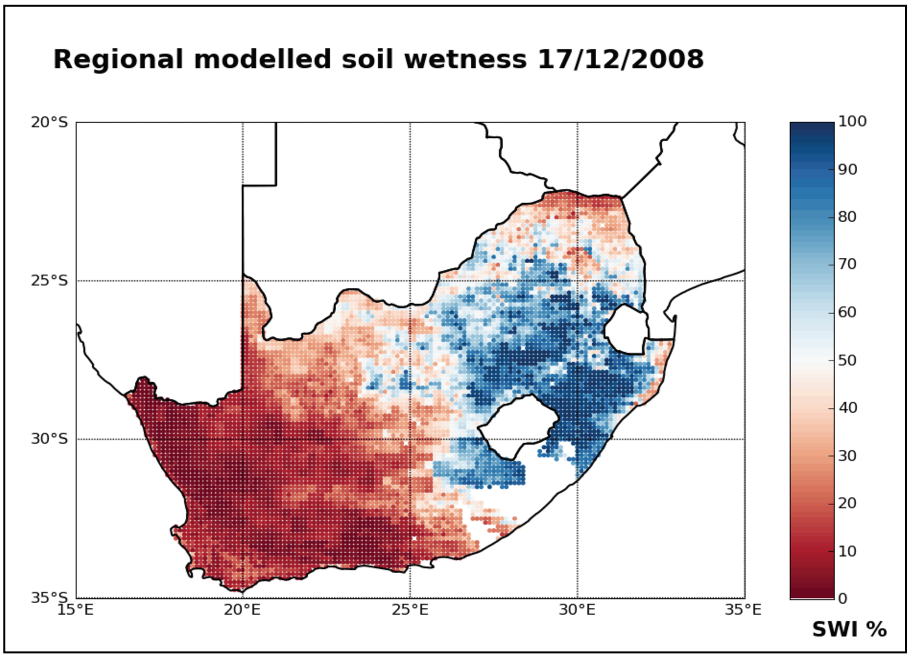


Figure 4.16 Image of estimated Soil Saturation Index (SSI) at 6989 sites at midnight on 17 December 2008. The full set of images appears later.

There are several features in Figure 4.16 which are worth discussing. These include: the colours, the missing information and the detail.

- The colours indicate that, by mid-December 2008, a large part of the Summer Rainfall region has wetted up nearly to saturation, while the Western part of the country has, by and large, remained dry.
- The blanks correspond to missing data in Swaziland, Lesotho and areas where there is currently no soil property information. As soon as this information is made available, the blanks can be filled in, because we have archived the driving information of actual evapotranspiration (ET_a) and rainfall.
- The third feature is the detail at which the SSI information is presented. Displayed here are the Soil Saturation Indices (SSI) at 6989 sites where the calculations have been made, covering just over 1 million sq km of RSA, where there are no blanks.

To derive the data for Figure 4.16, we started the modelling procedure on 1st August 2008. To initialise this collection of 6989 sites, we had two options. We could have started every pixel at 50% SSI (say) and moved on with the calculation, but, although we were sure that after the first rainy season the SSI would have become independent of the starting condition, decided against this procedure as there was a better alternative. We had access to the information collected by Julia Pfeffer (MSc student ex Grenoble) while she was working with us in early 2008 on comparisons of remote sensing estimates of SSI (she was in SAHG at UKZN for 4 months supported by the WRC-managed SAFeWater project). Her investigation showed that the best currently available daily estimate of SSI was the NSIDC product from Boulder. She had determined that the correspondence between their daily estimate and the weekly sampled SWI product of Vienna University of Technology (which we have available under the ESA SHARE project) was the best to hand. We chose the data from 1 August 2008 from the NSIDC website to initialise the TOPKAPI model cells for our start-up.

In Chapter 3 of this report a means of calculating reference evapotranspiration ET_0 is described, over each 12 km square pixel in the region from the meteorological data forecast by the Unified Model. It was found that applying this directly using a simple crop factor to the TOPKAPI cells without modification removed far too much water from the soil store. There was a need to devise a technique to achieve a proper balance to achieve a very realistic modification to the ET_0 by calculating ET_a at each site. The total accumulated ET_a estimated for the period at each of the 6989 sites is displayed in Figure 4.17. Note the areas (blue) where there was little soil water available to evaporate – quite different from the potential (reference) ET_0 as compared for a day with ET_a in Figure 4.18. The red areas in Figure 4.17 correspond to places that have experienced heavy rainfall as well as high NDVI later in the season, during the 4 ½ month period.

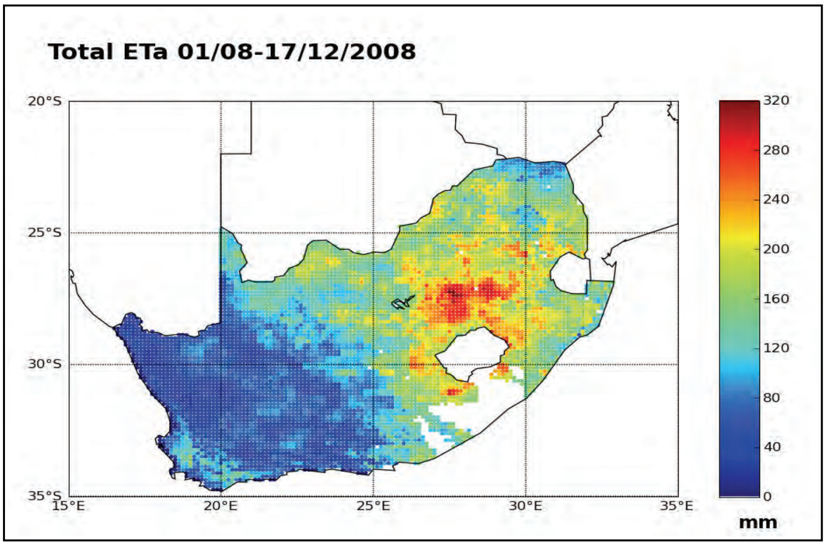


Figure 4.17 Total accumulation, over the 4 ½ months, of estimated Actual Evapotranspiration ET_a at the 6989 sites of the TOPKAPI cells.

How we calculated each SSI image

The next sequence of six images in Figure 4.18 gives a presentation of the technique used to produce the current day's SSI map from the previous day's map at each TOPKAPI cell treated as a Land Surface Model (not connected by overland flow or stream channels). The SSI is shown here as a 24 hour jump – in reality the operation is done in 8 steps of 3 hours, corresponding with the interval between the arrivals of the TRMM rainfall estimates.

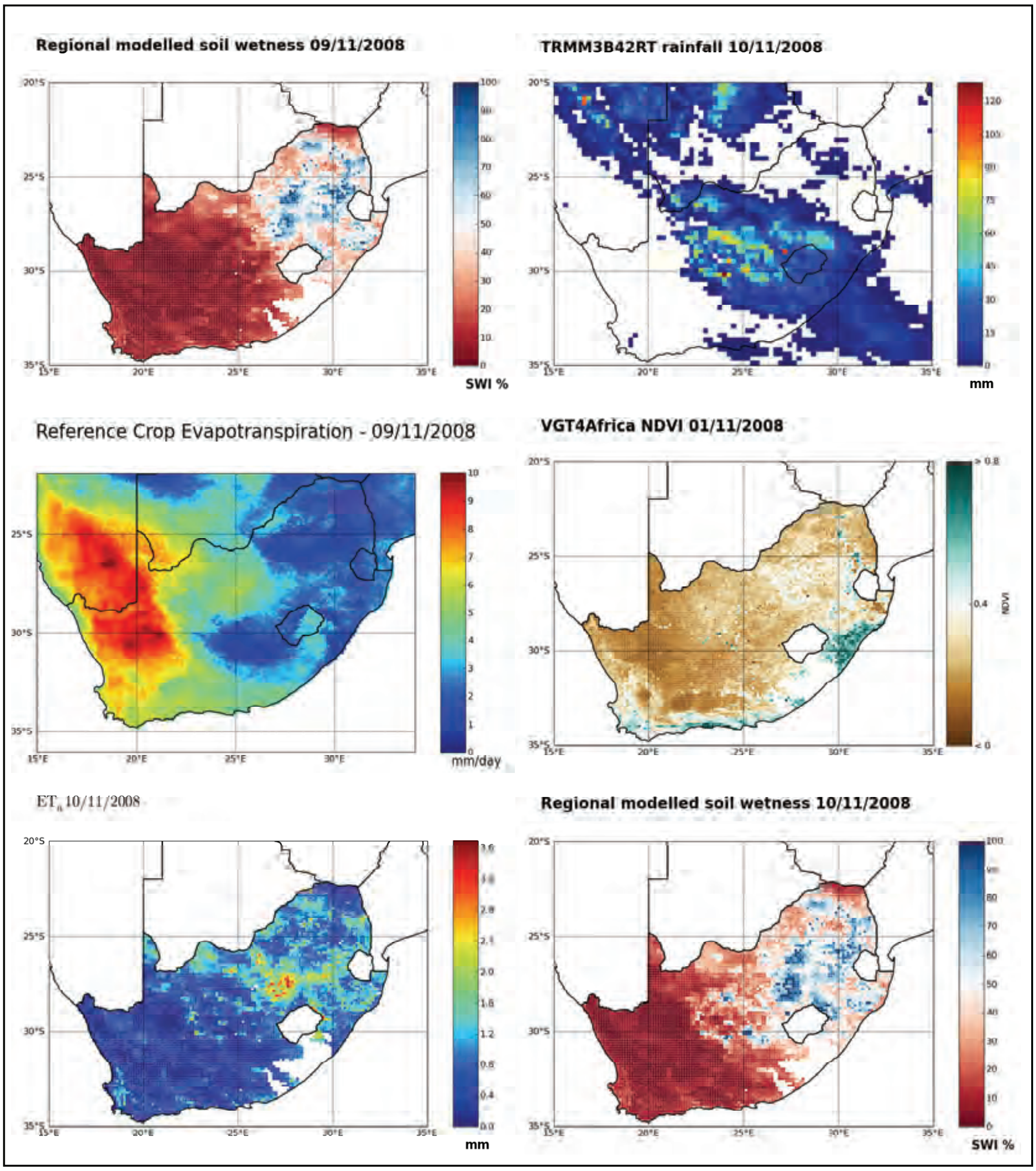


Figure 4.18 One day in the development of the regional SSI product. From top left: (i) the starting values of SSI at 0000h on 9 Nov; (ii) the total accumulated rainfall during the day of 9 Nov (up to 0000h on 10 Nov); (iii) the total ET_0 during 9 Nov; (iv) the most recent NDVI image available on 9 Nov (acquired on 1 Nov); (v) the total ET_a during the day 9 Nov (as recorded at 0000h on 10 Nov); (vi) the resulting regional SSI at 0000h on 10 Nov.

The six panels in Figure 4.18 show the following variables at every one of the 6989 cells: initial SSI (%); TRMM estimate of rainfall (in mm, accumulated to a daily total), available every 3 hours from NASA; Reference ET_0 (mm) calculated from the Penman-Monteith formula; NDVI image (these are available at decadal intervals, so are good for natural vegetation but will not reflect intermediate farming practice); the Actual Evapotranspiration (ET_a) calculated using our modified crop factor K_c and the final SSI map for the "current" day. The ET and rainfall figures are accumulations of the eight 3-hourly images to daily totals. The SSI images are snapshots at midnight. We again note the remarkable difference between ET_0 and ET_a.

Note that the colour scales differ from image to image. We have chosen this 24 hour period because there was reasonably heavy rainfall over the summer rainfall region and thus an obvious change in the SSI over the day, especially over the North-eastern Cape. The NDVI is well developed along the South-eastern coast and one can judge that there will be moderate ET_a there. In the interior, the NDVI is low as we are in the beginning of the wet season and vegetation has not yet recovered from the dry season. Our modification of the algorithm to calculate K_c (developed below), to allow evaporation of water from wet bare soil, has been successful, resulting in plausible SSI estimates.

The initialisation of the SSI over the region

The initialisation of the SSI over the region was accomplished by using the NSIDC product, shown by Julia Pfeffer to be the best daily SWI remote sensing product available. To indicate why, the following figures are taken from the report she compiled while with us in Durban sponsored by SAFeWater. Previously (Chapter 2) it was shown that there was very good correspondence between the TOPKAPI-based estimates of SSI and the ENVISAT product SWI as computed by our SHARE partners in Vienna (Vischel et al. 2007, 2008). We concluded that this was the best product (even though it has a cycle time of 7 to 10 days) against which to compare the two daily estimates of SWI, one from NSIDC in Boulder Colorado and the other from the Free University of Amsterdam (VUA). The results showed that, on balance, the NSIDC product was the better of the two.

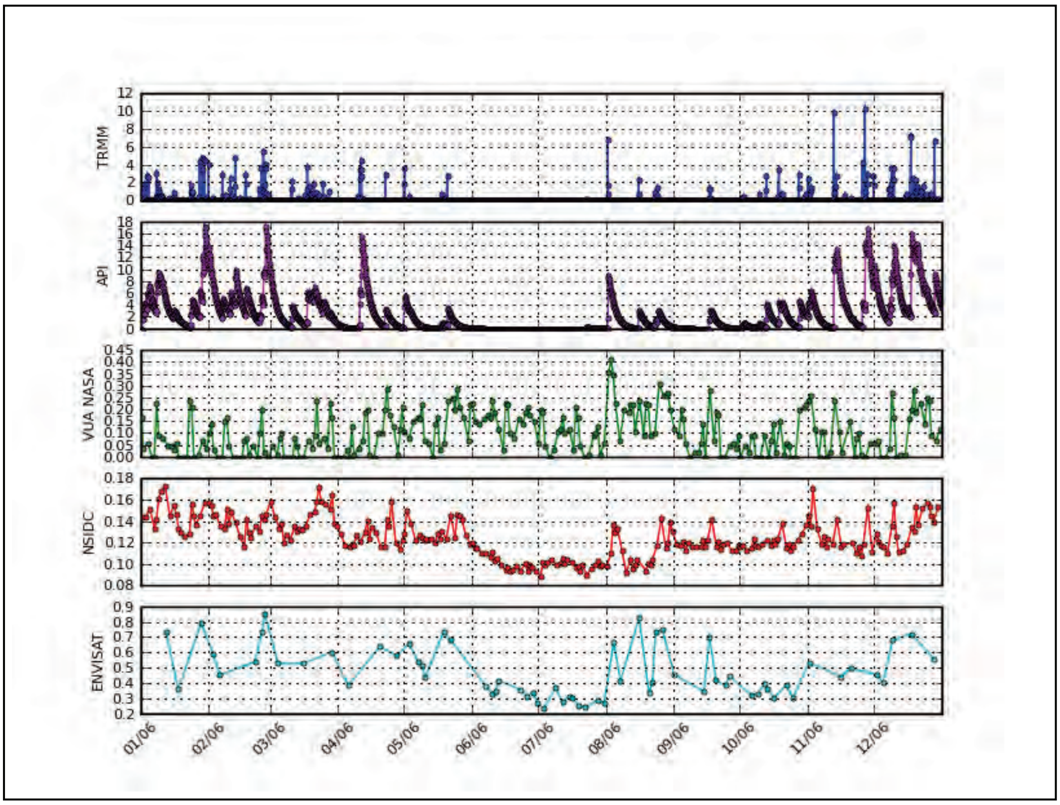


Figure 4.19 Time series of descending remote sensing estimates over Liebenbergsvlei area, for year 2006, showing good correspondence between the NSIDC product and the ENVISAT series.

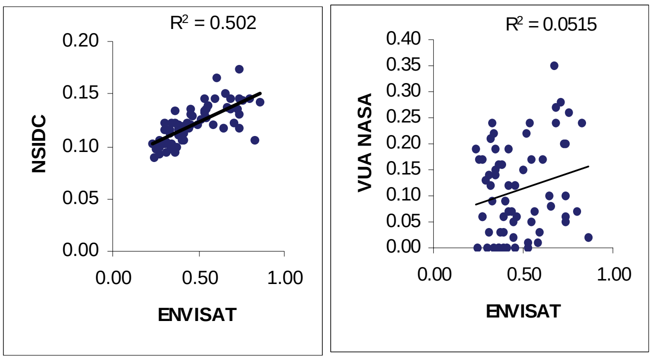


Figure 4.20 Scatterplots and determination coefficients computed for each pair of remote sensed estimates over Liebenbergsvlei site for year 2006 , with NSIDC and VUA NASA descending, versus ENVISAT. In summary, it was decided to use the NSIDC estimate on 1 August as the initial SSI estimate for this Deliverable. It is shown overleaf in Figure 4.21.

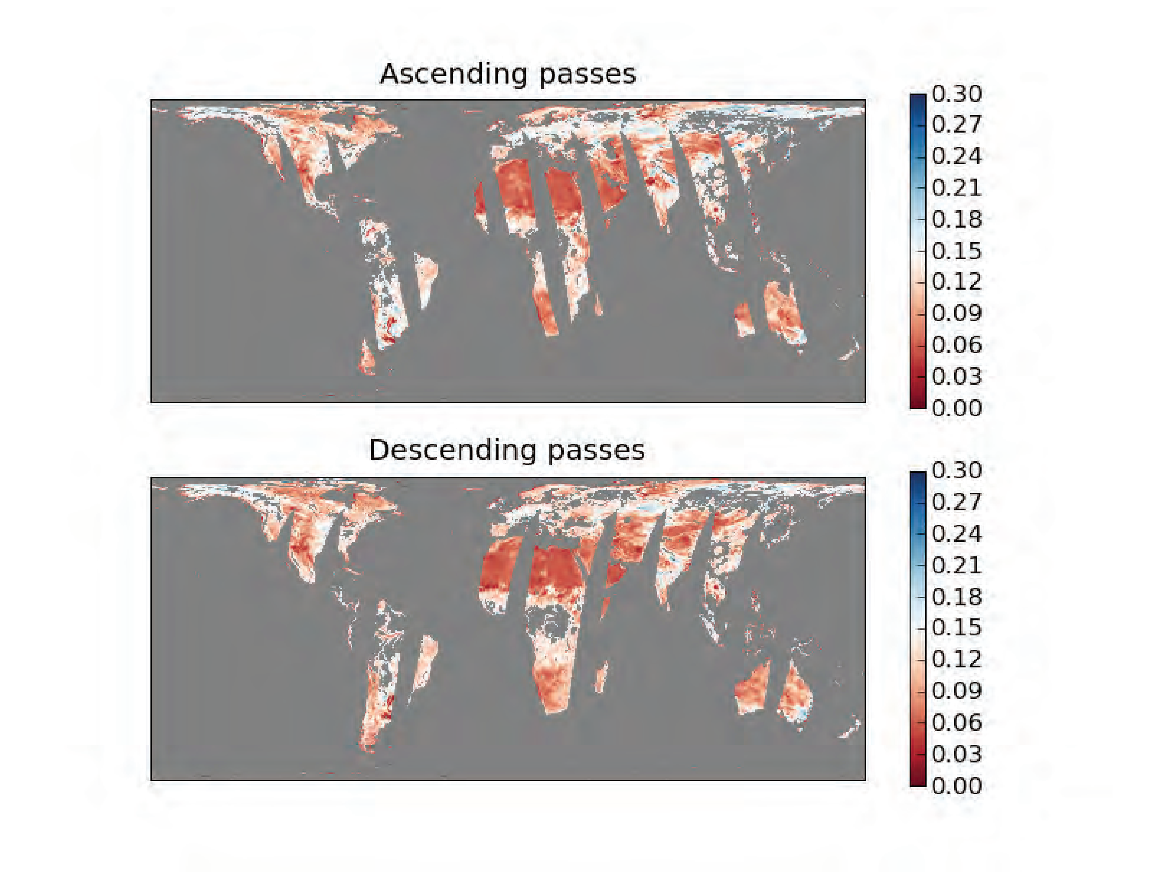


Figure 4.21 The information used for the initialization of the SSI using NSIDC soil moisture estimates (g/cm^3) which we converted into SSI values using local soil properties.

Derivation of the dynamic Adjustment Factor K_c for modifying the ET_0 estimates to ET_a.

We had already developed the algorithms to estimate Reference Evapotranspiration (ET_0) using the Penman-Monteith formula devised by Rick Allen of the University of Idaho (Allen et al., 1998) as described in Chapter 3. We needed a means of factoring this on a day-by-day basis to obtain ET_a and decided that the static Crop Factor (K_c) method traditionally used in agricultural practice would be too complicated. That static method uses a fixed ET_a ~ time graph for each vegetation type and cannot adjust to dynamic changes such as drought, late planting, fire etc. Such curves are not easily available for natural vegetation which covers a large proportion of the country. The alternative was to use a remote sensing product based on Leaf Area Index or similar as an indicator of vegetation demand for transpiration.

We were looking for a useful surrogate for K_c and were fortunate to have the opportunity to discuss the estimation of ET_a person-to-person with Rick Allen at a workshop on Remote Sensing and Evaporation run by Caren Jarmain for the WRC in Pietermaritzburg on 28 October 2008. He indicated that a good surrogate for K_c was NDVI as long as the vegetation was fairly well developed and transpiring. We studied the paper he co-authored (Tasumi et al., 2005) and adapted the formulation to allow for evaporation from wet soil when vegetation (hence NDVI) was low. The following two images from that paper in Figure 4.22, show the derived relationships between K_c (which they define as the ratio ET_a/ET_0) and NDVI for two of many crops they studied. Rick Allen recommended that we use NDVI as a surrogate for K_c. As can be seen in both figures, the K_{cb} line, if used by itself, produces a progressively smaller K_c as NDVI (vegetation cover) reduces. Also shown in the figures is a set of data for both crops for May 2000, where for very low NDVI, K_c can vary from 0 to 1, associated with the actual evaporation from wet bare soil. The K_{cb} curve is good for values of NDVI above about 0.6, but below that one needs to allow for wet soil transpiration.

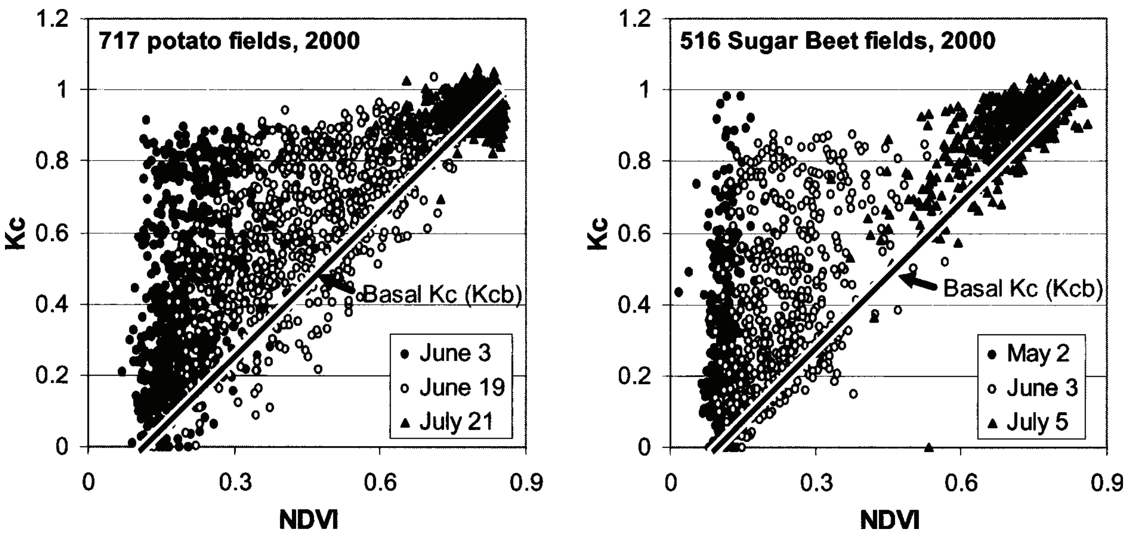


Figure 4.22 Crop coefficient versus normalized difference vegetation index for 717 potato fields (left) and 516 sugar beet fields (right) in Magic Valley area of Idaho, for three LANDSAT dates in crop developing periods during 2000 [Tasumi, et al., 2005]

The base-line relation, K_{cb} in Figure 4.22, appears in Figure 4.23.

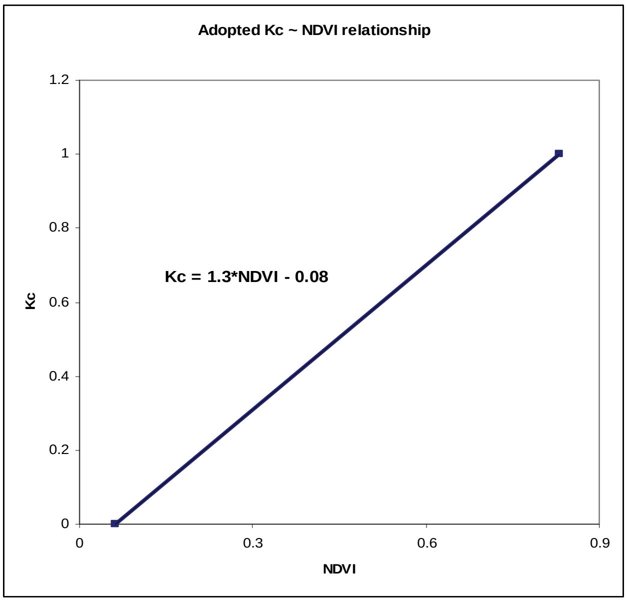


Figure 4.23 The adopted NDVI~Kc relationship, based on Figure 4.22

We started with the K_{cb} base-line and then adjusted K_c to accommodate a wet bare soil when vegetation is not active. As can be seen in the examples in Figure 4.22, we have to allow evaporation from wet soil even when vegetation is sparse. The formulation we adopted was the concept of a virtual store EV which we call the "available water for evapotranspiration" (we will likely make this into a real surface store in a future adaptation of the TOPKAPI model). We allow EV to experience carry-over during a rainy period using a simple correlation ***R***, modified with a limited amount of rainfall (it cannot exceed ET_0) with up to ET_0 being removed on the previous day. The ET_a on the current day cannot exceed the value of EV ("available water for evapotranspiration") nor K_c*ET_0 if there is well developed vegetation. The formulation is as follows, enabling us to calculate ET_a on each day:

$$EV(i) = \min(ET_0,\max(\boldsymbol{R}*EV(i-1)+\min(RAIN(i),ET_0)-ET_0,0))$$
$$ETa(i) = MAX(EV(i),Kc*ET_0) \tag{4.3}$$

In summary, the formula allows evaporation of some of the rainfall up to a maximum of the current ET_0 at times when there is no vegetation and also allows removal of soil water by active vegetation as soon as NDVI dominates. The following figure provides a spreadsheet comparison of ET_a traces at two different levels of K_c: 0.2, being at the very low end (corresponding to NDVI of 0.22) and 0.6, which is at the lower threshold of leafy crops, where NDVI is 0.52, both values linked through the relationship in Figure 4.23. The blue line in Figure 4.24 is rainfall (negative losses from the ground!) plotted against the right-hand axis.

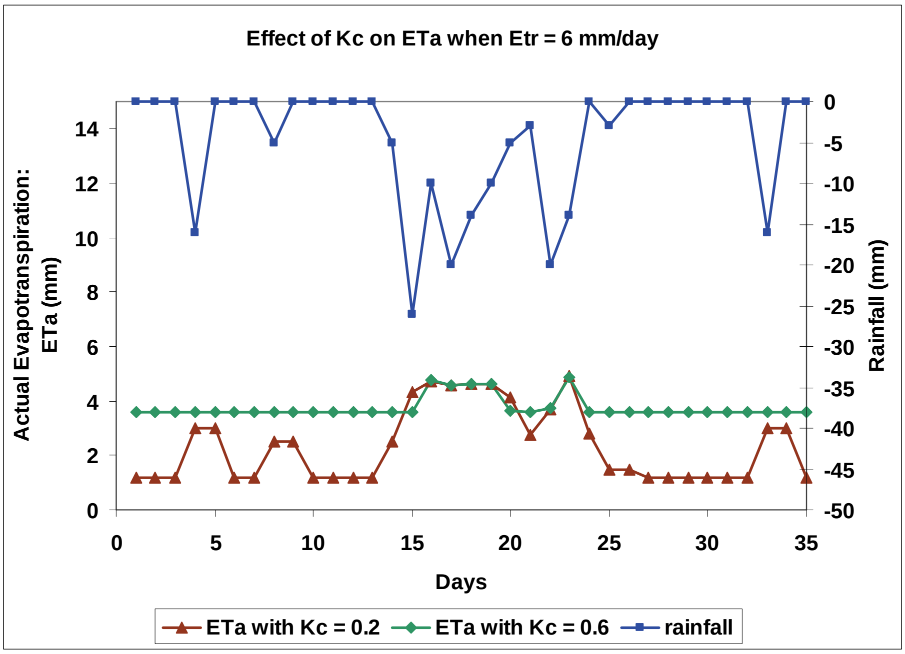


Figure 4.24 Traces of ET_a, calculated with the above formula, for 2 choices of K_c and other parameters and rainfall fixed.

For partial vegetation cover (K_c = 0.6 – the green line), the transpiration dominates at times when it is not raining and 0.6×6 = 3.6 mm is lost from the soil store each day. When it rains and there is available surface water, soil evaporation contributes to the losses, increasing the total ET_a to about 4.6 mm per day. When there is less vegetation with the same rainfall and ET_0 scenario, soil evaporation reacts whenever there is rain, removing some water from the soil store, reaching the maximum of 4.6 mm when it is raining heavily for a few days. It is only when the K_c is above 0.8 in this scenario that all the ET_a is passed through the plants, as indicated in the diagrams of Tasumi et al. (2005) in Figure 4.22.

On the following page, a sample of NDVI maps used in this study are displayed.

The NDVI maps used in this study

The following images collected in Figure 4.25 show the NDVI maps at the beginning of each of the 6 months, 1 Jul to 1 Dec 2008, sampled from the decadal series. Note that only in November does the NDVI in the Summer rainfall region increase to 0.5; it then increases further to 0.7 in December. In the mean time it has been raining quite heavily in the region in October. If we had not modified the K_{cb} curve of Figure 4.23, the ET_a would have been too low.

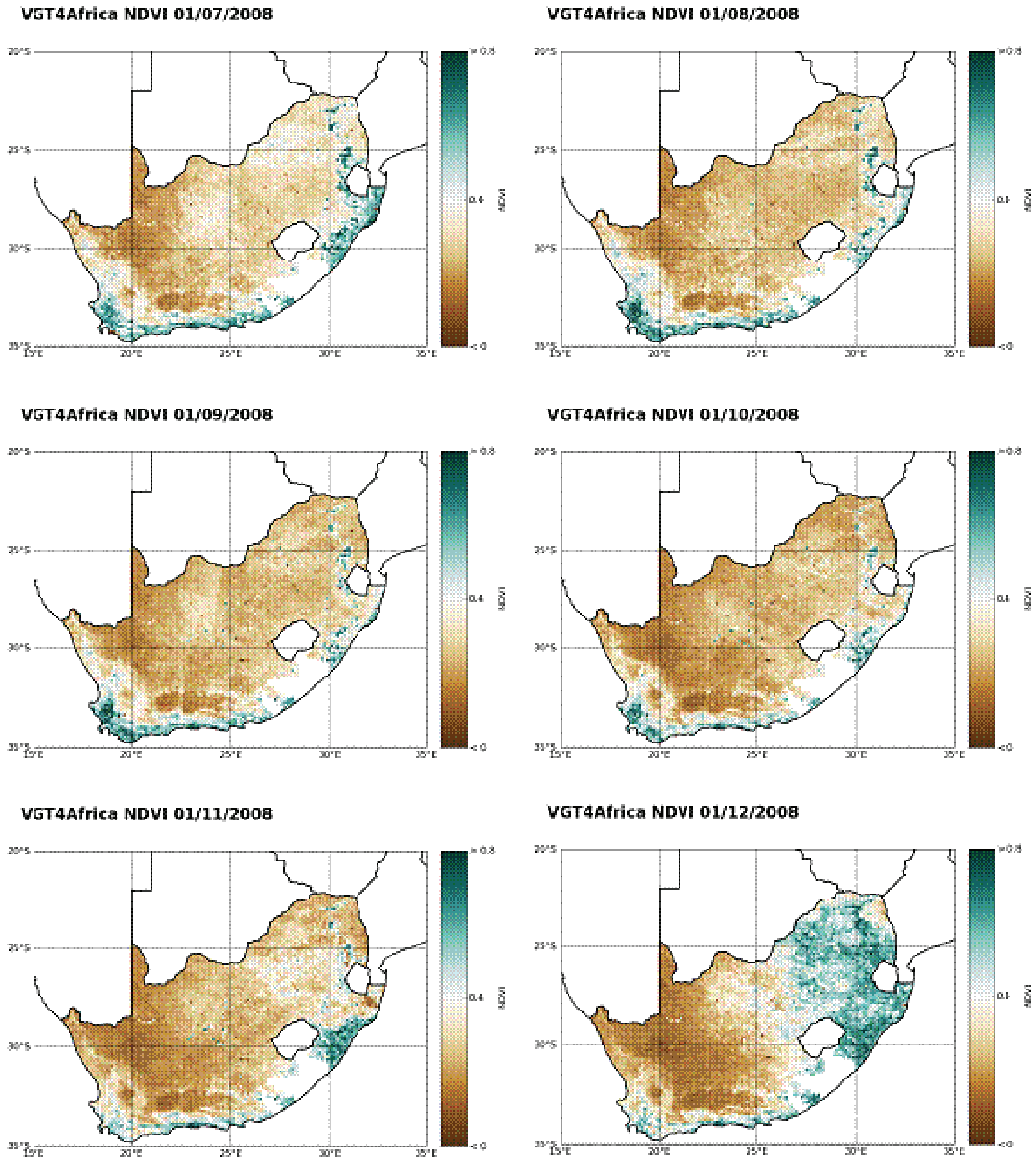


Figure 4.25 Time history of the VGT4Africa decadal NDVI product for consecutive months in 2008. The data have been thresholded if they lie outside the range 0-0.8, pixels lying outside this range are assigned the nearest boundary value.

On the following 5 pages, we show screen capture extracts from our website for the 4 ½ months from the beginning of August, 2008. Each page has 2 panels displayed, as we want to compare the daily rainfall and SSI spatial variability. The upper panel on each page is the daily accumulated TRMM rainfall product and the lower panel the SSI estimates at midnight at the start of each day. In August and September it will be seen that the TRMM transmissions were contaminated and had to be dealt with by our online error correction routine. The SSI is shown drying out over the 2 ½ month period until mid October when some substantial rainfall fell in the interior, wetting up the soils in that region. The Cape interior continued to dry out until well into November, while the Eastern part of the country wetted up considerably.

In more detail:
Figure 4.26 is the first of the set of 5 pairs of images, representing daily rainfall and the estimated SSI at the start of each day at midnight. This page is for August. On 5 August we had our first bad rainfall image which forced us to devise a means of dealing with this online. In the lower panel, the first day was initialised with the NSIDC SM estimate shown in Figure 4.21. The smoothness of the first image gradually changes to more textured SSI as time progresses, due to the spatial variation in soil properties and the actual evapotranspiration. There is a progressive darkening of the images during this time of low rainfall as the soil continues to dry out.

Figure 4.27 shows the pair of rainfall and SSI images for September. During this month we experienced three more days with bad TRMM rainfall estimates over 13 to 15 September. There is a continuing progressive darkening of the images during this time of low rainfall as the soil continues to dry out over the country.

Figure 4.28 shows the pair of rainfall and SSI images for October. During this month and onwards to December, we did not encounter any bad rainfall estimates. There is a continuing progressive darkening of the images during this time of low rainfall as the soil continues to dry out over most of the country, with the exception of the Free State. The heavy rainfall on 6 October wets up the ground, causing a light patch to appear in the SSI image just north of Lesotho, which persists for a week.

Figure 4.29 shows the November set. The summer rains have started in earnest and wet up the ground over the eastern part of the country. Even though the natural vegetation is only now starting to grow, indicated by the increase of NDVI to about 0.4 over this region as indicated in Figure 4.25, ET_a is abstracting water from the wet ground as indicated in Figure 4.27.

Figure 4.30 gives the pair of rainfall and SSI images for December. Heavy rainfall wets some parts of the eastern region of the country to full saturation, even though the vegetation has developed so that the NDVI reaches 0.7 and 0.8 over these parts as indicated in Figure 4.25. The Western part of the country (mostly the Cape) is still dry and will have to wait for winter rains. ET_a is high in this region as indicated in Figure 4.27, Figure 4.28 and Figure 4.29 at sites in Limpopo, Mpumalanga and Free State.

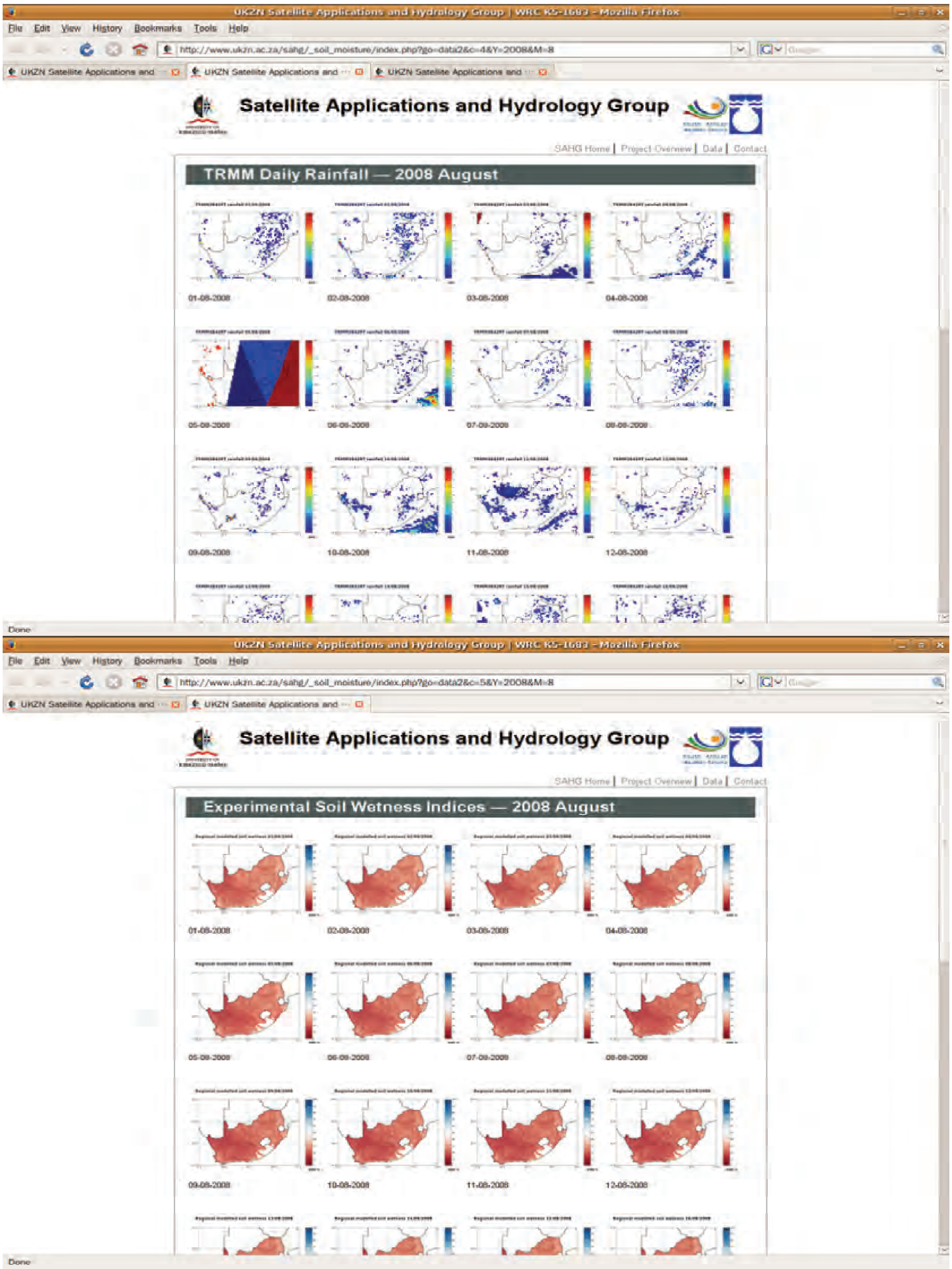


Figure 4.26 The pair of images for August. On 5 August we had our first bad rainfall image which forced us to devise a means of dealing with this online. In the lower panel, the first day was initialised with the NSIDC SM estimate shown in Figure 4.21. The smoothness of the first image gradually changes to more textured SSI as time progresses, due to the spatial variation in soil properties and the actual evapotranspiration. There is a progressive darkening of the images during this time of low rainfall as the soil continues to dry out.

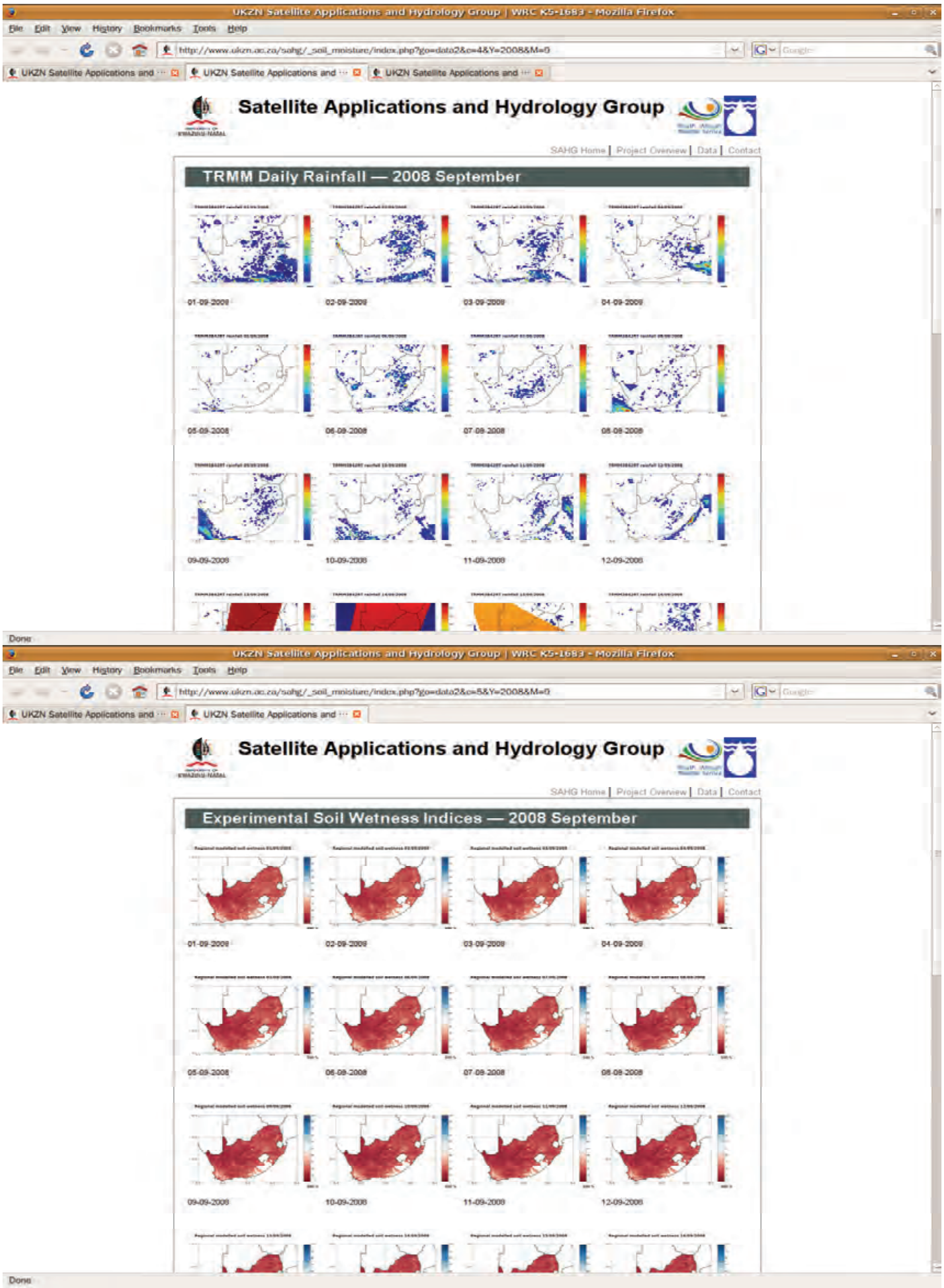


Figure 4.27 The pair of rainfall and SSI images for September. During this month we experienced three more days with bad TRMM rainfall estimates at various times over 13 to 15 September. There is a continuing progressive darkening of the images during this time of low rainfall as the soil stores continue to dry out over the country.

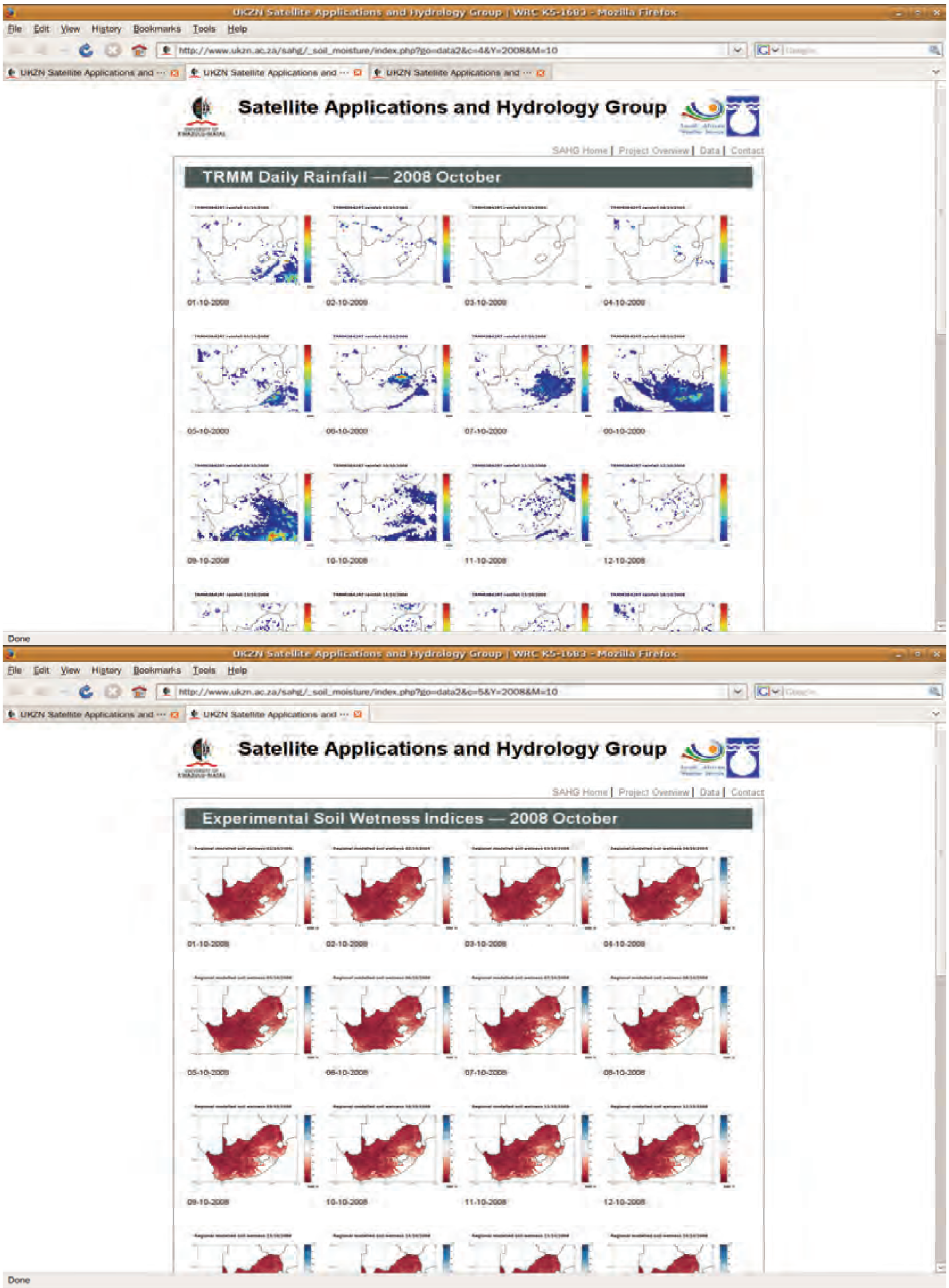


Figure 4.28 The pair of rainfall and SSI images for October. During this month and onwards to December, we did not experience any bad rainfall estimates. There is a continuing progressive darkening of the images during this time of low rainfall as the soil continues to dry out over most of the country, with the exception of the Free State. The heavy rainfall on 6 October wets up the ground, causing a light patch to appear in the SSI image just north of Lesotho, which persists for a week.

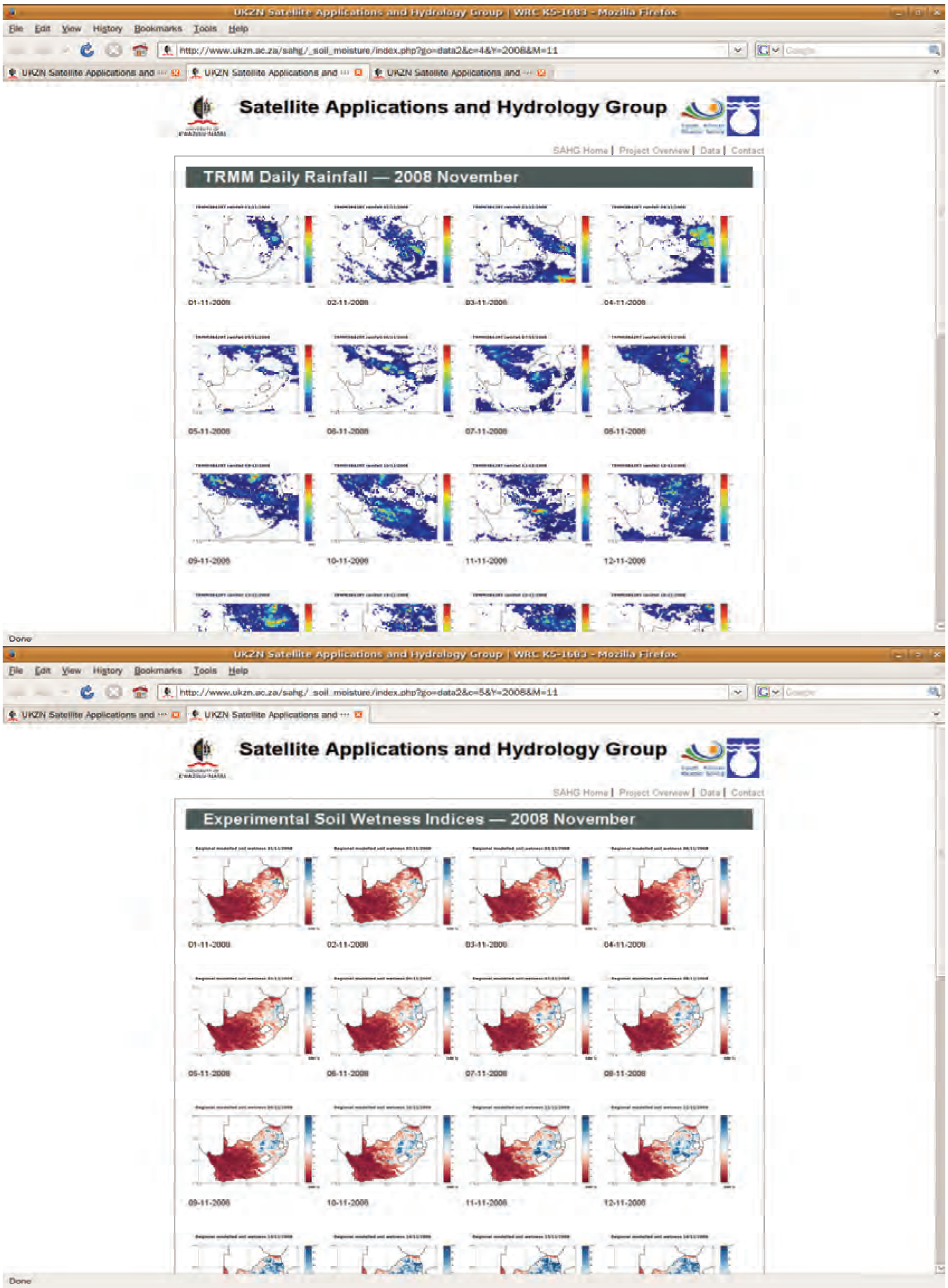


Figure 4.29 The pair of rainfall and SSI images for November. The summer rains have started in earnest and wet up the ground over the eastern part of the country. Even though the natural vegetation is only now starting to grow, indicated by the increase of NDVI to about 0.4 over this region as indicated in Figure 4.25, ET_a is abstracting water from the wet ground as indicated in Figure 4.33.

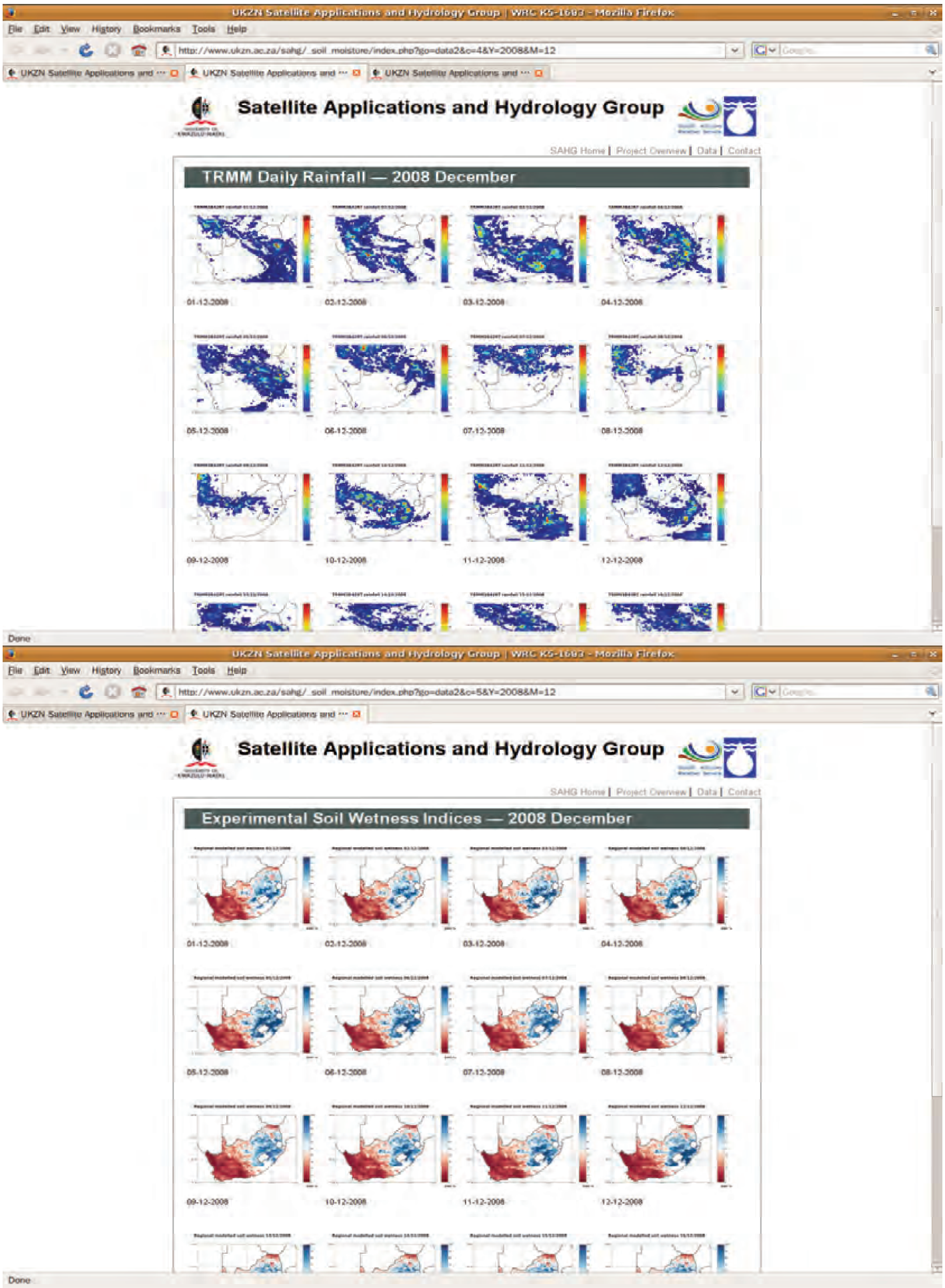


Figure 4.30 The pair of rainfall and SSI images for December. Heavy rainfall wets some parts of the eastern region of the country to full saturation, even though the vegetation has developed so that the NDVI reaches 0.7 and 0.8 over these parts as indicated in Figure 4.25. The Western part of the country (mostly the Cape) is still dry and will have to wait for Winter rains. ET_a is high in this region as indicated in Figure 4.33Figure 4.34 and Figure 4.35 at sites in Limpopo, Mpumalanga and Free State.

Time series of the variables at selected sites.
The following images in Figure 4.32 to Figure 4.37 provide a temporal summary of the SSI histories at some selected sites, which represent a variety of climates in the country from humid to dry. These time series give 3-hour detail to the daily snapshot images in Figure 4.26 through Figure 4.30 exhibiting spatial variability. Figure 4.31 below indicates the locations of those sites we selected, numbered by their pixel locations in our system, starting at the most northern part of RSA. We give them names for easier reference:

- 19 – Far North
- 763 – Limpopo
- 1024 – Mpumalanga
- 3050 – Free State
- 4693 – KZN
- 5001 – Cape

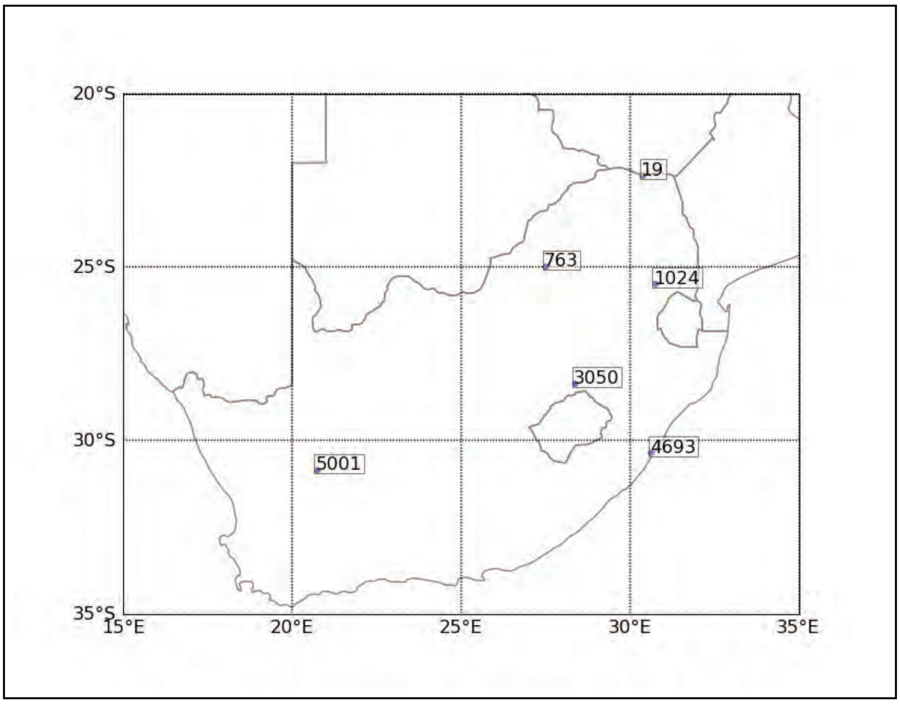


Figure 4.31 Locations of 6 TOPKAPI cells in different climate regions selected for time series plots to follow.

These six sites' records follow in Figure 4.32 to Figure 4.37. They are split into "input" to and "output" from the cells. The upper figures show SSI, rainfall and ET_a and the lower figures show SSI, Soil outflow (drainage) and overland flow from the cells. Only the KZN site, pixel 4693, exhibits overland flow after saturation and with the persistent high humidity, slightly lower ET_a on average than in the interior.

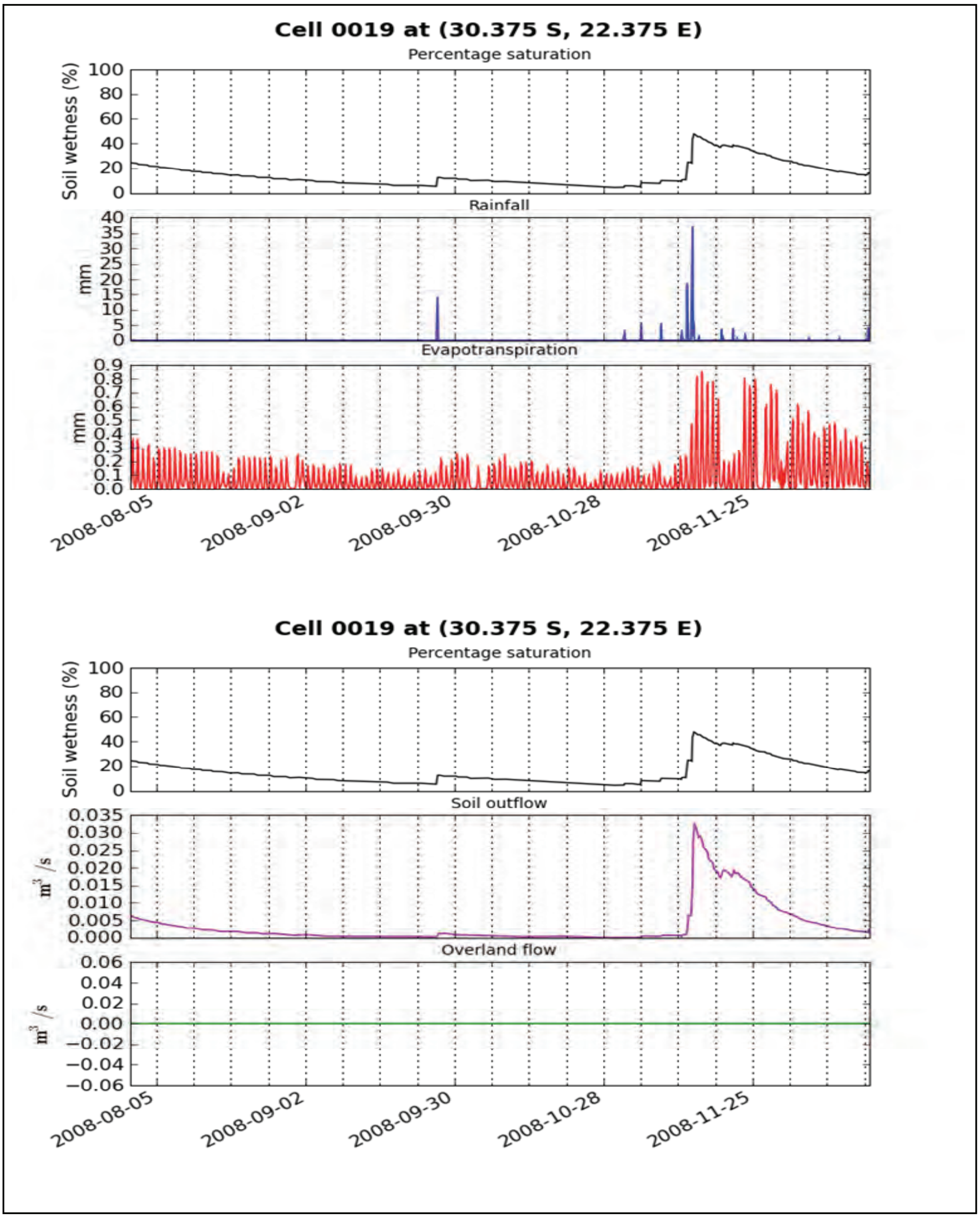


Figure 4.32 Far North. The SSI climbs to 50% after the October rains, but then experiences strong ET_a reducing it to 20% in 5 weeks. Because the ground does not saturate, there is no overland flow in the model. The red time series of ET_a (at 3-hour intervals) is spiky because it is high in the day and drops at night. ET_a also reacts to cloudy days showing a marked reduction at those times.

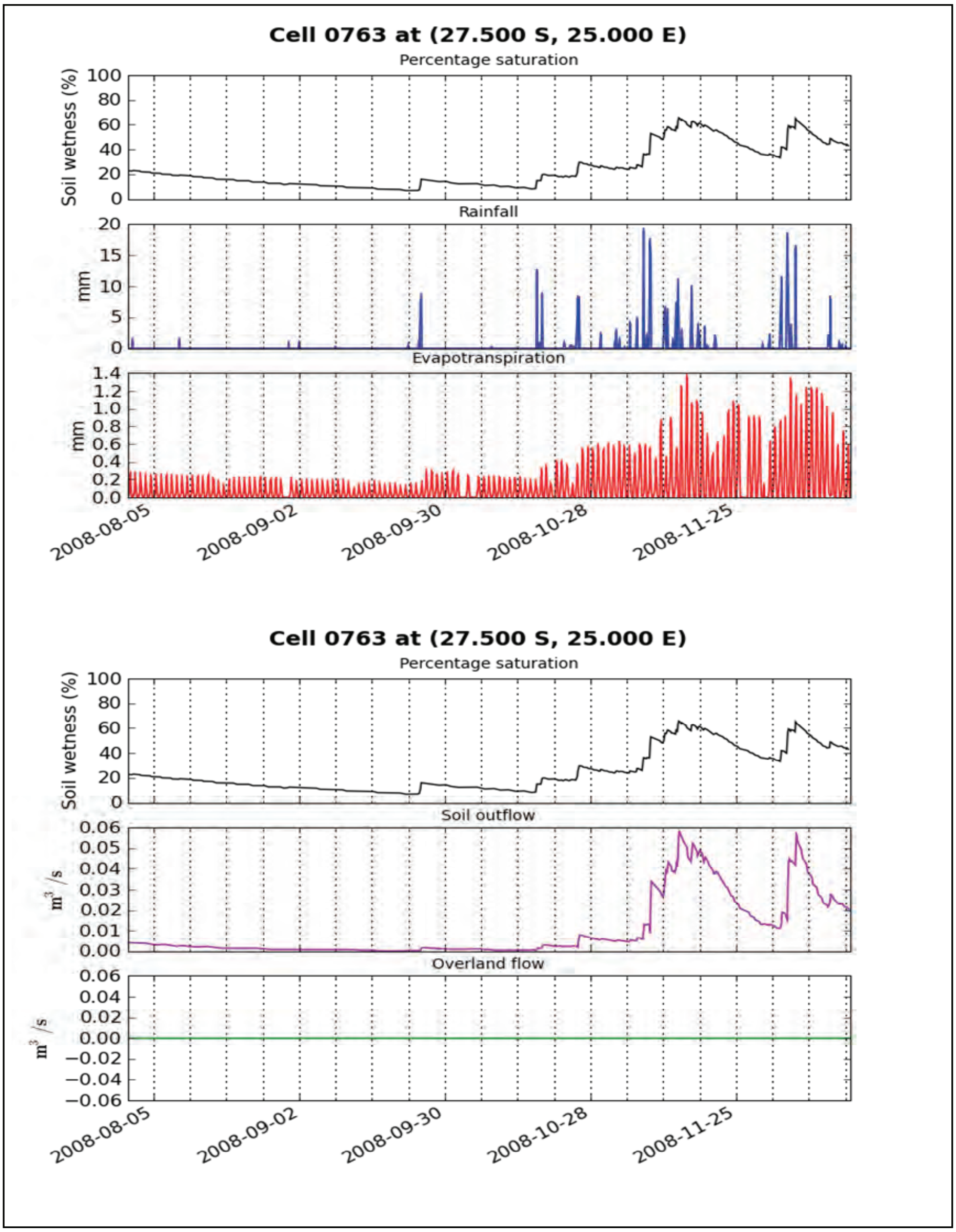


Figure 4.33 Limpopo. This site experienced earlier rains than the Far North in Figure 4.32. It also has higher ET_a (nearing maxima of 1.4 mm/3h compared to 0.8). Note that ET_a is low before the rains, dropping as the soil dries out and only increasing as the soil wets up (top panel). Cloudy and raining days lower ET_a as expected.

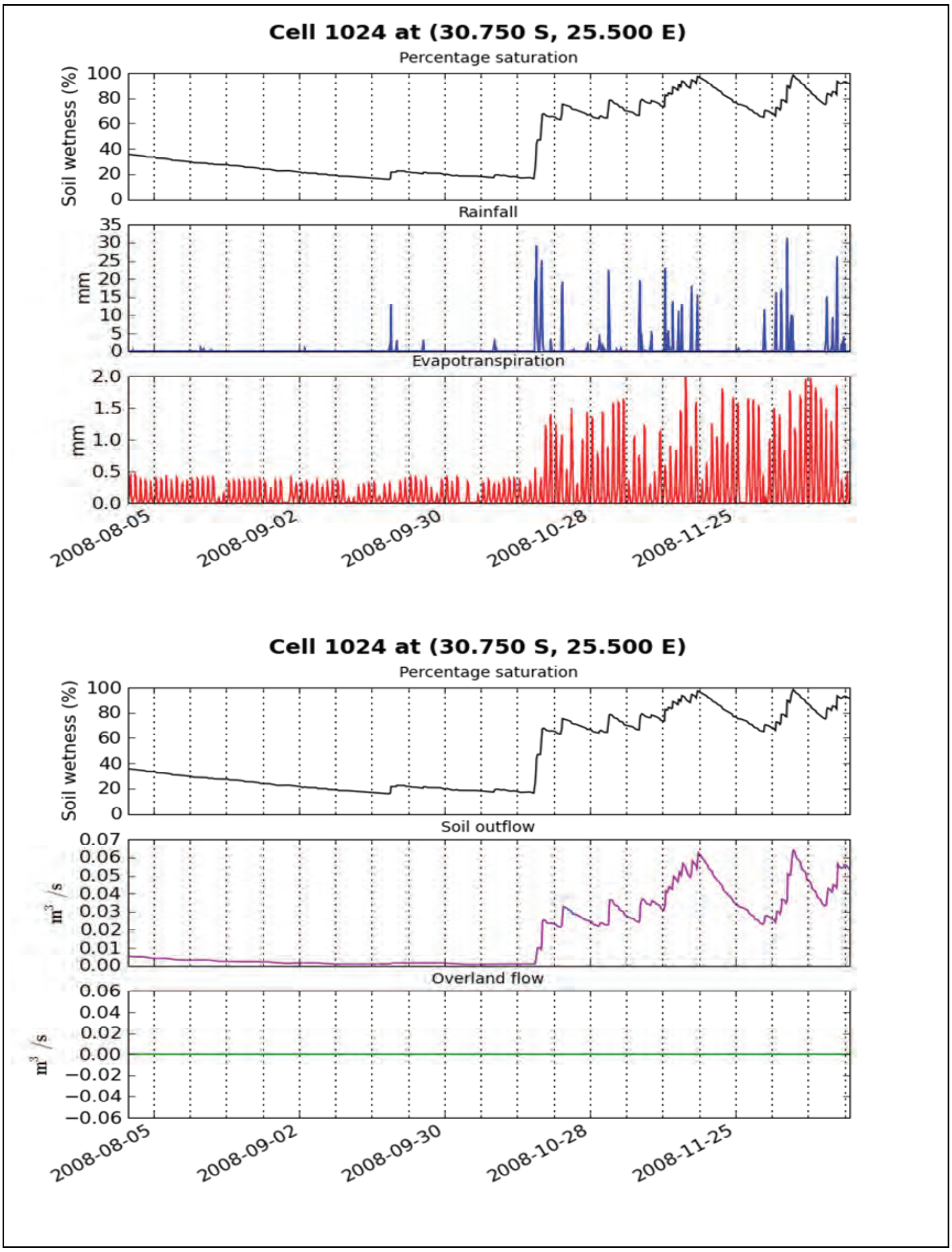


Figure 4.34 Mpumalanga This site experienced earlier rains even than Limpopo in Figure 4.33. It also has higher ET_a (nearing maxima of 2 mm/3h) starting earlier with the earlier rains. Note that ET_a is low before the rains, dropping as the soil dries out and only increasing as the soil wets up (top panel). A diminishing SSI, cloudy and raining days lower ET_a as expected.

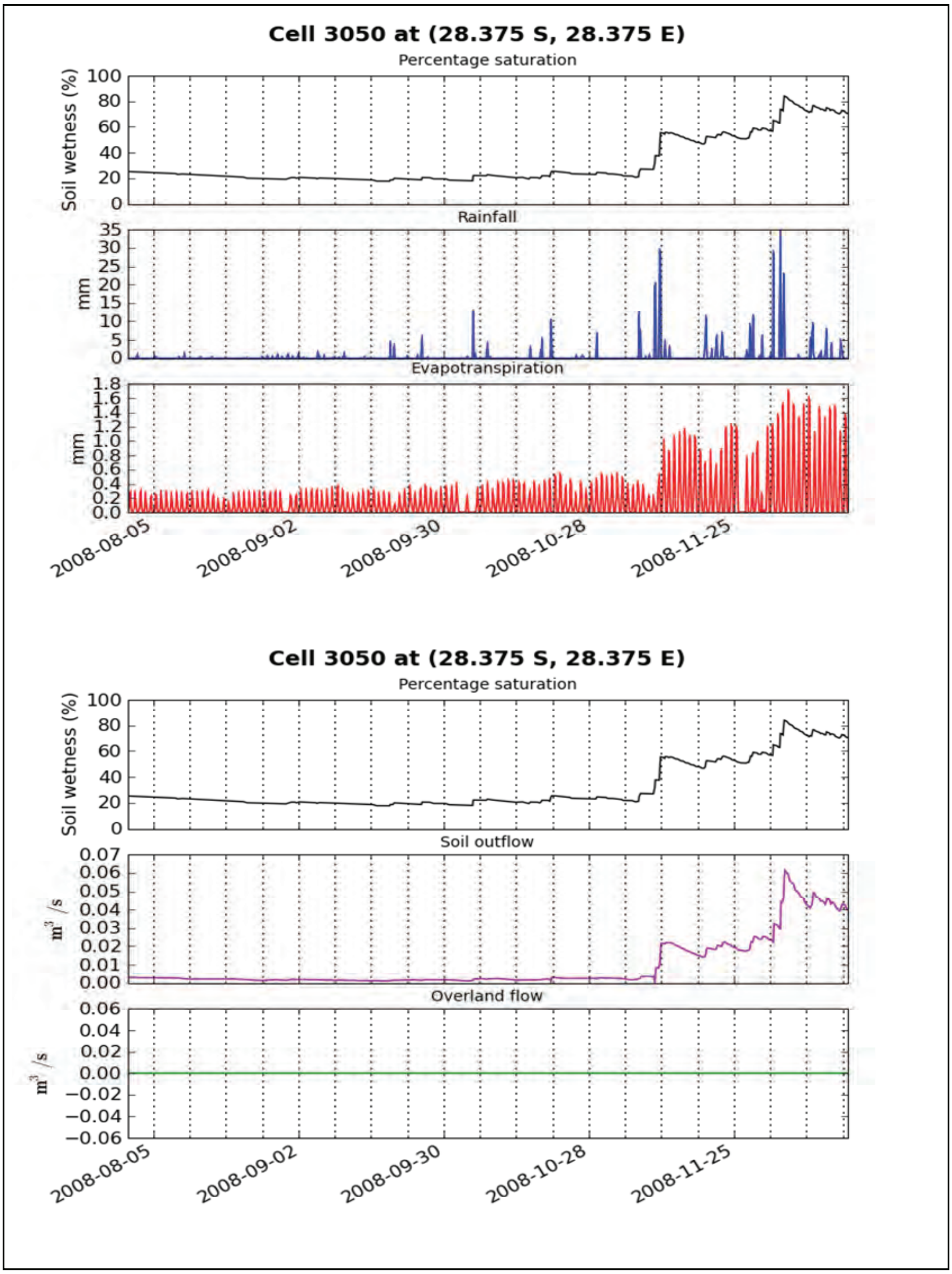


Figure 4.35 Free State The performance of this site is between those at Limpopo and Mpumalanga, and the comments made there are appropriate here

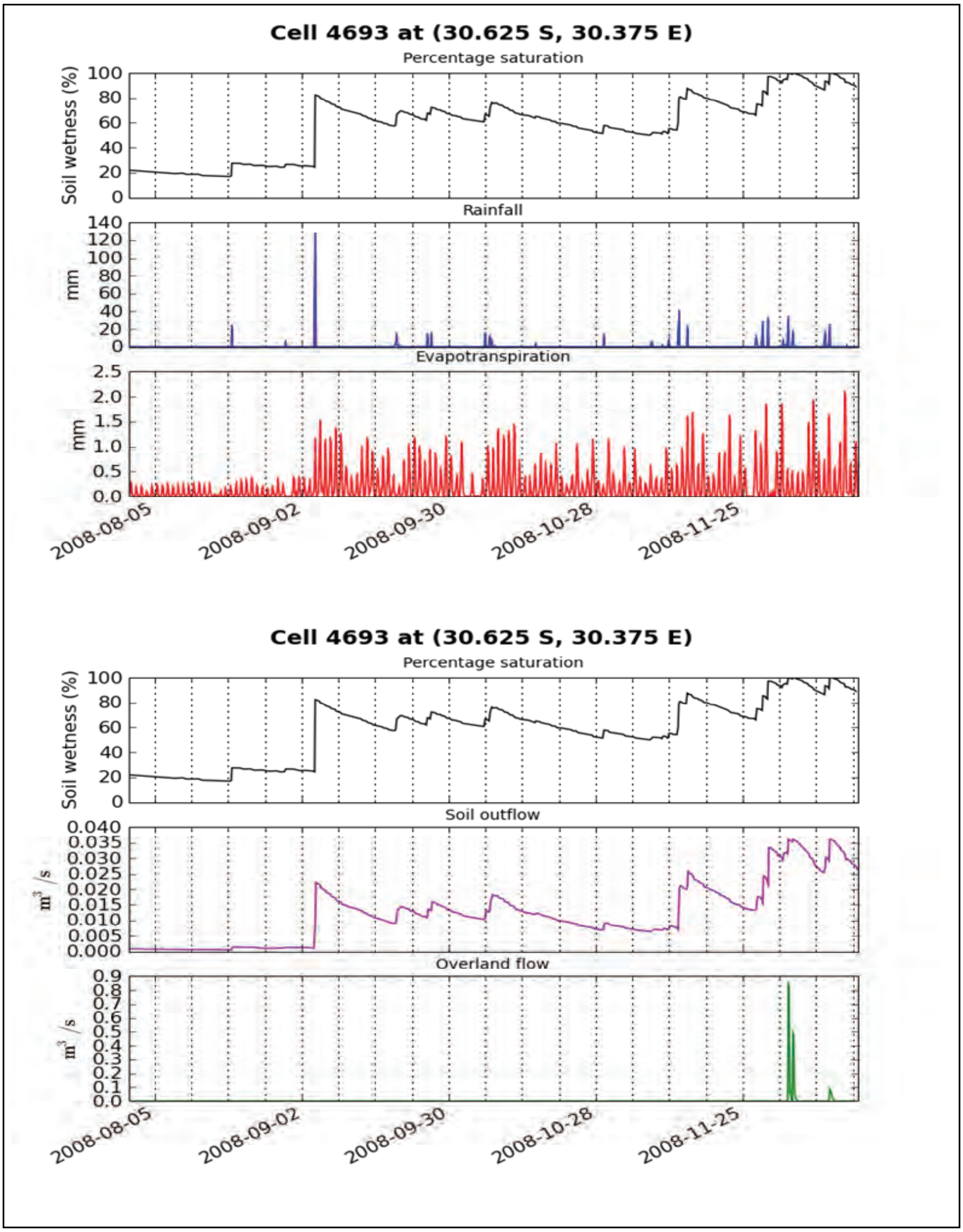


Figure 4.36 KZN Occasional rainfall starting in September kept the soil wet at this site which has all year round vegetation cover as indicated by the relatively high NDVI shown in Figure 4.25. This is the only site that exhibits overland flow after saturation. Due to the persistent higher relative humidity, there is slightly lower ET_a on average than in the interior.

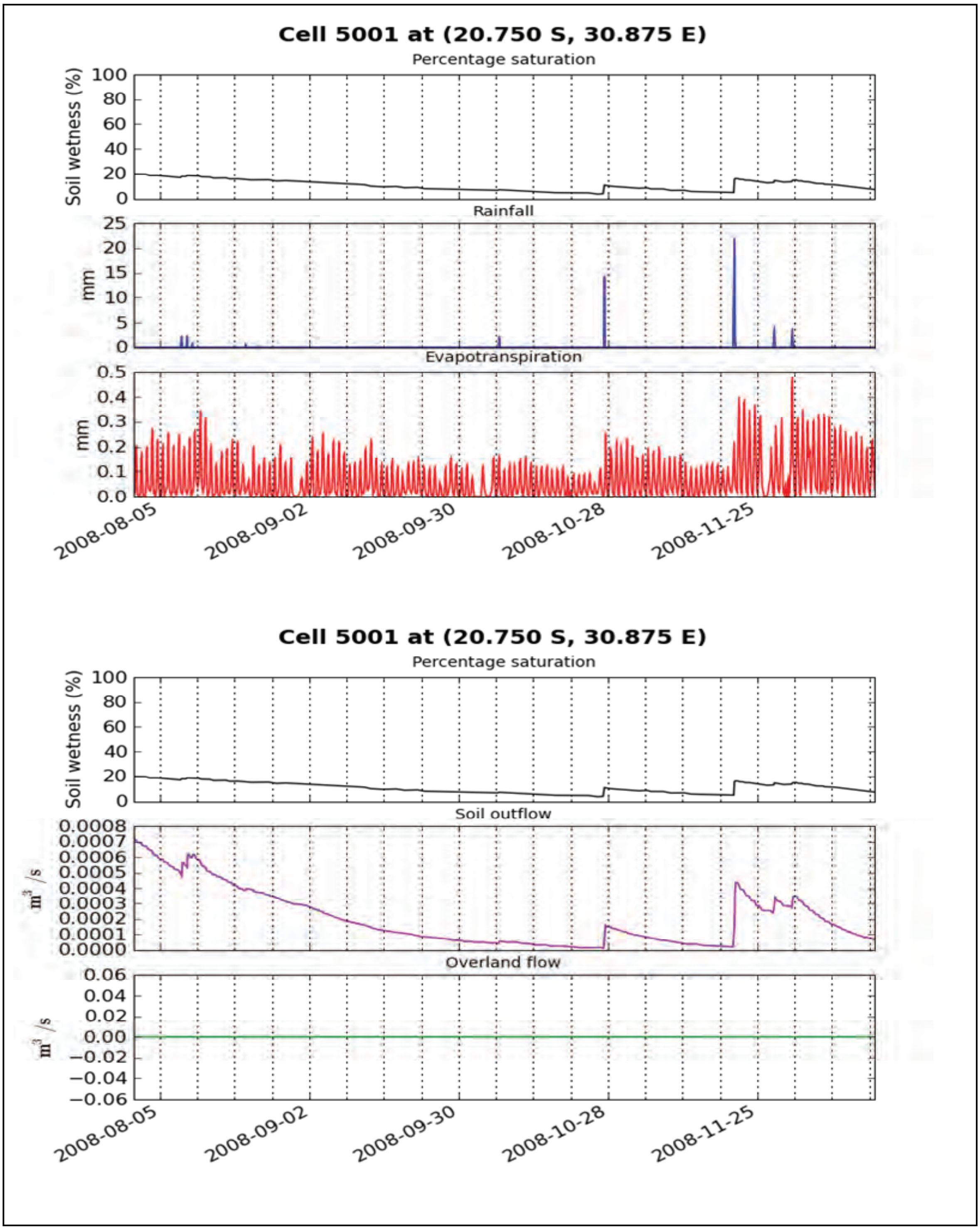


Figure 4.37 Cape This is an example of a dry area, experiencing occasional low rainfall and concomitant low ET_a (maximum 3-yourly rate of 0.5 mm after rain). This ET_a is driven by evaporation from the wet ground, not by vegetation, as can be confirmed in the NDVI panels in Figure 4.25.

SUMMARY

To our knowledge, this work has produced the first detailed country-wide estimate of spatial Soil Moisture at frequent (3 hourly) intervals over South Africa in real time. The methods used to achieve this are a combination of Meteorological variables, assisted by remote sensing estimates of rainfall, energy and vegetation. These are used to force a Land Surface Model (a TOPKAPI cell) at each one of 6989 sites covering over 1 million square kilometers of the country. Novel techniques were used to combine these data streams into the information product.

5. DETERMINING SOIL WETNESS FROM GROUND SURFACE EARLY MORNING TEMPERATURE GRADIENTS

Obtaining Surface Temperature from Brightness Temperature

The first of METEOSAT Second Generation satellites (MSG) was launched in August 2002 and it has become fully operational in geostationary orbit at 0º Longitude and 0º Latitude as indicated in figure 1.

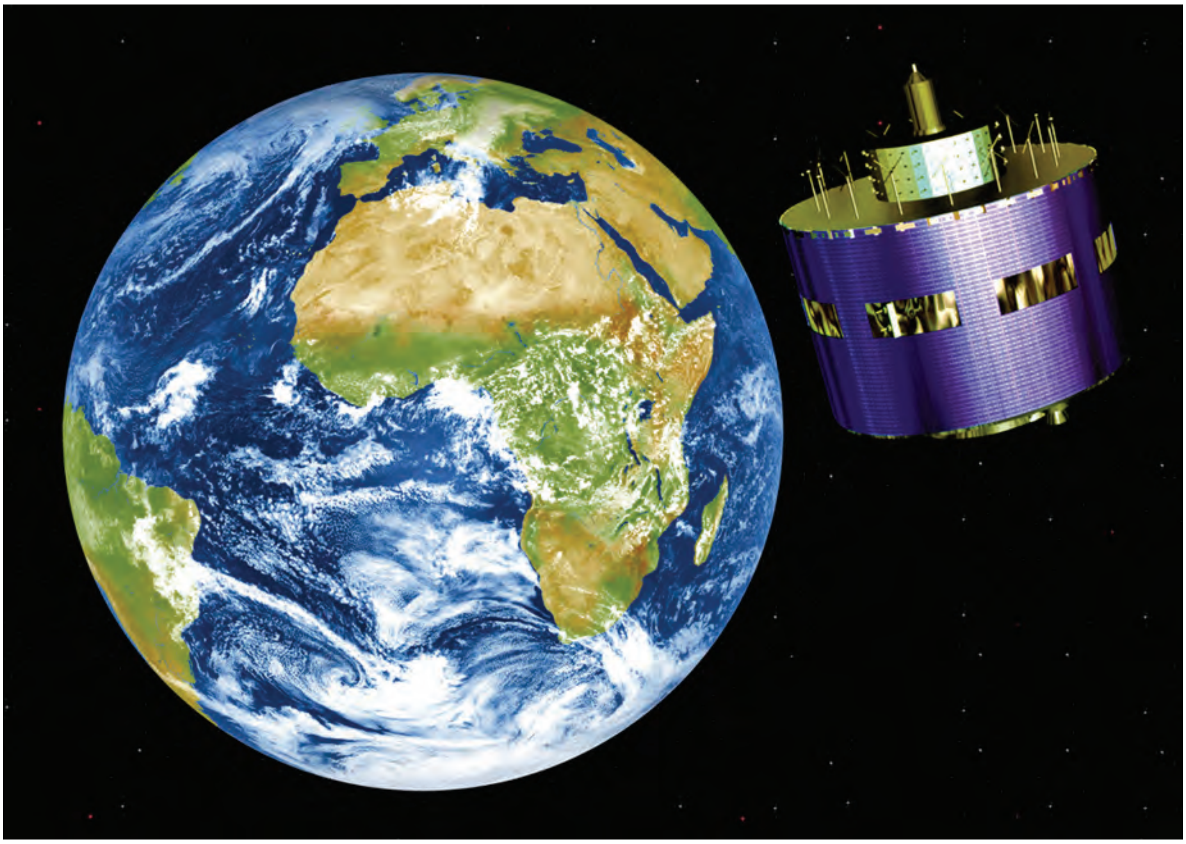


Figure 5.1 Impression of MSG in stationary orbit over 0° Longitude and 0° Latitude, 35 000 km above the earth's surface

The Spinning Enhanced Visible and Infrared Imager (SEVIRI) is the main payload on board MSG1. This optical imaging radiometer consists of three visible and near infrared channels with wavelengths centred at 0.6, 0.8 and 1.6 μm, eight infrared channels centred at 3.9, 6.2, 7.3, 8.7, 9.7, 10.8, 12.0 and 13.4 μm and finally one visible broadband channel at 0.5-0.9 μm called the High Resolution Visible channel (HRV). The spatial resolution for HRV is 1 km and for the other infrared and visible channels is 3 km. This compares with 5 km for first generation METEOSAT satellites. MSG has four visible channels that study the sunlight reflected from the Earth's surface and clouds, in comparison with the previous METEOSAT imager that has only one visible channel. The eight infrared channels on MSG are an improvement compared to the two thermal infrared bands on the first METEOSAT. Four of them are absorption bands of total atmospheric water vapour (WV), CO_2 and O_3, and the remaining four allow the land surface, sea surface, and clouds temperature to be determined (Schmid, 2000).

From the data available, methods of obtaining surface temperature were required. This was achieved by using a two-channel algorithm (also known as a split window algorithm).

Two-Channel Algorithms in Remote Sensing

Each of the satellite's radiometers measures radiation over a finite wavelength band. For any radiance measured within a specified wavelength window, there is an associated temperature such that at that temperature a blackbody would emit the same radiation. This temperature is known as the brightness temperature. If the emissivity of a body being observed was known, and the emitted radiance could be measured, then the body's true surface temperature could be determined.

However, there are numerous limitations in using passive radiance measurements to estimate land surface temperature. One limitation is that there are clouds, which block infrared radiation emitted by the surface from reaching the satellite sensors. If cloudy regions are not appropriately identified, measurements of cloud temperature can be incorrectly associated with land surface temperature. Also, the atmosphere absorbs some of the radiation emitted by the surface and it also emits its own radiation, some of which goes directly to the satellite radiance sensor and some of which is reflected from the surface back up to the satellite. The problems associated with cloud are circumvented by procedures which identify cloudy regions (cloud masking) and the problems associated with atmospheric effects can be remedied by observing the surface radiation at different wavelengths. This is because the effects of the intervening atmosphere are wavelength dependent.

Choosing the appropriate channels transmitted from MSG for the two-channel algorithm was done by Sobrino and Romaguera (2004), who examined the behaviour of the atmospheric transmittance and surface emissivity observed by the SEVIRI thermal infrared channels. With five channels to choose from, there is a total of 10 likely combinations. These are presented in Table 5.1.

Channel *i*	Channel *j*
IR 8.7	IR 9.7
IR 8.7	IR 10.8
IR 8.7	IR 12.0
IR 8.7	IR 13.4
IR 9.7	IR 10.8
IR 9.7	IR 12.0
IR 9.7	IR 13.4
IR 10.8	IR 12.0
IR 10.8	IR 13.4
IR 12.0	IR 13.4

Table 5.1 Likely Two-Channel Combinations from SEVIRI, with the chosen pair highlighted.

The bands IR 9.7 and IR 13.4 are absorption bands of ozone and CO_2 respectively and these channels are basically designated to study these atmospheric components. Therefore, the use of these for a two-channel surface temperature algorithm is not appropriate. Moreover, the band IR 8.7 is situated in a region of the spectrum where the emissivity is hard to determine because of its variability, as shown in the work presented by Sobrino and Romaguera (2004). They found that in this band the emissivity presents the lowest values with higher indeterminate errors. In conclusion, the IR 8.7, IR 9.7 and IR 13.4 SEVIRI bands were discarded as candidates for the two-channel algorithm and IR 10.8 and IR 12.0 SEVIRI bands constitute the best combination for the two-channel algorithm, called the split-window algorithm.

Other than being the most recently published T_S retrieval algorithm, the split window algorithm proposed by Sobrino and Romaguera (2004) was chosen based on several reasons. Firstly, this algorithm was specifically developed for retrieving surface temperature from SEVIRI data in two thermal infrared bands (IR 10.8 and IR 12.0). Also, the proposed algorithm takes into account the SEVIRI angular dependence, meaning that the effects of atmospheric water are taken into account. This is important, because the greater the viewing angle away from the zenith, the greater the atmospheric effects.

This split window algorithm has also been tested with simulated SEVIRI data over a wide range of atmospheric and surface conditions. Adapted from the one given by Sobrino and Raissouni (2000), the land surface temperature, T_S, is given as follows:

$$T_S = T_1 + a_1(T_1 - T_2) + a_2(T_1 - T_2)^2 + a_3(1-\varepsilon) + a_4W(1-\varepsilon) + a_5\Delta\varepsilon + a_6W\Delta\varepsilon + a_0$$

where T_S is the surface temperature (in K), T_1 and T_2 are the at sensor brightness temperatures of two SEVIRI thermal channels $IR_{10.8}$ and $IR_{12.0}$, ε is the mean effective emissivity $\varepsilon = (\varepsilon_1 + \varepsilon_2)/2$, $\Delta\varepsilon$ is the emissivity difference $\Delta\varepsilon = \varepsilon_1 - \varepsilon_2$, W is the total atmospheric water vapour (g/cm^2) and a_i is the angle dependent numerical coefficient of the two-channel algorithm, determined from look up tables provided by Sobrino and Romaguera (2004). Although look-up tables give the coefficients for specific observation angles, a single algorithm that explicitly takes into account the SEVIRI observation angle was preferred.

The algorithm, given by Sobrino and Romaguera (2004) interpreted with their suggested constants, is:

$$\begin{aligned} T_S = {} & T_{IR10.8} + [3.17 - 0.64\cos\theta](T_{IR10.8} - T_{IR12.0}) \\ & + \left[-0.05 + \frac{0.157}{\cos\theta}\right](T_{IR10.8} - T_{IR12.0})^2 \\ & + \left[65 - \frac{4}{\cos^2\theta}\right](1-\varepsilon) + \left[-11.8 + \frac{5.1}{\cos\theta}\right]W(1-\varepsilon) \\ & + \left[-180 + \frac{24}{\cos\theta}\right]\Delta\varepsilon + [-4 + 34\cos\theta]W\Delta\varepsilon + 0.6 \end{aligned} \qquad (5.1)$$

where $T_{IR\ 10.8}$ and $T_{IR\ 12.0}$ are brightness temperatures obtained from the IR 10.8 and IR 12.0 SEVIRI bands and θ is the viewing zenith angle obtained using the method outlined below.

The Determination of the Inputs of the Split Window Algorithm Proposed by Sobrino and Romaguera (2004)

The Satellite's Viewing Zenith

For each pixel on an MSG image, the location of the centre is known. Using this information, the satellite's viewing zenith can be determined. Shown in Figure 5.2 and Figure 5.3 are representations of the spherical angles involved.

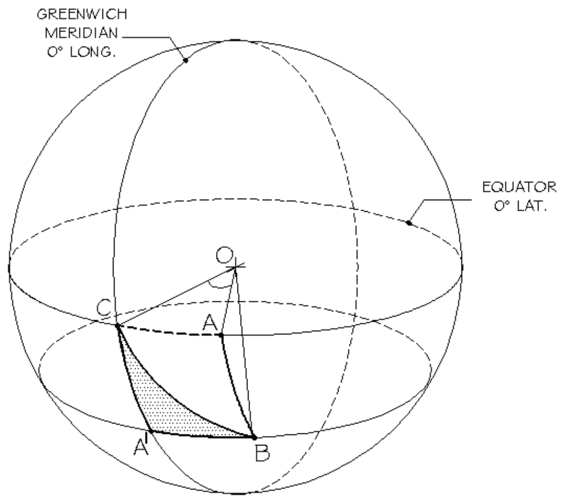
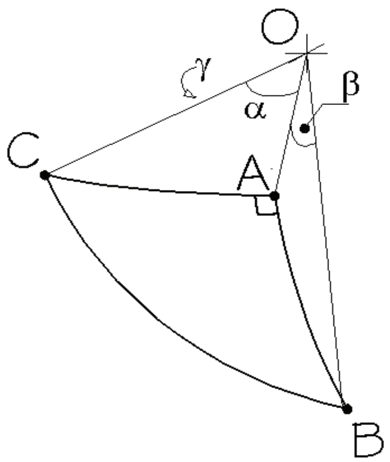


Figure 5.2 Illustration Showing the Spherical Angles Involved.

In Figure 5.2 and Figure 5.3, the point 'C' (0° Longitude, 0° Latitude) is the satellite's nadir[1]. The side A'B of the spherical triangle is at known latitude (that of the pixel of interest), hence using spherical geometry, the angle difference (viewed from the origin O) between points C and B can be determined.



α is the angle subtended at 'O' by arc CA (longitude)
β is the angle subtended at 'O' by arc AB (latitude)
γ is the angle subtended at 'O' by arc CB

Applying Spherical Geometry:

$$\cos\gamma = \cos\alpha.\cos\beta + \sin\alpha.\sin\beta.\cos A$$

$$\cos\gamma = \cos\alpha.\cos\beta \quad (since\ \ A = 90°)$$

$$\underline{\gamma = ar\cos(\cos\alpha.\cos\beta)} \qquad (5.2)$$

Figure 5.3 Detail of the spherical triangle shown in Figure 5.2.

Having determined the angle γ, the solution to the problem is obtained by planar trigonometry as presented in Figure 5.4.

[1] Nadir – point directly beneath a satellite.
(Verbyla *et al.* 1995)

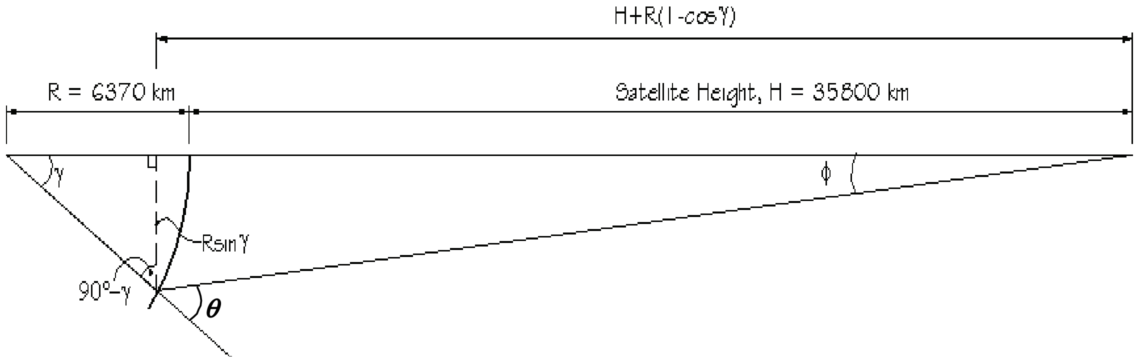


Figure 5.4   The ‘Normal’ trigonometric problem.

From the layout presented in Figure 5.4, the satellite viewing angle ϕ is given by the following expression:

$$\phi = \arctan\left[\frac{R\sin\gamma}{H + R(1-\cos\gamma)}\right] \tag{5.3}$$

where, R is the radius of the Earth taken as 6 370 km, γ is an angle dependent on the latitude and longitude of the pixel’s centre and H is the height of the satellite above the earth’s surface given as 35 800 km. The satellite zenith (given in radians) is hence found using the following relation:

$$\theta = \pi - (\gamma + \phi) \tag{5.4}$$

This value is used in equation (3.1).

The Emissivities of the Two Channels

The emissivity values used in the algorithm were obtained with the help of work presented by Sobrino and Romaguera (2004) who used a database by Salisbury and D’Aria (1992) to select a total of 75 spectral values that are representative of the Earth surface. The mean emissivity and standard deviations per channel were then obtained for the five IR SEVIRI channels and the results are as shown in Figure 5.5.

Otherwise, the channel emissivity can be estimated using the TISI method according to the procedure given by Dash et al. (2002b).

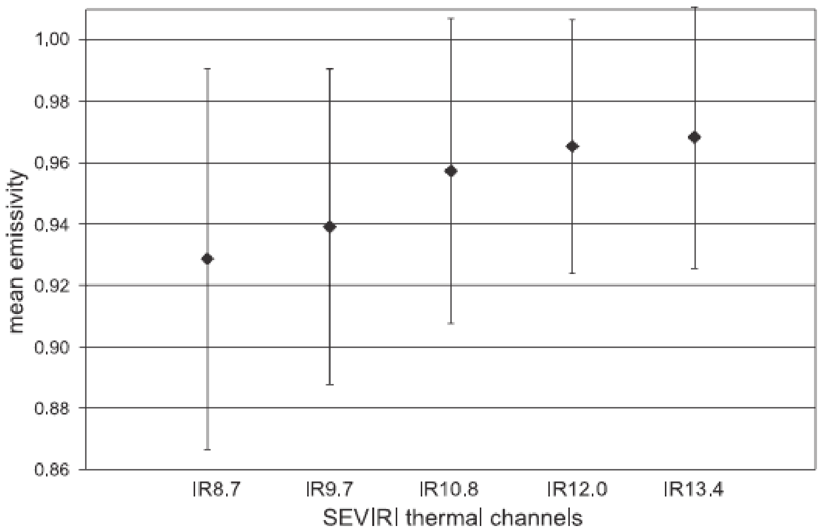


Figure 5.5 Figure 32 Mean Emissivity and Standard Deviations Obtained from 75 samples of the Salisbury and D'Aria (1992) database used by Sobrino and Romaguera (2004) for the Five IR SEVIRI Channels.

The Atmospheric Path Water Vapour, W

The total atmospheric water vapour, W in g/cm^2 is dependent on the satellites viewing angle and the retrieval of this parameter can be carried out by means of a split-window technique according to Li et al. (2003).

Estimating Soil Saturation Index (SSI) from Early Morning Temperature Gradients at 1 arc-minute resolution

The estimation of SSI is to be done on an area measuring 1024 by 1024 minutes of arc over Southern Africa as shown in Figure 5.6. This area matches the area covered by SIMAR, a product of previous WRC projects K5/1151, 1152 & 1153 currently maintained by SAWS. Currently, MSG1 IR data useful for extracting gradients are archived daily for this whole area (there is a complete set of early morning surface temperatures from mid June 2005 to the present archived on the SAHG server) and a typical archived brightness temperature image is shown in Figure 5.6.

At this stage when this work was started, we analysed test areas within this 'square', but it was subsequently realised that it would be beneficial to expand the archived region to include Africa South of the Equator.

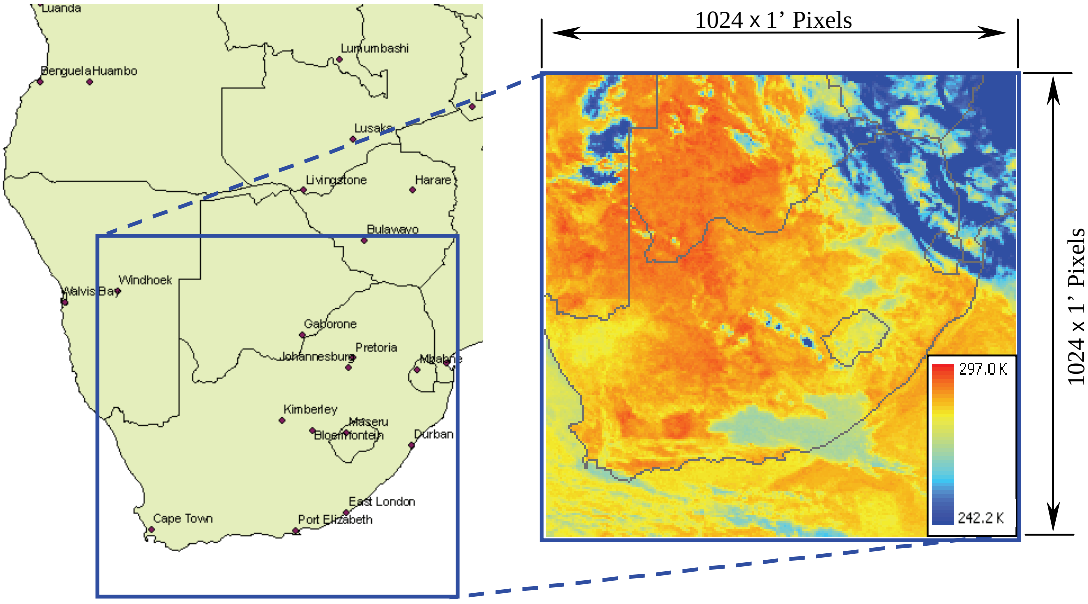


Figure 5.6 A typical brightness temperature image stored for a 1024 × 1024 pixel area over Southern Africa, an area matching the SIMAR (Kroese, 2004; Deyzel et al., 2004; Pegram, 2004) map.

The algorithm

Surface Temperatures T_S are calculated from Brightness Temperatures at each pixel every day, every 15 minutes from 8 am to 10 am local time. These are archived in addition to the original IR data. After enough time had elapsed (1½ years) for a reasonable range to have been recorded, these are ranked and their frequency distribution obtained. Outliers are excluded, then the range is compared with the range of soil moisture for given soils, whose properties are already known and mapped. The figures that follow indicate how the process evolved.

First the temperature gradients were observed daily for a given time interval. Figure 5.7 shows the warming up of the pixels on a test-area within the Liebenbergsvlei catchment over 2 hours at 15-minute intervals. The mosaic of images sampled at 15 minute intervals on the left of Figure 5.7 was obtained after a dry spell, while that on the right was obtained two days later rafter rain – note the cold (blue) clouds in the second figure (blue means cold, red means hot) moving over successive clips.

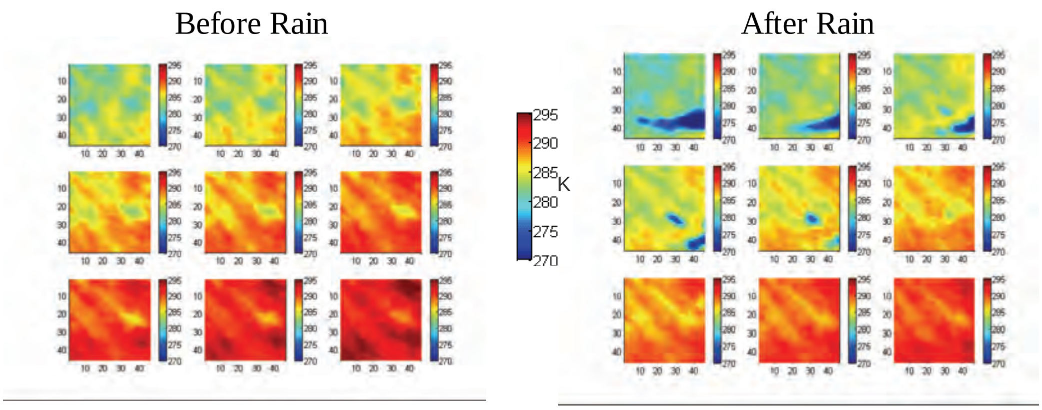


Figure 5.7 Surface temperatures at 15 minute intervals between 08:00 and 10:00 local time over a test area, two days apart, with intervening rainfall. Note generally lower temperatures and the cold (blue) clouds on the right image (after rain) moving through the area.

The temperature gradients measured on individual pixels over small "hydrologically homogeneous" areas as shown in Figure 5.8 were used for testing the assumptions and examining the outlier gradients.

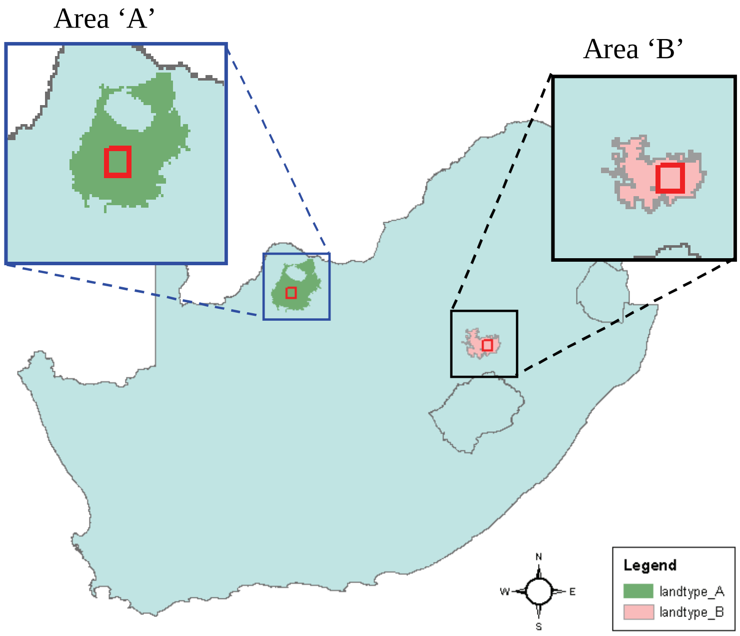


Figure 5.8 Chosen test sites A & B measuring 15′ square; the projection is in lat/long.

The first series of investigations were carried on site 'B', measuring 15′ square situated in the Eastern part of South Africa and this site was chosen to fall entirely within an area with known soil properties as shown in the insert in Figure 5.8. Site 'B' can be described as follows:

Location of Top Left Corner: 28º 23′ 33″E, 27º 47′ 18″S
Size: 15′ square i.e. 225 pixels in total
Top Soil Properties: W_{P1} = 0.146 m/m, F_{C1} = 0.226 m/m, soil depth = 0.24 m.
Sub-Soil Properties: W_{P2} = 0.203 m/m, F_{C2} = 0.273 m/m, soil depth = 0.41 m

Relating Temperature Gradient to Soil Saturation Index and Soil Moisture

A steep temperature gradient is expected for drier soils and the converse for wet soil, so the lowest recorded temperature gradient can be said to correspond to the drier of the known soil properties, the Wilting Point. The converse can be said to hold true for Field Capacity. This relationship (choosing a linear one as a prior assumption) is illustrated in Figure 5.9. [**NB** In this chapter, the terms Wilting Point and Field Capacity are informally equivalent to θ_r and θ_s respectively. The reason is that, historically, this work predates that on TOPKAPI reported in Chapter 2 and the Appendices. The distinctions are minor and do not affect the conclusions concerning the use of the method appearing at the end of this chapter.]

The relation between soil moisture θ and temperature gradient T_G for a given soil type is given by the following expression:

$$\theta = \frac{F_C - W_P}{T_{GL} - T_{GU}} \cdot T_G + C \qquad (5.5)$$

where FC and WP are the known soil parameters (Field Capacity and Wilting Point respectively), T_{GL} and T_{GU} are the lowest and highest observed temperature gradients for the soil type respectively. TG is the observed surface temperature gradient ($T_{GL} < T_G < T_{GU}$), which corresponds to a soil moisture of θ and C is a constant. Of more use for comparison purposes, and from which θ can be recovered, is the Soil Saturation Index:

$$\mathrm{SSI} = (\theta - W_P)/(F_C - W_P) \qquad (5.6)$$

which is also dimensionless and which lies in the interval (0, 1).

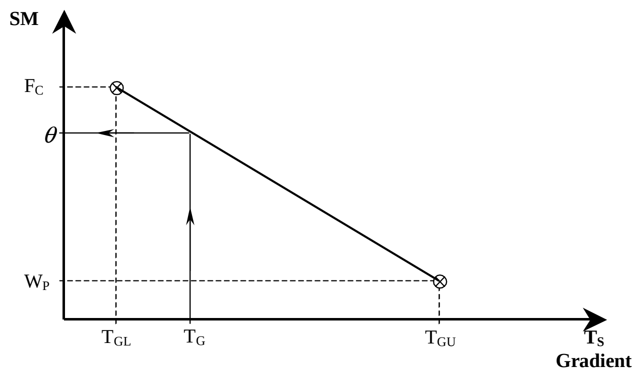


Figure 5.9 Relating Surface Temperature to Soil Moisture Conditions

It is important to identify the upper and lower bounds of temperature gradients with reasonable accuracy. The following section describes our strategy.

Outlier Detection

In the context of the temperature gradient dataset that was derived from MSG data at UKZN-SAHG, an outlier is an observation that is an abnormal distance from other values in the TS gradient distribution. In a sense, this definition leaves it to the analyst (or a consensus process) to decide what will be considered abnormal (Adamson, 1989). Before abnormal observations can be singled out, it is necessary to characterise what are normal observations. For the investigations, an exploratory set of calculations was done and these were later refined.

Figure 5.10 is the histogram of the average gradients recorded on Test Area A from mid-June to the end of December 2005. The observations that might be considered abnormal are identified roughly, but the choice of the cut-off level of 16°K/h is arbitrary; in this case it might need to be as low as 13. Robust statistics were used based on the interquartile range to define these limits.

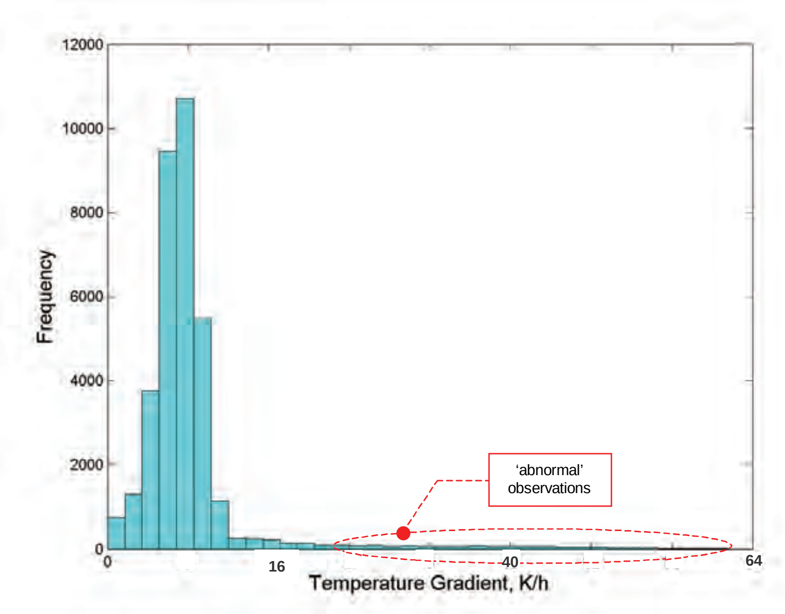


Figure 5.10 T_S Gradient Distribution Plot for test area A showing obvious outliers (circumscribed in red). Data shown are the daily averages from 14th of June to 31st of December 2005.

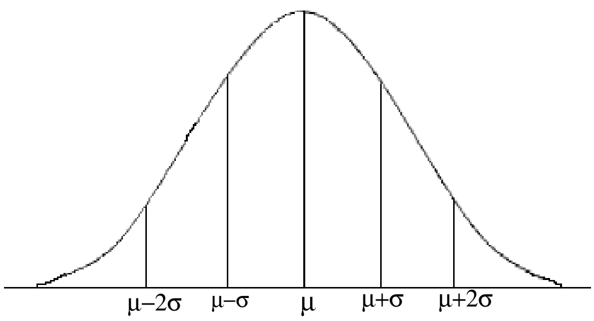


Figure 5.11 The Normal distribution density.

For a standard normal curve, as shown in Figure 5.11, the mean, μ = 0 and standard deviation, σ = 1. If a dataset follows a normal distribution, then it is known that 68% of the observations will fall within one standard deviation of the mean (or median, which coincides with the mean in this case), 95% of the observations will fall within 2 standard deviations of the mean and 99.7% of the observations will fall within 3 standard deviations of the mean.

Hence for the dataset of T_S gradients, to minimise the effect of the outliers which contaminate moment estimates of statistics, the standard deviation can alternatively be estimated using the following relation if there is near symmetry in the distribution:

$$\sigma = \frac{Interquartile\,Range}{2 \times [z_{0.75} - z_{0.5}]} \qquad (5.7)$$

where the *Interquartile range* is the difference between the upper and lower quartile values of the observations. The distance in the standard normal from the median to the upper quartile is $z_{0.75} - z_{0.5}$ = 0.6745, so the denominator becomes 1.349.

In contaminated data the mean is not used as the measure of central tendency, but the median, which is not influenced by the outliers.

The data giving the histogram in Figure 5.12 (which appear to be approximately symmetrical near the mean/median) has a median of 7.164 an interquartile range of 2.816. Hence using the relation given in equation 3.7, the estimate of the standard deviation is 2.088.

In this context, it is suggested that the data points that lie beyond 3 standard deviations from the median can be considered very unlikely, hence they are flagged initially as outliers. This idea was applied to the dataset where the lower and upper cut-offs for 3 standard deviations from the median are 0.904 and 13.42 respectively. The resulting distribution plot, modified from Figure 5.10, is shown in Figure 5.12. Of course, if the lower cut-off were to be zero or below, this would be nonsensical in the context of soil being warmed by solar radiation, so this bound would need to be modified to a sensible positive value. This effect might also be due to lack of symmetry, in which case the upper and lower quartiles would be individually estimated and used for defining the relevant bounds.

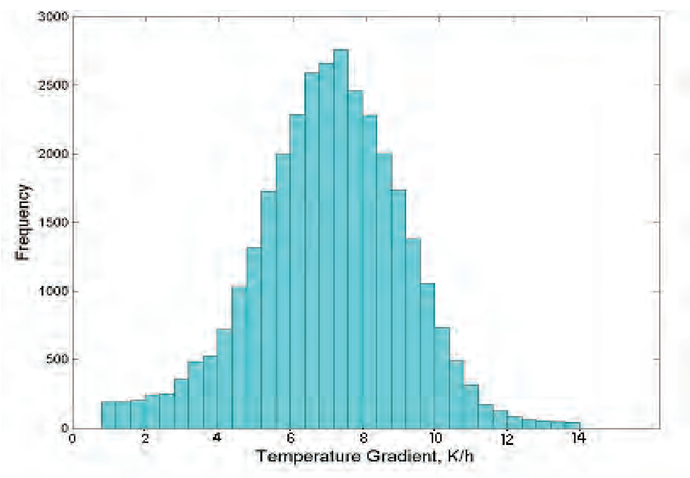


Figure 5.12 Trimmed data from 6 months of gradients

The next step is to identify those events which caused the outliers and hence obtain a physical explanation for why they should be excluded from the data-set. It is important to define the upper and lower bounds with some precision, because they are to be matched with the soils' Wilting Point and Field Capacity respectively. However, a note on precision is relevant here.

If a different value of cut-off is chosen instead of 3, say 2.7 standard deviations either side of the median, the range of T_S values will be reduced by 10% at the extremities. This will likely affect about 1% of the observations (about 2 or three days a year), but will not modify the body of the relationship which is stretched or shrunk around the median. It is also worth noting that the upper end of the T_S distribution is associated with dry soils and the outliers are likely to be due to cloud clearing and exposing a warmer ground surface or perhaps to fires. The problem for flood forecasting, however, is at the lower end of the distribution of gradients. Here, contamination is likely to be due to cloud cover, water or missing data. These need to be identified and excluded piecemeal.

Characterisation of Outliers

The trimmed data were further investigated in order to give a physical explanation of the cause of the outliers. That is: the data that lie outside the cut-off values were further classified as being due to partial cloud coverage, say, for the lower end of the T_S distribution and as having fire perhaps for the upper end.

For this purpose, the days that had T_S gradients outside the specified range were first flagged then the percentage of pixels that had gradients outside the specified range was determined per day. From this, the satellite images were revisited to determine if there indeed was cloud accounting for the predominately negative gradients and to explain what could have caused the excessively positive gradients at the upper end.

Classification of Negative T_S Gradients

For the days with negative T_S gradients, it was found that the major cause was the presence of cloud, as expected, at a particular interval (08h00, 08h15 etc) or throughout the viewing period. An example of the different cases is illustrated in Figure 5.13 to Figure 5.16. Figure 5.13 shows the test area starting off with clear skies and as time progresses it becomes covered by cloud. A typical pixel temperature sequence was then plotted over the viewing period and the results are shown in Figure 5.14.

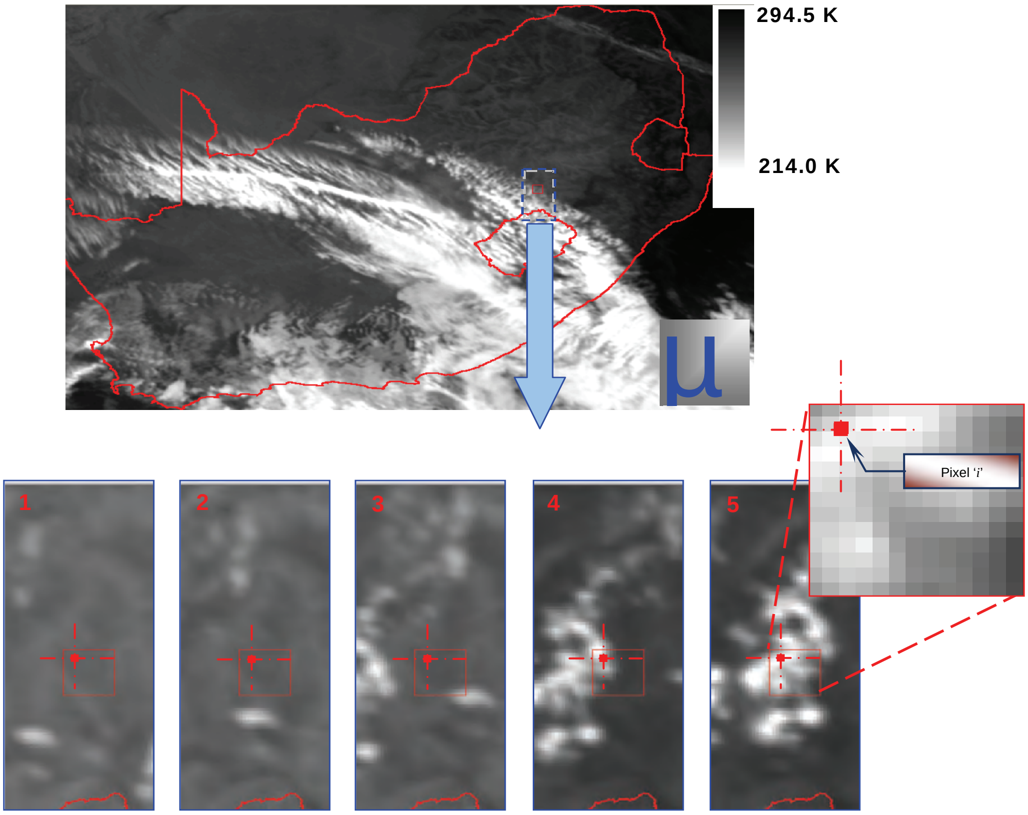


Figure 5.13 The upper image is one of the many IR images which have been archived and which were analysed to give the gradients, also been archived; note that white is colder than black in this interpretation. The lower sequence of images is a set of enlargements of the region at intervals of 15 minutes showing clouds moving Easterly over the test area during the 1-hour interval 0800 to 0900 on 5th of July 2005. The further enlargement on the right shows the location of the pixel '*i*' whose data were extracted to make Figure 5.14.

It was initially thought that a lower temperature threshold could be identified to exclude data which must be cloud. This "guess" of 265°K is shown in Figure 5.14 and is seen not to be a good choice of outlier detection method. It was decided that the fitted r^2 values be used as indicators instead.

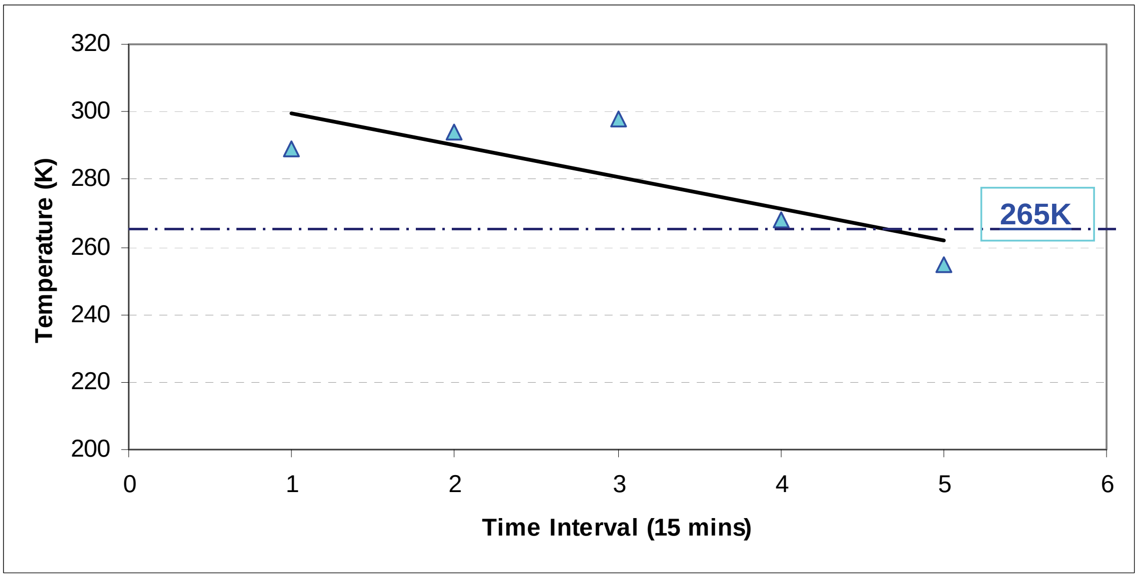


Figure 5.14 Graph illustrating T_S gradients change for the observation period shown in Figure 5.13 for pixel '*i*'. The fitted trend-line has an r^2 of 0.692

Classification of Positive T_S Gradients

It was found that the major cause of high positive T_S gradients, was temperature variation caused by presence of cloud at given time intervals. Unlike for the days with negative T_S gradients, unusually high positive T_S gradients are caused by the pixels starting off cold (covered by cloud) then as the clouds move away from the area there is a sudden jump in temperature. This is then calculated by the linear regression as a high positive T_S gradient. An example of this case is illustrated in Figure 5.15 and Figure 5.16. Figure 5.15 shows the test area initially covered by cloud and clearing with the progression of time.

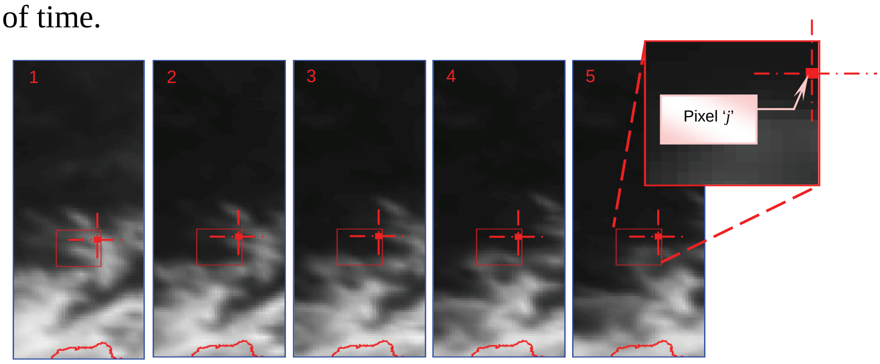


Figure 5.15 Five image sequence over test area 'B' showing cloud (white) cover over the area gradually moving away (17th October 2005).

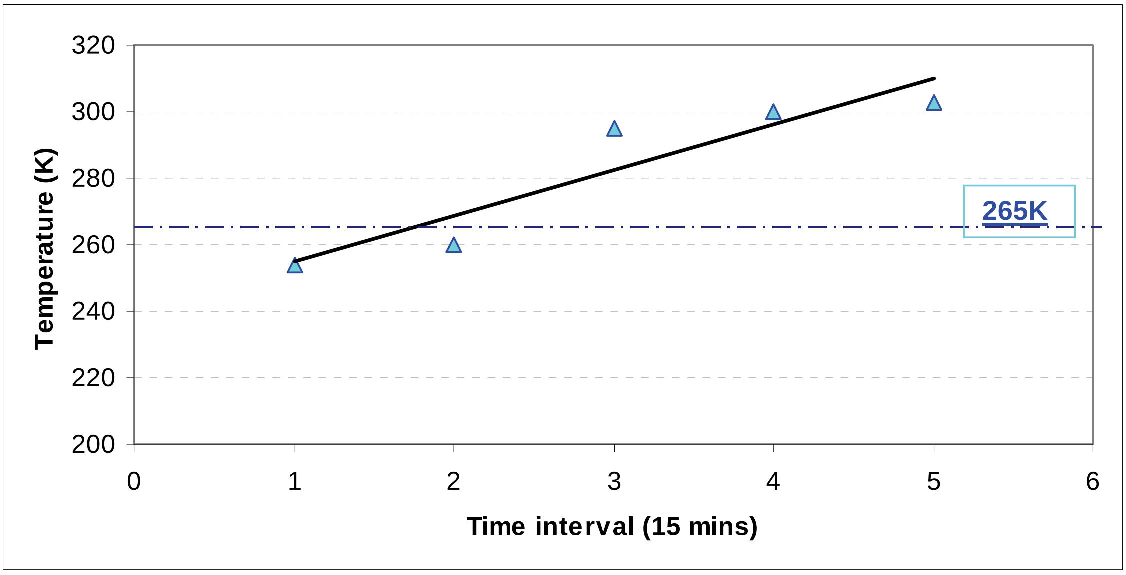


Figure 5.16 Graph illustrating T_S gradients change for the observation period shown in Figure 5.15 for pixel '*j*'. The fitted trend-line has an r^2 of 0.817.

An alternative approach to outlier detection

Based on the observed trends shown in Figure 5.14 and Figure 5.17, an alternative means of detecting outliers was investigated. This method involves the determination of the correlation between temperature gradient and time for the 5 data points observed for each pixel for a particular day. The r^2 was hence calculated per pixel per day and the findings are presented in what follows.

For the plot shown in Figure 5.14, the r^2 value was calculated for full five observations for pixel '*i*' (5 points). A separate trend line was then fitted through the data expected to be characteristic of the heating of a land surface. Then, the r^2 value was recalculated using only these values and compared to the first r^2 value. The modifications are shown in Figure 5.17. In summary, it was decided to work solely with the r^2 as a discriminator to define the upper and lower cut-offs from the archived data.

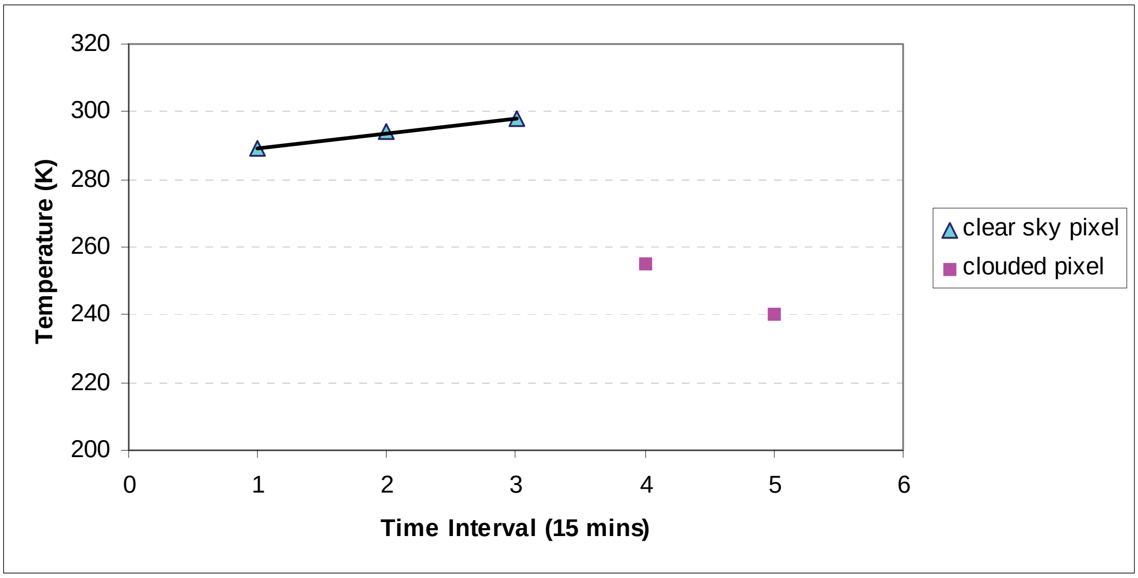


Figure 5.17 The data points shown in Figure 5.14 with a trend line fitted to the data relating to clear skies for pixel '*i*', $r^2 = 0.996$.

From these observations, the use of daily r^2 values as a way of detecting outliers was further investigated. Firstly, a histogram for daily r^2 values for the same area was plotted. The resulting plot is shown in Figure 5.18.

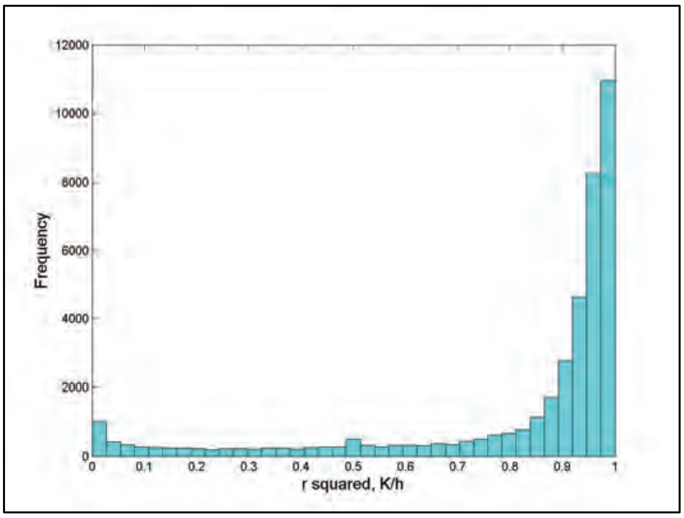


Figure 5.18 The r^2 Histogram Plot for the test area. Data shown is from the 14th of June to the 31st of December 2005.

More careful trimming and the seasonality factor

Gradients as a function of r^2 were examined in detail for all 12 months (June was split in 2005 and 2006 and these data were compared then pooled once they were found to be consistent). For example in the month of December the values of gradient as a function of r^2 are given in Figure 5.19.

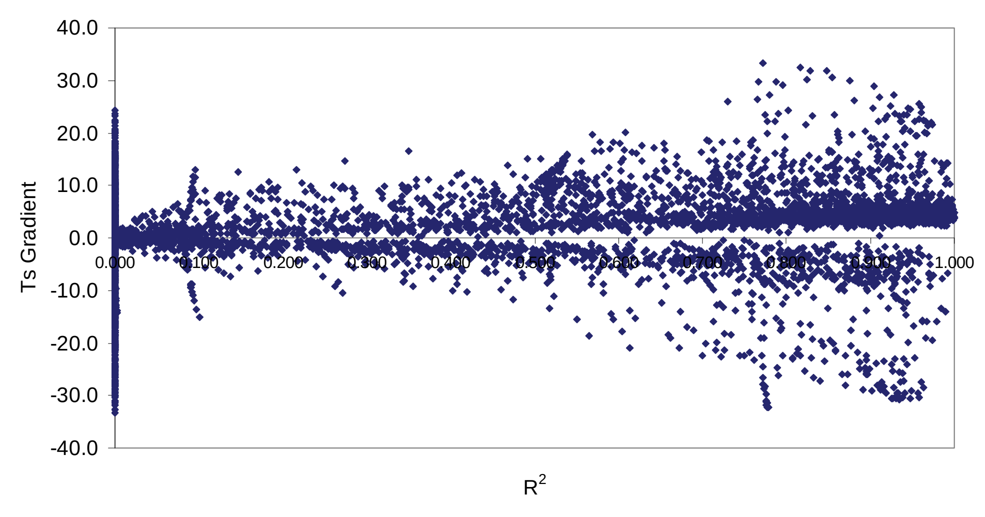


Figure 5.19 Temperature Gradient – R^2 plot for January 2006 Over Test Area 'B'

If data with $r^2 > 0.925$ are selected, then the plot appears as in Figure 5.20.

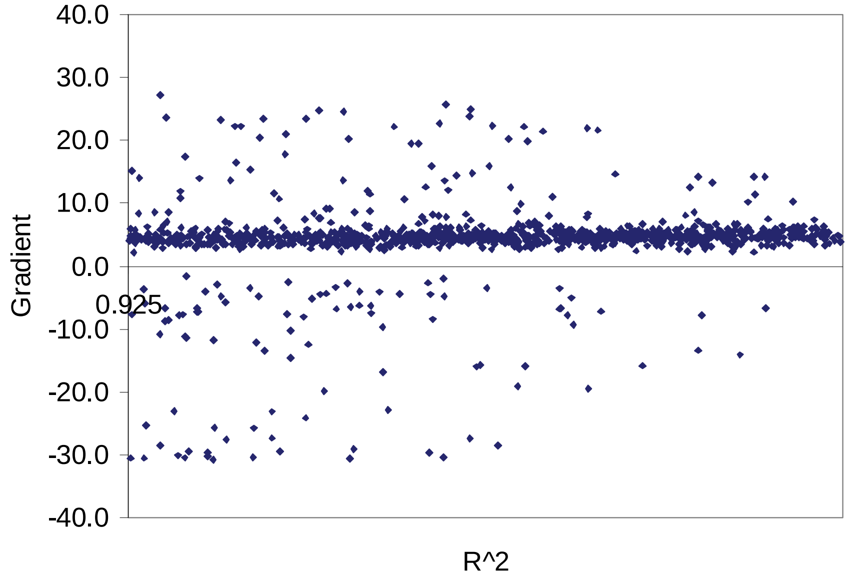


Figure 5.20 Temperature Gradient – r^2 for January 2006, with data censored at $r^2 > 0.925$.

It will be noted that there are no points with *positive* gradients below 2 and that the sparse points above the dense cluster lies above a threshold of about 8 degrees per hour. When plotted in a histogram, these data appear as in Figure 5.21.

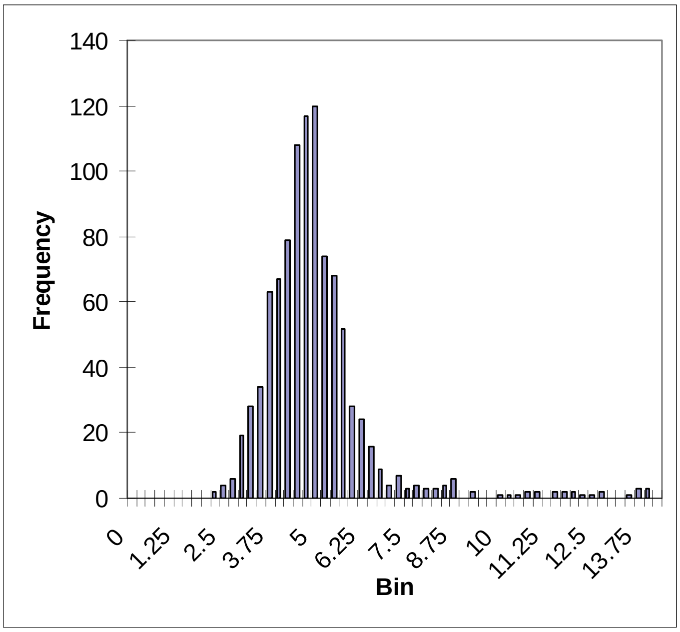


Figure 5.21 Data from Figure 5.20 rearranged in a histogram, after trimming the gross outliers

Outliers are still evident above about 7.5 and below 2.5. These points were refined using the quartile argument introduced earlier and the values calculated for all the months. Finally, the values of T_{GU} and T_{GL} were obtained using robust statistical arguments to exclude outliers. The smallest number of data in a month surviving the initial pruning were in February 2006 (267) and the most in a month were in July 2005 (5036), suggesting that these were the wettest and driest months, respectively, in this particular year. Their behaviour over the year, with a fitted Fourier series (2 harmonics) appears in Figure 5.22.

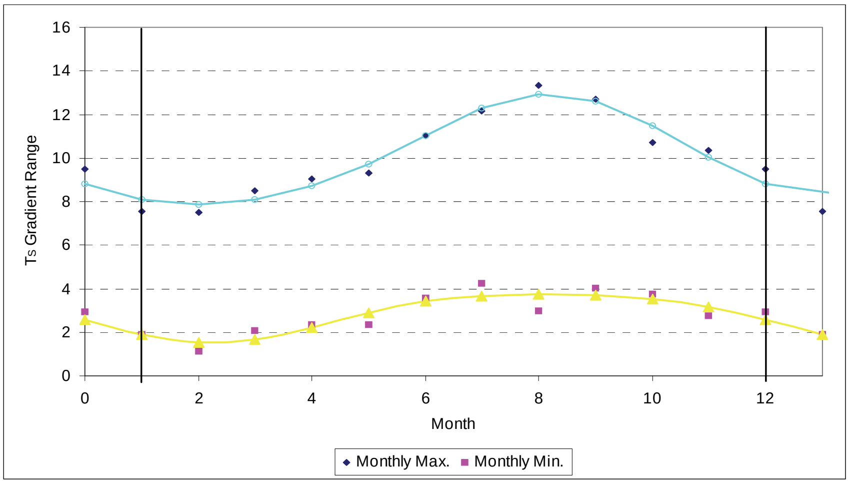


Figure 5.22 Upper and lower threshold values of temperature gradients obtained from the one year of data over Test Area 'A'. The data are fitted by Fourier series of 2 harmonics each and the first and last months are indicated by vertical lines in a diagram designed to reflect the periodicity.

Further work done was with those data whose regressions yield coefficients of determination greater than 0.925 and which lie between the thresholds given in Figure 5.21. These ideas were developed and tested on the rest of the region later in this section.

In 2006, it was thought that the estimation of temperature gradients is meaningful for interpreting the Soil Saturation Index. The derived data were archived and a method developed from a Test area with homogeneous soils which shows promise for performing two important tasks: selecting those data which are meaningful and defining the upper and lower bounds for SSI determination. It is easy to derive the estimate of Soil Moisture content on each pixel on each morning, because we know (in RSA) the soil properties in detail; for example we have (among others) Wilting Point and Field Capacity mapped over the whole country at the same resolution as METEOSAT-8 as shown in Figure 5.23. It turns out that our optimism was to be dampened drastically by the high frequency of cloud cover.

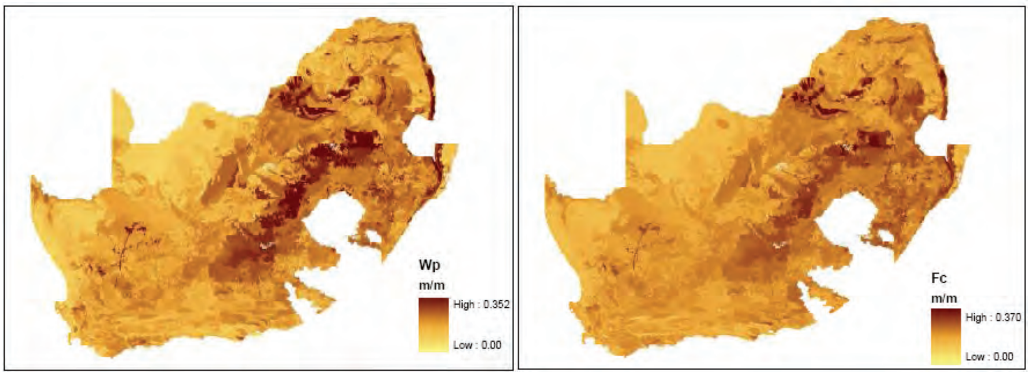


Figure 5.23 Maps of soil properties in RSA (Wilting Point and Field Capacity) at the METEOSAT-8 pixel scale

Working with the Meteosat projection

One of the important lessons learned in this work on temperature gradients was the handling of data and the transfer of data sets from one projection to the other. This rather technical subsection gives a summary of the work than enabled the above results. These computer intensive studies are easily hidden by the need to communicate pictorially, but are the very nerves which make the body of this work come to life.

It is well known (e.g. Gotway and Young, 2002) that changing the scale and shape of data elements in spatially averaged data will affect the properties of the observed quantities. However, the exact nature of these changes is not well understood, as evinced by the numerous statistical procedures available to account for such changes of support, some of which are introduced in the following section.

In order to minimise the effects discussed above and set the stage for a more robust soil moisture estimate, in 2007 the project team elected to reduce re-sampling and re-projecting of the data as far as possible. The Meteosat based temperature gradients were therefore be produced at the original projection of the Meteosat data before being resampled onto the product grid, an extension of the SIMAR grid (Kroese, 2004; Deyzel et al., 2004; Pegram, 2004). An SSI is then calculated for each pixel in the product grid and subsequently estimates of soil moisture produced as a final product.

Figure 5.24 and Figure 5.25 are provided as a motivation for taking this approach. Figure 5.24 shows the regular grid formed by the Meteosat-8 pixel centres (shown as black crosses) at a spacing of 100 pixels when superimposed on a Geostationary projection of the earth (reference Longitude 0°). For comparison, Figure 5.25 shows a portion of the earth mapped onto a square lat-long grid. The Geographic co-ordinates of the Meteosat-8 pixel centres are shown as black crosses (again with a spacing of 100 pixels in the Geostationary projection). A large degree of distortion (in the size and shape of the originally square pixels) is evident as the location becomes more distant from the origin of the Geostationary projection (at 0° Latitude, 0° Longitude).

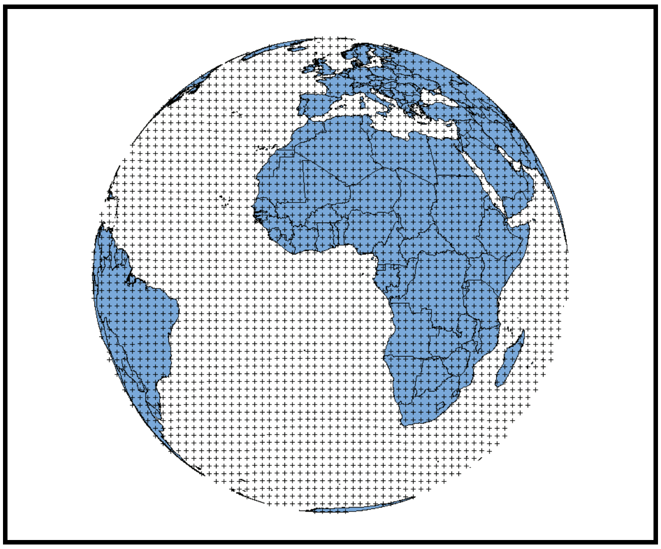


Figure 5.24 Location of the Meteosat-8 pixel centres in the Geostationary satellite projection.

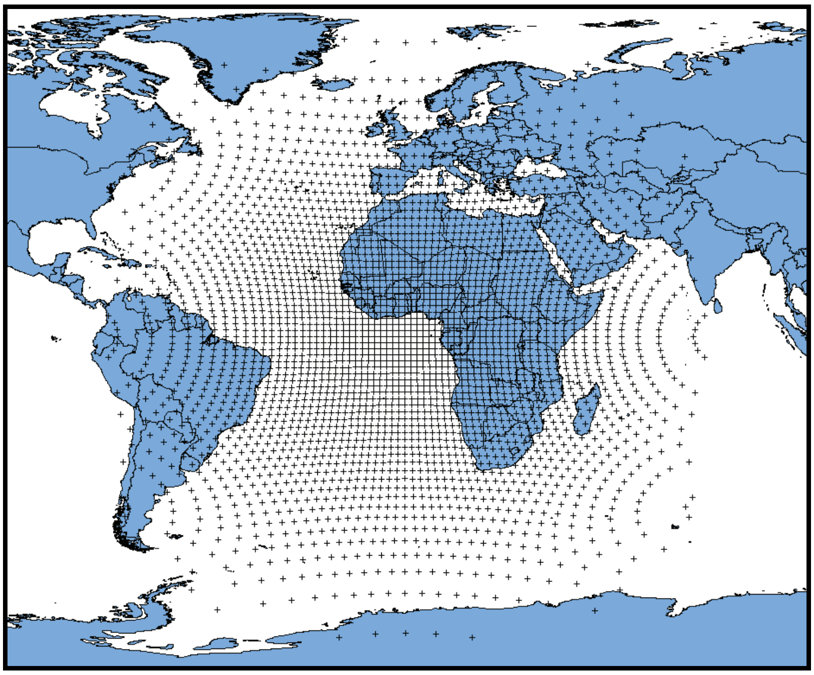


Figure 5.25 Location of the Meteosat-8 pixel centres on a square Lat-Long grid. The black crosses represent the location of the pixel centres at a spacing of 100 pixels in the Geostationary projection. The black crosses are at the same geographic locations as those in Figure 5.24 for comparison.

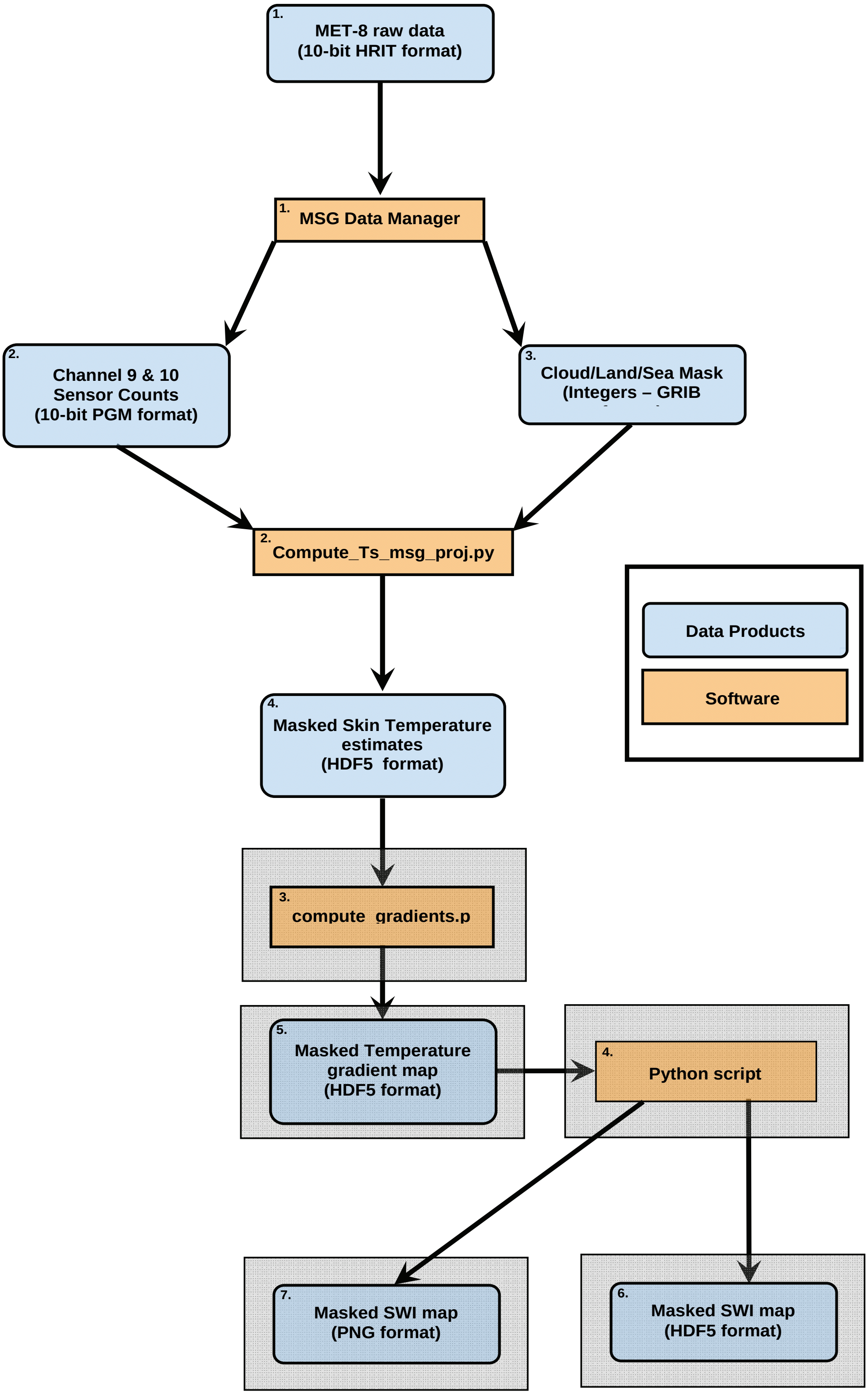
The work reported on in the first part of this chapter was carried out as part of the initial investigation to determine the feasibility of the methodology for soil moisture estimation. In order to minimise the technical obstacles to getting started with this research, the data used were a resampled and re-projected version of the original Meteosat-8 data. A software tool developed by a consultant for another project was used to resample the original data from the Geostationary projection to a square latitude-longitude grid which matches the SIMAR grid.

As the project progressed it became increasingly important that the project team free itself from closed source solutions that place limitations on the manner in which the data must be processed. In a research environment it is essential that changes in the envisioned structure can be made easily in response to the acquisition of new knowledge and skills. For this reason, the project team developed techniques and software in house to process and manipulate the data. It was felt that the time and effort put into this development was well balanced by the increased flexibility achieved since external consultants are no longer required to produce customised data processing solutions. This acquired skill has been a boon in both data management and information data sharing, to be discussed in Chapter 7 of this report.

Step by step procedure for computing the gradients of surface temperature

The process for producing a Soil Saturation Index (SSI) is described in this subsection. Soil Moisture will subsequently be computed directly from the SSI. The processing chain is summarised in the following chart and a detailed description of the products and processes given in the text.

The processing chain for producing the Met-8 SSI product.



Descriptions of the data products

Here follows a description of the format of each data product in the processing chain shown in the flowchart on the previous page. The algorithms for processing the data are also described. The functioning of the processing software that implements the algorithms is discussed separately in the section which follows.

Product 1 – MET-8 raw data:

The data is in HRIT format and contains sensor counts from the SEVIRI instrument at 10-bit precision (i.e. the integer count value, c, must satisfy $0 <= c <= 1023$). The data are stored as a rectangular array, in a geostationary satellite projection, with a nominal sub-satellite Longitude of 0°. These data are obtained in real-time via the EUMETCast receiving station. Data files arrive at a frequency of 15 minutes.

Product 2 – Channel 9 & 10 Sensor counts (16-bit PGM format):

The MSG Data Manager software is configured to convert the HRIT format data for channels 9 and 10 into a Portable Grey Map (PGM) raster format. The data remains at 10-bit resolution and the calibration coefficients for converting from the sensor count to radiance values are stored in the PGM file. The radiance may be computed from the integer sensor values using the following linear relation:

$$R = ax + b \qquad [\mathrm{mWm^{-2}sr^{-1}(cm^{-1})^{-1}}]$$

where R is the pixel radiance, a is the calibration slope and b is the calibration offset. Figure 5.26 shows images of the sensor counts for Met-8 channels 9 and 10 from a single scan time (12:00 UTC on 21/02/2007). A zoomed version of the same data is shown in Figure 5.27, corresponding to the SAfr region defined by the LSA-SAF. In each of the figures, the pair of top panels show the sensor count for all data pixels which represent a location on the earth (space pixels are masked out in black). The two bottom panels show everything apart from cloud-free land surfaces masked out. The EUMETSAT cloud mask (Product 3 shown in Figure 5.30) has been used to do the masking.

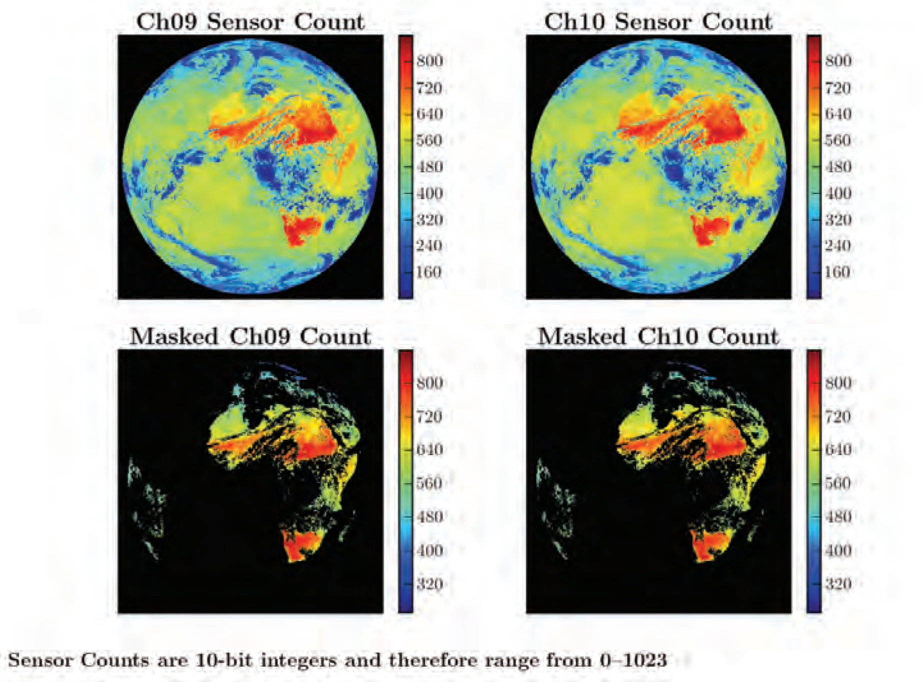


Figure 5.26 Meteosat-8 channel 9 and 10 raw sensor counts. Data from 21 February 2007 at 12:00 UTC.

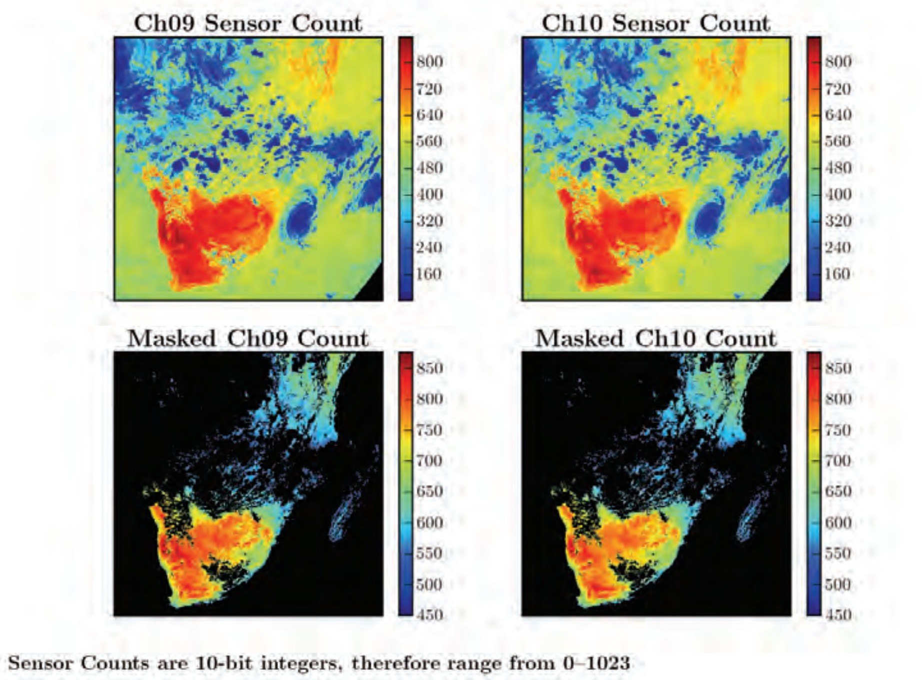


Figure 5.27 Meteosat-8 channel 9 and 10 raw sensor counts. Data from 21 February 2007 at 12:00 UTC. The region displayed matches the SAfr region defined by LSA-SAF for their data products.

After applying the conversion from sensor count to radiance, the data in Figure 5.28 and Figure 5.29 are obtained. The regions displayed and masking procedure applied for producing these figures are the same as those for Figure 5.26 and Figure 5.27.

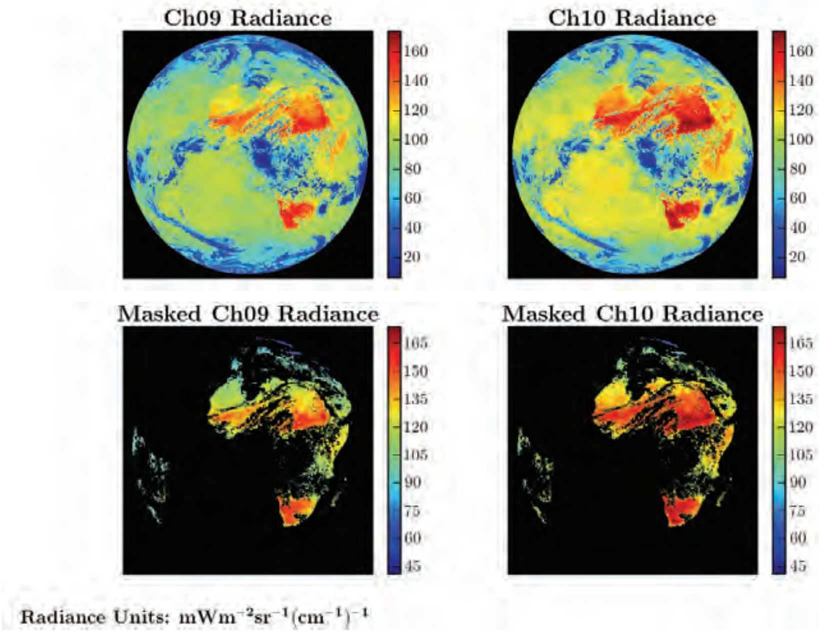


Figure 5.28 Meteosat-8 channel 9 and 10 radiance. Data from 21 February 2007 at 12:00 UTC.

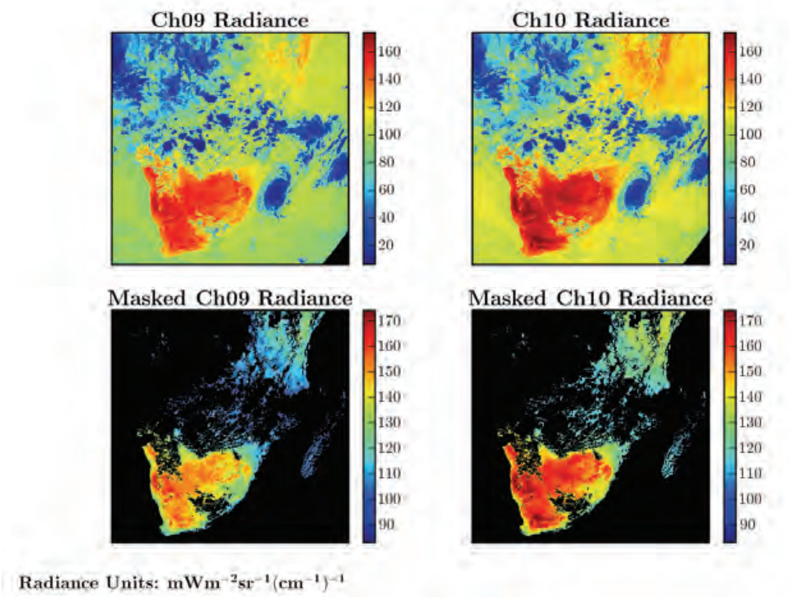


Figure 5.29 Meteosat-8 channel 9 and 10 radiance. Data from 21 February 2007 at 12:00 UTC. The region displayed matches the SAfr region defined by LSA-SAF for their data products.

<u>Product 3 – Cloud/Land/Sea mask (GRIB2 format):</u>

Since the surface temperature gradient maps can only be sensibly defined over cloud free land surfaces, the EUMETSAT cloud mask product disseminated via the EUMETCast system is used to assign a “no data” mask to the portions of the data which do not satisfy this criterion (See Figures 53 to 56). Software has been developed in-house to read the cloud mask data files.

EUMETSAT's cloud mask product is distributed in GRIB2 format at a temporal resolution of 15 minutes and the standard geostationary projection with sub-satellite longitude of 0°. A unique cloud mask is produced for each repeat cycle and the mask distinguishes between, (i) no-data, (ii) cloud, (iii) cloud-free sea surfaces and (iv) cloud-free land surfaces. Figure 57 shows an example of a typical cloud mask; note that the data region doesn't cover the entire portion of the globe visible from the satellite.

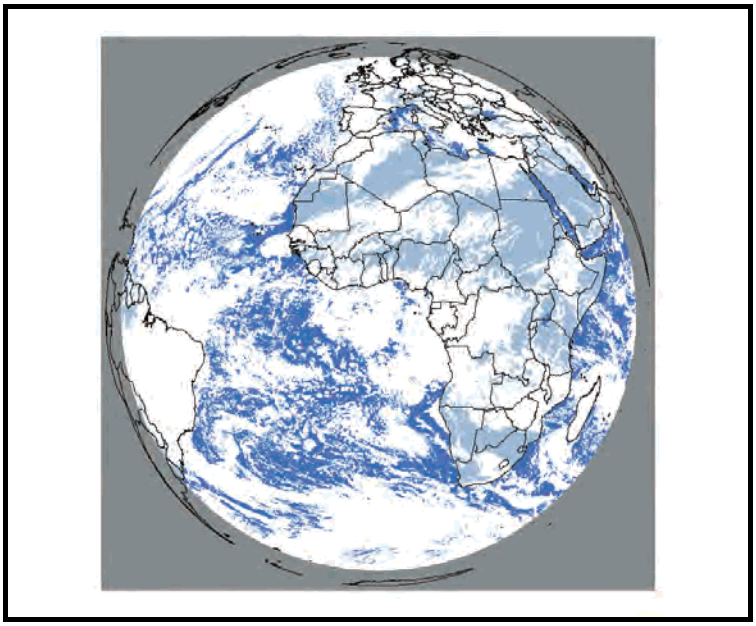


Figure 5.30 Meteosat-8 cloud mask. The mask is for the 11:15 UTC repeat cycle on 20 February 2007.

The data contained in the cloud mask files consist of four integer values, each of which represents one of the four data classes (Table 5.2).

Table 5.2 Meteosat-8 cloud mask pixel classifications.

Value	**Class**
0	Cloud free ocean or water body
1	Cloud free land surface
2	Cloud
3	Space/No data

The cloud masks are copied once daily from the EumetCAST working directory (3 day rolling buffer) to a separate archive directory in the SAHG computer cluster; this action is controlled by the batch file "cloud_mask.bat", which is scheduled to run daily at 12:00 SAST. All the GRIB2 files in the archive directory are compressed using the gzip program, this compression process is scheduled to run weekly and controlled by the batch file "compress_archives.bat". The archive of cloud masks contains data received since 07/02/2006 to the present and the archiving of cloud masks is automated and ongoing. Data for repeat cycles from 05:30 GMT to 08:30 GMT are collected. Average compressed file size is around 400 Kb (depending on cloud cover since this will affect the efficiency of the compression algorithm) and uncompressed size is 3365 Kb.

Product 4 – Masked Skin Temperature estimates (HDF5 16-bit unsigned integer format):

The Brightness Temperature values for each pixel in Geostationary coordinates are computed using the following relationship:

$$T_B = \left[C_2 v_c \Big/ \log\left(\frac{C_1 v_c^3}{R} + 1 \right) - B \right] / A$$

where:

T_B is the brightness temperature, R is the radiance, C_1 = 1.19104x10^{-5} mWm^{-2}sr^{-1}(cm^{-1})$^{-4}$ and C_2 = 1.43877 K(cm^{-1})$^{-1}$. The other unknown parameters are channel dependent and are given in Table 5.3 for channels 9 &10.

Table 5.3 Parameters for channel 9 and 10 Brightness temperature estimates.

Channel No.	**Channel ID**	v_c	A	B
9	IR 10.8	930.659	0.9983	0.627
10	IR 12.0	839.661	0.9988	0.397

For data projected on the SIMAR (lat-long) grid, the sensor count data is processed to a re-projected GeoTIFF format (brightness temperature) for repeat cycles from 08:00 to 10:00 SAST. The process of producing the GeoTIFF's is controlled from the batch file "current_GeoTIFFs.bat", which is scheduled to run every 15 minutes from 08:17 to 10:17 SAST. The resulting GeoTIFF's are archived and compressed using gzip on the server. The weekly batch file "compress_archives.bat" controls this. The entire process is automated and ongoing. GeoTIFF brightness temperature data have been collected since 14 June 2005 with only limited periods of missing data. Uncompressed data files are 2049 Kb and compressed files are roughly 550 Kb each.

For the product in Geostationary projection at the Meteosat-8 pixel resolution, brightness temperatures are computed as an intermediate step towards producing skin temperature estimates. Figure 5.31 and Figure 5.32 show examples of computed brightness temperature estimates using the algorithm described above. The skin temperature estimates computed from the split window algorithm are subsequently archived in compressed HDF5 format on a daily basis.

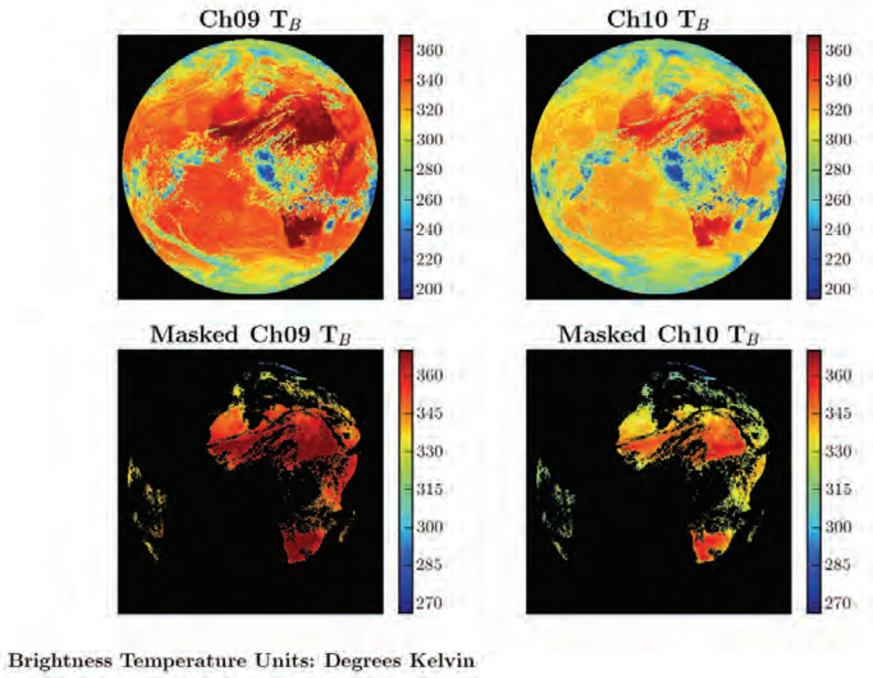


Figure 5.31 Meteosat-8 channel 9 and 10 brightness temperatures. Data from 21 February 2007 at 12:00 UTC.

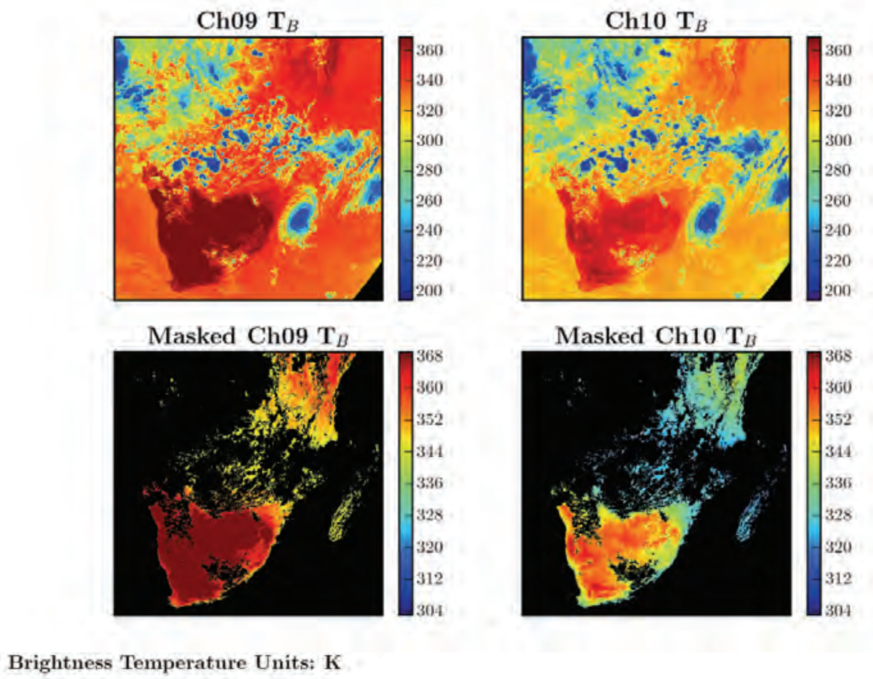


Figure 5.32 Meteosat-8 channel 9 and 10 brightness temperatures. Data from 21 February 2007 at 12:00 UTC. The region displayed matches the SAfr region defined by LSA-SAF for their data products.

Product 5: Skin temperature gradients (HDF5 16-bit integer format)

For the data projected on the SIMAR grid, the HDF5 files are archived on the server and compressed weekly to reduce space requirements. The file size before gzip compression is 2051 Kb and post compression roughly 1350 Kb.

Process Descriptions

Process 1: MSG Data Manager

The primary purpose of the MSG Data Manager software is to copy the raw data received in HRIT format from the receiver PC to the data server, the raw data is then deleted from the receiving PC. There is a host of possible choices regarding what data to store and what formats to choose, however only the current configuration relevant to the SSI product will be discussed.

Data from each of the 12 MET-8 channels and the cloud mask product are stored in a 3 day rolling buffer on the server. The channel data are stored as greyscale JPEG images. In addition, the raw sensor count data for channels 9 and 10 are stored in 10-bit PGM format (product 2). The cloud mask product is stored in GRIB2 format (product 3). No raw HRIT data are retained.

Process 2: compute_Ts_msg_proj.py

This Python script computes skin temperature estimates from Met-8 data using a split window algorithm. Input data are sensor count data from Channels 09 and 10 in a 10-bit PGM format. The pre-computed cosines of the satellite viewing angle at each pixel are also required. The script works on the assumption that these are pre-computed and contained in an HDF5 file. Only the current day's data are processed and an estimate of the surface temperature for each scan time between 06:00 and 07:00 UTC is stored in a single HDF5 file for the day (product 4).

Process 3: compute_gradients.py

For data projected on the SIMAR grid, this script computes the gradient of a 'best fit' straight line through the skin temperature estimates from Met-8 data at each pixel for the day in question. The 5 data points between 06:00 and 07:00 GMT are used. The skin temperature estimates are assumed to be in an HDF5 file (product 6), which is then deleted by this script. The resulting gradients (and their R^2 values) are stored in an HDF5 file as 16-bit signed integers scaled by 100 to give two decimal points of precision (product 7). An image of the gradient field is also produced (product 8).

Process 4: Python script

Compute SSI estimates for each pixel – see the next subsection.

The Daily Temperature Gradient Maps over the country.

After a considerable interlude (work on temperature gradients halted in December 2007 with the disappearance of Ntoko), from late 2009, the temperature gradients were calculated from surface temperatures estimated by the LSA-SAF group in Lisbon working for EUMETSAT.

Here follow some images of estimated surface temperatures and the gradients that were derived from them, extending the earlier work to the subcontinent. It is clear that the data-handling capabilities of the group have grown considerably over the period.

Calculating soil moisture from temperature gradients

This section presents a sub-continental extension and development of the very locally focussed methodology reported in the first part of this chapter, dating from 2006, whose details will not be repeated. Suffice it to say that if soil warms up more quickly on one day than it does on another, the quicker warming is likely to be associated with drier soil. We want to calculate the rate at which each detailed area of land heats up in the 1 hour between 8 and 9 am (local time) at ¼ hour steps (the sampling rate of the SEVRI instrument on board METEOSAT-9). If the values of temperature are increasing at a reasonable rate (between 1 and 14 degrees Centigrade per hour) and if they lie on a reasonably straight line (we previously determined that an r^2 of more than 0.925 would be the determinant) we would record that gradient. Once enough daily gradients have been collected (more than a year) their range would indicate the range of behaviour of the soil wetness. We postulated an inverse linear relationship connecting gradient and SSI.

Figure 5.33 shows 6 successive days (the rows) in the ‘dry’ period (depending on location) in August for Surface temperatures calculated by the EUMETSAT LSA-SAF group and disseminated via EUMETCast. The columns are snapshots taken by METEOSAT-9 at 15 minute intervals starting at 8 am SA time. We chose this period because (i) any mist/fog has usually burnt off and (ii) it is the quickest warming period of the day. Although the detail is masked by the scale of the figures, it is clear that there is considerably more difference between days at the same time than between successive observations on the same day. Figure 5.34 is an expansion, and therefore gives more detail, of the 4 images in the red rectangle.

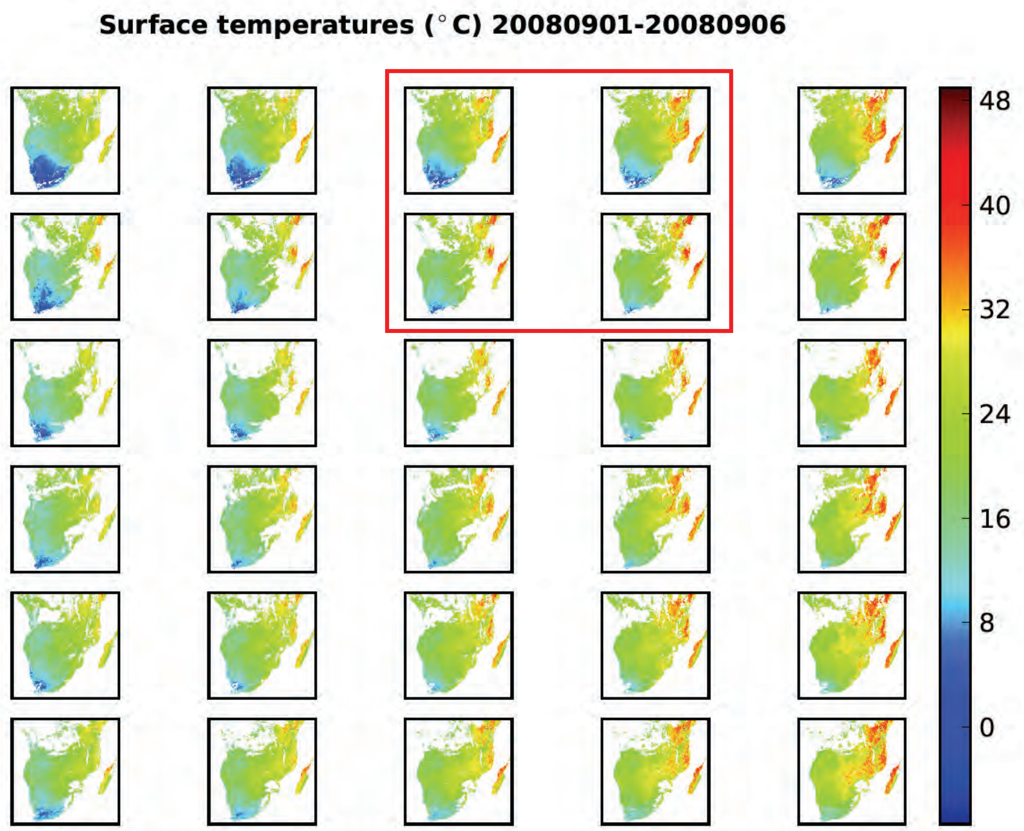


Figure 5.33 Sequence of surface temperature maps used in the gradient calculation. Each row in the figure shows a sequence of surface temperature snapshots at 15 minute intervals. White regions in the each block show areas of "no data" that are caused by sea or cloud.

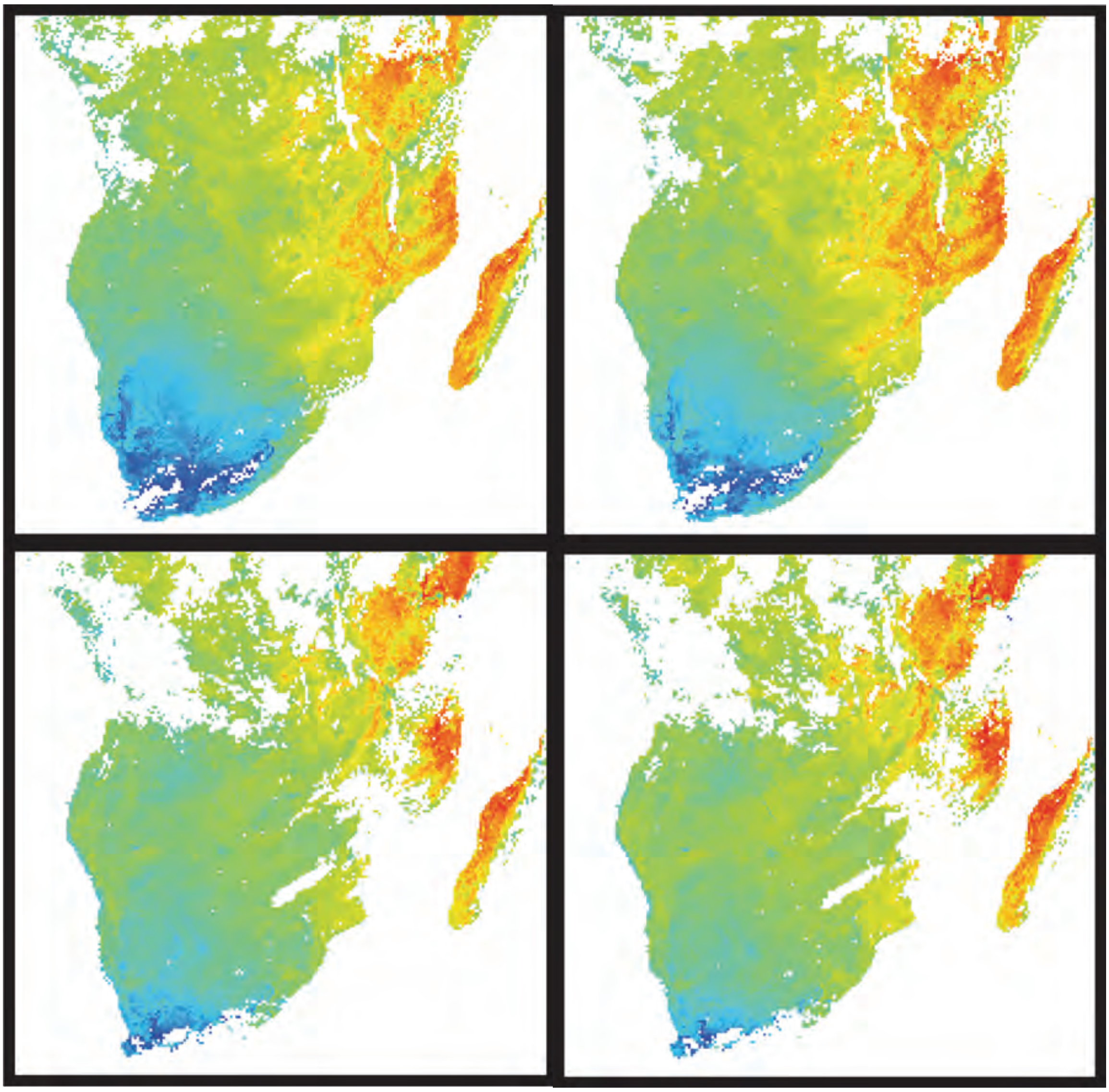


Figure 5.34 Two pairs of surface temperature maps used in the gradient calculation, expanded from those in the red rectangle in Figure 5.33. Each row in the figure (1st and 2nd Sept 2008) shows a surface temperature snapshot at 0830 and 0845 SA local time. White regions in each block show areas of “no data” that are caused by sea, freshwater bodies or cloud.

It is clear that there is a gradual change along the row, but that the patterns of temperature change markedly from one day to the next, hardly retaining the gross features. Figure 5.35 shows another sequence from January 2008 (with warmer temperatures), where large tracts of land are covered in cloud and one whole row of transmissions is missing on 5th January.

In spite of these difficulties, we persevered with the analysis.

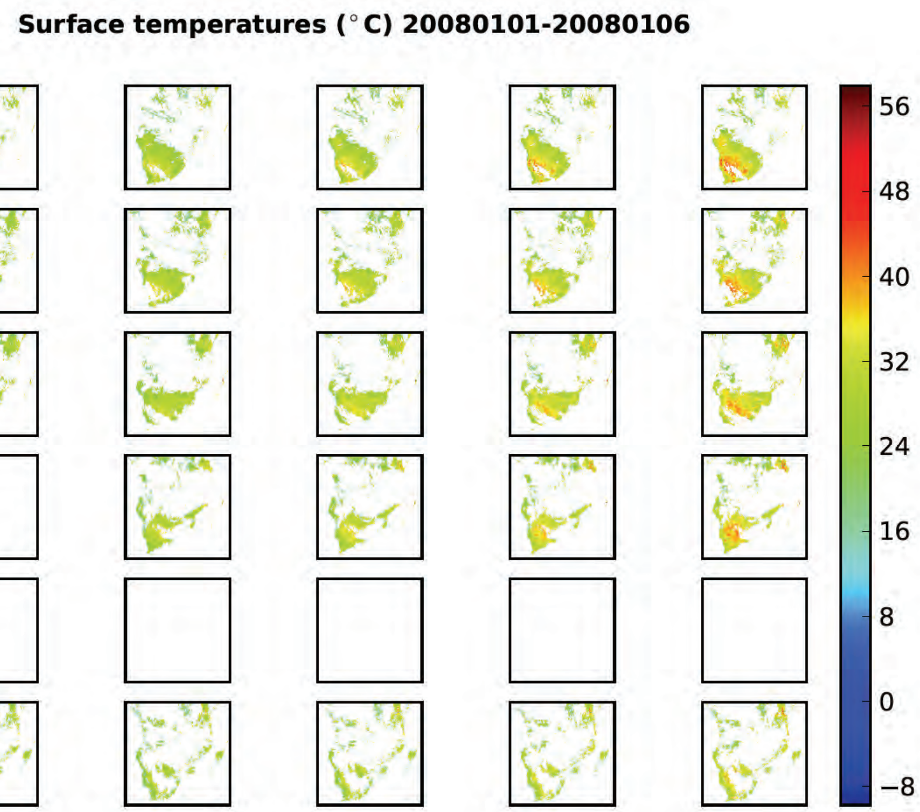


Figure 5.35 Sequence of surface temperature maps used in the gradient calculation. Each row in the figure shows a sequence of surface temperature snapshots at 15 minute intervals. White regions in the each block show areas of “no data” that are caused by sea or cloud. Note the significant effect of cloud in this case.

Figure 5.36 shows a comparison of the count of valid temperature gradients collected in months January and July. It is evident that the North-Western Cape is the only region in January south of the equator where useful gradients are available and is (paradoxically) an area of low population density.

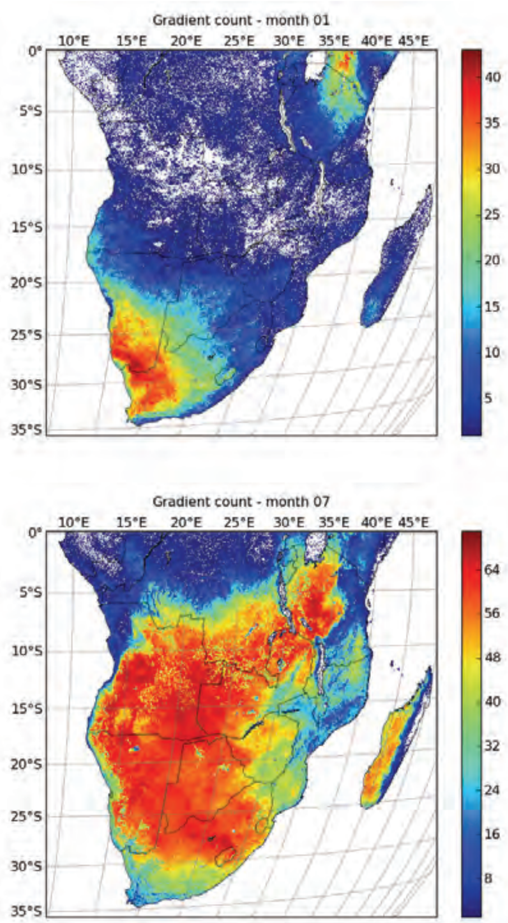


Figure 5.36 Total number of valid gradient estimates available on each pixel during the analysis period. Seasonal cloud and rainfall effects are clearly evident. It's worth noting that month 01 (January) represents data from 2 years (2008-2009), while month 07 (July) has data from 3 years (2007-2009). Also note the different colour scales.

Figure 5.37 shows the count for the whole 30-month period of 900 days. The legend indicates the number of useful days where the sampling at each pixel resulted in acceptable gradients; the result is that less than half are useful, almost everywhere except in the Namib Desert.

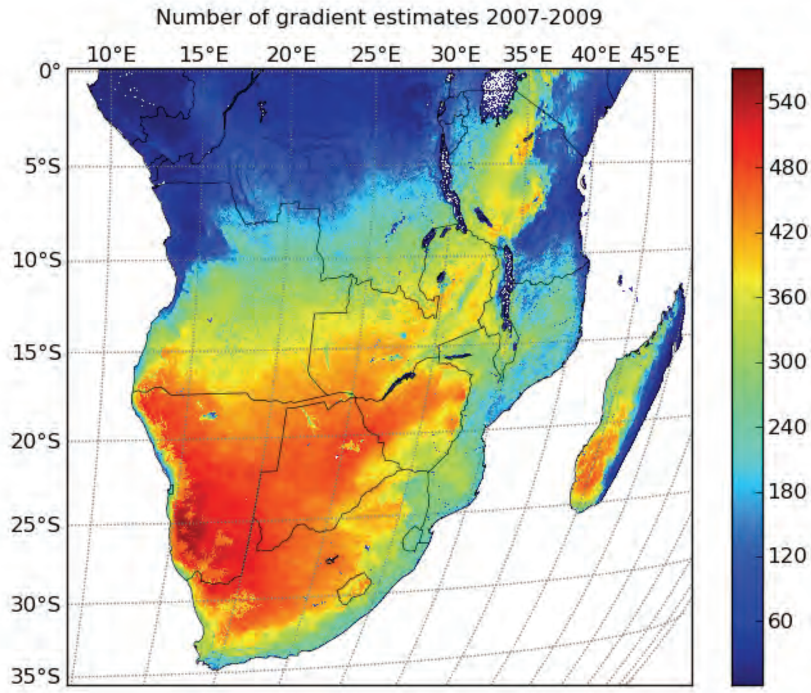


Figure 5.37 The total number of gradient estimates for available for each pixel during the 30 month analysis period.

Figure 5.38 shows the temperature gradients calculated for a pixel in Central Africa for 6 successive days in September 2008. All have passed the $r^2 > 0.925$ test. The starting temperature of the sequence of gradients, which begins on 5 Sept (red line) in the lower part of the figure, declines gradually over 3 days and then jumps by 7 degrees to the black line and gradually cools. In addition, in the bottom 3 gradients, the fastest rise is rather unexpectedly from the lowest temperature. Thus the gradients are not generally associated with the starting temperature, a behaviour which has no obvious explanation. We suspect that wind and humidity are likely factors but, because we already take those into account in our ET calculations, there is no point in adding complexity to what was intended to be a simple device. The result was that we decided to concentrate on the gradients themselves for the purpose of analysis.

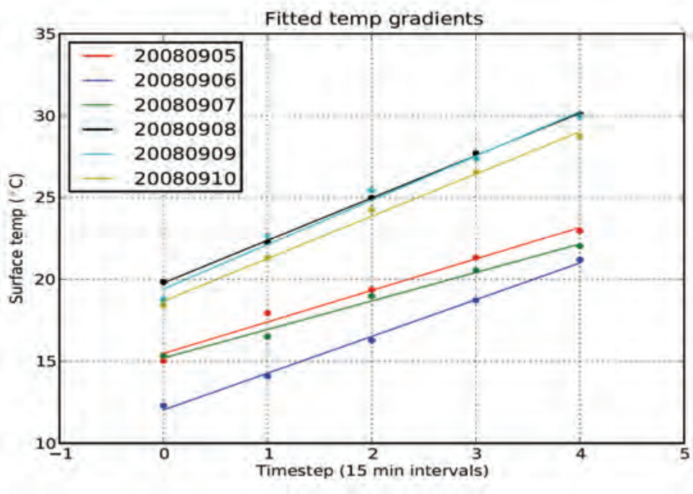


Figure 5.38 Examples of fitted temperature gradients at a single location, for six consecutive days. The r^2 value for each of these linear regressions is greater than 0.925 (this value is used to exclude data where the fit is poor).

The three panels of Figure 5.39 show (i) the temperature gradients for a selected pixel in the Liebenbergsvlei, where gradient observations on contiguous days are joined by lines – the lack of these is an indication of missing data, (ii) the interpreted SSI series and (iii) the inverse linear transformation used, picking out the actual points and hence the range of the transformation.

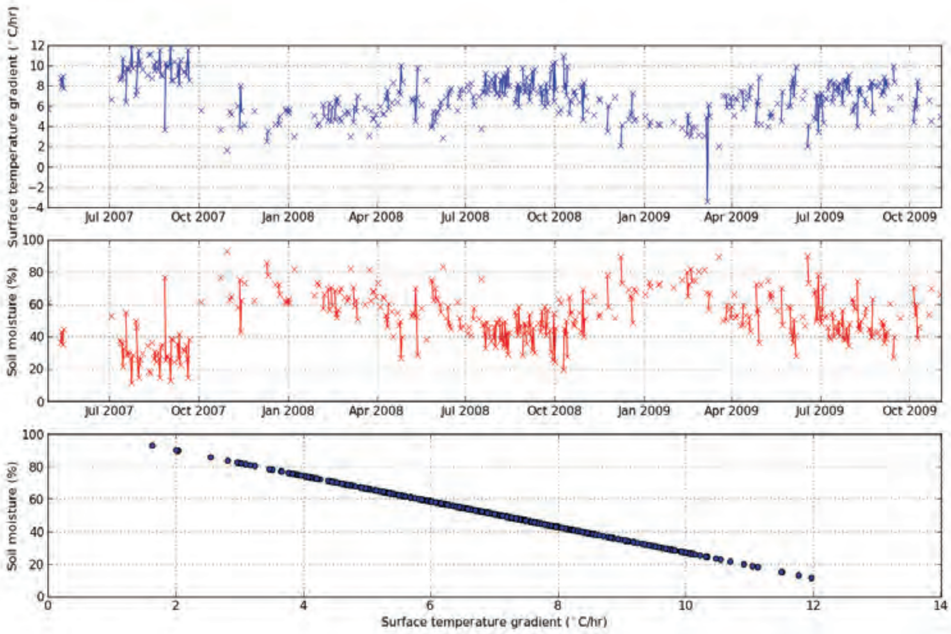


Figure 5.39 Time series of surface temperature gradients for a single pixel located in the Liebenbergsvlei. The blue/red lines (panels 1 and 2) connect observations if these are available on consecutive days i.e. are not masked out due to cloud contamination. The seasonal variations evident in the figure match the wet and dry cycles for this region.

In Figure 5.40, the upper and lower bounds observed by month are fitted by trigonometric series, emphasizing the strong intra-annual variation in the gradients. We originally thought (as described in the earlier part of this section) to perform linear transformations based on these limits, but it was found that we ended up with almost constant SSI over the year. As a result, the idea was abandoned and instead, constant upper and lower bounds were computed as detailed in Chapter 6.

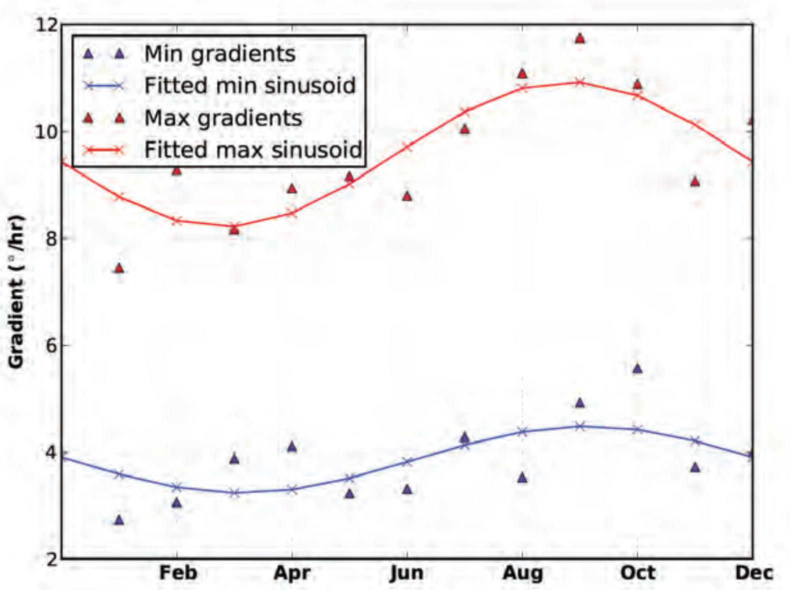


Figure 5.40 Sinusoids fitted to the time series of monthly maximum and minimum temperature gradients at a single location. The fitting procedure was carried out for all land pixels in the 1191-square LSA-SAF Southern African grid.

Summary

It was thought at the inception of the project that early morning temperature gradients measured at the earth's surface using remote sensing might be a useful indicator of Surface Soil Wetness, which might then be linked to Soil Moisture (SM). This chapter documents the effort put into this idea. It turns out that the observations are occluded by cloud cover to such an extent that the inference on SM is not as useful as was first hoped. Nevertheless, the work done (in the early days of the project before we had made linkages with the influential people in the remote sensing world) was extremely useful in getting our minds around the remote sensing paradigm and very good training in the management of large quantities of data.

6. Comparing TOPKAPI/LSM estimates against two remote sensing based estimates

Introduction

The focus of this Chapter is to compare three independent estimates of soil moisture and attempt to determine whether a single estimate or a combination of the estimates is best for the purpose of providing the best available soil moisture estimate on the web, which is the objective of the work reported in Chapter 7.

The first part of the work reported in this chapter was to take the previous successfully completed work reported in Chapter 4:

(i) Using TOPKAPI in land surface model mode to estimate SM & Actual ET

and

(ii) The ASCAT remotely sensed estimates of Relative surface soil moisture (SSM) that we acquired from EUMETSAT

and compare the two against each other. This comparison resulted in a publication in an international journal: Sinclair and Pegram (2009).

It is appropriate at this juncture to repeat the definition of the TOPKAPI SM estimate, which is the Soil Saturation Index, SSI = $(\theta - \theta_r)/(\theta_s - \theta_r)$; here θ_s = porosity (direct from the SIRI data base) and θ_r is obtained from table 5.3.2 of Maidment (1993) based on WR90 (Midgley et al., 1994) texture classes. The results were very encouraging and we expect that anticipated improvements to the ASCAT remote sensing algorithm will make them even better in the future.

As recommended by the reference group at the May 2009 meeting, the next task in the deliverable was to revisit the earlier work done using local early morning surface temperature gradients, as a surrogate for Soil Saturation Index (SSI) and reported in Chapter 5, and extend it country-wide, in order to compare these estimates with the other two spatial estimates. By now we had a long enough record of surface temperatures acquired from METEOSAT to understand the behaviour of the data and thus to infer the SSI.

The outcome was that the SSI estimates based on temperature gradients were disappointing for the following reasons:

(i) Cloud cover occludes measurement at the ground resulting in severely interrupted estimates, particularly in the rainy season over populous areas when the estimates are most needed for Flash Flood Forecasting. It had been thought that we would be able to accommodate this deficiency by filtering recent observations, however we were bedevilled by the fact that the METEOSAT measurement of the IR band cannot penetrate clouds, so that for much of the region less than half the days allow ground temperature measurement.

(ii) The temperature gradients over much of the region exhibit unexpectedly large day-to-day and spatial variability (likely due to wind and humidity changing daily) which necessitated considerable spatial and temporal smoothing of the depleted observations, in order to extract any meaningful information, so that day-by-day observations are far too noisy to provide useful contemporaneous estimates.

Nevertheless, there are obvious correspondences between the three estimates of the soil wetness status. **Our conclusions are that**

(i) We found TOPKAPI estimates in Land Surface Mode to be consistent and sensible, giving us a continuous, low noise estimate of Soil Saturation Index (SSI), the ratio of the volume of water to the volume of voids in the soil matrix

(ii) We are cautiously comfortable with the remote sensing ASCAT estimates of SSM when temporally filtered to SWI, in spite of their noisiness and

(iii) We found that, although an interesting concept, the temperature gradient estimates of soil wetness are not practically useful, so will be discontinued.

This Chapter is mostly pictorial, with some short comments fleshing out the above. It starts with the TOPKAPI/ASCAT comparisons which were published in Sinclair and Pegram (2009) and, because they are described in previous chapters, will not be discussed in detail.

Comparisons of 2 estimates – TOPKAPI SSI with ASCAT SWI

In this section we compare two independent soil moisture estimates over South Africa. The first estimate is provided by automated runs of the TOPKAPI hydrological model. The model has been adapted to run as a collection of independent 1 km cells located on a grid with a spatial resolution of 0.125°, using 3 hourly rainfall estimates and evapotranspiration forcing calculated at 1 h intervals. The rainfall forcing used is NASA's TRMM 3B42RT product, while the evapotranspiration forcing is based on a modification of the FAO56 reference crop evapotranspiration (ET_0), which accounts for vegetation health and the availability of surface and soil water, as limiting factors on the potential rate of evapotranspiration.

Using the rainfall and evapotranspiration forcing data, the percentage saturation of the TOPKAPI soil store (SSI) is computed, for each of the 6989 uncalibrated TOPKAPI cells at 3 h time-steps, and compared with estimates of surface soil moisture from the ASCAT instrument onboard the METOP polar orbiter. The comparisons indicate a good correspondence in the dynamic behaviour of the TOPKAPI SSI estimates against an exponentially filtered time series (SWI) of the ASCAT surface soil moisture (SSM) estimates, for several climatic regions in South Africa. The linear agreement in dynamic behaviour for these independent soil moisture estimates suggests that both are correctly capturing the soil moisture dynamics for a significant proportion this region, and could potentially be combined to produce a "best estimate" soil moisture field.

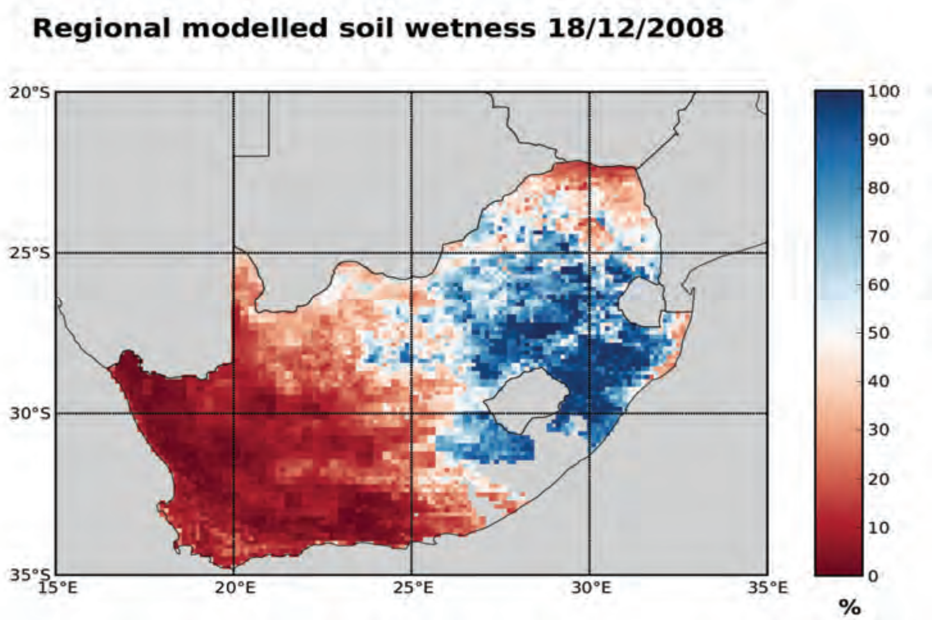


Figure 6.1 An example of the country-wide soil moisture estimates produced by TOPKAPI in Land Surface Model (LSM) mode. The colour scale represents Soil Saturation Index (SSI), the percentage of soil void space filled by water. The grey areas are places where insufficient data are available to parameterise the TOPKAPI model.

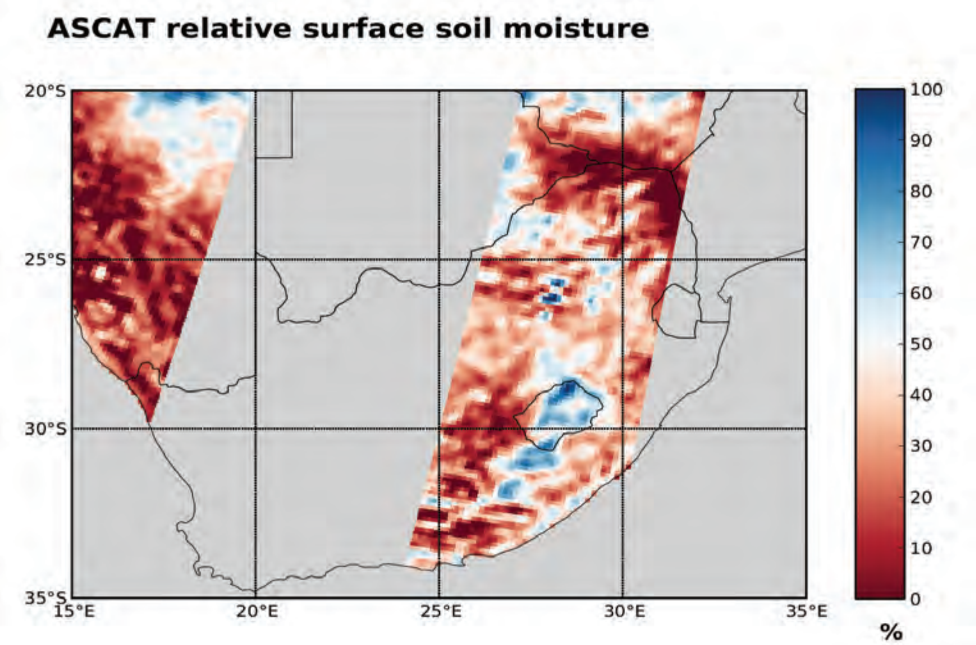


Figure 6.2 Relative surface soil moisture (SSM) retrieval from ASCAT. The time of the sampling of the twin swaths of the downward morning satellite overpass is approximately eight hours after the TOPKAPI-LSM SSI estimates shown in Figure 6.1

We selected 4 sites A, B, C & D, with a spread of climates, from the 6989 available. Their locations appear in Figure 6.3

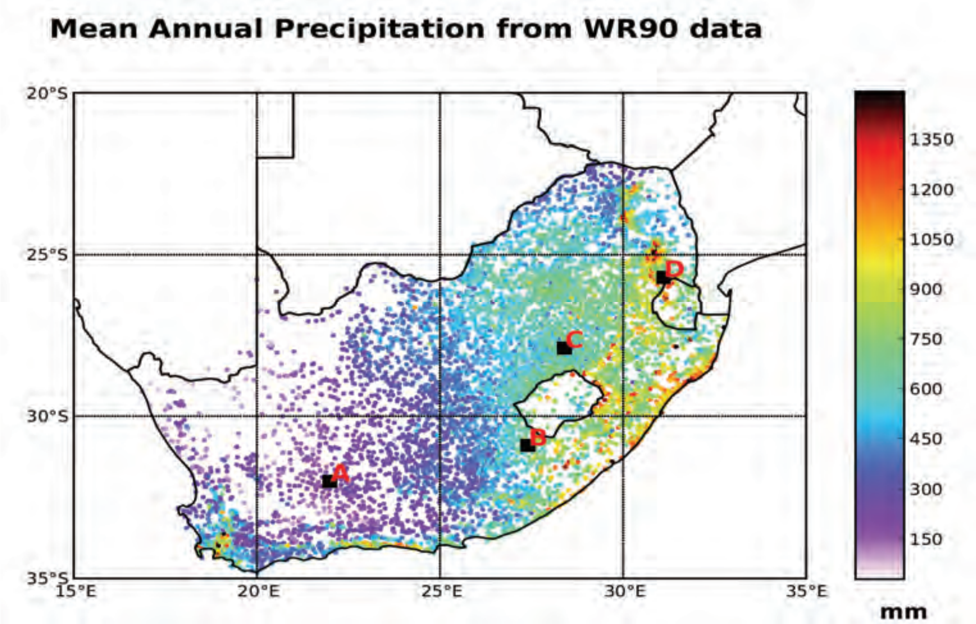


Figure 6.3 Locality of the four different regions considered, plotted over Mean Annual Precipitation (MAP) data obtained from WR90 (Midgley et al., 1994). The general trend is for MAP to increase from West to East. An important exception is the southern coastline, which receives significant winter rainfall associated with frontal systems.

We were able to acquire only 5 months (1 August to 31 December 2008) of ASCAT data off-line from EUMETSAT, covering the whole region. These were sampled at the four sites indicated in Figure 6.3 and are compared with the TOPKAPI SSI estimates which we have been routinely computing for the last 18 months. The comparisons appear in Figure 6.4 through Figure 6.7 in the order B, C, D & A; details are in the captions.

In the following Figure 6.4 through Figure 6.7, the top panel of each figure shows the median TOPKAPI-LSM SSI averaged over a block (blue line), with the inter-quartile range of the estimates within the block shown by the grey fill. The range of ASCAT SSM estimates (they fluctuate between 2 and 6 on a given day in a 0.25° square) within the area at each overpass time is shown by the box and whisker plots, while the red line shows the filtered time series of the mean of the ASCAT SSM estimates to give the SWI. The bottom panel in each figure shows a histogram of 3 h rainfall totals.

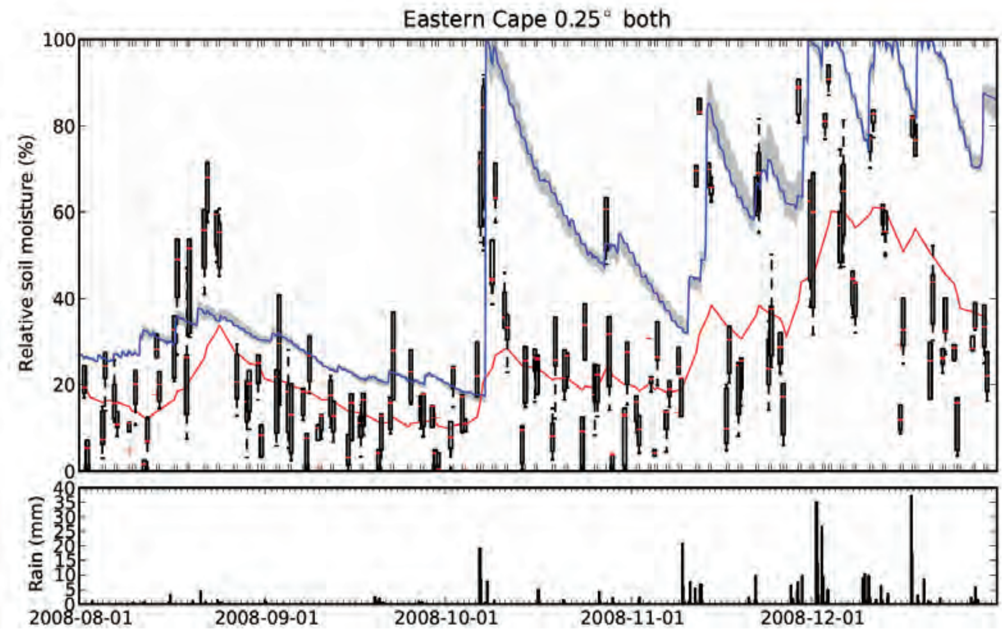


Figure 6.4 Site B: TOPKAPI SSI and ASCAT derived soil wetness (both for descending/morning and ascending/evening passes), and TRMM 3B42RT 3 h accumulated rainfall for site B (Eastern Cape). The size of the block where data are collected is 0.25° square. There are 4 TOPKAPI cells in the block.

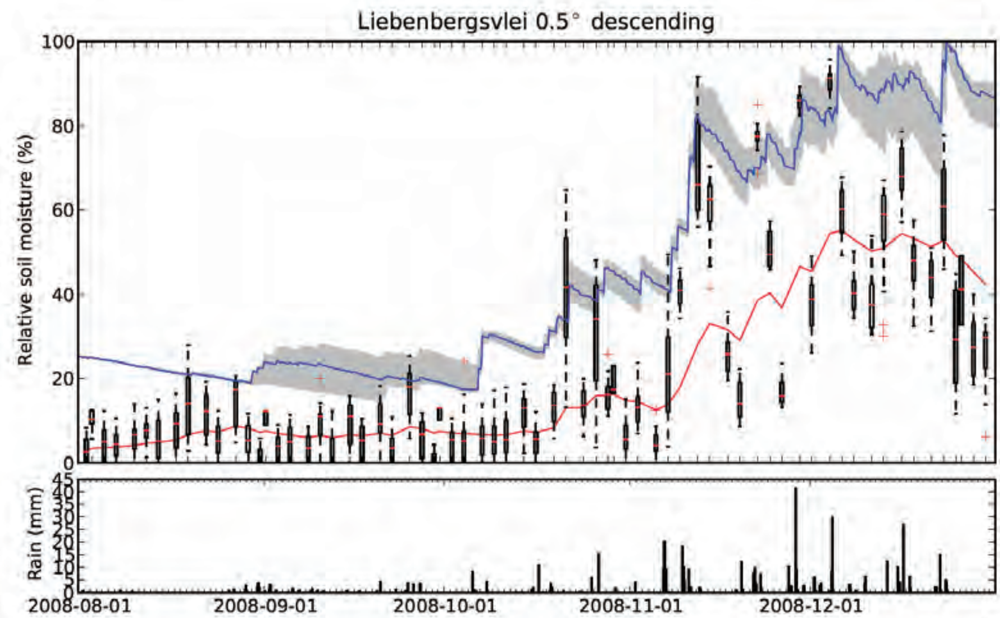


Figure 6.5 TOPKAPI-LSM and ASCAT derived soil wetness (descending pass only), and TRMM 3B42RT 3 h accumulated rainfall for site C (Liebenbergsvlei). The description follows that of Figure 6.4; the size of the area is now 0.5°, approximately 50 km square, so there are 16 TOPKAPI cells and 8 to 24 ASCAT observations in the area.

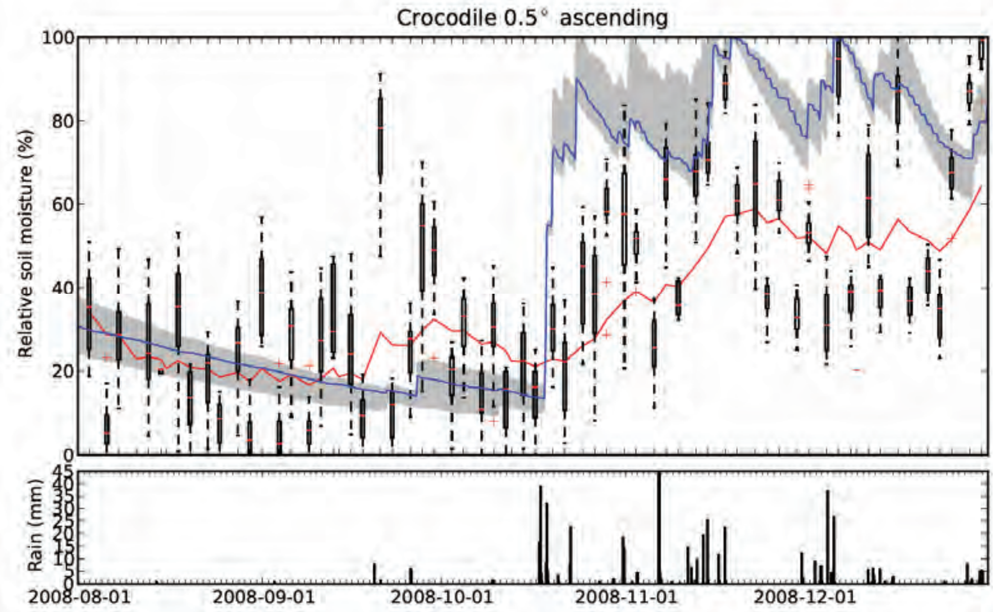


Figure 6.6 TOPKAPI-LSM and ASCAT derived soil wetness (ascending passes), and TRMM 3B42RT 3 h accumulated rainfall for site D (Crocodile catchment). The description follows that of Figure 6.4, except that the size of the area is 0.5°, approximately 50 km square.

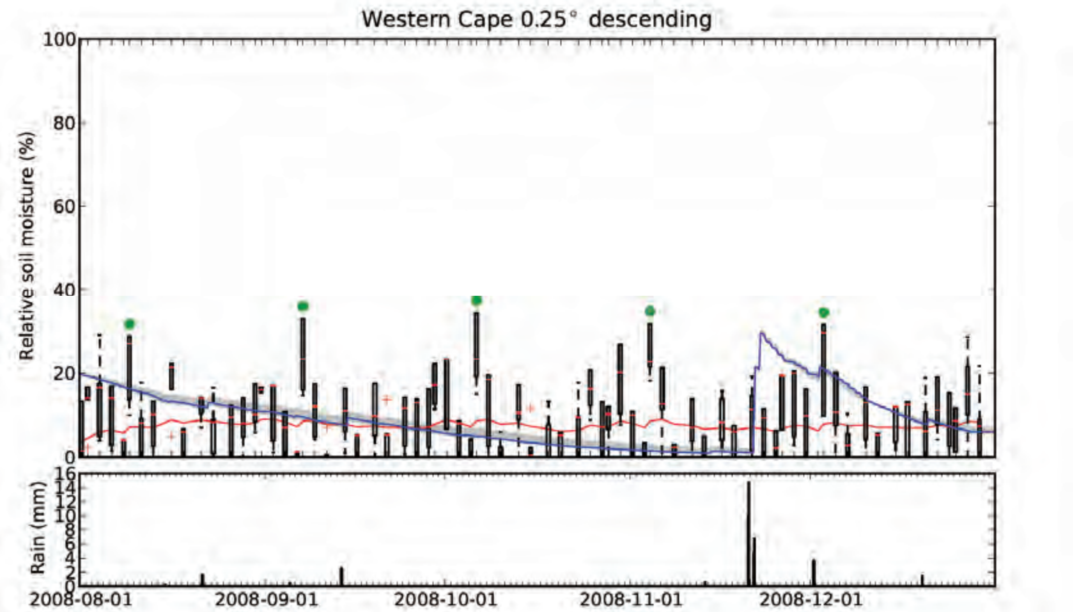


Figure 6.7 TOPKAPI-LSM and ASCAT derived soil wetness (descending pass), and TRMM 3B42RT 3 h accumulated rainfall for site A (Western Cape). The description follows that of Figure 6.4, with the added comment that the sequence of green dots highlights a marked increase in the ASCAT estimated soil moisture, which has a periodicity that matches the 29 day repeat cycle of MetOp (Figa-Saldaña et al., 2002). The size of the area is 0.25°, approximately 25 km square.

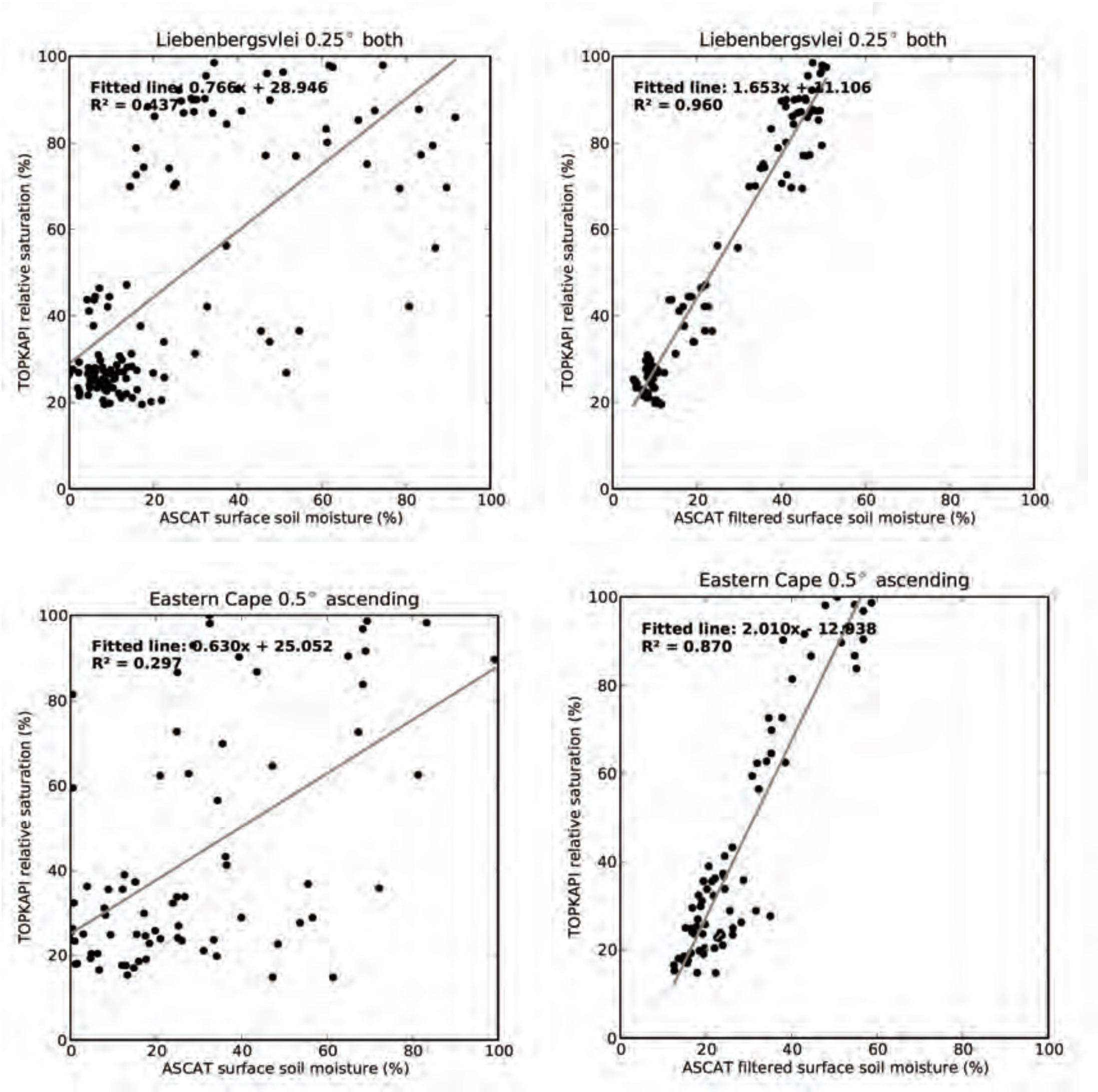


Figure 6.8 Scatter plots of the block mean ASCAT soil moisture and closest (in time) block mean TOPKAPI SSI, showing the fitted linear regression line and R^2 values. The two left panels show the relationship between SWI and ASCAT's Surface Soil Moisture (SSM); the right panels show SSI plotted against Soil Wetness Index (SWI), the filtered series of SSM estimates.

The comparisons in Figure 6.8 indicate that there is a considerable useful linear relationship between the TOPKAPI SSI and ASCAT estimates, provided that appropriate smoothing to SWI is conducted. The R^2 values in the right panels are very useful. We note that these are computed at two sites that exhibit a strong relationship between the variables (see Figure 6.9).

This becomes very clear in the comparison between the two images shown in Figure 6.9 following, where the upper image shows the R^2 values of the regressions between the TOPKAPI and unsmoothed ASCAT SSM values, computed (as in the left panels of Figure 8) at each 0.25° square in the region. The highest R^2 value in the upper panel is in Mpumalanga with a value nudging 0.7. In contrast, in the lower image of Figure 6.9 which shows the R^2 values for the smoothed ASCAT estimates, a large region in the Summer rainfall region (and some of the Western Cape) exhibits R^2 values near 0.9 and above, whereas the very dry (< 300 mm/a) regions yield remarkably low values.

Fortunately these areas are not flood prone and, by and large, are not densely inhabited.

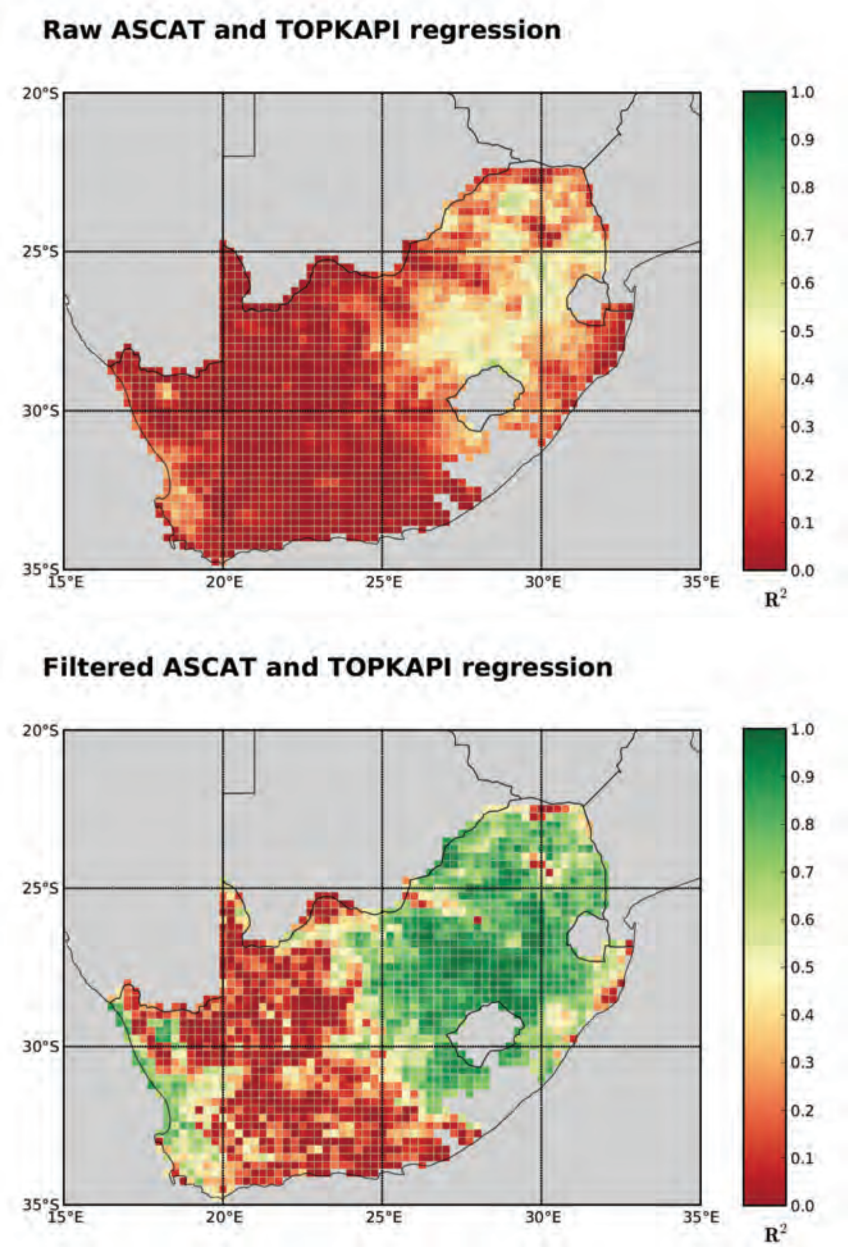


Figure 6.9 Maps of R^2 computed for the block mean (0.25°) ASCAT soil moisture and closest (in time) block mean TOPKAPI-LSM SSI – the upper and lower panels show the SSI comparisons against unfiltered SSM and filtered ASCAT SWI estimates, respectively. In comparing this figure with Figure 6.3, it will be seen that the areas with R^2 < 30% generally lie in the area where MAP < 300 mm and where the gauge records are sparse, indicating low population density. There are exceptions in the coastal zones and areas of high urbanisation and irrigation.

A three-way comparison of the Soil Water Estimates: TOPKAPI & Temperature Gradient-based SSI with ASCAT SWI

In the following four pairs of figures, three-way comparisons are made between (i) the TOPKAPI-SSI, (ii) the smoothed ASCAT SWI and (iii) the smoothed Temperature Gradient ($\Delta T/\Delta t$) estimates of SSI, at each of the four earlier chosen sites: Eastern Cape, Liebenbergsvlei, Crocodile and Western Cape (B, C, D & A). In each pair of figures, the upper figure shows the histograms of all acceptable ($\Delta T/\Delta t$) values in the vicinity of the TOPKAPI cell. The lower figure makes the comparisons over available contemporaneous periods.

At the 4 sites A to D shown in Figure 6.3, all acceptable gradients were pooled over the period of observation. Gradients (i) less than 1°C/hr, (ii) with fewer than 5 data points, or (iii) with R^2 less than 0.925, were excluded from analysis. Limits (red lines in the upper figure on each page) were calculated using robust statistics based on the median and interquartile range of the surviving data at each site, described in Chapter 5. These were used as the bounds of the transformation at the site concerned. We recall that the methodology of normalising the gradients seasonally over the year, using Fourier Series to determine lower and upper bounds as shown in Figure 5.22 in Chapter 5, was abandoned. This decision was based on the observation that the seasonal variation of Soil Moisture as estimated by TOPKAPI-LSM was absent in the ($\Delta T/\Delta t$) estimates and recovered to some extent in the example of Figure 5.21 in Chapter 5.

The four $\Delta T/\Delta t$ SSI estimates are compared against the TOPKAPI SSI and ASCAT SWI estimates (taken from Figure 6.4 to Figure 6.7) in the lower of the pairs of figures on the following 3 pages. The order of the sites is: B, C, D & A, (Eastern Cape, Liebenbergsvlei, Crocodile and Western Cape) as before.

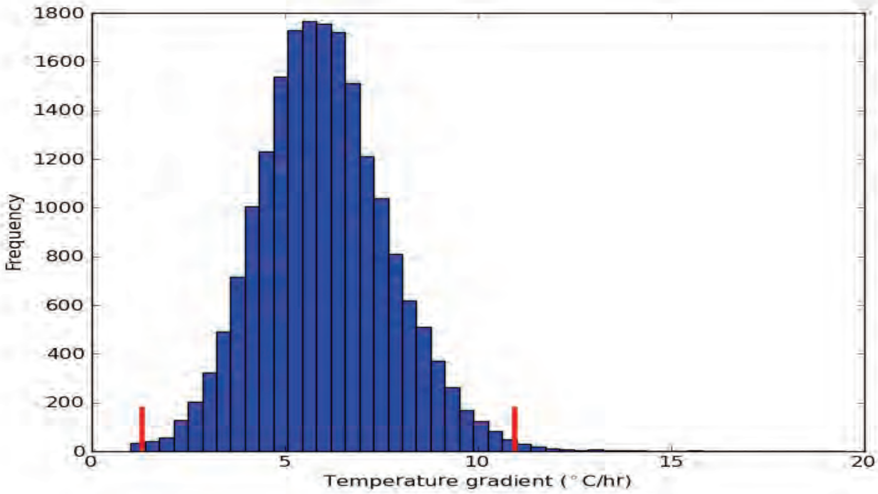


Figure 6.10 Distribution of temperature gradients for the Eastern Cape site, B. The data shown here are all temperature gradients pooled in a 7x5 pixels region (roughly a 0.25° square when projected on a regular Latitude/Longitude grid) for each day in the 30 month period analysed.

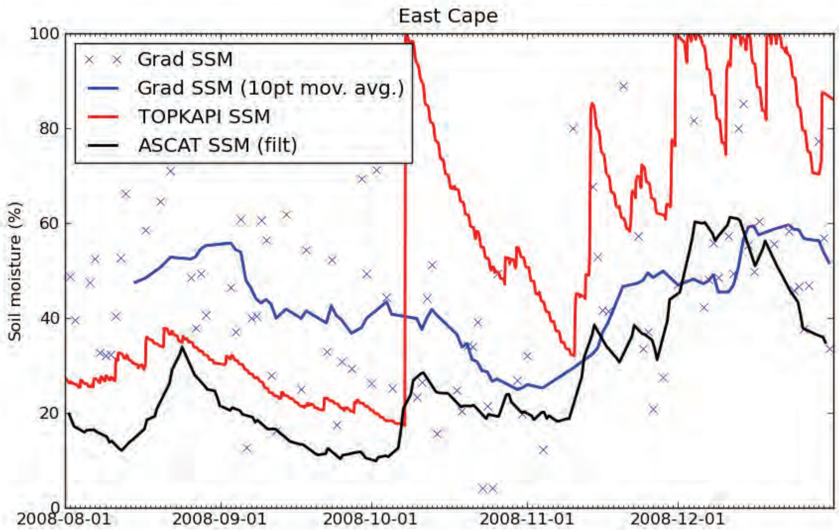


Figure 6.11 Site B: Eastern Cape. Time series comparison of the 3 soil moisture estimate over the period where matching data are available. The blue crosses are the calculated SSI soil moisture estimates from temperature gradients (due to the large number of contaminated data, the time step is variable). The solid blue line shows a 10 point moving average of the $\Delta T/\Delta t$ based estimates. The red line shows the TOPKAPI SSI averaged over the same 0.25° block. The black line is the time filtered sequence of the averaged ASCAT SWI estimates over the 0.25° block. These last two estimates are using the data composing Figure 6.4.

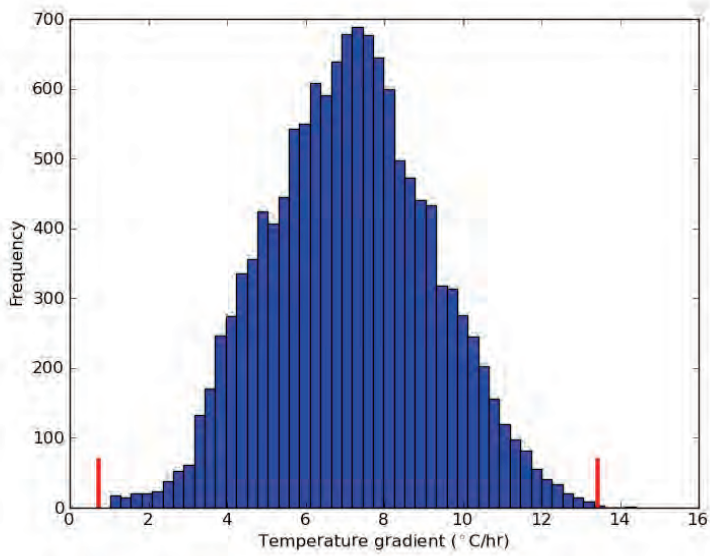


Figure 6.12 Distribution of temperature gradients for the Liebenbergsvlei site, C. The data are all temperature gradients pooled in a 7x5 pixels region (nearly a 0.25° square when projected on a regular Lat/Long grid) for each day in the 30 month period analysed. Gradients less than 1°C/hr, or with fewer than 5 data points, or an R^2 values less than 0.925 were excluded before trimming at the red limits.

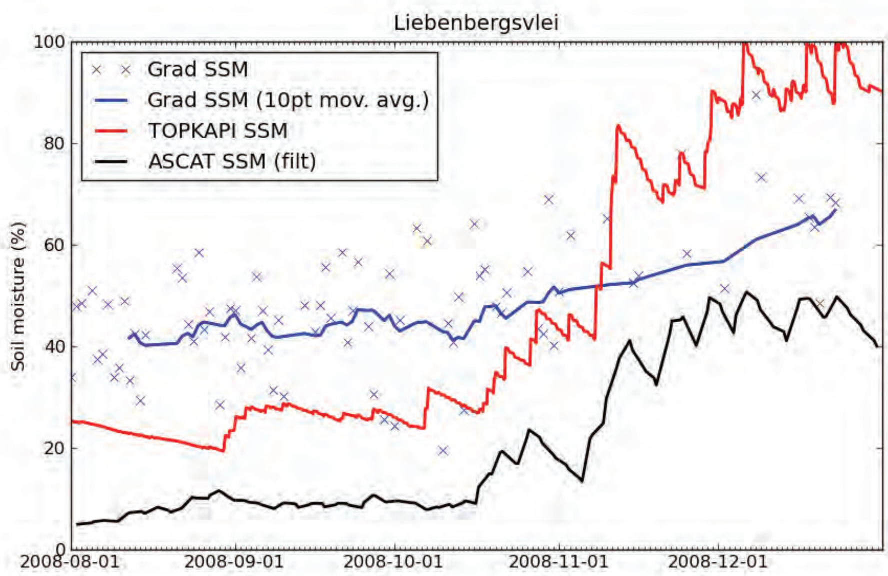


Figure 6.13 Site C: Liebenbergsvlei. The description of the points and curves follows that of Figure 6.11. The estimates are over a 0.25° block surrounding the TOPKAPI-LSM cell.

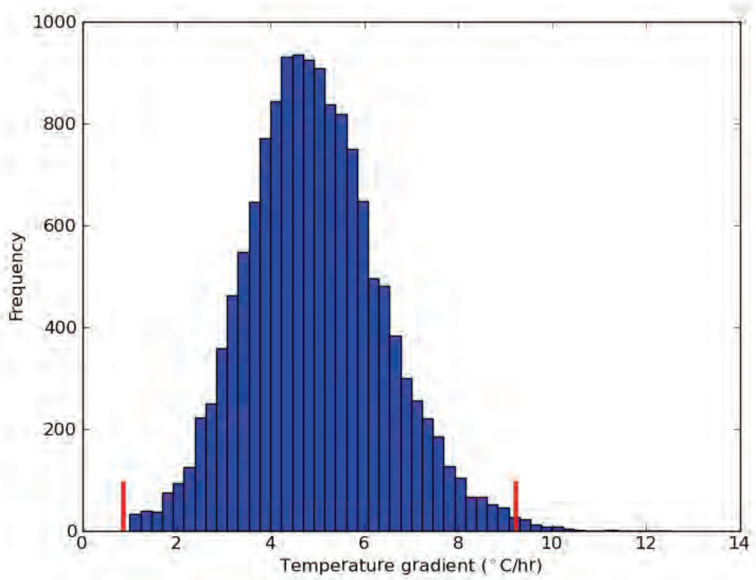


Figure 6.14 Distribution of temperature gradients for the Crocodile site, D. The description is the same as that in Figure 6.12.

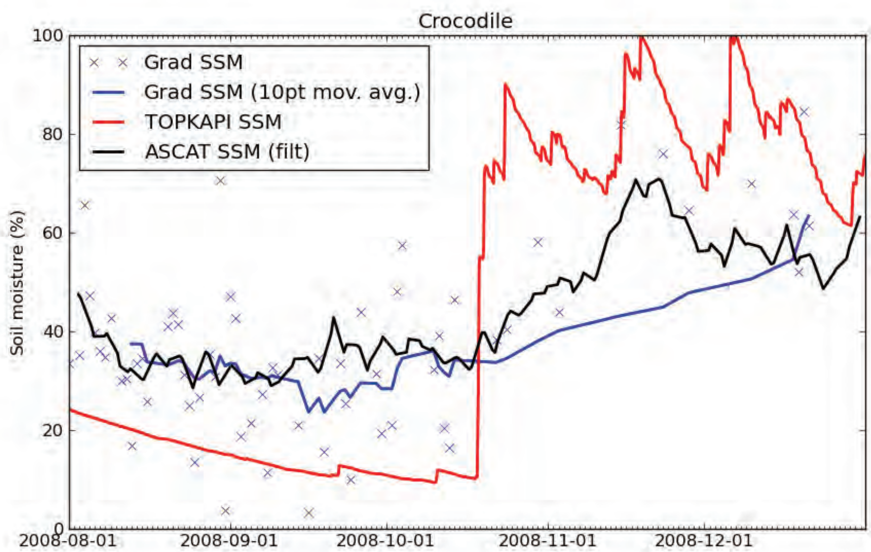


Figure 6.15 Site D: Crocodile. The description of the points and curves follows that of Figure 6.13. The estimates are over a 0.25° block surrounding the TOPKAPI-LSM cell.

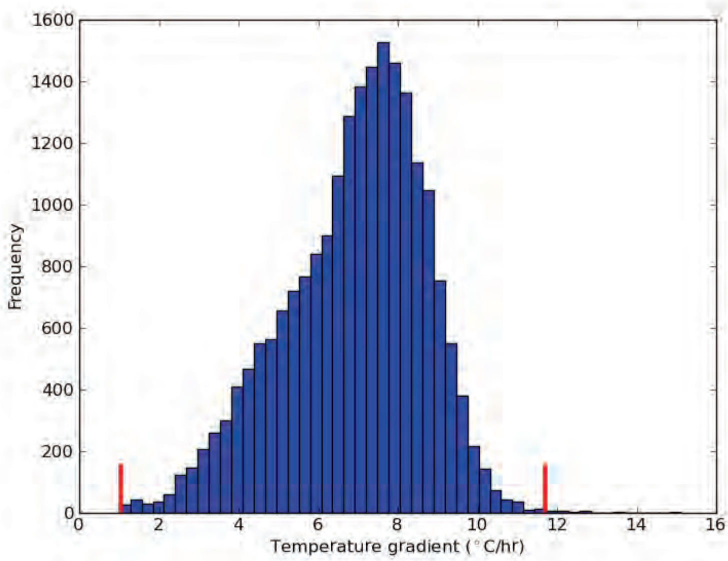


Figure 6.16 Distribution of temperature gradients for the Western Cape site, A. The description of the points and curves follows that of Figure 6.12. The estimates are over a 0.25° block surrounding the TOPKAPI-LSM cell. Note the crowding of the higher gradients, suggesting that dry periods predominate.

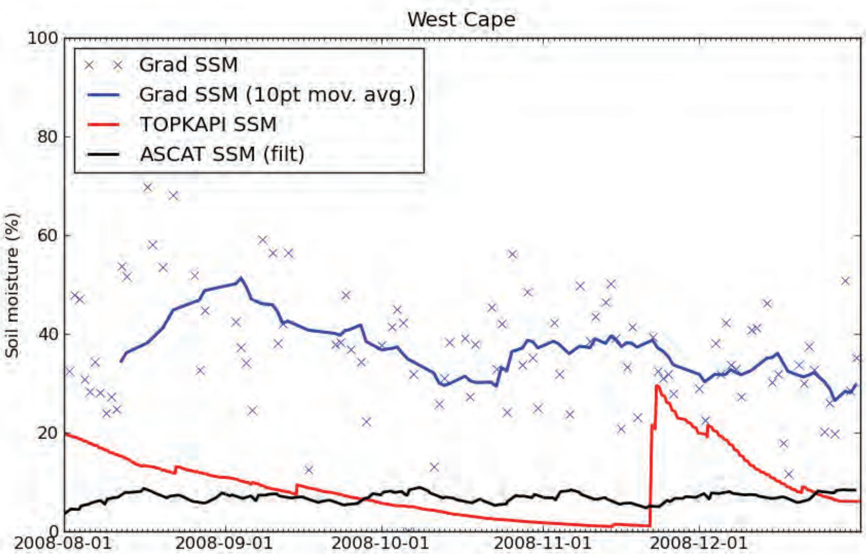


Figure 6.17 Site A: Western Cape. The description of the points and curves follows that of Figure 6.13. The estimates are over a 0.25° block surrounding the TOPKAPI-LSM cell.

The ΔT/Δt SSI estimates for all four sites are bounded by 40% to 60% wetness, with much less variability and greater intermittency than the other two estimates. The interim conclusion is that this is not a useful analysis method.

Summary

The temperature-gradient ($\Delta T/\Delta t$) based SSI estimates, after filtering with a moving average, loosely follow the TOPKAPI and ASCAT estimates in a general sense (but not in magnitude), so the idea had some merit. However, the method is not useful operationally, specifically because it is based on Infra Red observations of the ground from METEOSAT, which are severely affected by cloud, reducing observations during critical periods. Therefore the method will be abandoned as a means of SSI estimation.

The TOPKAPI estimates of SSI are available every three hours, using rainfall and other meteorological variables to compute SM estimates. The ASCAT SSM observations are available at a given site every 3 days on average and its C-band radar penetrates cloud to the ground; thus both estimates are unaffected by cloud cover and (barring loss of transmission) give reliable estimates on time. Once the ASCAT SSM observations are filtered to SWI, they have a very good linear relationship ($R^2 > 90\%$) with the TOPKAPI SSI values over a large proportion of the country but do not match over the dry Western areas, for reasons yet to be determined. However, we note that there is an expectation of an update of the ASCAT product which will enable us to make some more calculations in the future. This may help to explain the poor comparisons over the very dry areas of the country, but we do not expect there to be much improvement of the already high R^2 values for a large portion of the country, except in improving the slopes of the regression lines.

We are comfortable with the association between these two estimates and look forward to two other methods of corroboration: hydrological validation of TOPKAPI-based estimates and ground truthing. At this stage, we judge that the most promising near real-time estimate of SM is the TOPKAPI model in Land Surface Mode which uses the appropriate remote sensing and meteorological data streams.

7. PROVIDING THE 'BEST SOIL MOISTURE ESTIMATE ON THE WEB'

INTRODUCTION

The purpose of this project is to make the best available estimate of Soil Moisture over the whole country, in near real time, in as much detail as feasible, on a daily basis, freely accessible on the web. Users of this information include Agriculturalists, Flood Forecasters and Drought and Catchment Managers. The time-scales will likely be different for each. Some will want up-to-date information as soon as possible; others may want a longer perspective. What we have made possible with this deliverable is that either a highly technical or a more informal enquirer should be able to access not only the images, but if desired, the data which were used to create them. This has been made possible in an unsupervised manner, by the design of the web-site which this report will describe.

The best soil moisture estimate available, via this project at this juncture, is undoubtedly the Soil Saturation Index (SSI) product calculated every 3 hours from the TOPKAPI hydrologically-based cells in Land Surface Model (LSM) mode. We reiterate that SSI = $(\theta - \theta_r)/(\theta_s - \theta_r)$. The alternatives we envisaged were (i) temperature gradient estimates (ii) probe data and (iii) remote sensing estimates. As noted in the last paragraphs of Chapter 6, it turned out that temperature gradient estimates are intermittent and contaminated by cloud cover and the currently available probe data are insufficient for our purposes. The most promising alternative to the LSM estimates is the remote sensing product, but the observations are relatively infrequent. The ASCAT SWI estimates were useful in corroborating the TOPKAPI SSI outputs, but are not useful in raw form as they require filtering to remove noise and extract trends.

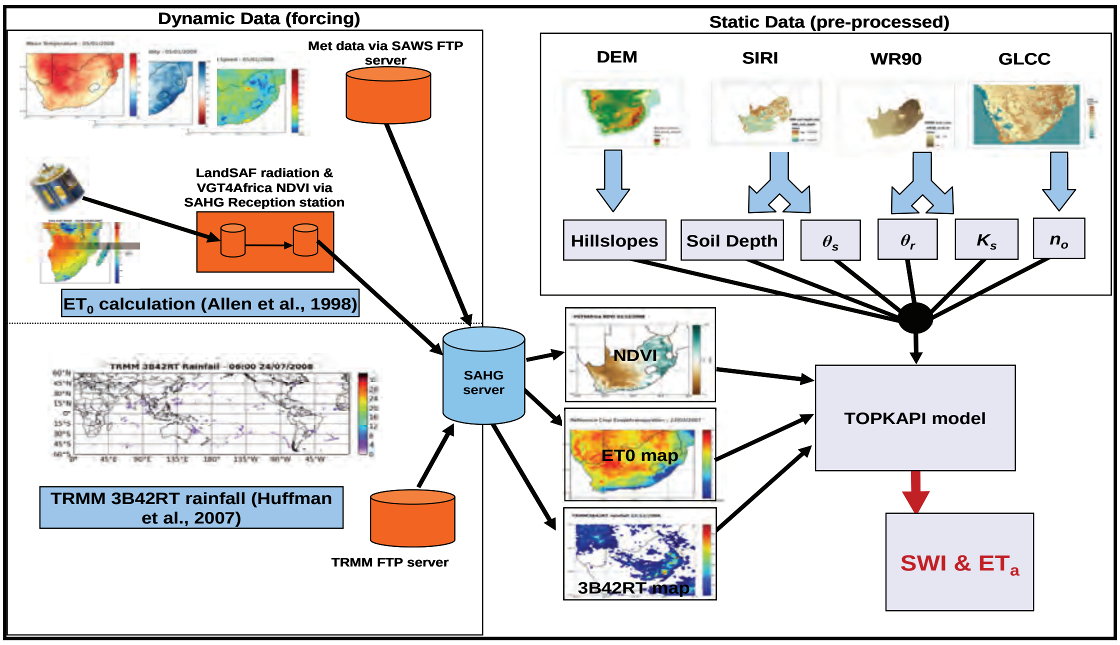


Figure 7.1 Flow chart summarising computation of Actual Evapotranspiration (ET_a) and Soil Saturation Index (SSI).

What then constitutes the data set producing the "best SM estimates available on the web"? The data are created as displayed in the flowchart in Figure 7.1, repeated from Chapter 4.

Each day at 12:40 SAST, our computers at UKZN-SAHG digest the data streams of various forms of information, already concentrated by our file servers:

- Unified Model forecasts from SAWS provide temperature, humidity and wind speed at hourly intervals over the subcontinent on a 0.11° grid.
- LSA-SAF (via EUMETCast) provides us, at 30 minute intervals, estimates of energy reaching the Earth's surface at a spatial resolution of about 3 km. The VGT4AFRICA project produces SPOT-5 NDVI estimates of vegetation activity (at 10-day intervals, delivered via EUMETCast) .
- NASA gives us TRMM 3B42 RT estimates of spatial rainfall on a 0.25° grid (roughly twice that of the Unified Model) at 3 hour intervals.

Our static data sets include:

- the Shuttle Radar Topography Mission (SRTM) 90 m Digital Elevation Model which yields our local slopes for drainage of the TOPKAPI cells,
- SIRI's soil depth and conductivity estimates
- USDA's global estimate of Land Cover

The Reference Crop Evapotranspiration (ET_0) is calculated at hourly time-steps on the 0.11° UM grid. The computation of the Soil Water content of each of the near 7000 TOPKAPI cells starts 37 hours earlier at midnight of the previous day and is performed at each cell, having its own physical characteristics, with the data available at 3-hour intervals. These 1 km square cells are 'located' on the nodes of the (approximately 16 km square) UM grid and are forced by the already calculated ET_0 and the TRMM rainfall. The reason for the delay in starting the computation only at noon is that the TRMM data are only available from NASA typically 12 hours after observation.

The results of these computations are stored in HDF5 (compressed binary) file format. Further computations convert these numbers to geo-referenced raster (gridded) data in GeoTIFF format, which can then be used to create maps of ET_0 and SSI. GeoTIFF and ASCII files of the numbers making up the maps are available for download on our FTP site.

Thus, the basic data store available to an enquiry from about midday onwards on any day will be maps of ET_0 (accumulated depths in mm at 1 hour intervals) and SSI (as snap-shots every 3 hours) up to midnight 12 hours before; the most recent maps will automatically load if an enquiry is made to the web-site. The enquirer will be able to browse any previous time interval to access maps of interest. An enquirer desiring access to the data directly for their own purposes (Flash Flood Forecasting, Agriculture, Drought Monitoring or Catchment Management) will find on each image, a pointer to GeoTIFF and ASCII files storing these data. These files can be downloaded either informally, or in blocks using FTP software. We note that the basic HDF5 files are not publicly available, as we expect that this format will be unfamiliar to most users. The system is automatic and a data user does not require any permission or human intervention by us to obtain the model results.

DESCRIPTION OF THE PRODUCT

The essence of the work is to make the results of our modelling available to anyone who is interested in either viewing the images or downloading the numbers from which the images are produced. This includes the serious user (for the purposes of Flash Flood Forecasting, Agriculture, Drought Monitoring and Catchment Management) as well as the more casual enquirer.

As described in the introduction, the idea is that the web-site should be friendly, easy to use and accessible without needing permission in asking for data. The enquirer should be able to comfortably browse the available data (displayed directly in a web browser) and easily download the desired information in a GIS ready format.

On the following two pages are Figure 7.2 and Figure 7.3 which show the web based data browsing interface. The arrow buttons above the display area can advance (or reverse) through the set of images in order. On first access to the site, the most recent image available is loaded by default. If desired, entering a date and time in the appropriate text box will load the required frame from the data-base.

The first of the examples we display here in Figure 7.2 (top panel) starts with the standard background of the Natural Earth relief (Natural Earth, 2010). In the dialog box in the top-right corner it will be seen that the 'country borders' have been checked and the currently available 'Reference Crop ET' products are not checked. We have not provided access to the ET_a product at this juncture as it has yet to be validated by hydrological back-calculation, to be conducted in the follow-on project. Also, we are not in a position to share the core data from EUMETSAT, NASA & SAWS, such as rainfall, meteorological and energy data etc. We may only share information where we have processed the data to add value.

In the second image in Figure 7.2 (bottom panel), the 'Reference Crop ET' block has been checked and the date and time selected in the 'Current file date' box are shown as 2010-01-07T15:00. This shows the ET_0 data for the 15:00 time-slot on 2010-01-07.

The third and fourth images displayed in Figure 7.3 are accessed by clicking the arrow-keys to the right of the 'Current file date' box. Navigation is that easy. Note that these images are separated by an hour. The 'Soil Saturation Index' is calculated at 3-hour intervals, constrained by the frequency and availability of TRMM rainfall data coming to us from NASA.

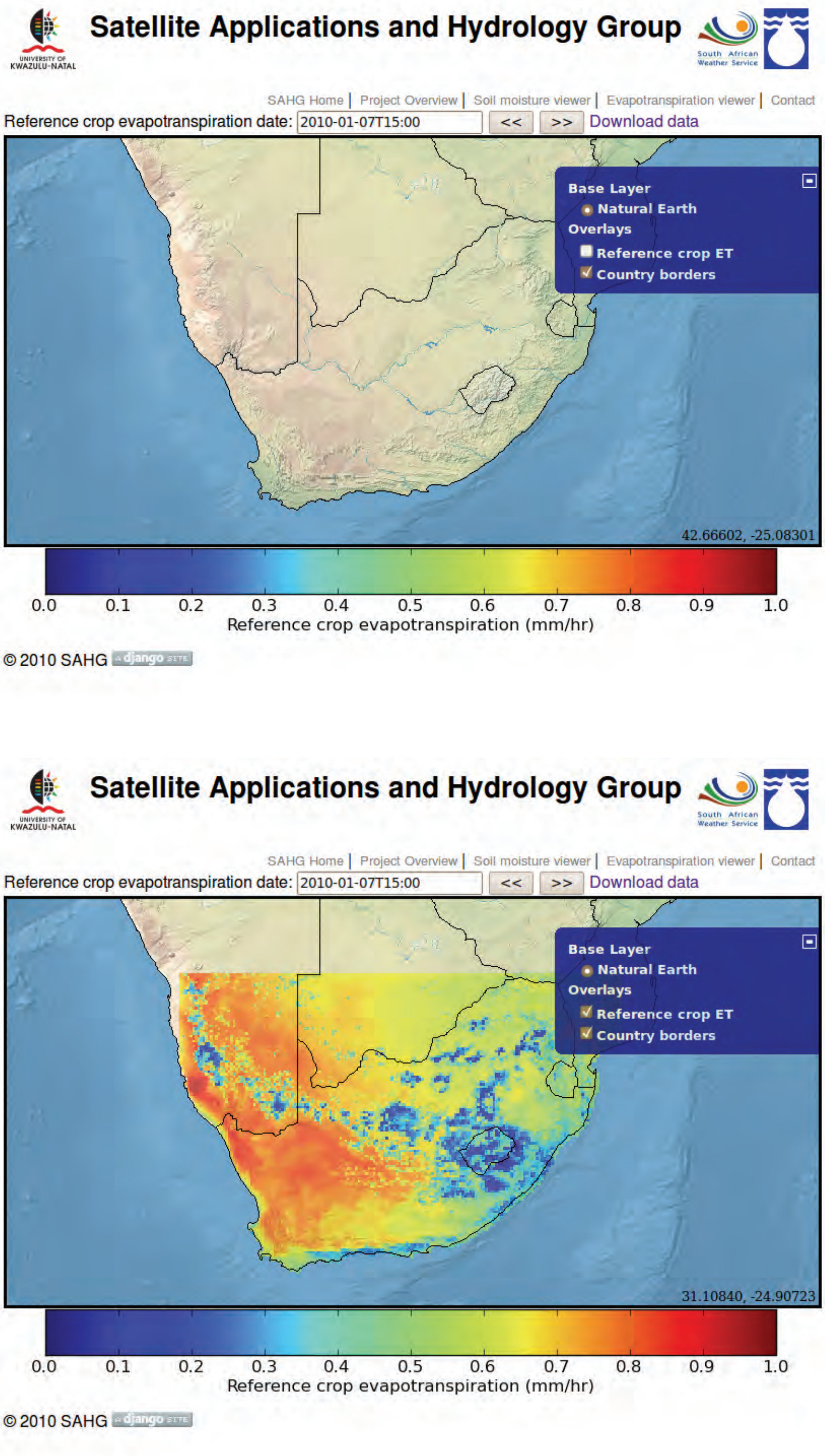


Figure 7.2 The background of relief and country boundaries, then overlaid by ET_0

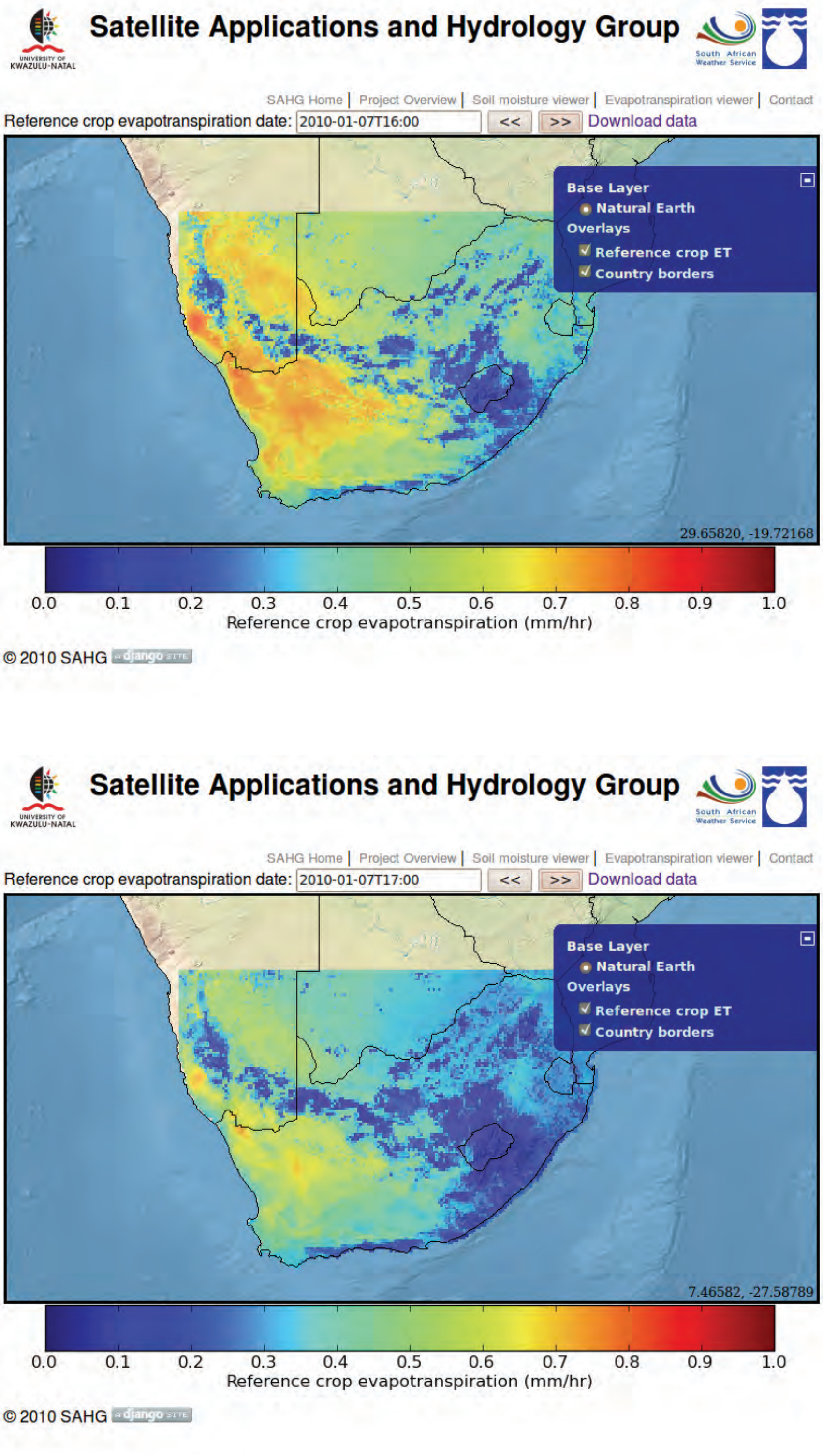


Figure 7.3 ET_0 overlay snapshots for the 2 hours following that at 1500 on Figure 7.2

The initial design of the web-site was to provide as much convenience as possible to the enquirer in a 'one size fits all' package. We first envisaged that access to the data and images would be achieved in the way described in the following flowchart in Figure 7.4.

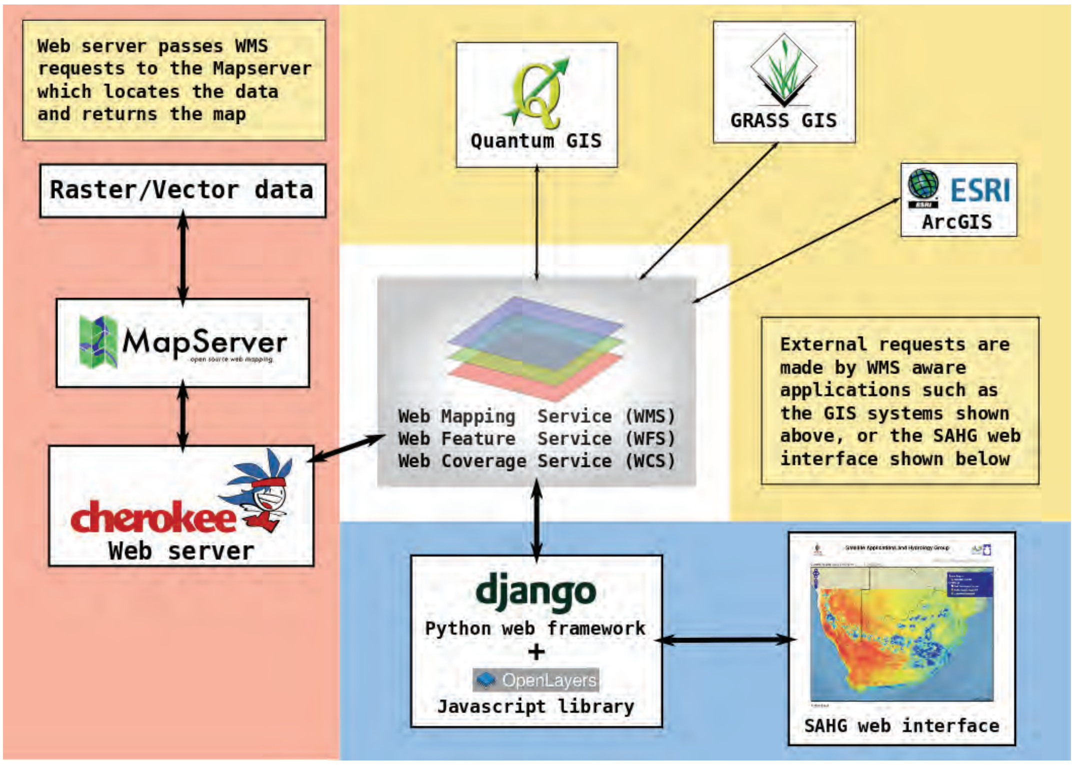


Figure 7.4 Outline sketch of the Web Mapping Service (WMS) providing the SAHG data products demonstrated in Figure 7.2 and Figure 7.3.

The core concept which this flowchart conveys is that the SAHG web server would expose a Web Mapping Service (WMS) interface to an enquiry. Clients (GIS software in the hands of its users) were to make requests for information or data to the WMS using a standard syntax. The web server would pass the requests to the Mapserver, which, after collecting the appropriate data, would respond in a standard manner (via the web server) to the user. The client software would then display the results to the user. This design favours the more serious user.

In development of the system envisaged in Figure 7.4, we designed a more streamlined approach. The alternative architecture is much simpler to implement and easier for any enquirer to use, so we have decided to share the information with enquirers through the architecture which appears in the flowchart in Figure 7.5. The initial enquiry will display the images; an enquiry for the original data will easily be directed to the FTP site, whence the relevant files may be downloaded individually or in blocks. The enquiry procedure is described on the following pages.

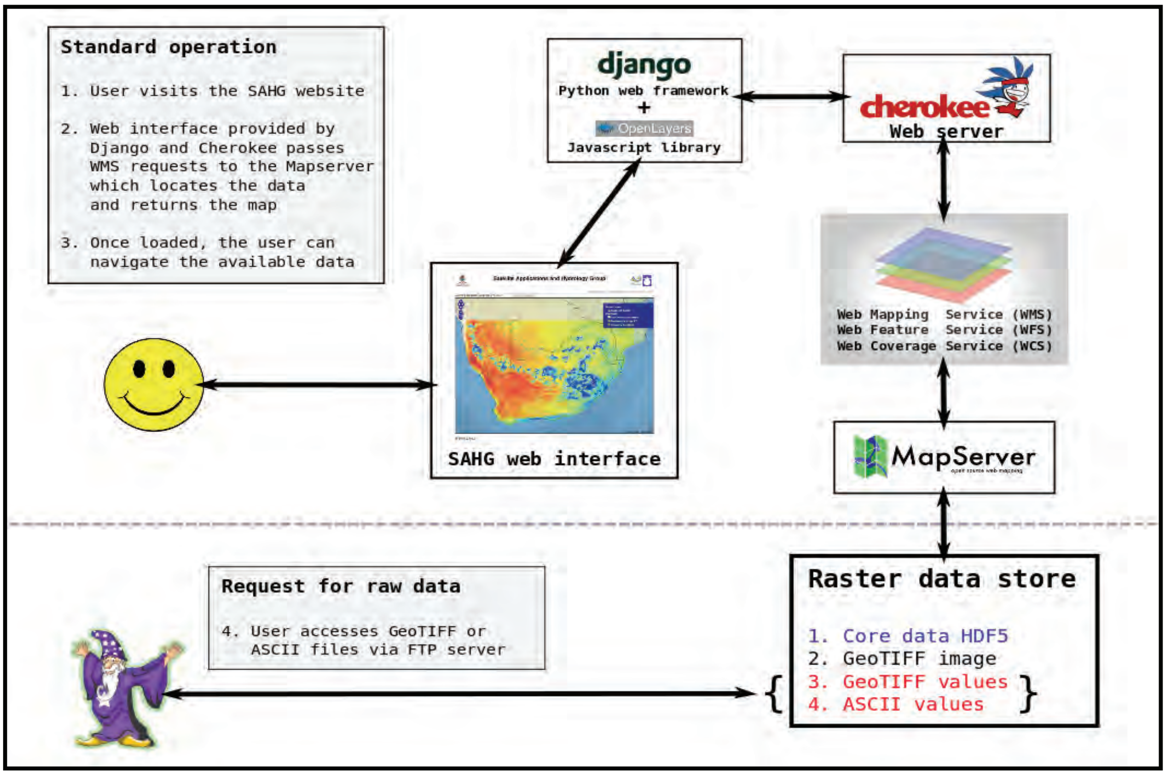


Figure 7.5 Two paths to the information on the web-site: the initial interrogation and one requesting data. The former provides images of ET_0 and SSI; the latter allows the same enquirer, via a few clicks of a mouse, to make use of FTP software to download either GeoTIFF or ASCII files which can then be processed by the enquirer's own software.

Figure 7.6 shows some archived SSI data. Where there were gaps in the data due to lack of soil properties, the space has been infilled using nearest neighbour modelling. Figure 7.7 demonstrates the initiation of an enquiry for the background data. This demonstration shows how easy it is for an enquirer to access the raw data.

Hidden behind this description is the programming effort which required several months of coding approximately 5000 lines of original Python code (calling standard libraries) to make it all happen. Thus this created system ensures that all the data streams are handled and archived as efficiently as possible and allow access by external users as fluently as is feasible. We hope that the system will be taken over in due course by either SAWS or ARC, so that it continues to supply valuable information after the end of the new 3-year follow-on project commencing in April 2010.

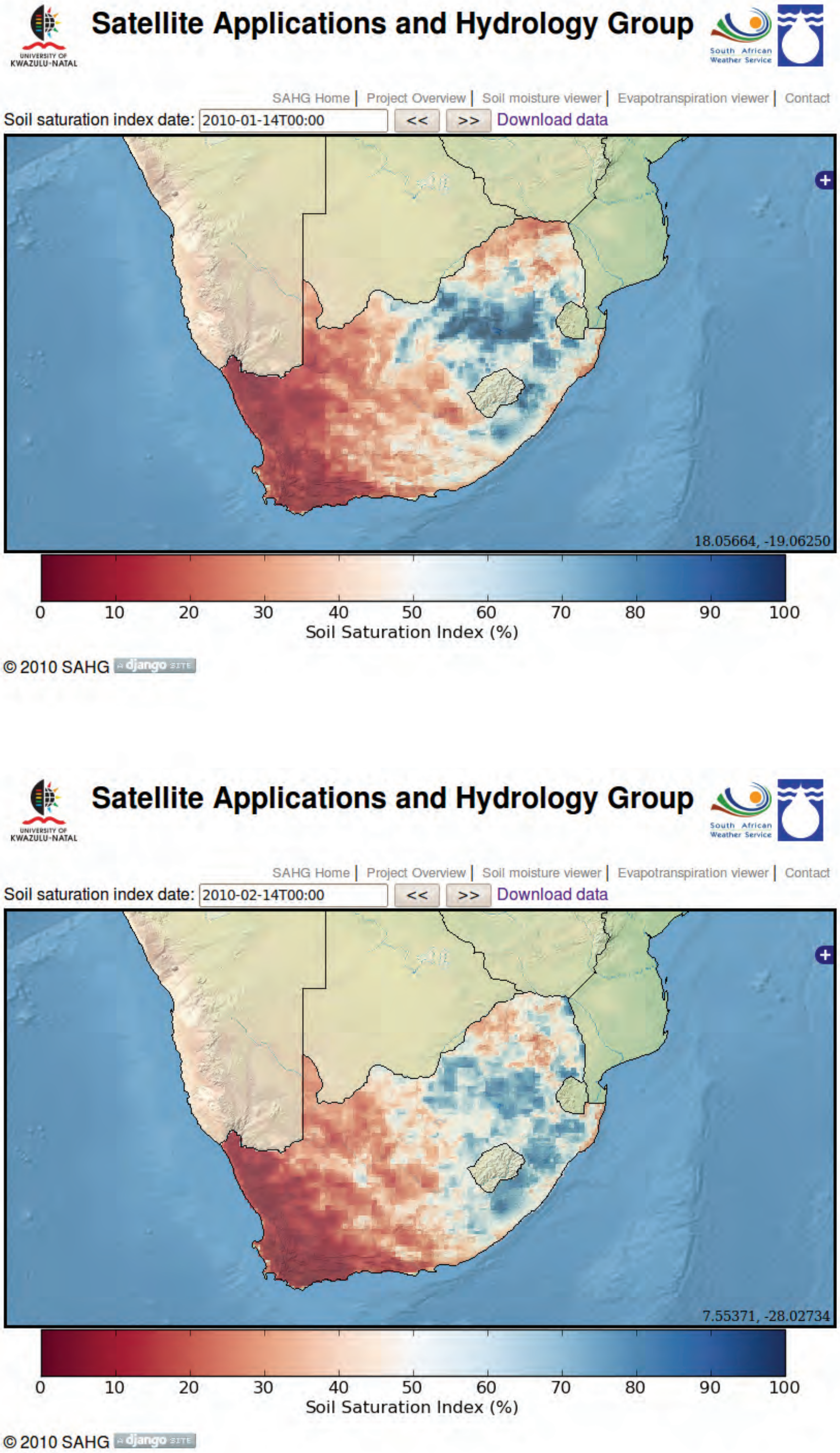


Figure 7.6 Soil Saturation Index maps at midday on 15th of January & February 2009

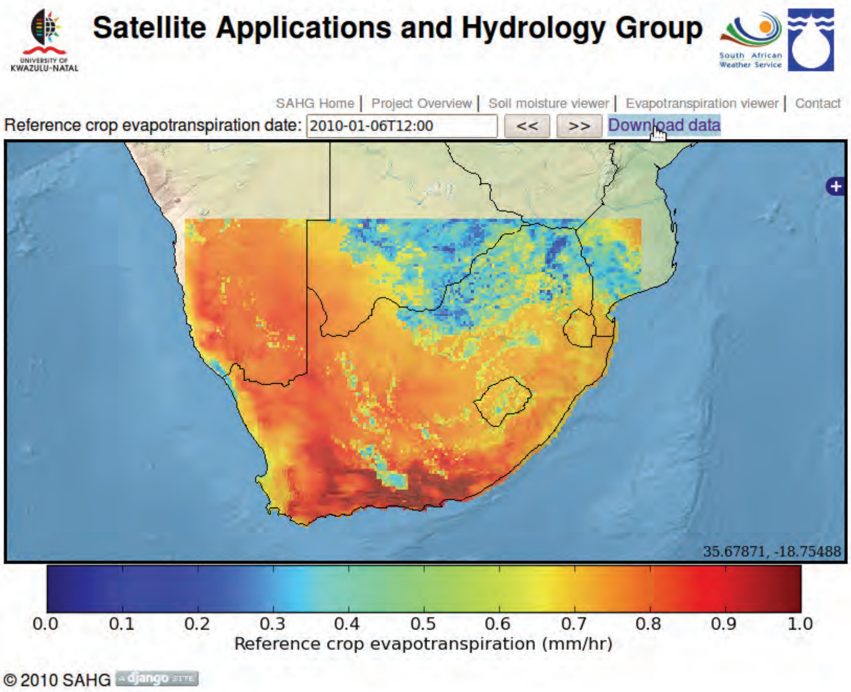


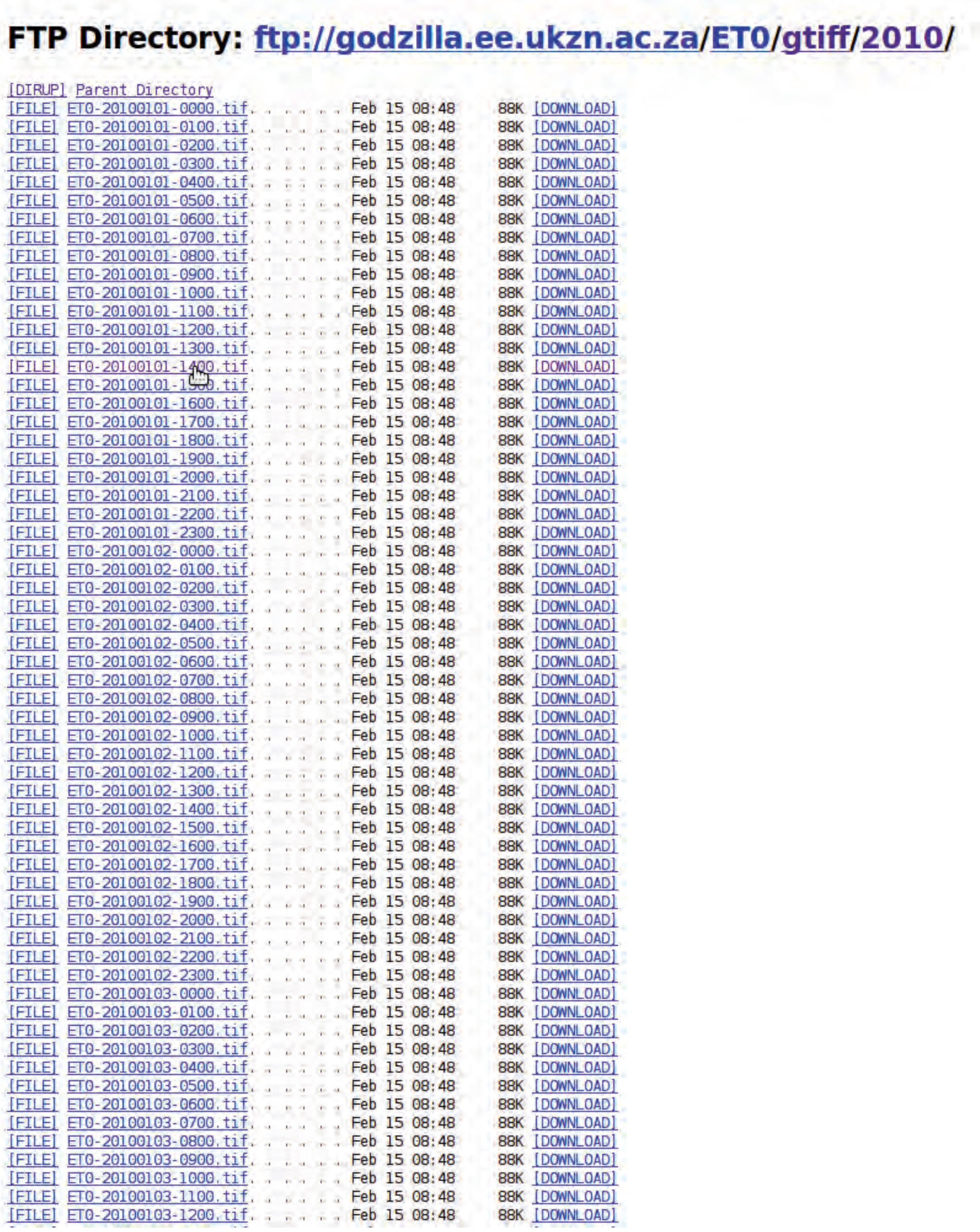
Figure 7.7 Map of ET_0 at midday on 6 Jan 2010, demonstrating link to data download.

The ‘Download Data’ block has been clicked which launches the window shown in Figure 7.8. Clicking on the ‘gtiff’ directory opens a new window shown in Figure 7.9.

Figure 7.8 FTP directory to data files.

Figure 7.9 The GTIFF directory containing files collected in years.

Clicking on the 2010 directory opens a new window as in Figure 7.10. Clicking on an individual file will open it; alternatively clicking on 'DOWNLOAD' opens a dialog box asking where to download the data. Note that these ET_0 data are archived every hour.



FTP Directory: ftp://godzilla.ee.ukzn.ac.za/ET0/gtiff/2010/

[DIRUP] Parent Directory
[FILE] ET0-20100101-0000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0400.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0500.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0600.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0700.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0800.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-0900.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1400.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1500.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1600.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1700.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1800.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-1900.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-2000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-2100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-2200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100101-2300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0400.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0500.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0600.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0700.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0800.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-0900.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1400.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1500.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1600.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1700.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1800.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-1900.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-2000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-2100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-2200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100102-2300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0300.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0400.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0500.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0600.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0700.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0800.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-0900.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-1000.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-1100.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]
[FILE] ET0-20100103-1200.tif. . . . . . Feb 15 08:48 88K [DOWNLOAD]

Figure 7.10 List of files in the 2010 GTIFF directory, available for opening or downloading.

TECHNICAL SUMMARY

We have chosen to move towards distributing our data products using the Open Geospatial Consortium's (OGC) standards for Web Mapping Service's (WMS). Using WMS brings many advantages, but the most important are (i) flexibility in the data formats and (ii) dissemination methods available to the data user.

Software applications that are WMS aware (e.g. any well known GIS packages) can make a request to a WMS over the internet. The first request will usually be for the capabilities of the particular WMS. Capabilities requests return a list of available data-sets and formats, which the service is set up to provide. Further requests are typically for a particular combination of the available data-sets, which will be sent to the software application.

A concrete (but simplified) example would be to access the SAHG web interface. When an internet user types the correct web address into their web browser's address bar, a request for a web page is sent to the SAHG web server. This request is interpreted dynamically and the result will depend on the currently available data (i.e. a slightly different web page will be displayed in the user’s web browser, based on the data available). The request for a web page includes a WMS request (produced on the fly by the Django web framework and the OpenLayers Javascript library used to build the web site). The web server recognises the request as a WMS request and passes this to the Mapserver. It is the Mapserver that interprets the request and returns the required data or information (via the web server) to be handled in the users web browser.

In addition to the WMS services we are making this data set available on an FTP server, as this vehicle is more familiar and convenient for many users, and enables us to simplify our conversion of data from the core HDF5 format files.

Summary

The work reported in this chapter is the culmination of the entire endeavour that is this project. It is designed to make these valuable data readily available to practitioners in Agriculture, Catchment Management Drought Monitoring and Flood Forecasting. The information will also find use in initialising hydrological models of catchments. In addition, we suggest that the archive will be valuable for monitoring the reactions in the country’s soil water and evaporation due to changes in weather and climate.

8. CONCLUSION

We open this conclusion by summarising what we believe to be the original technical contributions made by this project

TOPKAPI	- (Ch 2) modification to the basic algorithm to (i) ensure continuity (ii) maximise speed of computation using a hybrid computation (iii) conversion of 4D flow to 8D flow on the grid of cells (iv) modification to the ET algorithm to ensure saturation at the end of an interval under appropriate conditions (the original model always produced a deficiency even if there was water available on the surface in the overland store).
ETa	- (Ch 3) utilizing Unified Model Forecast 'data' and LSA-SAF radiance data to drive the Penman-Monteith model, hence to interpolate this information spatially over the country at 1 hour intervals in near real time.
Temperature Gradients	– (Ch 4) learning to manage huge data streams in various projections and knowing when to quit.
SAHG	- (Ch 5 & 7) our Satellite and Hydrology Group website – real-time spatial monitoring of ET_a and SM over the country and dissemination of data; we designed the web-site specifically to disseminate this work

As will be evident from the work done in this project, one of the most difficult hydrological variables to collect and interpret is Soil Moisture. It is more spatially variable than rainfall because of the spatial variability of soil type and depth, local slope of ground and vegetation. Local estimates of Soil Moisture are valuable in the vicinity of the measurement. However to interpolate this information spatially is difficult. It was the purpose of this project to find a way to do it and share the results. This conclusion pulls together the summaries made at the end of each of the 6 Chapters that make up the body of the report and ties them into a coherent whole.

The purpose of Chapter 2 was to compare, for the purpose of corroboration, not validation, two independent approaches used to estimate soil moisture at the scale of a region-sized catchment (Liebenbergsvlei, 4625 km^2, South Africa). The first estimation was derived from physically based hydrological modelling of the catchment using the TOPKAPI model (Liu and Todini, 2002) and the second was derived from the remotely sensed observations of the scatterometer on board the ERS satellite. A calibration procedure of the TOPKAPI model was carried out consisting of the adjustment of the four most sensitive parameters of the model according to runoff production and routing. A good agreement was found between observed and modelled hydrographs of the Liebenbergsvlei catchment for both the calibration and the verification period. The comparison between the modelled Soil Saturation Index (SSI) and the remotely sensed Soil Wetness Index (SWI) soil moisture estimates was made over catchment and footprint scales. As the satellite only provides soil moisture for the topsoil layer (first 5 centimetres), a conceptual infiltration model developed by Wagner et al. (1999c) was applied to the remotely sensed surface soil moisture estimates in order to estimate an SWI. The comparison between the modelled SSI and remotely sensed SWI was shown to be good with regression coefficients between 0.678 and 0.923. Even if a constant bias of around 19% is identified, the dynamic of the soil moisture behaviour is very coherent between the two approaches.

The results obtained at this stage (the first year of the project) were very encouraging for (i) hydrological modelling and the possibility of using remotely sensed soil moisture to validate the models and also to initialise them; assimilation of the remotely sensed soil moisture data into hydrological models during simulations is also an exciting possibility, (ii) remote sensing and the possibility of using physically based hydrological models to validate and disaggregate the soil moisture estimations down to fine spatial scales.

Further research will aim to improve the modelling of the vertical fluxes that explicitly represent the vertical water transfers in the soil and will allow direct comparison between the remotely sensed soil moisture at the surface (first 5 centimetres of soil) without being dependent on the conceptual infiltration model used in the present study to infer the soil moisture profile from the surface remotely sensed soil moisture. Such a complete physically based model should help to better understand the processes that control the soil moisture patterns at regional scale and will be applied as a physically based soil moisture back-calculation and disaggregation tool.

Chapter 3 presented a brief explanation of the methodology employed for producing spatially distributed estimates of reference crop evapotranspiration using the guidelines set out by Allen et al. (1998). Novel use was made of remotely sensed radiation estimates and forecast fields from the SAWS numerical weather prediction model, the Unified Model. Initial results show that the correspondence in ET_0 estimated from the model forecast fields are strongly correlated with ET_0 calculated from weather station observations at both the hourly and daily timescales. This builds significant confidence in the Unified Model forecasts produced by SAWS, as well as suggesting that spatially distributed ET_0 estimation is viable. Spatial estimates of ET_0 were then produced routinely and maps were made available on the internet since 2008.

In Chapter 4, an estimate of the actual evapotranspiration ET_a was obtained from the ET_0 fields through the application of appropriate crop and water stress coefficients using individual cells of the TOPKAPI model. The methods used to achieve this are a combination of Meteorological variables, assisted by remote sensing estimates of rainfall, energy and vegetation. These are used to force a Land Surface Model (a TOPKAPI cell) at each one of 6989 sites covering over 1 million square kilometers of the country. Novel techniques were used to combine these data streams into the information product.

It was thought at the inception of the project that early morning temperature gradients measured at the earth's surface using remote sensing might be a useful indicator of Surface Soil Wetness, which might then be linked to Soil Moisture (SM). Chapter 5 documents the effort put into this idea. It turns out that the observations are occluded by cloud cover to such an extent that the inference on SM is not as useful as was first hoped. Nevertheless, the work done (in the early days of the project before we had made linkages with the influential people in the remote sensing world) was extremely useful in getting our minds around the remote sensing paradigm and very good training in the management of large quantities of data.

Chapter 6 documents the crucial comparisons between the three estimates of Soil Moisture (SM). The temperature-gradient ($\Delta T/\Delta t$) based SSI estimates, after filtering with a moving average, loosely follow the TOPKAPI SSI and ASCAT SWI estimates in a general sense (but not in magnitude), so the idea had some merit. However, the method is not useful operationally, specifically because it is based on Infra Red observations of the ground from METEOSAT, which are severely affected by cloud, reducing observations during critical periods. Therefore the method was abandoned as a means of SSI estimation. The TOPKAPI estimates of SSI are available every three hours, using rainfall and other meteorological variables to compute SM estimates. The ASCAT SSM observations are available at a given site every 3 days on average and its C-band radar penetrates cloud to the ground; thus both estimates are unaffected by cloud cover and (barring loss of transmission) give reliable estimates on time. Once the ASCAT SSM observations are filtered to SWI, they have a very good linear relationship ($R^2 > 90\%$) with the TOPKAPI SSI values over a large proportion of the country but do not match over the dry Western areas, for reasons yet to be determined. However, we note that there is an expectation of an update of the ASCAT product which will enable us to make some more calculations in the future. This may help to explain the poor comparisons over the very dry areas of the country, but we do not expect there to be much improvement of the already high R^2 values for a large portion of the country, except in improving the slopes of the regression lines. We are comfortable with the association between these two estimates and look forward to two other methods of corroboration: hydrological validation of TOPKAPI-based estimates and ground truthing. At this stage, we judge that the most promising near real-time estimate of SM is the TOPKAPI model in Land Surface Mode which uses the appropriate remote sensing and meteorological data streams.

The work reported in Chapter 7 is the culmination of the entire endeavour that is this project – making our products freely available on the web. The web-site is designed to make these valuable data readily available to practitioners in Agriculture, Catchment Management Drought Monitoring and Flood Forecasting. The information will also find use in initialising hydrological models of catchments. In addition, we suggest that the archive will be valuable for monitoring the reactions in the country’s soil water and evaporation due to changes in weather and climate.

We conclude by suggesting that what is exciting about this product is that, for the first time, it provides a dynamic monitoring tool to flood forecasters, planners and agriculturalists, who will be able to see what is happening to the wetness of the ground, in spatial and temporal detail, up to date, on a daily basis. The WRC should be justly proud of sponsoring this valuable research.

REFERENCES

Allen R.G., Pereira L.S., Raes D., Smith M. 1998. Crop evapotranspiration – Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56. Rome.

Allen R.G., Pruitt W.O., Wright J.L., Howell T.A., Ventura F., Snyder R., Itenfisu D., Steduto P., Berengena J., Yrisarry J.B., Smith M., Pereira L.S., Raes D., Perrier A., Alves I., Walter I., Elliott R. 2006. A recommendation on standardized surface resistance for hourly calculation of reference ET_0 by the FAO56 Penman-Monteith method. Agricultural Water Management, 81, 1-22.

Bartholomes J and Todini E (2005) Coupling meteorological and hydrological model for flood forecasting. *Hydrol. Earth Syst. Sci*. **9**(4) 333-346.

Beven K (1997) TOPMODEL: a critique. *Hydrol. Process.* **11** (9) 1069-1086.

Beven K and Germann P (1982). Macropores and waterflow in soils. *Water Resour. Res.* 18, 1311-1325.

Beven K. and Kirkby M.: A physically-based variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43-69, 1979.

Biftu, G. F. and Gan, T. Y.: Semi-distributed, physically based, hydrologic modeling of the Paddle River Basin, Alberta, using remotely sensed data, J. Hydrol., 244, 137-156, 2001.

Blyth, E.: Modelling soil moisture for a grassland and a woodland site in south-east England, Hydrol. Earth Syst. Sci., 6(1), 39-47, 2002.

Brooks RH and Corey A.T (1964) Hydraulic properties of porous media. *Hydrology Papers*, No. 3. Colorado State University, Colorado, USA.

Bruckler L. and Witono H.: Use of remotely sensed soil moisture content as boundary conditions in soil-atmosphere water transport modeling. 2. Estimating soil water balance, Water Resour. Res., 25, 2437-2447, 1989.

Cai J., Liu Y., Lei T., Pereira L.S. 2007. Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. Agricultural and Forest Meteorology, 145, 22-35.

Chow VT, Maidment DR and Mays LR (1988) *Applied Hydrology*. McGraw-Hill, New York.

De Lannoy, G. J. M., Verhoest, N. E. C., Houser, P. R., Gish, T. J., Meirvenne, M.: Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE3), J. Hydrol., 331, 719-730, 2006.

DLSI (1996) A digital elevation model for South Africa. Directorate of Land Surveys and Information. South Africa.

Dunne T (1978) Field studies of hillslope flow process. In: M J Kirkby (ed.) *Hillslope Hydrology*. p. 227-293. Wiley, New York.

Eisenhauer J.G. 2003. Regression through the origin. Teaching Statistics, 25(3), 76-80.

Entekhabi, D., Rodriguez-Iturbe, I., and Castelli, F.: Mutual interaction of soil moisture state and atmospheric processes, J. Hydrol., 184, 3-17, 1996.

Figa-Saldaña, J., Wilson, J., Attema, E., Gelsthorpe, R., Drinkwater, M., and Stoffelen, A. (2002): The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers, *Can. J. Remote Sens*., 28, 404-412.

Gerald CF and Wheatley PO (1994) *Applied Numerical Analysis*. (6th edn.) Addison-Wesley, New York, United States of America.

GLCC (1997) Global Land Cover Characteristics. United States Geological Survey (USGS). http://edcdaac.usgs.gov/glcc

GLCC: Global Land Cover Characteristics. *United States Geological Survey (USGS)*, http://edcdaac.usgs.gov/glcc, 1997.

Gordon H.A. 1981. Errors in computer packages. Least squares regression through the origin. The Statistician, 30(1), 23-29.

Gradshteyn IS and Ryzhik IM (1994) *Table of Integrals, Series, and Products.* Alan Jeffrey (ed.) Associated Press, New York, United States of America 1 204 pp.

Grayson R. B, Western, A. W., and Chiew F. H. S.: Preferred states in spatial soil moisture patterns: Local and nonlocal controls, Water Resour. Res., 33(12), 2897-2908, 1997.

Hébrard, O., Voltz, M., Andrieux, P., and Moussa, R.: Spatio-temporal distribution of soil surface moisture in a heterogeneously farmed Mediterranean catchment, J. Hydrol., 329, 110-121, 2006.

Hodnett, M. G. and Bell, J. P.: Soil moisture investigations of groundwater recharge through black cotton soils in Madhya Pradesh, India, Hydrol. Sci. J., 31(3), 361-381, 1986.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, Water Resour. Res., 34(12), 3405-3420, 1998.

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

Jarvis A., Reuter H.I., Nelson A., Guevara E.: Hole-filled seamless SRTM data V4, *International Centre for Tropical Agriculture (CIAT)*, available from http://srtm.csi.cgiar.org, 2008.

Jury MR (2002): Economic Impacts of Climate Variability in South Africa and Development of Resource Prediction Models. *J. Appl. Meteor.*, 41, 46-55.

Kleinschmidt I, Sharp BL, Clarke GPY, Curtis B and Fraser C (2001) Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu-Natal, South Africa. *Am J Epidemiol*, 153:1213-21.

Kostov K. G. and Jackson T. J.: Estimating profile soil moisture from surface layer measurements – a review. Proceedings of the international Society for Optical Engineering (SPIE), Orlando, 11-14 April, 125-136, 1993.

Liu Z and Todini E (2002) Towards a comprehensive physically-based rainfall-runoff model. *Hydrol. Earth Syst. Sci*. **6**(5) 859-881.

Liu Z, Mario LV and Todini E (2005) Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian catchment. Hydrol. Earth Syst. Sci., 9, 347-364.

Liu Z., Martina M.L.V. and Todini E.: Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian catchment. Hydrol. Earth Syst. Sci., 9, 347-364, 2005

Liu, Z. and Todini, E.: Towards a comprehensive physically-based rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 6(5), 859-881, 2002.

LSA-SAF. 2006. Down-Welling Surface Shortwave Flux Product User Manual. Version 1.4.

Maidment DR, (1993) *Handbook of hydrology*. McGraw-Hill, New York.

Mark DM (1988) *Network models in geomorphology. Chapter 4, modelling Geomorphological Systems.* M.G. Anderson (ed.). John Wiley & Sons, Ltd. p.73-97.

Martina MLV (2004) *The distributed physically based modelling of the rainfall-runoff process.* Ph.D. Thesis. University of Bologna, Italy.

Martina MLV, Todini E and Libralon A, 2006. A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrol. Earth Syst. Sci.*, 10, 413-426.

Mckenzie RS and Craig AR (1999) Evaporation losses from South African rivers. WRC Report No.638/1/99.: Water Research Commission, Pretoria, South Africa.

McNamara, J. P., Chandler, D., Seyfried, M., and Achet, S.: Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment, Hydrol. Process., 19, 4023-4038, 2005.

Merz R and Blöschl G (2004). Regionalisation of catchment model parameters. *J. Hydrol.* 287, 95-123.

Merz, B. and Plate, E. J.: An analysis of the effect of spatial variability of soil and soil moisture on runoff, Water Resour. Res., 33, 2909-2922, 1997.

Midgley DC, Pitman WV and Middleton BJ (1994) Surface Water Resources of South Africa 1990. WRC Report No. 298/2.2/94. Water Research Commission, Pretoria, South Africa

Nash JE and Suttcliff JV (1970), River flow forecasting through conceptual models part I — A discussion of principles, *J. Hydrol.*, 10 (3), 282-290.

Natural Earth, "A Public Domain map dataset", http://www.naturalearthdata.com/, last accessed 2010-02-17.

Ottlé, C. and Vidal-Madjar, D.: Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY region, J. Hydrol., 158(3-4), 241-264, 1994.

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., and Scipal, K.: Assimilating scatterometer soil moisture data into conceptual hydrological models at the regional scale, Hydrol. Earth Syst. Sci., 10, 353-368, 2006.

Parak M (2007). Flood estimation for roads, bridges and dams. M.Sc. Eng. Thesis. University of KwaZulu-Natal.

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., De Troch, F. P., and De Troch, P. A.: Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, Hydrol. Process, 16, 995-1013, 2002.

Pegram G.G.S., Sinclair S., Wesson S. 2006. Daily Rainfall Mapping over South Africa: Modelling. WRC Report No. 1425/1/06. Water Research Commission, Pretoria, South Africa.

Pegram GGS, Nxumalo N and Sinclair S (2006) Soil Moisture Mapping from Satellites. Report K8/598 on a consultancy to the Water Research Commission, Pretoria

Pegram GGS, Sinclair S, Parak M, Sakulski D and Nxumalo N (2007) National Flood Nowcasting System: Towards an integrated mitigation strategy, Water Research Commission Report No. 1429/1/06, Pretoria, South Africa.

Pellenq, J., Kalma, J., Boulet, G., Saulnier, G.-M., Wooldridge, S., Kerr, Y., and Chehbouni, A.: A disaggregation scheme for soil moisture based topography and soil depth, J. Hydrol., 276, 112-127, 2003.

Python (2008). http://www.python.org/

Reichle, R. H., Mc Laughlin, D. B., and Entekhabi, D.: Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications, IEEE Trans. Geosci. Remote Sensing, 39(8), 1708-1718, 2001.

Rodell M., Houser P.R., Jambor U., Gottschalck J., Mitchell K., Meng C.-J., Arsenault K., Cosgrove B., Radakovich J., Bosilovich M., Entin J.K., Walker J.P., Lohmann D. and Toll D.: The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, 85(3), 381-394, 2004.

Rodriguez-Iturbe I.: Ecohydrology: a hydrologic perspective of climate-soil-vegetation dynamics, Water Resour. Res., 36, 3-9, 2000.

Savage M.J., Odhiambo M.G., Mengistu M.G., Everson C.S., Jarmain C. 2005. Theory and practice of evaporation measurement, with special focus on surface layer scintillometry as an operational tool for the estimation of spatially-averaged evaporation. 12th SANCIAHS symposium. 5-7 September, Eskom convention centre, Midrand, Johannesburg.

Sinclair S., Pegram G.G.S. (2005). Combining radar and rain gauge estimates using conditional merging. Atmospheric Science Letters, 6(1), 19-22.

Sinclair S. and G. G. S. Pegram (2009): A comparison of ASCAT and modeled soil moisture over South Africa, using TOPKAPI in land surface mode, *Hydrol. Earth Syst. Sci*. Discuss., 6, 7439-7482

Sinclair S, GGS Pegram and T Vischel (2007). Spatial conditioning of Evapotranspiration potential for distributed hydrological modelling in Southern Africa. 13th SANCIAHS Symposium, Cape Town, September.

SIRI (1987) *Land Type Series. Department of Agriculture and Water Supply, Soil and Irrigation Research Institute.* Memoirs on the Agricultural Natural Resources of South Africa. Pretoria.

SIRI: Land Type Series. Department of Agriculture and Water Supply, *Soil and Irrigation Research Institute*, Memoirs on the Agricultural Natural Resources of South Africa, Pretoria, 1987.

Strahler AN, 1957. Quantitative analysis of watershed geomorphology. *Trans. Amer. Geophys. Union*, 38, 913-920.

Tarboton DG, Bras RL and Rodriguez-Iturbe I, 1991. On the extraction of channel networks from digital elevation data. *Hydrol. Process.*. Vol. 5, p. 81-100.

Tasumi M., Allen R.G., Trezza R. and Wright J.L.: Satellite-based energy balance to assess within-population variance of crop coefficient curves. *Journal of irrigation and drainage engineering*, ASCE, 131(1), 94-109, 2005.

Teuling A. J., Uijlenhoet, R., van den Hurk, B. and Seneviratne, S. I.: Parameter sensitivity in LSMs: An analysis using stochastic soil moisture models and ELDAS soil parameters. Accepted for publication in *Journal of Hydrometeorology*, 2008.

Todini E (1996) The ARNO rainfall-runoff model. *J. Hydrol.*, 175, 339-382.

USGS DEM (2006). http://eros.usgs.gov/products/elevation/gtopo30/hydro/index.html

Vieux BE, CUI Z and Gaur A, (2004) Evaluation of a physics-based distributed hydrologic model for flood forecasting. *J. Hydrol.*, 298,155-177.

Vischel T., Pegram G.G.S, Sinclair S., Wagner W., and Bartsch A.: Comparison of soil moisture fields estimated by catchment modelling and remote sensing: a case study in South Africa , *Hydrol. Earth Syst. Sci.*, 12, 1-17, 2008a.

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

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, Nordic Hydrology, 38(1), 1-20, 2007.

Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sens. Environ., 70, 191-207, 1999c.

Wagner, W., Lemoine, G., Borgeaud M. and Rott, H.: A study of vegetation cover effects on ERS scatterometer data, IEEE Trans. Geosci. Remote Sens., 37(2), 938-948, 1999b.

Wagner, W., Noll G., Borgeaud M. and Rott, H.: Monitoring soil moisture over the Canadian prairies with ERS scatterometer, IEEE Trans. Geosci. Remote Sens., 37(1), 200-216, 1999a.

Wagner, W., Scipal, K., Pathe, C., Gerten, D., Lucht,W., and Rudolf, B.: Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data, J. Geophys. Res. Atmos., 108(D19), 4611, doi:10.1029/2003JD003663, 2003.

Walker, J. P., Willgoose, G. R., and Kalma, J. D.: One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms, Adv. Water Resour., 24, 631-650, 2001.

Wechsler S. P.: Uncertainties associated with digital elevation models for hydrologic applications: a review, Hydrol. Earth Syst. Sci., 11. 1481-1500, 2007.

Wesson SM and Pegram GGS (2006) Improved radar rainfall estimation at ground level, *Nat. Hazards Earth Syst. Sci.*, 6(3), 323-342.

Western A. W. and Blöschl G.: On the spatial scaling of soil moisture, J. Hydrol., 217, 203-224, 1999.

Western, A. W., Grayson, R. B., and Blöschl, G.: Scaling of soil moisture: a hydrologic perspective, Ann. Rev. Earth Planet. Sci., 30, 149-180, 2002.

Appendix A: Towards a Real Catchment

Mass balance continuity and ODE solving methods

Evaluation of the mass conservation

The mass conservation in the TOPKAPI model is evaluated by computing the mass balance for the total simulation time.

An absolute error is calculated by comparing the mass balance of the input with the output and storage, as follows:

$$Err = \sum_{t=0}^{nb\_time\_step} Input_t - \left(\sum_{t=0}^{nb\_time\_step} Output_t + V(nb\_time\_step) - V(0) \right) \quad in\ m^3 \qquad \text{A.1}$$

A relative error is also computed, as follows:

$$\mathrm{Re}l\_Err = \frac{Err}{\sum_{t=0}^{nb\_time\_step} Input_t} \times 100 \quad in\ \% \qquad \text{A.2}$$

These two criteria can be computed for the different reservoirs (soil, overland and channel) at each cell, as well as for the total catchment.

The mass balance in the model formulation

The outflows from each reservoir are computed with an equation of the type:

$$Q_i^{out} = Q_i^{in} - \frac{V_i(t+\Delta t) - V_i(t)}{\Delta t} \qquad \text{A.3}$$

This equation is the original equation proposed by Liu and Todini (2002) to compute the reservoir outflows. However, Parak (2006) proposed to compute the outflow through the following equation:

$$Q_i^{out} = \frac{1}{2}\left(b_i V(t)^{\alpha} + b_i V(t+\Delta t)^{\alpha} \right) \qquad \text{A.4}$$

Both equations A.3 and A.4 are justifiable, but they can have a different impact on the model simulations and particularly on the mass balance conservation.

The main advantage of equation A.3 is that, the mass continuity is always preserved because equation A.3 is directly extracted from the initial continuity equation A.1. Equation A.4 however considers that the outflow between *t* and *Δt* can be considered as the average of the instantaneous outflows at *t* and *Δt*. Due to its independence to the continuity equation A.1, the mass continuity is not straightforwardly conserved.

Since it forces the continuity of the simulations, equation A.3 seems to be the most convenient. But one must be aware that it may hide some potential numerical imperfections in the model. For instance, an error in the solving of the ODE equations will not be visible by analyzing the continuity of the model. By contrast, if equation A.4 is used, an error in the ODE solution will be rapidly identified. As an illustration, a systematic underestimation error of 10% was intentionally imposed at the end of each interval on the soil reservoir volumes computed from the Runge-Kutta-Fehlberg method (RKF), which is an accurate method for solving the ODE. The continuity errors, associated with the simulations are reported in Table A.1. The simulations are also compared by plotting the channel discharge at the catchment outlet (Figure A.1).

As expected, when the continuity is forced by equation A.3, there is no continuity error, although Figure A.1 shows a strong difference between discharges simulated with and without error. In contrast, when equation A.4 is used to compute the outflows, the error in the ODE solving is reflected by a strong continuity error of more than 100%, and the associated discharge is markedly underestimated at the outlet.

It is worth noting that when the solution of the ODE is accurate (as provided by the RKF method used herein without artificial error), the discharges simulated by using equation A.4 are quite similar to those obtained by using equation A.3 (coefficient of determination of 0.992%). The continuity error associated with the use of equation A.4 is only 0.35% which is reasonable. This means that despite a small continuity error, the equation A.4 proposed by Parak (2006) can be used to compute outflows. The use of this equation will be especially useful to evaluate the potential numerical biases in the tested ODE solving method, since it makes the model sensitive to discontinuities in mass balance. The use of equation A.3 is however allowed only if an accurate method is used to solve the ODE.

Table A.1 Continuity error computed for simulation with and without error in the ODE solution.

		Continuity error at catchment scale Rel_err (%)
Outflows computed From equation 2.3	Without error	0.
	With a 10% error on computed soil storage	0.
Outflows computed From equation 2.4	Without error	0.35
	With a 10% error on computed soil storage	101.3

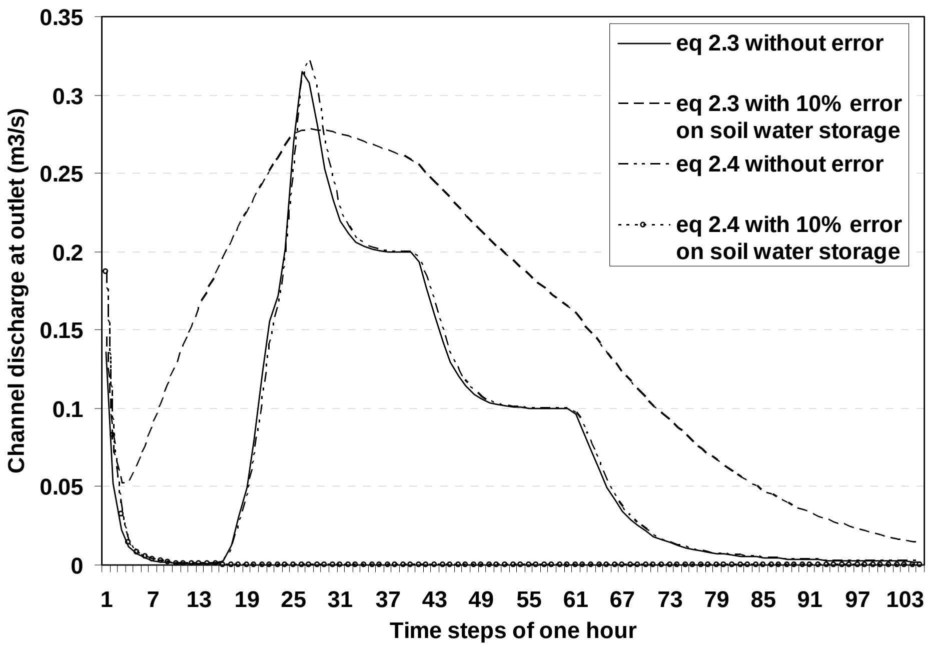


Figure A.1 Channel discharge at outlet of the 4 cells generic catchment. Comparison between the discharge without error in the ODE solving and discharge with 10% of error in the computed volume of water in soil store.

Test of different ODE solving methods

The ODE solution is a key operation in the TOPKAPI model. The method must be chosen to be (i) accurate – in order conserve the mass balance in the model- and (ii) fast – in order to reduce the computing time to a minimum.

Three methods are tested and compared to solve the ODE in the TOPKAPI model (with input term Q_i^{in} different from 0, since an analytical solution is known in this particular case):

1. the Runge-Kutta methods (see e.g. Gerald and Wheatley, 1992)
2. the quasi-analytical solution (Liu and Todini, 2002)
3. the analytical solution

In order to evaluate the accuracy of the tested solving methods, the criteria of continuity are computed locally at the cell scale for each reservoir (for each reservoir the maximum values among all the cells is kept) and globally at the catchment scale. The channel discharges at the catchment outlet are also compared.

In order to evaluate the computation speed, the computation time of 50 runs (loop of 50 similar simulations) is evaluated on the 4-cell model.

Solving of ODE with Runge-Kutta

Principle

The Runge-Kutta (RK) methods are very popular ways of finding the numerical solution of the ODE because of their good accuracy (Gerald and Wheatley, 1992). The most used formulations are the fourth- or fifth-order method (RK4 and RK5), or the Runge-Kutta-Fehlberg that combines RK4 and RK5 to evaluate the error committed by using the RK method and allows one to adapt the simulation time step during computation.

After an unsuccessful application of the RK4 method that led to unexpected divergence in the simulated volumes, it was decided to directly apply the adaptive time-step Runge-Kutta-Felhberg method. This method computes the estimation error by comparing the solution obtained from the fourth- and fifth-order Runge-Kutta formulas. One can thus increase or decrease the simulation time-step depending on this error: (i) while the error is greater than a fixed maximum error value, the time step is decreased by a factor two; (ii) while the error is smaller than a fixed minimum error value, the time step is increased by a factor two. The evaluation of the error means that the accuracy of the RKF method is always ensured and that the RKF method can be reliably used as a reference method in terms of accuracy.

For our application case, good results were obtained by fixing the error limits as a percentage of the initial value (meaning the volume at time step t while one seeks to determine the volume at time $t+\Delta t$) by setting: max_error = 10e-3% and min_error=10e-7%.

Results

Simulated discharge

As an illustration of the simulations, Figure A.2 shows the rainfall-runoff relationship at the channel outlet of the 4-cell generic catchment with the calculation performed by RKF. The comparison between the input of the model and the simulated outputs for each reservoir at each cell can be seen in Appendix B.

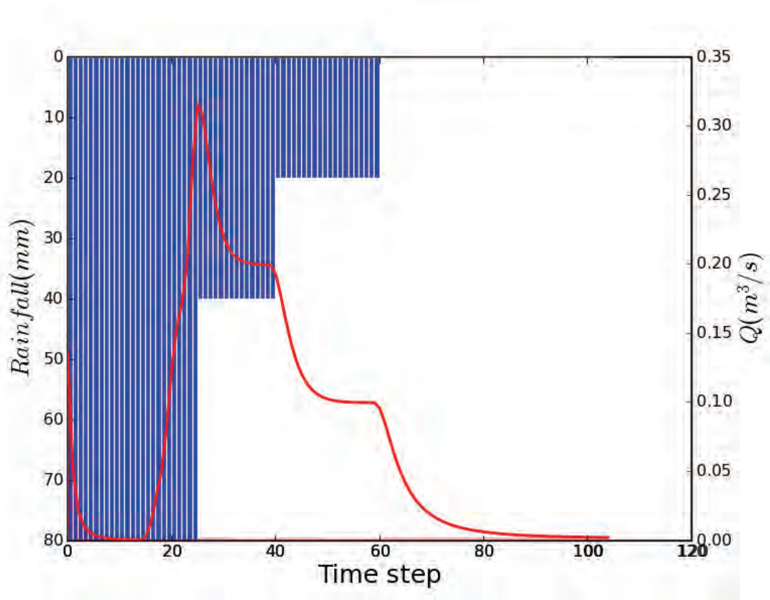


Figure A.2 Rainfall over the entire catchment and the simulated discharge at the channel outlet as a function of time.

It can be seen in Figure A.3, that the time-step division required by the RKF method depends on the character of the input values within the ODE equations. Particularly, the sudden input gradient from a time step to another require finer resolution time steps, especially for the Manning's equation (overland and channel). Note that when the input into a cell is zero during an interval (this was only applied to the uppermost cell within a cascade) the solution to the ODE has a known analytical form as it is in pure drainage mode and is directly integrable. This default is used in the RKF model and explains why the time-step goes to zero as shown in the bottom left-hand graph in Figure A.3.

It will also bee seen in Figure A.3 that the number of calculations per major time step varies considerably and increases (or reduces) dramatically when there are sudden changes in the input to any one of the stores. This adaptability optimises the computational load of the integration routines while minimising the errors in calculation, as is indicated by the closure of the continuity equation for each time step and globally.

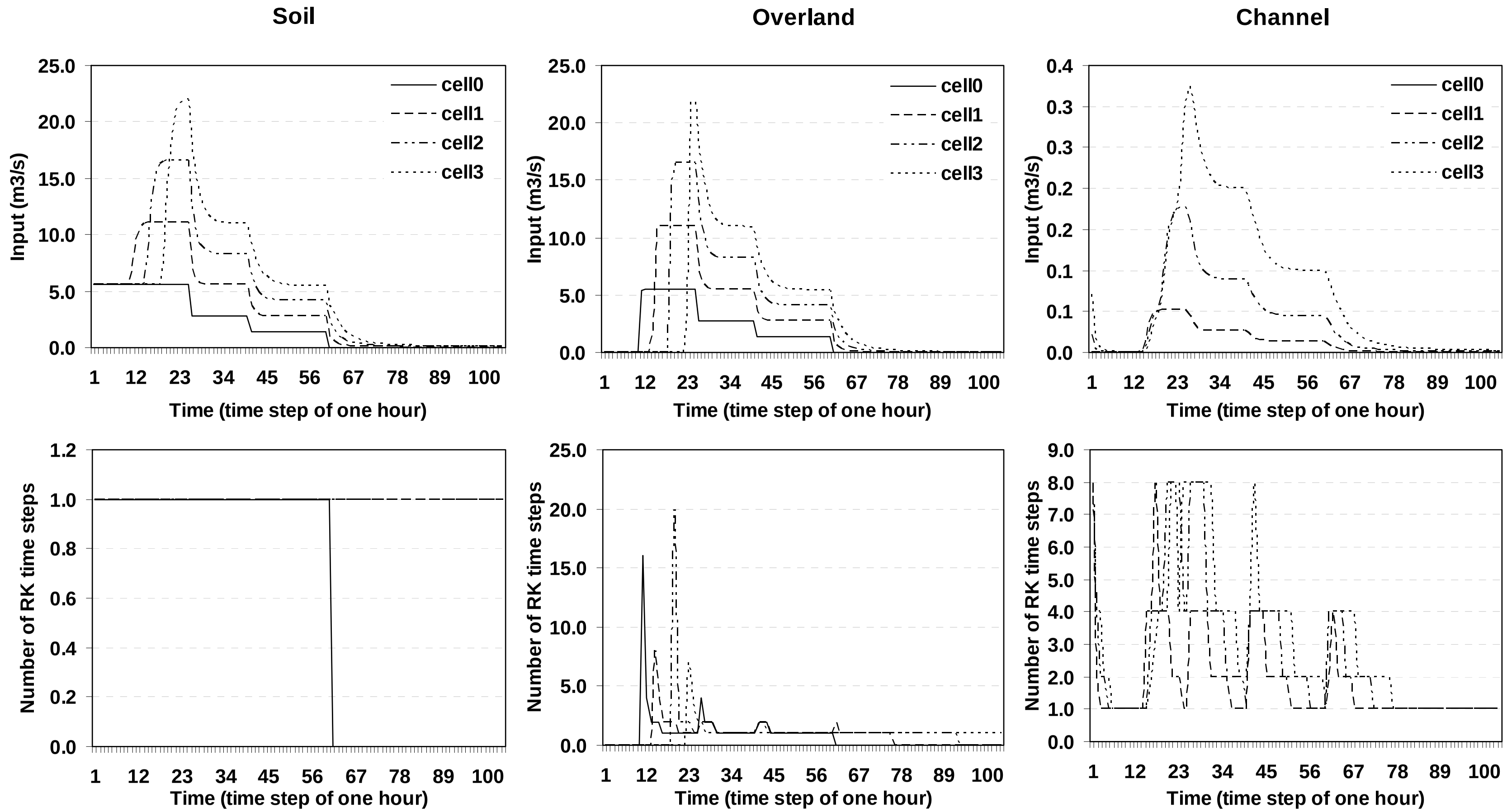


Figure A.3 The upper row of graphs show the input (of the ODE) into each sub-store. The lower row of graphs shows the number of sub-time steps used (during each of the time-steps of the corresponding upper graphs) by the RKF method to estimate the ODE solution

Continuity

The continuity of the reservoirs and the whole catchment is evaluated from the values calculated and reported in Table A.2. The relative error never exceeds more than 0.35% at the catchment scale, which is relatively satisfactory.

Table A.2 Reservoir continuity maximum errors among the 4 cells and continuity at catchment scale, over the whole simulation for the pure RKF method.

	Maximum Relative Error (%)
Soil	5.2e-6
Overland	2.9e-3
Channel	1.2e-2
Catchment	3.5e-1

Computing time

The time required to simulated 50 runs was **22.5 s**

Quasi-analytical solution

Principle

The quasi-analytical solution (QAS) is due to Liu and Todini (2002). For an ODE of the form $\frac{dy}{dt} = a - by^c$ with $1 \le c \le 2$, they proposed to approximate the function, y^c, by a piecewise second order polynomial, i.e. $y^c = \beta y^2 + \alpha y = y(\alpha + \beta y)$, where the parameters α and β are fitted locally by a least squares method. The integration of this second order polynomial can be then solved analytically.

The QAS can be applied directly for the Manning's equation that governs the overland and channel reservoir, since the c exponent is equal to 5/3. In the case of the soil reservoir for which c lies between 2 and 4, a simple variable substitution $u = y^{-(c-1)}$ allows one to retrieve an equation with an exponent between 1 and 2 and thus use the QAS to solve the ODE.

Liu and Todini (2002) suggest that the QAS can reduce substantially the computing time (by an order of magnitude), while improving the integration accuracy (Liu and Todini, 2005). The critical point however is to correctly adjust the polynomial to the power function y^c. But, except for the vague information of a 'least squares' adjustment, no supplementary prescriptions are available in the whole TOPKAPI literature that could help one to understand how this adjustment is practically done.

Two approaches are thus tested here, starting from the point that adjusting $y(\alpha + \beta y)$ to y^c is equivalent as adjusting $(\alpha + \beta y)$ to the function $f(y) = y^{c-1}$.

For the first approach, the assumption is made that a good line adjustment of a curve at the neighbourhood of the given value *y(t)* could be the tangent of $f(y) = y^{c-1}$ at *y(t)*. The slope β of the tangent being the value of the derivative function f' of the function f at the initial value *y(t)*.

For the second proposed approach, the suggestion is to find a rough estimation of the ODE solution $y^*(t+\Delta t)$ and to adjust a line passing through the pair of points $[y(t), f(y(t))]$ and $[y^*(t+\Delta t), f(y^*(t+\Delta t))]$. Note that this approach is certainly closed to what might be called a 'least squares' adjustment…

Results

Simulated discharge

In Figure A.4 the channel discharges at the outlet of the catchment, computed from the two proposed QAS approaches, are compared to the discharge computed from the RKF methods.

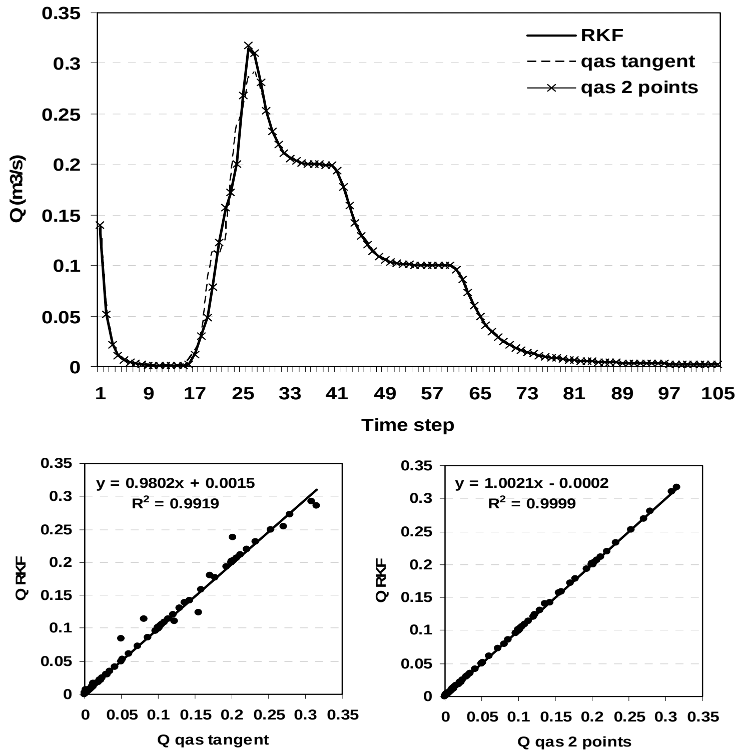


Figure A.4 Comparison of the discharge simulated at the channel outlet of the 4 cells generic catchment from the two approaches of quasi analytical solution (qas) and the Runge-Kutta-Fehlberg (RKF) method.

Since the RKF method allows one to explicitly evaluate the error committed by the numerical approximations, it can be taken reliably as a reference. Figure A.4 thus shows that the QAS can give similar results to the RKF method when the line is adjusted with two points. However the tangent line approach is less appropriate according to the differences of simulated discharges compared to the RKF method;

the error is particularly noticeable due to the large jump in the input sequence – the solution has difficulty anticipating the next (future) point without iteration.

Continuity

A good result for continuity is obtained for the 2 point adjustment approach (0.49% at catchment scale) while the tangent adjustment approach shows a too large a continuity error (higher than 1%) to be accepted as an accurate approach (see Table A.3).

The QAS with the two point adjustment approach can thus be retained as a fair numerical solution. But it is important to note that some unexpected numerical divergences can occur. Such divergences were observed for the soil equation only, for which sometimes the QAS overestimates the output which can become higher than the input. A by-pass was created in the program to switch into the RKF solution when this error occurs. As can be seen in Figure A.57 when the input gradient is high, due to the divergence of the QAS, the program switches into the RKF solution as designed.

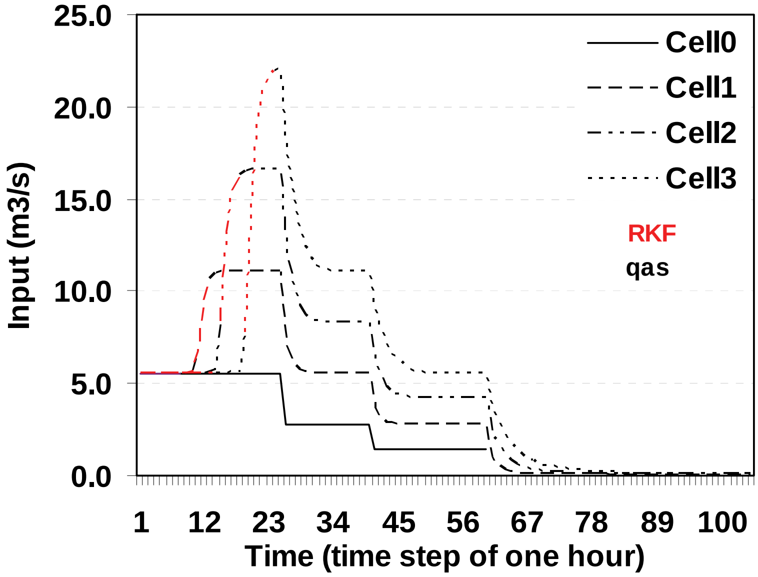


Figure A.5 Link between the input gradient and the ODE solving method for the soil stores.

Computing time

By taking into account the possibility of the program to switch between RKF and QAS, the computing time of 50 runs was evaluated at **11.9s**, meaning a reduction of almost 50% of the computing time required by the pure RKF method.

Table A.3 Reservoir maximum continuity errors among the 4 cells and continuity at catchment scale, over the whole simulation for the QAS method (tangent adjustment and 2 points adjustment approaches).

		Maximum Relative Error (%)
QAS with tangent adjustment	**Soil**	2.0e-4
	Overland	5.8e-3
	Channel	9.6e-3
	Catchment	-5.97
QAS with 2 point adjustment	**Soil**	1.9e-4
	Overland	4.4e-3
	Channel	1.2e-2
	Catchment	4.9e-1

Analytical Solution

Principle

The possibility of finding an analytical solution for the integration of the function of the type $f(y) = \dfrac{1}{a - by^c}$ is often rejected although some formulations in the primitive of this function can be found. Primitive formulations can be computed from tables in books (e.g. Gradshteyn and Ryzhik, 1982) or from mathematical software (e.g. *Mathematica*).

If a primitive function F of the function f exists, the resolution of the ODE in equation A.1 becomes:

$$F\left(y(t + \Delta t)\right) = F\left(y(t)\right) + \Delta t \qquad \text{A.5}$$

where the right hand side term $F\left(y(t)\right) + \Delta t$ is known and constant. The problem is reduced to the finding of the root $y(t + \Delta t)$ that satisfies equation A.5.

An efficient method to numerically find roots is the bisection method, with the condition that two initial values bracketing the root value can be found.

Mathematica gives a primitive for the function in the form of Manning's equation:

$$\int \frac{1}{a - b\,x^{5/3}}\,dx =$$

$$\frac{1}{20\,a^{2/5}\,b^{3/5}}\left(3\left(-2\sqrt{10-2\sqrt{5}}\,\tan^{-1}\left(\frac{4\sqrt[5]{b}\,\sqrt[3]{x}-\left(-1+\sqrt{5}\right)\sqrt[5]{a}}{\sqrt{2\left(5+\sqrt{5}\right)}\,\sqrt[5]{a}}\right)+2\sqrt{2\left(5+\sqrt{5}\right)}\,\tan^{-1}\left(\frac{4\sqrt[5]{b}\,\sqrt[3]{x}+\left(1+\sqrt{5}\right)\sqrt[5]{a}}{\sqrt{10-2\sqrt{5}}\,\sqrt[5]{a}}\right)-4\log\left(\sqrt[5]{a}-\sqrt[5]{b}\,\sqrt[3]{x}\right)+\left(1+\sqrt{5}\right)\log\left(b^{2/5}\,x^{2/3}-\frac{1}{2}\left(-1+\sqrt{5}\right)\sqrt[5]{a}\,\sqrt[5]{b}\,\sqrt[3]{x}+a^{2/5}\right)-\left(-1+\sqrt{5}\right)\log\left(b^{2/5}\,x^{2/3}+\frac{1}{2}\left(1+\sqrt{5}\right)\sqrt[5]{a}\,\sqrt[5]{b}\,\sqrt[3]{x}+a^{2/5}\right)\right)\right) \qquad \text{A.6}$$

And for some values of *c* between 2 and 4, a primitive also exists e.g. for *c*=2.5:

$$\int \frac{1}{a - b\,x^{5/2}}\,dx =$$

$$\frac{1}{10\,a^{3/5}\,b^{2/5}}\left(2\sqrt{10-2\sqrt{5}}\,\tan^{-1}\left(\frac{4\sqrt[5]{b}\,\sqrt{x}-\left(-1+\sqrt{5}\right)\sqrt[5]{a}}{\sqrt{2\left(5+\sqrt{5}\right)}\,\sqrt[5]{a}}\right)-2\sqrt{2\left(5+\sqrt{5}\right)}\,\tan^{-1}\left(\frac{4\sqrt[5]{b}\,\sqrt{x}+\left(1+\sqrt{5}\right)\sqrt[5]{a}}{\sqrt{10-2\sqrt{5}}\,\sqrt[5]{a}}\right)-4\log\left(\sqrt[5]{a}-\sqrt[5]{b}\,\sqrt{x}\right)+\left(1+\sqrt{5}\right)\log\left(b^{2/5}\,x-\frac{1}{2}\left(-1+\sqrt{5}\right)\sqrt[5]{a}\,\sqrt[5]{b}\,\sqrt{x}+a^{2/5}\right)-\left(-1+\sqrt{5}\right)\log\left(b^{2/5}\,x+\frac{1}{2}\left(1+\sqrt{5}\right)\sqrt[5]{a}\,\sqrt[5]{b}\,\sqrt{x}+a^{2/5}\right)\right) \qquad \text{A.7}$$

There is thus potentially an analytical solution for these two cases that are tested here.

Results for analytical solution

Overland and Channel ODE (Manning's type equation)

Although it is not mentioned in the *Mathematica* outputs, the primitive function F defined by equation A.6 is only defined in the interval $\left[0,\left(\frac{a}{b}\right)^{\frac{3}{5}}\right]$ where a and b are coefficients in equation $f(y)=\frac{1}{a-by^{c}}$.

For example at a given step of iteration t_i of overland of the last cell, $a \approx$ 8.38e-8 and b ≈ 2.51e-5, the function curve is shown in Figure A.6.

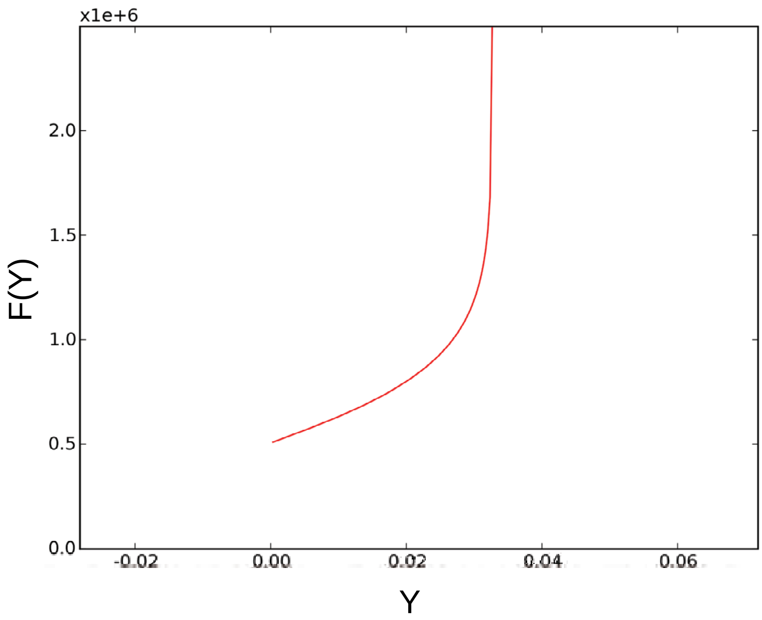


Figure A.6 Curve of the primitive function of equation 1.3 for a ≈ 8.38e-8 and b≈ 2.51e-5

This means that a value of the surface store V(t) due to 1 mm of excess rainfall is 1000 m^3, and the typical value of a and b yield $\left(\frac{a}{b}\right)^{\frac{3}{5}} \approx 0.03267$, thus the primitive is not defined for V(t). There is thus no possibility of using an analytical method to solve the ODE of the overland and channel reservoirs.

Soil ODE: example of equation 2.7for c=2.5, applicable to a soil store

There is no obvious problem of continuity intervals for equation A.7 which can be thus used to solve the equation. The problem of the method is more to find a direct means to adjust a short interval comprising the root $y(t+\Delta t)$ of equation A.5.

A lower value of the root *low_y* is obtained by solving the ODE equation without input, an upper value *up_y* is obtained by adding the input to the initial volume at t. The interval is then taken as:

$$[low\_y, y(t)] \quad \text{if} \quad F(y(t)) \times F(low\_y) < 0$$

or

$$[y(t), up\_y] \quad \text{if} \quad F(y(t)) \times F(up\_y) < 0$$

The tolerance of convergence was chosen to obtain a relative error of continuity at catchment scale lower than 1%. To obtain such a precision, the computing time is 3 times longer than the time required to compute the RKF method, although a solution is feasible.

Despite its good accuracy, the analytical method is thus not as fast as the RKF or QAS method in this limited case. Moreover it is only applicable for an exponent coefficient equal to 2.5, which is not very convenient since the exponent coefficient α_s will certainly be calibrated (Benning, 1995) in practical applications of TOPKAPI.

Towards a real catchment: a 32 cell generic catchment

The program created with the Python language has been designed in terms of maximum modularity and adaptability. In order to test the adaptability of the model to a more complex case that introduces a 2-D cell connectivity, a 32 cell generic catchment was created and used as a new case study.

Catchment characteristics

The catchment geometry and cell types are shown in Figure A.7. The catchment is composed of 32 cells of 1 km² with 8 cells having a channel. The width of the channel network was adjusted using the formula proposed by Liu and Todini (2002):

The physiographic and physical characteristics of the generic catchment are plotted in Figure A.8. The cell slopes were fixed to realistic values inspired by the Liebenbergsvlei catchment. The soil depth was fixed arbitrarily with a steep increase towards the outlet. The soil saturated hydraulic conductivity, the overland manning roughness and the channel roughness were fixed at constant values (respectively: Ks=1e-3, n_overland=0.1 and n_channel=0.25).

$$W_i = W_{max} + \left[\frac{W_{max} - W_{min}}{\sqrt{A_{total}} - \sqrt{A_{threshold}}}\right]\left(\sqrt{A_{drained_i}} - \sqrt{A_{total}}\right) \ [\mathrm{m}]$$ with W_{max}=10 and W_{min}=1.

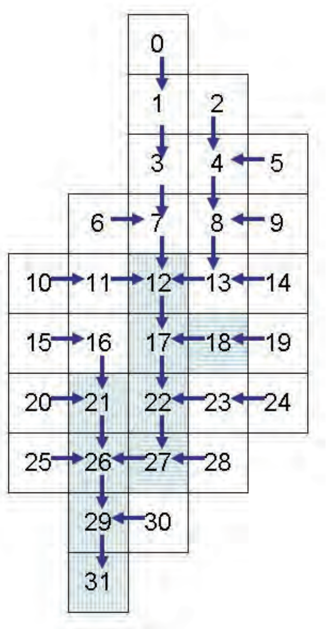


Figure A.7 32 cell generic catchment. White cells are hillslope cell (without channel drainage), blue cells are channel type cells. The arrows show the drainage connectivity between cells.

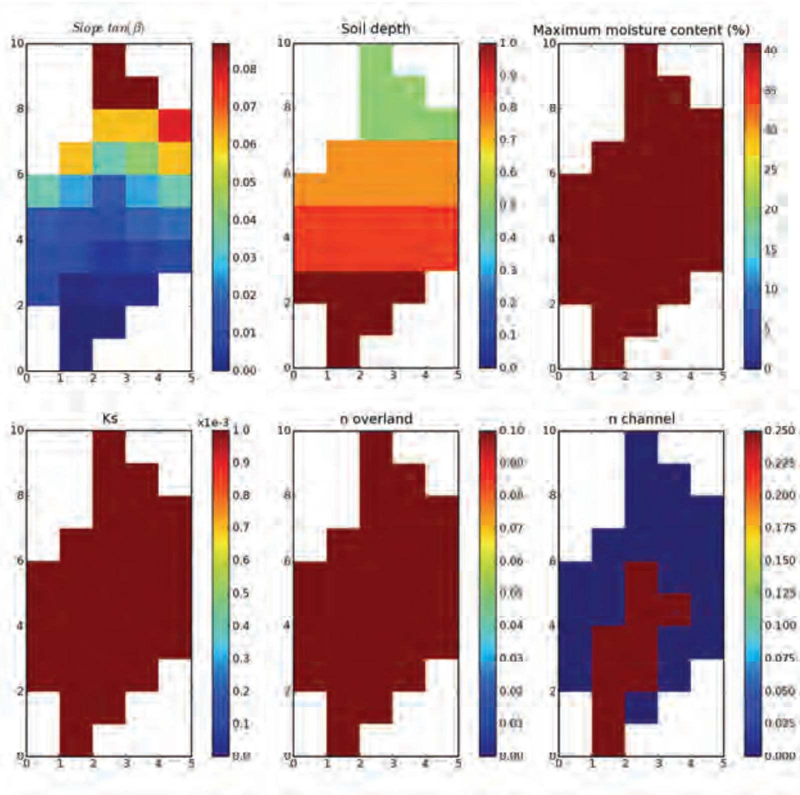


Figure A.8 Catchment characteristics.

Forcing rainfall

A basic hyetograph applied on each cell is used to force the hydrological model. In order to see the influence of initialisation of the model, the program was designed to allow one to concatenate the hyetograph into a longer hyetograph.

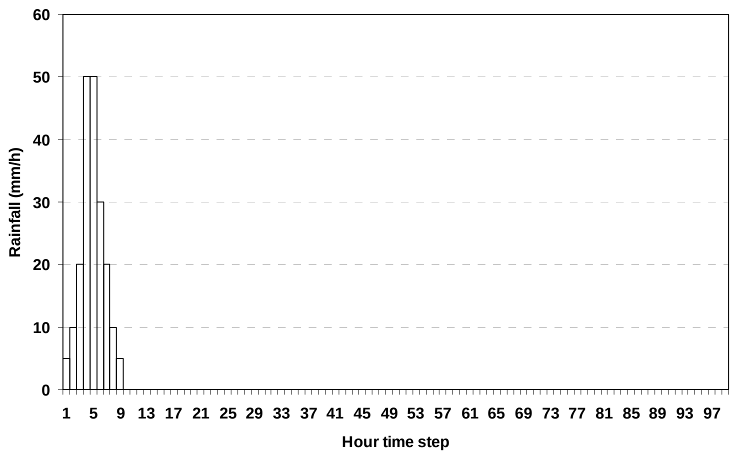


Figure A.9 Basic hyetograph applied on each cell.

Analysis of the simulations

Numerical aspects

In order to be sure that the results obtained in the preceding section with the simple 4-cell case study are valid for a more complex case study, the continuity and computing-time are again evaluated.

The hyetograph is concatenated 15 times in order to have a reasonably robust evaluation of the computing time. Four cases are studied: simulations using RKF and QAS ODE solving method and for each the outflows of the reservoirs are computed either by equation A.3 or by equation A.4.

Table A.6 shows the values obtained in term of computation time and relative error in the continuity at the catchment scale.

Table A.6 Computing time and continuity of the simulations on 32 cells.

	Outflow equation	Computing time	Relative Error of continuity (%)
RKF	2.3 *(Liu and Todini, 2002)*	46.8	0.
	2.4 *(Parak, 2006)*	47.1	3.1
QAS	2.3	36.2	0.
	2.4	37.9	4.97

It was found that the QAS method reduces the computing time by about ¾ while conserving a good agreement with the RKF method (see Figure A.10).

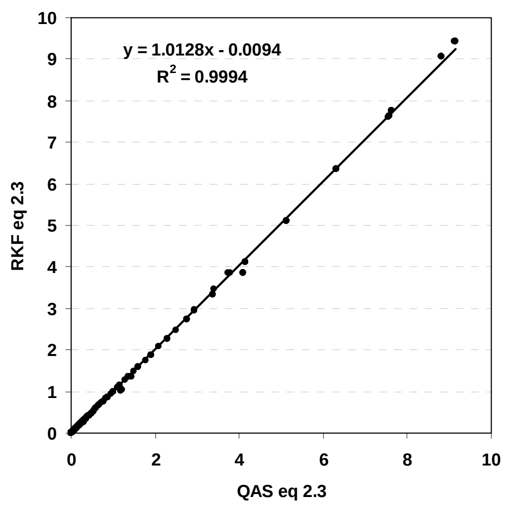


Figure A.10 Correlation between discharges at the end of each interval at the outlet of the catchment obtained from RKF and from QAS methods.

It is worth noting however that the use of equation A.4 (Parak, 2006) produces a continuity error at the catchment scale of 3.1% and almost 5% when combined with the RKF or QAS methods respectively, which is thus higher than a reasonable limit value of 1%. By contrast, Equation A.3 yields negligible error in both cases. As can be seen in Figure A.11, for 3 of the 15 simulated peak discharges, this continuity error seems to mainly affect the peaks of discharge, for which, although the magnitudes are similar, the timing is different.

According to these results, it is certainly preferable to compute the outflows by using equation A.3 proposed by Liu and Todini (2006) in combination with the hybrid RKF/QAS method.

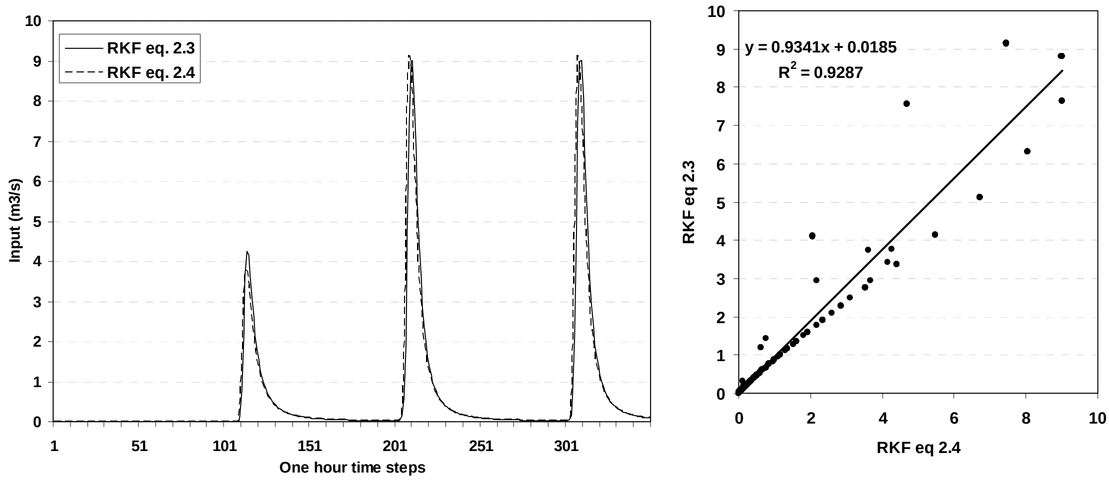


Figure A.11 Comparison using the RKF method between the discharges obtained by using the equation A.3 or A.4 to compute the outflows.

Contribution of the subsurface flows to discharge at the outlet

In order to evaluate the role of the subsurface storage in the contribution of the discharges, it is interesting to evaluate the amount of flows in the channel coming respectively from subsurface flows and overland flows.

A means to evaluate the relative contribution of overland and subsurface flow is to assume that at each time step the simulated channel outflows contain the same proportion of overland and subsurface water that feed the channel input (see a simple example in Figure A.12)

.

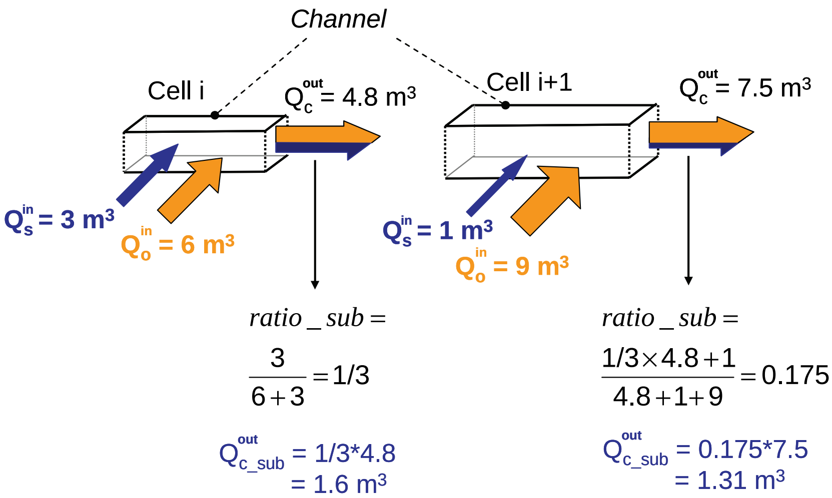


Figure A.12 Simple example of the estimation of the subsurface contribution to channel outflows

In a generalised case the rate of subsurface flow in a channel cell outflow can be expressed as:

$$ratio\_sub_i = \frac{\frac{W_i}{X} Q_{s_i}^{out} + \sum_{k \in \{cell\_up\}} ratio\_sub_k \times Q_{c_k}^{up}}{Q_{c_i}^{in}} \qquad \text{A.8}$$

Applied to the present 32-cell case study, it is possible to assess the contribution of subsurface flows to channel flows at the outlet of the generic catchment. Figure A.13 illustrates the results obtained for the simulations. It can be seen that for the initial soil conductivity Ks value (1e-3), the subsurface contribution is quite negligible compared to the overland contribution. However if the Ks is globally multiplied by a factor 100 then 1000, the contribution of subsurface to channel outflow is logically higher.

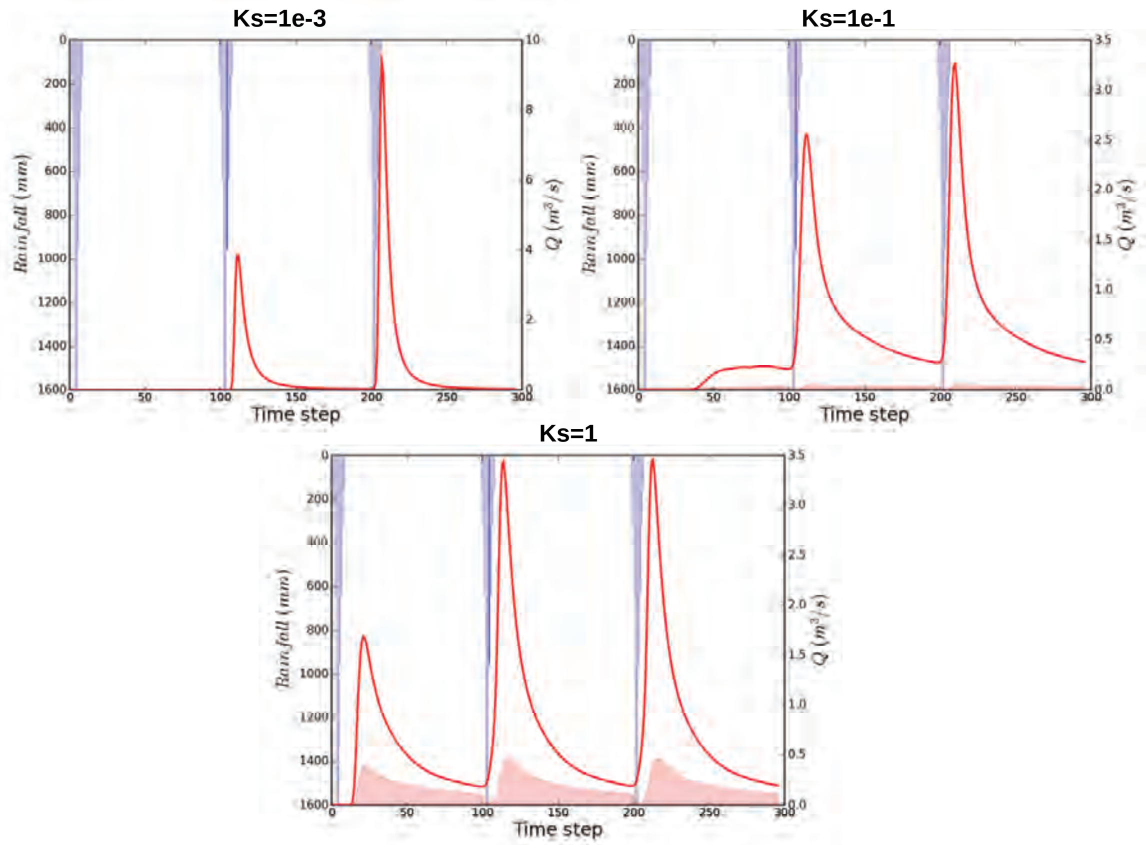


Figure A.13 Channel discharges at the outlet of the generic catchment for three values of Ks. The pink filled curve is the assessed contribution of subsurface flows to discharge.

It is worth noting however that the contribution of subsurface flow plotted in Figure A.13, behaves unrealistically during rainfall events. This must be due to the strong assumption considering that the proportion of subsurface input contributes to the outflows at the same time step. Actually because of the delay due to the kinematic wave equations, when a high input is applied at a given time step, the damping effect of storage is important and the outflow does not increase immediately. This means that the computed ratio is low and that the contribution of subsurface flows to channel flows is no more realistic.

Definition of a simple Soil Saturation Index SSI

The soil moisture as simulated by the model can be considered from two points of view:

i. the first one refers directly to the soil layer as represented in TOPKAPI and defines the soil moisture as the water volume contained in each soil cell of the catchment
ii. the second one refers to the soil moisture as it could be viewed only from the surface (picture of surface moisture) and defines the soil moisture as the saturated areas.

The first definition allows one to directly have an idea of the soil moisture patterns by plotting the water content of cells (see an example in Figure A.14a). It also provides a convenient framework to define a Soil Saturation Index (SSI) as a relative saturation of the soil layer (Figure A.14b). The second definition of surface wetness gives only binary information since surface soil moisture is viewed as saturated/non saturated (Figure A.14c), but can be extended to a more gradual view of moisture as a function of the overland runoff depth (see Figure A.14d).

The soil moisture as it is measured by the satellites is certainly an intermediate case between the two definitions given above, since the satellites see the surface and the first centimetres of the soil layer. The ideas must thus be pursued to find a suitable definition of soil moisture in TOPKAPI in order to be able to fairly compare modelling and remotely sensed soil moisture estimations.

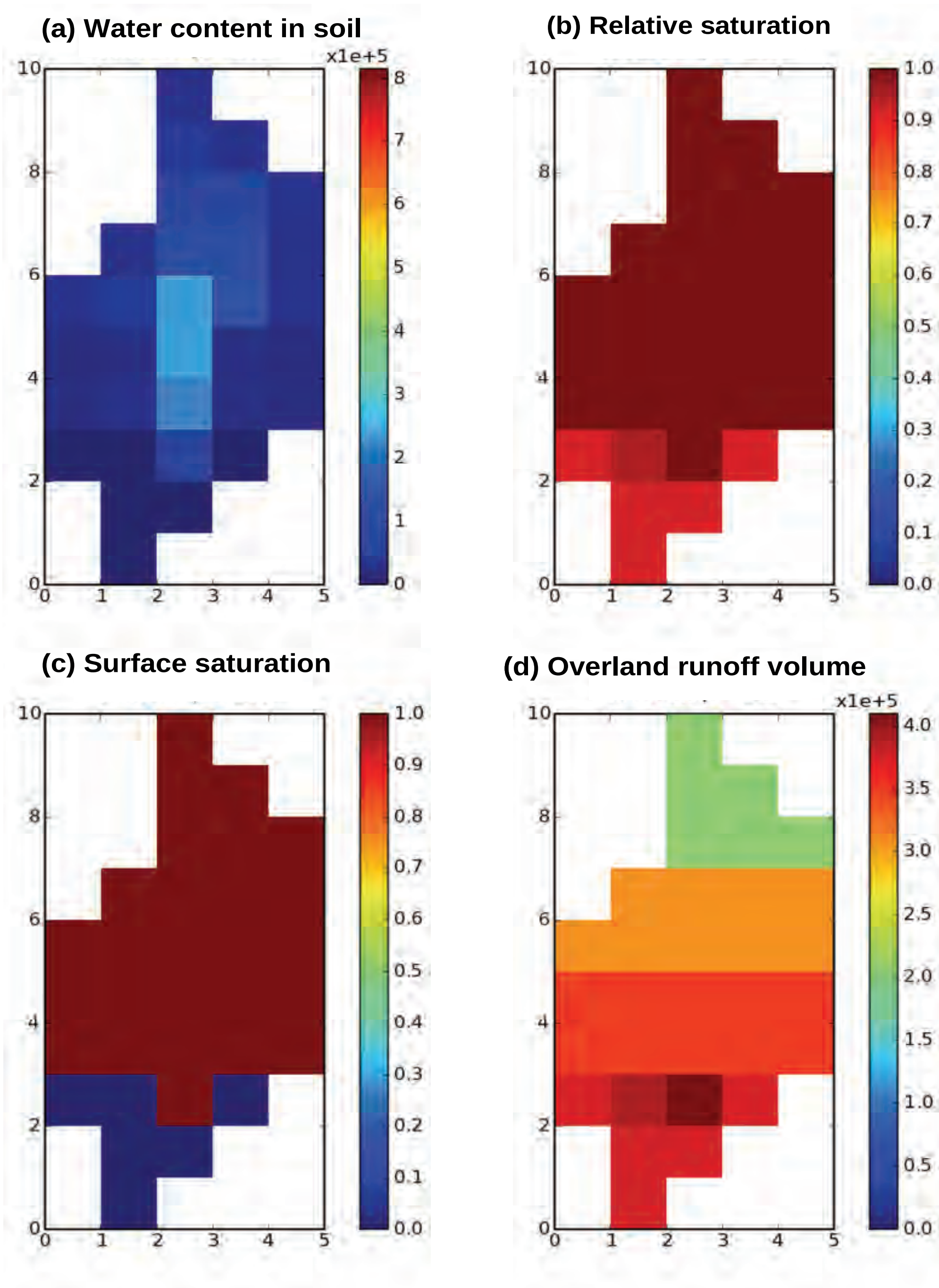


Figure A.14 Characterisation of the soil moisture in the TOPKAPI model. An example of fields at the time step t=105

Summary

The continuity problems of the TOPKAPI model such was implemented by Parak (2006) was solved, especially by finding a good method to solve the ODE equations with efficient algorithms. It was shown that the quasi-analytical solution proposed by Liu and Todini (2002) was faster than the robust method of Runge-Kutta-Fehlberg. However it less accurate and can sometimes show unexpected divergences in the ODE solution. For that reason the solution retained is to use a hybrid method comprising the fast QAS with the possibility of switching to the RKF method in case of divergence.

Since the 4-cell catchment case study presented here is simple and the simulation time was short, it is important to keep in mind the potential problem of instability of QAS for further more complex applications and thus to continue to check the continuity in the simulations.

This was successfully checked on a more realistic, complex idealised 32-cell catchment, thus validating the capacity of the program to easily adapt to complex spatial connections between cells.

The next objective was to apply the model on the Liebenbergsvlei catchment. Comparison between simulated and observed discharges, simulated and remote sensed soil moisture should help us to continue the ideas initiated about the definition of the soil moisture in TOPKAPI as a base to be compared to remotely sensed soil moisture estimation. This will also help us to see if further improvements are required in the modelling strategy, like adding soil water losses to a deep aquifer or adding an upper thin soil layer that could represent the infiltration processes in the first centimetres of the soil. We must be aware that such improvements have already been proposed by Liu and Todini (Liu et al., 2005) and can be used to update the model.

APPENDIX B: 4-CELL MODEL INFLOWS AND OUTFLOWS

Runge-Kutta-Fehlberg graphics for each reservoir in each cell in the 4-cell model.

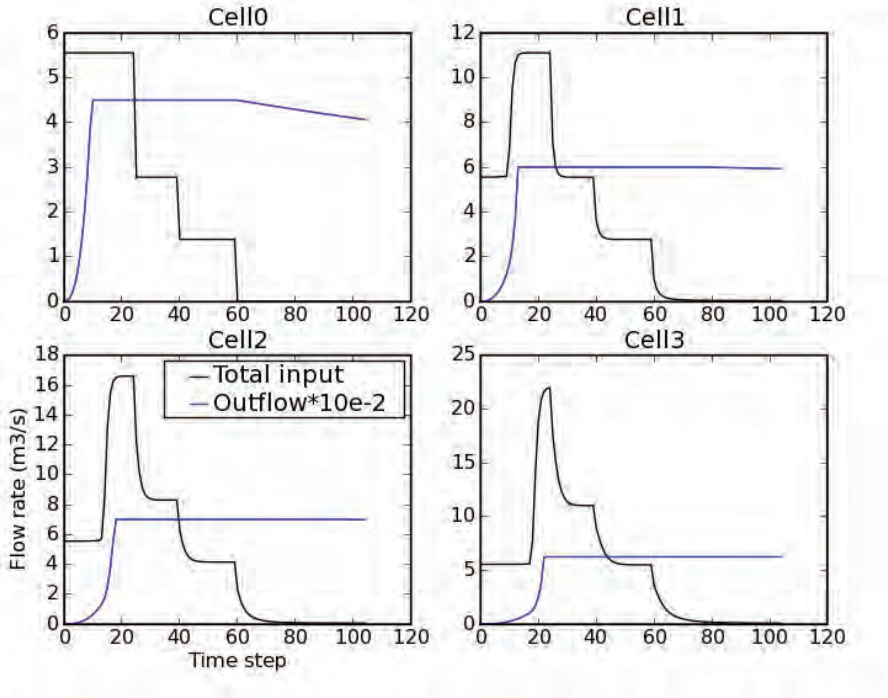


Figure B.1 Soil reservoir

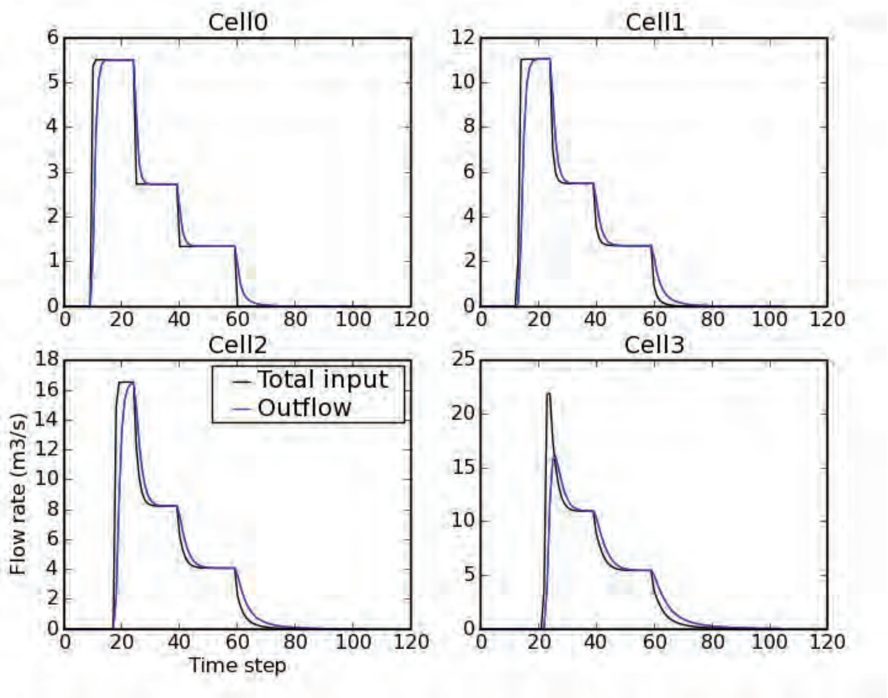


Figure B.2 Overland reservoir

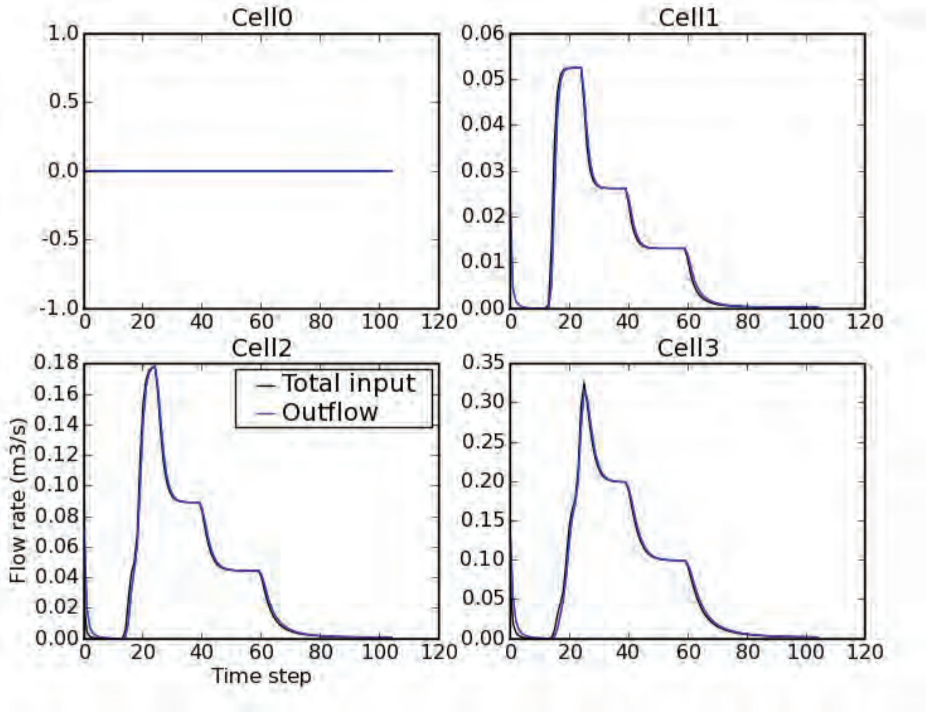


Figure B.3 Channel reservoir.