A Strategy for Managing Deficiencies of SWMM Modeling for Large Undeveloped Semi-Arid Watersheds

Hanlie Pretorius, William James and Julian Smit

For this study, the aim was twofold: to determine whether sufficient data is freely available on the internet for input to an open source hydrological model whose purpose is streamflow management; and, if so, to suggest a methodology whereby any deficiencies may be managed. Chosen for study was the C83A quaternary catchment (746 km²) in the Northern Free State, part of the Vaal primary catchment in South Africa. Because the catchment has a response time of ~13 h, sub-daily calculation steps are required in the hydrological model.

Data readily available on the internet were: catchment boundaries, river lines, a digital elevation model, soil texture, land use, evaporation rates, flow rates and rainfall estimates from the Tropical Rainfall Measurement Mission (TRMM) 3B42 data product (satellite rain). After determining the requirements for the model and evaluating a number of models, the USEPA Storm Water Management Model (SWMM) was chosen. Only 2 y observed rain and runoff were suitable for analysis. Coarseness of the available input data necessitated a coarser model than is the norm for SWMM. Nevertheless, the calibrated model performed well overall, having Nash-Sutcliffe (N-S) correlations of 0.80 and 0.73 for the calibration and evaluation periods respectively.

A systematic methodology is proposed for simultaneously increasing model complexity while reducing model error. Denoted UCTPCSW-MMCC83A.1, the derived model-and-data system performed best during dry
periods in the observed record, but inadequately during the rainy periods in the observed record (N-S relation of 0.31). Further investigation into the derived model-data system and the TRMM 3B42 rainfall estimates confirmed that the model-data system can compute total flow volumes to within 0.31% of observed values using physically realistic parameters.

It was concluded that poor model-data system performance during wet periods is most likely caused by the accuracy, and coarse time and spatial resolution, of the 3B42 rainfall estimates relative to the size of the catchment. Such smaller catchments do need to be studied, however, because it is the under-resourced authorities in charge of small catchments who could benefit from a streamflow management system that uses freely available data and software.

1.1 Introduction

Water resources of the Ash River, in the Free State in South Africa, are used for multiple purposes: water supply, irrigation, recreation, fishing, white water rafting and hydroelectricity. In addition to local runoff, trans-basin flows are imported from the Lesotho Highlands Water Project, and water is supplied to the town of Bethlehem from the Saulspoort Dam, and, via the Vaal Dam, downstream to the entire Witwatersrand region in Gauteng. (Further downstream the Vaal River flows rejoin the Orange River but are again partly diverted at Gariep.) Farmers also extract irrigation water from the river. In addition recreational use is made by cold water fly fishermen, and quite heavily by white water rafters. Several small scale hydroelectric facilities have been built and more are planned. Conflicts have arisen and the Save the Ash River advocacy group has launched a public petition opposing the hydroelectric developments.

Floods are also of concern. As the climate continues to change and rainstorms become more intense, costs to clean up and repair damaged infrastructure will increase. Floods can take several hours to a day or more to develop, depending on the size of the catchment and the extent of the rainfall, and so a flood warning system, that can use a model to compute in a few minutes what may happen in the next few hours, would be very useful to limit deaths and flood damage.

Clearly the desiderata for a hydrological-hydraulic water management model would include the computation of a wide range of flows and flow depths, from droughts to floods.

Many municipalities and river stakeholders in South Africa lack the human and financial resources to implement expensive software and data.
collection systems to manage their watersheds. Thus an aim of this study was to test whether data which is freely and readily available on the internet (at least for dynamic data, such as rainfall and flow), together with free and open source software, can be used to construct a hydrological model of a mid-sized watershed. Such a model would input real time rainfall data and compute the flow at the watershed outlet in a few minutes, giving approximately 3 h advance warning of impending flows. Satellite based rainfall estimates (SBRE) become available approximately 9 h (at best; the time varies) after the sensors on the satellites observe their parameters (Huffman et al., 2007). A streamflow management system may be said to be real time if the computation takes place while the rain and resulting flows are yet taking place.

1.2 Recent Studies

Many researchers have used SBRE and other internet data in hydrological models to compute stream flows, but the time steps are usually >1 d and the study areas are usually thousands of square kilometres in area. Using the Xinanjiang rainfall–runoff model, Li et al. (2008) modeled runoff in the Nzoia River basin (12 696 km²) in East Africa with Tropical Rainfall Measurement Mission (TRMM) SBRE as rainfall input. These modelers achieved Nash-Sutcliffe (N-S) correlations of 0.67 in the evaluation period and 0.84 in the calibration period. They used a model time step of 1 d and calibrated the model separately using ground based rain gauge data and SBRE before evaluating the model using satellite rainfall. They found that the model performed better when they calibrated it with the SBRE. For our purposes a time step of 1 d would be too coarse, even though one would expect a finer time resolution to incur higher error.

Lacroix et al. (2000) used the Semi-Distributed Land Use Runoff Process Rainfall–Runoff Model in their model of the Küçük Menderes Basin (3 617 km²) in Turkey. They used only data available from the internet and a model time step of 1 d. To evaluate their model, they compared the computed monthly flows to observed monthly flows for 4 y of data. The first three years gave a mean flow error that was 73% of the observed mean. Adding the last year, the error increased to 121% of the mean. They attribute the poor results in the last year to changes in weather patterns or to that fact that the precipitation stations no longer adequately represented the basin precipitation. Again, for our purposes, monthly totals would be too coarse, even though one might expect higher error for finer time scales.

To develop a flood monitoring system for the Limpopo basin (400 000 km²), Asante et al. (2007) used the Geospatial Streamflow Model
(GeoSFM), an extension of ArcView, and TRMM rainfall data. They used the SBRE as input to the rainfall–runoff model to establish baseline climatology for identifying extreme events (events in which the flow was 1.5 standard deviations above the short term mean annual flow). They found that the climatology stabilised after 5 y of data. Since they did not verify the model results against stream flow measurements, they used the stream flow forecasts primarily as a trigger to verify a forecasted extreme event from field sources.

Hong et al. (2007) used the Natural Resources Conservation Service Curve Number method to estimate runoff from rainfall for the whole world between 50°N and 50°S. They used only data available on the internet: TRMM rainfall, Food and Agricultural Organization (FAO) soil data, NASA’s moderate resolution imaging spectroradiometer (MODIS) land classification, the University of New Hampshire (UNH)/Global Runoff Data Centre (GRDC) composite runoff data, and data from HYDRO1k for routing information. (HYDRO1k is a geographic database that provides global coverage of topographically derived datasets, including streams, drainage basins and ancillary layers derived from the United States Geological Survey’s 30 arc-second DEM.) They calculated the runoff/day and achieved a 0.75 correlation between the calculated runoff and GCRD values. Artan et al. (2007) also assessed satellite rainfall to compute runoff in large basins: the Nile river basin in Africa (6 000 km²) and the Mekong river basin in Asia (22 000 km²). They use the GeoSFM model and achieved a best N–S correlation of 0.81. One day is again too coarse for our purposes.

No studies were found that used a combination of the stormwater management model SWMM and SBRE, and in this sense this study is original. Of course, various researchers have applied SWMM successfully with gauged rainfall and achieved correlation between observed and calculated hydrographs above 0.8 (Jun et al., 2010; Chen and Adams, 2007; Tsihrintzis and Sidan, 2008) or to within 15% of a observed value (Xie et al., 1997). Study areas in these cases ranged from <100 km² to thousands of square kilometres. Time steps used in these studies were typically relevant to our analyses here.

Many studies have evaluated the accuracy of SBRE. To do so, researchers use ground based rainfall data (gauges or radar) and convert point data to areal values with interpolation methods before comparing the two values. In general, researchers find that the correlation between the ground based data and SBRE increases as the number of gauges available for comparison increases (Sorooshian et al., 2000) and that SBRE is more sensitive to rainfall >2 mm/h (Hossain and Anagnostou, 2004; Villarini and Krajewski, 2007). Also, validation indices improve as the time intervals over which they are aggregated increase (Li et al., 2008; Artan et al., 2007). Reported correlation
coefficients range from 0.30 (Huffman et al., 2007) to as high as 0.83 (Hughes, 2006). No hydrological studies available for review have used SBRE data finer than 1 d. In the present study no rain gauges were located within the watershed, and so good correlations are contra-indicated.

1.3 Theory

1.3.1 Hydrological Model

The model engine used in this study, SWMM5, is a mainly deterministic physics based model that divides the study area into a collection of smaller, homogeneous subcatchments, each with its own properties (James et al., 2008:2). Readers familiar with SWMM may skip the following paragraph.

Each subcatchment is assigned a hypothetical rain gauge that supplies rainfall (although more than one subcatchment can use the same rain gauge), and each catchment drains to an outlet node. Conduits (channels) convey water from the subcatchments to the hydraulic transport system (conduits and nodes). Nodes mark, for example, the start and end of channels to match drainage connectivity, such as confluences of rivers, to discretize channels for channel parameter averaging, and for significant hydraulic structures, such as diversions, dams, bridges and the catchment outlet.

SWMM treats each subcatchment as a nonlinear reservoir with inflows from precipitation (or designated upstream catchments). Outflows include infiltration, evaporation and surface runoff. Surface runoff occurs only when the depth of water on the subcatchment surface exceeds the maximum depression storage. SWMM establishes the nonlinear reservoir by coupling the continuity equation with Manning’s equation. For channel routing, SWMM optionally solves the Saint-Venant dynamic wave equations, a combination of continuity and momentum equations, for the flow in each conduit and the water surface elevation at each node, at each time step.

To evaluate model performance, one needs an objective function: a statistic that has observed values with which one can compare the calculated values, such as peak flow, time of peak flow, duration of exceedance, duration of deficit, and total flow volume. The choice depends on the purpose of the model. Performance evaluation functions used in this study are listed in Table 1.1. In all formulas $OOF_i$ denotes the observed objective function measurement $i$; and $COF_i$ denotes the calculated objective function measurement $i$. The observed and calculated time series consist of $n$ observations that are compared one with another.
Table 1.1: Three performance evaluation functions for evaluating models.

<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>[ \frac{1}{n} \sum_{i=1}^{n} \left</td>
<td>\frac{O_i - C_i}{C_i} \right</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>[ \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{O_i - C_i}{C_i} \right)^2} ]</td>
<td></td>
</tr>
<tr>
<td>Nash-Sutcliffe (N-S) correlatio n</td>
<td>[ 1 - \frac{\sum_{i=1}^{n} \left( \frac{O_i - C_i}{C_i} \right)^2}{\sum_{i=1}^{n} \left( \frac{O_i - \bar{C}}{C_i} \right)^2} ]</td>
<td>Ranges from –∞ to 1. One is a perfect match and a model with an N-S correlation of 0 has the same predictive power as the mean of the observed values. Has gained wide acceptance and is a good choice for a dimensionless measure of fit (Green and Stephenson, 1986).</td>
</tr>
</tbody>
</table>

1.3.2 TRMM 3B42 SBRE

The primary TRMM data product, TRMM multisatellite precipitation analysis (TMPA) (also called 3B42), is a merged passive microwave and infrared product, computed every 3 h on a 0.25° × 0.25° grid between the 50°N and 50°S latitudes. It combines estimates from various satellite systems and, where available, ground based gauges. This SBRE is available as a quasi real time (RT) product with a 9 h lag and as a research product computed 10 d to 15 d after each month end. The research product is corrected using ground-based data, while the real time product is not (Huffman et al., 2007).

1.4 Study Method

1.4.1 Study Area

This study focuses on South Africa’s quaternary catchment C83A (746 km²), shown in Figure 1.1, which is part of the Vaal primary catchment. This catchment was selected because its flow is well gauged by South African standards and it is in a rural area that a smaller municipality would typically manage. (Currently, 2012, evidently outflows are no longer properly gauged, as flows through a local hydroelectric power scheme are excluded.)

The land cover in the catchment is mostly farmland and grass land. The nearest town, Bethlehem, received on average 638 mm rain/y between 1993 and 2009, with most of it occurring during midsummer. (The average was calculated from data ranging from 1993 03 01 to 2010 03 31, received from the South African Weather Service.) The elevation in the catchment ranges
from about 1 600 m to 2 500 m above sea level, with a slope that ranges from very steep in the mountains close to the Lesotho border in the south (78.56%) to very flat on the plains near the outlet (0.22%) and an average of 10.22%. (All elevation and slope values are taken from the Shuttle Radar Topography Mission 90 m Digital Elevation Model.) Soil cover is extremely thin at high elevations, and base rocks (basalt, cave sandstone and shales) are widely exposed. Lower areas include animal and crop agriculture with numerous ponds supplied by wind driven pumps.

Figure 1.1 Study area.
Midgley et al. (1994:Map 1) give the mean annual precipitation for this catchment as 600 mm to 800 mm and they give the mean annual runoff as 20 mm to 50 mm (1994:Map 9.1). Therefore, the proportion of rainfall that becomes runoff in this catchment ranges from 0.033 to 0.063 (often called the runoff coefficient).

In the western part of the catchment is the Ash River, which receives water from the Katse dam in Lesotho through a tunnel transfer. At the outlet of the transfer tunnel the South African Department of Water Affairs (DWA) has constructed a concrete streamflow gauge, C8H036, which measures the inflow from Lesotho. This inflow has a marked daily and weekly cycle, because the water flows through a hydroelectric power station in Lesotho and the flow cycle reflects changes in Lesotho electricity demand through the work day and the work week (Lloyd, C, 2010 11 25, pers. comm.).

In the northwest corner of the catchment is the Saulspoort dam (surface area 443 ha and capacity 16 000 000 m³ at full service level), at the outlet to the catchment.

Variability of the Katse inflow is attenuated by the time it reaches the Saulspoort dam, where DWA gauge C8R004 measures the outflow from the dam. C8R004 also measures the water level in the dam. The flow from Lesotho ensures that significant amounts of water flow through the catchment outlet, even during the dry winter months.

1.4.2 Data and Study Period

Rainfall data available without cost on the internet for South Africa includes SBRE, that agencies such as NASA provide. This project used the 3 h TRMM 3B42 (Tropical Rainfall Measurement Mission Project, 2011) research product, which Huffman et al. (2007) describe in detail. Other SBRE data are available, such as CMORPH and PERSIANN, but Raymond and Sapiano (2010) concluded that no single SBRE dataset outperforms the others. Three TRMM grid cells that overlapped the study area were used in the hydrological model a shown in Figure 1.1 above.

GIS data of the catchment boundary, dams and river lines was obtained from the DWA web site (www.dwaf.gov.za/iwqs/gis_data/river). DWA also provided via email a storage curve for the Saulspoort dam (surface area vs elevation). Flow data at 12 min intervals for the two meters (inflow from the Katse and outflow from Saulspoort dam) are available from the DWA hydrology web site (www.dwa.gov.za/hydrology/). Daily evaporation data at Bethlehem is available at the same web site.

The Shuttle Radar Topography Mission (SRTM) 90 m digital elevation model (DEM; USGS, 2004) was used to determine slopes and to delineate
subcatchments. A MODIS land cover image (NASA, 2008) provided land cover classes to determine Manning’s \( n \) for overland flow in the catchment. Soil infiltration parameters were derived from the Harmonised World Soil database (HWSD) version 1.1 (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009).

Two factors determined the study period. Firstly, a hydroelectric turbine was installed at the Saulspoort dam during 2008 and most of its flow is now diverted through the turbine. The flow measurements through the turbine are not publicly available. Second, a period of statistically extraordinary rainfall and flow was needed. The chosen study period is 2004 11 01 to 2006 10 31 (2 y).

During January 2005, an event occurred that was above the 99th percentile in terms of rainfall and outflow at the Saulspoort dam. The rainfall event lasted for approximately 6 h, while the outflow at the Saulspoort dam remained above the 99th percentile value for \( \sim 50 \) h because a significant amount of water was already present in the catchment after weeks of good rain. The first 12 months of the study period served as the calibration period and the last 12 months as evaluation period.

1.4.3 Hydrological Model

Model Selection

To fulfil the aims of the project, the ideal rainfall–runoff model had to satisfy the requirements shown in Table 1.2 below.

A literature review and an internet search revealed some 24 models that were investigated further for their capabilities, data requirements and software availability. These included well known models, such as IHACRES (identification of unit hydrographs and component flows from rainfall, evaporation and streamflow data), KINEROS (kinematic runoff and erosion, SHE (système hydrologique européen), TOPMODEL, the USEPA’s storm water management model (SWMM) and the Hydrologic Engineering Center (HEC) HMS (hydraulic modeling system) and RAS (river analysis system) models. After evaluation only two possible models remained: SWMM and a coupled HEC–HMS and HEC–RAS model (HMS models overland flow, while RAS models channel flow).

The USEPA first developed SWMM in 1971 and they make its source code publicly available. Tens or perhaps hundreds of thousands of applications and modelers throughout the world use SWMM for hydrological modeling (monthly downloads currently exceed 2 000). For support, SWMM has an active email list (http://www.epa.gov/nrmrl/wswrd/wq/models/swmm/#Contact, retrieved 2011 06 24) and extensive user and developer documentation on the
A Strategy for Managing Deficiencies of SWMM Modeling ...

web. Computational Hydraulics International (CHI) provides in printed format an 945 page User’s guide to SWMM 5, now in its 13th edition. Support is essential for operational models in municipalities where modeling skills may be limited. Although SWMM is mostly used in urban settings, researchers have successfully modeled rural or predominantly rural study areas with SWMM (Chung et al., 2011; City of Austin, 1997; El-Sharif, 1998; Jang et al., 2007; Jun et al., 2010). Of course, urban areas usually include tracts of undeveloped land, which also have to be modeled.

Table 1.2 Model requirements.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Reason for requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide a free and, preferably, open source software implementation of the model.</td>
<td>Purpose of the study.</td>
</tr>
<tr>
<td>Take as input data available for free from the internet, especially for dynamic data such as rainfall.</td>
<td>Purpose of the study.</td>
</tr>
<tr>
<td>Model overland flow and channel flow.</td>
<td>Simpler than having to couple two separate models.</td>
</tr>
<tr>
<td>Model input flow from outside the catchment.</td>
<td>Study area includes a tunnel outlet that transfers water from the Katse dam into the catchment.</td>
</tr>
<tr>
<td>Model events and model continuously.</td>
<td>Satisfy different operational requirements.</td>
</tr>
<tr>
<td>Make the model results available in an external file so that another program can read it.</td>
<td>A control program must be able to access the results and use it in other parts of the management system.</td>
</tr>
<tr>
<td>Use a sub-daily computational and reporting time step.</td>
<td>The study area has a response time of at most 13 h and a real time stream flow system needs to make its information available as soon as possible.</td>
</tr>
<tr>
<td>Have a hotstart capability to process new data as it comes in, starting with the results of the previous model run.</td>
<td>Reduces the time it takes the model to run and so makes the results available earlier.</td>
</tr>
<tr>
<td>Include all relevant hydrological and hydraulic component processes.</td>
<td>Limit model errors.</td>
</tr>
<tr>
<td>Include a capability to implement of real time actions.</td>
<td>Applications such as power generation and online flood warning are anticipated.</td>
</tr>
</tbody>
</table>

Modeling natural channels in SWMM usually includes detailed cross sections of the rivers at regular intervals; this implies high resolution elevation data. Such data were not available for this study, but engineers expressed the opinion that SWMM would do as good a job as the HEC family of models, given the coarse data available for defining and running the model (James, R, 2010 10 22, pers. comm.; Townsend, D., 2010 09 27, pers. comm.). Furthermore, SWMM had the advantage that the hydrological and hydraulic models are already coupled in its code. Additionally, by adding the decision support
software PCSWMM, SWMM has been used in several locations for flow management and flood warning. Thus SWMM was selected for this project.

**Model Construction**

DEMs freely downloadable on the web are commonly rather coarse; the SRTM DEM used in this study has a ground pixel size of 90 m, while the rivers in the study area are seldom wider than 30 m. In the absence of a high-resolution DEM, the rivers were modeled as trapezoidal channels with configurable widths, depths and bank slopes. Initial values for these characteristics were measured from GIS data or estimated where these data were not available (e.g. depths). Rivers were split into sections no longer than 5 km to make them hydraulically as homogenous as possible without making the model too complex.

The study area was discretised into 22 subcatchments for hydrological similarity. Subcatchment boundaries were determined automatically using the GRASS GIS r.watershed module. Some manual delineation was also carried out to separate very steep from flatter parts of the study area.

Smaller farm dams were modeled with surface areas derived from GIS, but with estimated depths and bank slopes. The Saulspoort dam was however modeled using the storage curves obtained from DWA, comprising accurate surface areas at intervals of one metre from the bottom of the dam to its full service level. A transverse weir object modeled the dam wall, for which SWMM calculated flow at each time step and this could be compared to the observed outflow from the dam. An approximately 30 page detailed step-by-step procedure for developing a model is available from the second co-author of this chapter. Table 1.3 shows the complete model inventory for the derived model and data system, denoted UCTPCSWMMC83A.1.

<table>
<thead>
<tr>
<th>SWMM object</th>
<th>Number of objects in final model</th>
<th>SWMM object</th>
<th>Number of objects in final model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subcatchments</td>
<td>22</td>
<td>Junctions</td>
<td>93</td>
</tr>
<tr>
<td>Aquifers</td>
<td>22</td>
<td>Storage units</td>
<td>13</td>
</tr>
<tr>
<td>Rain gauges</td>
<td>3</td>
<td>Weirs</td>
<td>1</td>
</tr>
<tr>
<td>Conduits</td>
<td>111</td>
<td>Outlets</td>
<td>1</td>
</tr>
</tbody>
</table>

In UCTPCSWMMC83A.1 options selected included Green–Ampt infiltration, because it is the more accurate method (Hsu et al., 2002), and Dynamic Wave Routing, because it solves the complete one-dimensional Saint-Venant flow equations and therefore produces the most theoretically accurate results (Computational Hydraulics International, 2010). Routing was calculated every 300 s, runoff every 15 min and results were reported every 15 min.
Before calibrating UCTPCSVMMC83A.1, continuity errors were brought to within 5% by adjusting the widths and depths of conduits, junctions and storage units. UCTPCSVMMC83A.1 parameters were calibrated within their uncertainty limits.

### 1.5 Managing Model Deficiencies

For this study it was found useful to set up a model control system, to help manage projects, scenarios, data files, and run control parameters using Excel, as shown in Figures 1.2 and 1.3.

![Figure 1.2 Model control system, part 1.](image1)

![Figure 1.3 Model control system, part 2.](image2)

In the first part of the worksheet the user enters scenario descriptors and selects interesting run control parameters such as four time steps and
three dynamic wave control parameters (Figure 1.2). In the second part of the spreadsheet, the user enters key dependent variables, such as model run duration, model complexity (total number of uncertain input parameters), three continuity errors, and selected model error functions; in this case both the integral square error and the Nash-Sutcliffe error for (i) the full 26 month run, (ii) a selected dry season, and (iii) a selected wet season are entered (Figure 1.3).

Plotting some of these variables may help interpret their influence, for example complexity vs ISE model error (Figure 1.4), or complexity vs runtime (Figure 1.5) or show the reduction of model error with increase in run time (Figure 1.6).

Figure 1.4 Complexity vs ISE model error.

Figure 1.5 Complexity vs runtime.
We found that this model control system greatly sped up decision making. Incidentally the methodology is especially useful when returning to the project after some time, or when taking over a colleague’s earlier work.

1.6 Results and Analysis

During calibration UCTPCSWMMC83A.1 was found to be most sensitive to the three Green–Ampt infiltration parameters (suction head, conductivity and initial deficit) and channel roughness. Calculated flow was consistently lower than the observed flow and setting baseline flow at the seven starting junctions of streams high up in the mountains reduced this underestimation. Seeps from persistently wet areas or reduced evapotranspiration because of shading could be the sources of such flows. Furthermore, the imperviousness of the study area was set to 5% and adjusted during calibration to 4.1%. In the light of the coarseness of UCTPCSWMMC83A.1, these adjustments to the imperviousness and the objects in the model are considered to be justifiable.

Figure 1.7 shows observed flow at Saulspoort dam in black and the calculated flow from the calibrated UCTPCSWMMC83A.1 during the calibration period in grey, along with the rainfall in the TRMM grid cell that overlaps most with the study area. Overall, UCTPCSWMMC83A.1 performed well, with an N–S correlation of 0.80. UCTPCSWMMC83A.1 performed better during winter (dry season) than during summer (wet season). The N–S correlation of 0.30 during the three flow peaks in late January 2005 (the three biggest storms in Table 1.4) is not accurate enough for a real time stream flow
management system. In particular, the large peak flow error, even after calibration, is unacceptable (see Table 1.4). Further work on reducing this error is desirable.

Figure 1.7 Calculated and observed flow at Saulspoort dam (calibration period).

Table 1.4 Model error statistics.

<table>
<thead>
<tr>
<th>Error Statistic</th>
<th>Uncalibrated Nov 04–Oct 05</th>
<th>Calibrated Nov 04–Oct 05</th>
<th>Evaluation period Nov 05–Oct 06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runoff continuity error (%)</td>
<td>0.00</td>
<td>−0.02</td>
<td>−0.01</td>
</tr>
<tr>
<td>Channel flow continuity error (%)</td>
<td>−0.75</td>
<td>−0.02</td>
<td>−0.03</td>
</tr>
<tr>
<td>Peak flow error (%)</td>
<td>−28.08</td>
<td>−11.39</td>
<td>−33.07</td>
</tr>
<tr>
<td>Mean flow error (%)</td>
<td>−8.77</td>
<td>0.61</td>
<td>2.39</td>
</tr>
<tr>
<td>Total flow error (%)</td>
<td>−9.04</td>
<td>0.59</td>
<td>2.39</td>
</tr>
<tr>
<td>MAE (m$^3$/s)</td>
<td>7.01</td>
<td>1.21</td>
<td>2.11</td>
</tr>
<tr>
<td>RMSE (m$^3$/s)</td>
<td>552.12</td>
<td>255.04</td>
<td>441.00</td>
</tr>
<tr>
<td>N–S $R^2$ overall</td>
<td>0.43</td>
<td>0.80</td>
<td>0.58</td>
</tr>
<tr>
<td>N–S $R^2$ summer</td>
<td>0.18</td>
<td>0.68</td>
<td>0.51</td>
</tr>
<tr>
<td>N–S $R^2$ three biggest storms</td>
<td>−0.52</td>
<td>0.31</td>
<td>n.a.</td>
</tr>
<tr>
<td>N–S $R^2$ winter</td>
<td>0.44</td>
<td>0.87</td>
<td>0.73</td>
</tr>
</tbody>
</table>

During the evaluation period, UCTPCSWMMMC83A.1 performance deteriorated further. Table 1.4 shows the error statistics for the uncalibrated model and the calibrated model during the calibration and evaluation periods.
Figure 1.8 compares the N–S correlations to the standard set for this project: 0.80.

Inadequate performance in wet periods and relatively good performance during dry periods indicate that UCTPCSWMMC83A.1 simulates channel flow well but overland flow less well. This can be ascribed to several possible factors, each of which is covered below:

- coarse time and spatial resolution of the TRMM rainfall estimates;
- errors in the TRMM rainfall estimates;
- errors in streamflow observations;
- poor or missing process models in UCTPCSWMMC83A.1; and
- poor parameter estimation in UCTPCSWMMC83A.1.

A discussion of each of these factors follows.

**Coarse Time and Spatial Resolution of the TRMM Rainfall Estimates**

The temporal resolution (3 h) of the TRMM rainfall is coarse compared to the response time of the catchment (13 h at most) and to the type of rainfall that prevails in the area (convective thunderstorms) and that can move over the area of a TRMM grid cell in <3 h. The TRMM rainfall estimates do of course have errors, but absent knowledge about the size of these errors, they cannot be factored into the model. The lack of several ground based rainfall measurements inside the TRMM coverage prevents us from obtaining such an error estimate. (Comparing the South African Weather Service’s ground based gauge at Bethlehem to one of the TRMM grid cells gave a Pearson correlation coefficient of 0.23 only, but comparing a point value to an aerial value is not very meaningful).
Errors in the TRMM Rainfall Estimates
The errors in the SBRE are difficult to estimate, since only one ground based gauge was operational in the area during the study period: at the SAWS Bethlehem weather office. Comparing its readings to the SBRE resulted in the statistics shown in Table 1.5.

Table 1.5 : Rain gauge and SBRE three hourly comparison statistics.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation coefficient</td>
<td>0.23</td>
</tr>
<tr>
<td>Probability of detection</td>
<td>0.55</td>
</tr>
<tr>
<td>False alarm rate</td>
<td>0.65</td>
</tr>
<tr>
<td>Critical success index</td>
<td>0.27</td>
</tr>
<tr>
<td>Bias</td>
<td>29%</td>
</tr>
</tbody>
</table>

However, the value of these statistics are not clear since the rain gauge measures rainfall at a point, while the satellite sensors and data algorithm estimate rainfall for a 28 km \( \times \) 28 km area.

Errors in Streamflow Observations
Streamflow observations at Saulspoort dam are derived from a stage–discharge relationship, and a pressure transducer measures the stage. DWA has quality control procedures to check the data and all the observations used were marked as good quality data. Nevertheless, no instrument is perfect and the stage–discharge relationship may change with time as water wears the weir down or as silting changes the profile of the dam floor. Hershey (1999:368) estimates the standard error of the mean for stage–discharge relationships at 5%, provided that measurement practice follows the standards of the International Standard Organisation (ISO).

Poor or Missing Process Models in UCTPCSWMMC83A.1
All models contain errors. UCTPCSWMMC83A.1 is coarse because of the available input data, and the node–link (hydraulic) part of UCTPCSWMMC83A.1 is the coarsest because the least information was available about the rivers in the catchment. Yet the hydraulic part of UCTPCSWMMC83A.1 performed best, so it is possible that the first two sources of error mentioned above contribute the most to the inadequate performance of the model during wet periods rather than do the SWMM routines.

Poor Parameter Estimation in UCTPCSWMMC83A.1
Errors in the way that SWMM models overland flow and runoff seem unlikely given SWMM’s long and successful record of use in rural and urban studies.
Even if this was a factor, it would probably be small compared to the other three factors, and could be reduced during calibration.

Further work was done to test the hypothesis that the TRMM 3B42 SBRE is the largest source of error that affects the model results. This was done by calculating the study area’s runoff coefficient using the 3B42 SBRE and comparing it to values in the literature.

1.6.1 Runoff Coefficient

The surface runoff due to rain was estimated and used, along with the TRMM 3B42 SBRE to calculate a runoff coefficient. To estimate the surface runoff due to rain in the catchment, the observed Katse inflow and a mean loss due to evaporation and infiltration was subtracted from the observed Saulspoort inflow during a rain storm. The mean loss due to evaporation and infiltration in the catchment was estimated by subtracting the Katse inflow from the Saulspoort outflow during a dry period.

Dividing the surface runoff thus determined by the TRMM rainfall during the period gives a runoff coefficient of 0.83, which is much larger than the available runoff coefficient from the literature for this catchment: 0.033 to 0.063 (Midgley et al., 1994). The SWMM calibrated model has a runoff coefficient of 0.033, which is within the literature range. The recalibrated model, with the total flow volume closely approximating the observed flow volume (see previous section), has a runoff coefficient of 0.040, which is also within the literature range.

1.7 Conclusion

This study shows that while hydrological modeling with free and open source software and free input data is certainly feasible, accuracy for the full range of computed flows poses significant challenges. With the data currently available without cost on the internet or from other sources, hydrological modeling at the spatial and temporal scales used in this project is problematic, most probably because the input satellite rain data are too coarse. Smaller municipalities with limited resources will typically need to model smaller catchments; thus further research on smaller catchments with some means of improving rainfall input data is evidently advisable.

A model such a UCTPCSWMMC83A.1 may be improved by managing run control parameters, such as four time steps and three dynamic wave control parameters. Other parameters that could be varied are model run duration, model complexity (total number of uncertain input parameters), three
continuity errors, and selected model error functions for: the full 2 y run; a selected dry season; and a selected wet season. Plotting these variables helps to interpret their influence, for example complexity vs ISE model error or complexity vs runtime. Such a methodology may identify areas of the model that could be modified to improve its accuracy.

The amount of data available on the internet to use as input for hydrological models is impressive. Rainfall estimates and DEMs derived from sensors mounted on satellites have created new possibilities for applying these data to real world problems, such as disaster risk management. However, at the moment the spatial and temporal scales of these data, as well as their accuracy, probably limit such applications to larger study areas, unless ground truth rain data, or reliable radar rain data are available. Larger catchments have longer response times and therefore modeling at longer time steps might still be useful in such cases.

Tools used in this study indicate that automatic model development is achievable, but model reliability will remain an important concern.

Acknowledgments

The authors acknowledge financial assistance from the University of Cape Town and The Meraka Institute at the Council for Scientific and Industrial Research, and Computational Hydraulics International for a free licence of the PCSWMM software. The first author conducted this work for a graduate degree, advised by the third author. The second author oversaw model development and developed the model management methodology.

References


City of Austen (1997), *Barton Creek Watershed Model Study*, City of Austin Watershed Protection and Development Review Department, Environmental Resources Management Division, Austen, Texas.


Hsu, S, Ni; C-F and Hung, P-F (2002) ‘Assessment of Three Infiltration Formulas based on Model Fitting on Richards Equation’, *Journal of Hydrological Engineering* vol. 7 no 5 pp.373–379.


