Frankly my dear I give a dam!
Or Using satellite observation to determine water resource availability in catchments

In order to support their needs and improve their individual water security many farmers and water users have built their own dams to store water. Monitoring of these private assets until recently has been unfeasible, as the equipment to monitor water levels and volumes is relatively expensive and difficult to maintain. This has meant that a large portion of any nation's water resources are not effectively monitored. Recent development in satellite technology and cloud processing have now made it possible to monitor smaller dams in specific areas.

The application of these new technology developments has resulted in a new product being developed which enables the estimation of dam volumes at a quaternary catchment level. These include the assessment of the volume of all small dams and medium dams that are not monitored by DWS on a continuous basis. This paper assesses the application of this technology in short to medium term water resources management.

M Thompson[1]; Jens Hiestermann[2], Bruce Eady[3] and J Hallowes[4]

1. Director: GeoTerraImage
2. Technical Specialist: GeoTerraImage
3. Technical Specialist: EkoSource
4. Director: EkoSource

Introduction

Water is a critical and scarce resource in southern Africa and is likely to become even more so as the regional impacts of global climate change become more evident. Being able to accurately and repeatedly monitor available water resources across the entire landscape is a key information requirement for successful water resource management. Currently, water resources management activities across Africa are hampered due to lack of credible, reliable and consistent data. This means that it is difficult to perform assessments effectively to support major development initiatives. Satellite data has now become reliable and can provide information at a level of accuracy which can support many water resource assessments. The developments presented here are the first step in a set of improved monitoring and modelling services which will provide regular and reliable assessment of water availability across the SADC and African region.

GeoTerraImage in collaboration with EkoSource has developed a unique, web-accessible water resource monitoring service, that provides on a monthly basis, detailed, wall-to-wall countrywide inventory of all surface water features across South Africa. This service can also be easily expanded to cover international cross-border hydrological catchments across SADC and even globally.

The web-site provides the means for water resource managers, and other interested parties to be able to access highly detailed, reliable, and regular information on the status of water resources across the country, in support of activities such as compliance monitoring, water security, water licence applications and water resource inventory. In Figure 1 a presentation of the types of results presented from the service are shown.

The advantage of this system is that it provides extremely cost-efficient, desk-top based access to reliable monthly data for all surface water bodies across South Africa, without reliance on expensive in-situ based monitoring equipment; with the associated risks and costs of instrumentation installation, maintenance, and possible loss.

Cloud-based satellite image data archives, combined with big data processing capabilities are used to automatically generate spatially detailed information on the extent of all surface water features, from small farm dams and natural pans to large impoundments, across the South African landscape, on a monthly basis. The service is based primarily on the European Space Agency's (ESA) 20m resolution Sentinel2 satellite imagery, which allows all surface water bodies typically > 0.25 ha to be identified and mapped on a repeatable and accurate manner. The service is currently able to generate information on water surface areas. These will be updated shortly to dam volume estimates.

Modelling principles

The current month's total surface water represents the average water surface extent for the month under assessment, rather than the maximum extent that occurred within that month, as a result of the image data modelling approach used to minimise cloud obscured data losses. Based on the original ten-day overpass schedule of Sentinel-2A, a six-week period represents approximately five potential image acquisitions over the same location.

The image acquisition and overpass rate has increased to a five-day period in late 2017 when Sentinel-2B became operational – Sentinel-2A's tandem pair in the dual constellation. This means that it is now possible to acquire up to approximately ten image acquisitions in any six-week period, over the same area, cloud cover conditions permitting. The median value for each image pixel (per spectral band), from all image acquisition dates within the six-week period, is then used as the final value on which the presence or absence of water is modelled in that month. This is based on the assumption that even if cloud cover has obscured an image pixel on one acquisition date, it is highly unlikely that clouds will have obscured the same pixel on all dates. Hence, the median (rather than average) pixel values will either remove or minimise the occurrence of cloud or cloud shadow impacted pixel values being included in the water modelling calculations. Should a pixel be cloud affected over several acquisition dates so that it is not possible to extract a pixel value for that month, then that pixel is classified in the final output as a "cloud-loss" pixel, which is accounted for in the monthly surface water area calculations and reporting. Cloud problem effects are further minimised with the application of a cloud-top and cloud-shadow exclusion mask that is generated for each image date and which effectively masks out ± 95% of cloud affected areas. However, since the cloud masking process does not guarantee 100% exclusion of cloud and cloud shadow areas, it is deemed necessary to use this approach in combination with the median pixel value approach. The advantage of this approach is that no false positive water areas (resulting from cloud shadow areas) are included in the final water surface area output. The disadvantage is that any given month's water surface area representation is in reality the median surface extent for the month, and not necessarily the maximum, especially if the significant majority of rainfall occurred in, for example, the last quarter of the month and the preceding weeks had been dry. This will result in the real current, maximum surface water extent only becoming evident in the following month's water modelling update, which would include two weeks of image data from the preceding month.

Longer-term surface water areas, such as over a six or twelve-month window, can also be generated from the combined, cumulative individual monthly water area outputs. In such instances it is highly unlikely that these longer-term surface water representations will contain any cloud top and cloud shadow data loss issues, due to the high number of surface observations making up the long-term picture.

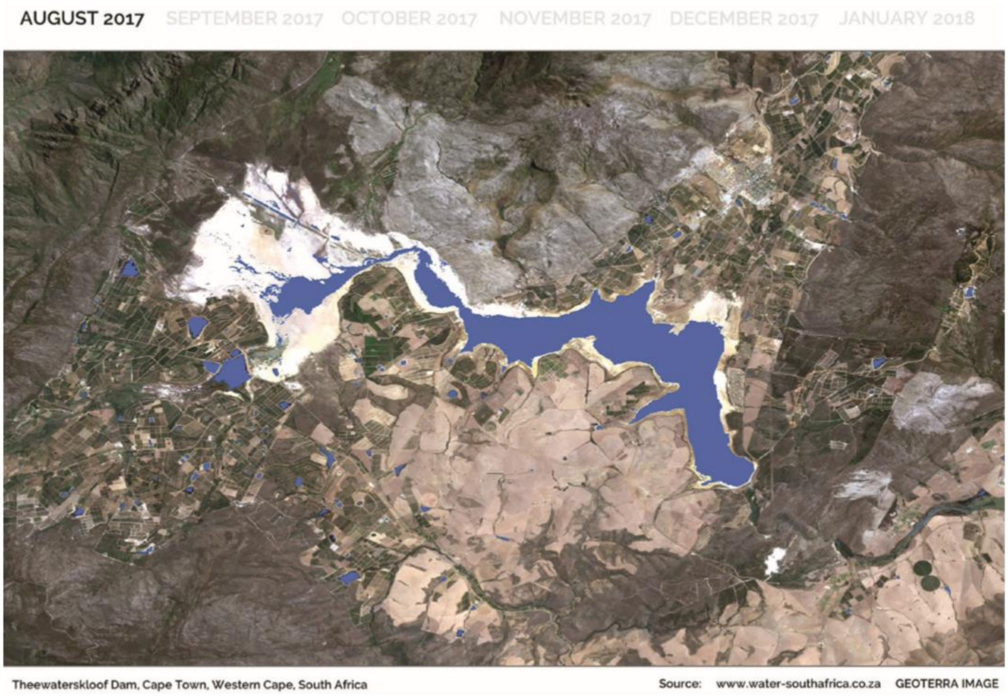

Figure 1: Changes in surface water extents between August 2017 and January 2018 in Theewaterskloof Dam, Cape Town.

Data modelling and automation

The core procedural objective has been the full automation of the image data access and subsequent water surface area detection procedures. This has successfully been achieved by utilising cloud-based global image data archives and associated big-data processing analytical capabilities, and removed the need for downloading, preparation and conventional modelling and analysis of large volumes of image data using office-based proprietary software. The result is significant enhancements in procedural efficiencies that requires minimal office-based support infrastructure.

Sentinel-2 Imagery The surface water extents are all modelled from ESA's Sentinel-2 imagery, sourced as 20 x 20 m resolution MSI Level 1C data from the Google Earth Engine cloud-based data platform. Level 1C imagery is all imagery precisely co-registered and provided in standardised Top-of Atmosphere (ToA) reflectance values, in Web Mercator projection format.

Detection algorithms using decision tree modelling. Decision tree classifiers are predictive modelling algorithms that can be used to generate explicit classification rules, and are ideally suited to developing generic modelling routines for standardised and repeatable classifications of satellite imagery. Typically, a set of training data (i.e. reference samples) are used to generate the ruleset, which can then be applied to larger data populations for repeatable and consistent classification outputs (Figure 2). As such decision tree classifiers are ideal tools for deriving standardised, threshold-based rules for image classification, they can be applied repetitively over time and/or space with the same output content and accuracy. The water surface modelling procedure is based on a set of decision tree generated rules that have been derived from a comprehensive set of water and non-water feature reference points distributed across South Africa. The reference points are all associated with a single 20 x 20 m image pixel, and represent a wide range of seasonal and geographic water and non-water surface characteristics across the country, which can be determined visually on Sentinel-2 imagery. The sample points represent the geographical positions at which spectral image characteristics are extracted from the cloud-based image archives in order to characterise and describe seasonally-defined spectral signatures for all land surface conditions. The specific rulesets for spectral water detection, including potential non-water confusion features, are generated using water and non-water spectral reference characteristics as inputs into the decision tree algorithm. The final water-only identification ruleset represents a comprehensive set of spectral threshold-based rules which can be applied to multi-seasonal Sentinel-2 imagery to determine the presence or absence of water in any given image pixel.

NDVI <= 0.481442 (branch 1)

| NHI <= -0.066 (branch 2)

| | GNDVI <= -0.084901 (branch 3)

| | | NDTI <= 0.07211 (branch 4)

| | | | NHI <= -0.248952 **WATER** (branch 5, final 'water decision' leaf)

| | | | NHI > -0.248952 (branch 6)

| | | | | NDTI <= 0.005025 (branch 7)

| | | | | | etc...

Figure 2: Decision tree hierarchical rule structure.

Training data A total of ±60 000 sample points across the South African landscape were identified and used to represent both water and non-water (but with similar spectral characteristics to water) landscapes. The distribution and location of these points covered a wide range of landscape types and associated seasonal conditions to ensure full representation of all spectral characteristics likely to be encountered during image-based water modelling. One example is differences in water colour as a result of depth and/or turbidity. All sample points were visually identified and defined on Sentinel-2 imagery (circa 2016-17) using manual, desktop mapping techniques. At each sample point a range of spectral values were extracted, based on a pre-defined set of potentially useful spectral indices and individual spectral band values. In some instances, spectral values were extracted for a sample point linked to a specific image acquisition date in order to ensure the correct representation of a seasonally dependent feature's characteristics, whereas in other instances, spectral values were extracted for the full seasonal range of feature characteristics. Both water and non-water sample points were used to ensure that the water identification ruleset generated by the decision tree algorithm was able to accurately extract water features, and exclude non-water features that had similar spectral characteristics to water, such as dark terrain or cloud shadow areas, dark non-vegetated surfaces from both natural and man-made environments, and temporary burn scars from wildfires.

Spectral indices

The list of suitable spectral indices for water and other landscape feature modelling was sourced from various publications, with the final selection based on proven usefulness with Sentinel-2 imagery, and to some degree due to similar spectral input data, the ability to potentially replicate the same processing (if ever required) on Landsat 8 imagery. The final selection of the most suitable combination of spectral indices for surface water area detection was determined solely from interim outputs generated during the decision tree rule modelling process. The decision tree classifier software used to identify both the optimal spectral input data and generate the final water detection rulesets was the Waikato Environment for Knowledge Analysis (WEKA) suite of open source machine learning software.

Machine learning and optimal rule generation The WEKA software includes the open source Java J48 version of the C4.5 algorithm. This algorithm, which is considered one of the top performing data mining algorithms, is used to generate decision trees which are ideally suited to spectral image classification applications. A decision tree consists of sets of hierarchically branches, each eventually ending with a leaf, which is the end of a particular ruleset. The size of a decision tree is defined by the number of hierarchically linked branches that collectively represent a single ruleset that defines a classification decision and final outcome. For example, the first five branches of the decision tree in Fig. 3 represent collectively the ruleset for classifying one instance of water, with branch five representing the end-point, i.e. leaf, of the ruleset for this water classification decision. Within a decision tree there will be many leaves that collectively describe all the rulesets required to classify, for example, all occurrences of water.

Calculating dam volumes

The automated detection of water surface from satellite is the initial step in the process. While, information relating to the dams surface are is very useful, the conversion of the information to volume provides a better perspective on the overall water availability in a region. This spatial information is now being further enhanced with the conversion of the information to dam volumes.

The process to convert dam the spatial results to actual volumes is influenced by a number of different factors but are primarily impacted by the dam shape and the valley topography. It is thus not possible to generate dam volumes in a universal manner and it is important to take into account these characteristics. Fortunately, the latest GIS modelling techniques enable the effective calculation of dam volumes with a reasonable degree of accuracy, enabling the individual calculation of dam volumes in automated processes. While it is not possible to calculate this information on the fly it is possible to generate individual volume area relationships for larger individual dams in a specific area and to obtain relevant generalised curves for the smaller dams in an area. Evidence suggests in the majority of cases the topographic attributes at least at a quaternary catchment scale seem to present a similar volume area relationships. It is possible to calculate a consistent generalised volume area relationships that represent the characteristics of the smaller dams in a specific area. Larger dams can present different results but it is possible to calculated these volumes on an individual basis. This information can then be combined with the existing surface area estimates to calculate the volumes into 3 categories at a quaternary catchment level, namely: Small farm dams lumped together, medium size and strategic dams with volume area information calculated from DEM information process discussed below and finally, larger dams with known surveyed volume area calculations.

The calculation of dam volumes is hampered in many circumstances by the difficulty in knowing the true bathymetry of a dam. Unfortunately, the majority of methods used to determine accurate topography presently, are not able to penetrate the water surface meaning that it is not possible to determine the bathymetry to generate dam volumes. A process has thus been developed to generate hydrologically corrected DEM which is then used to calculate the Volume Area estimates. The sequence used to generate this information is discussed below.

The image below shows the raw DEM received from Airbus WorldDEM, for dam in the Theewaterkloof catchment area. A dam shape which has been digitised from photography is indicated by the blue line. A good match over the DEM, where the dam is indicated by a flat surface area can be observed. This level of congruency appears to occur for all the dams analysed so far.

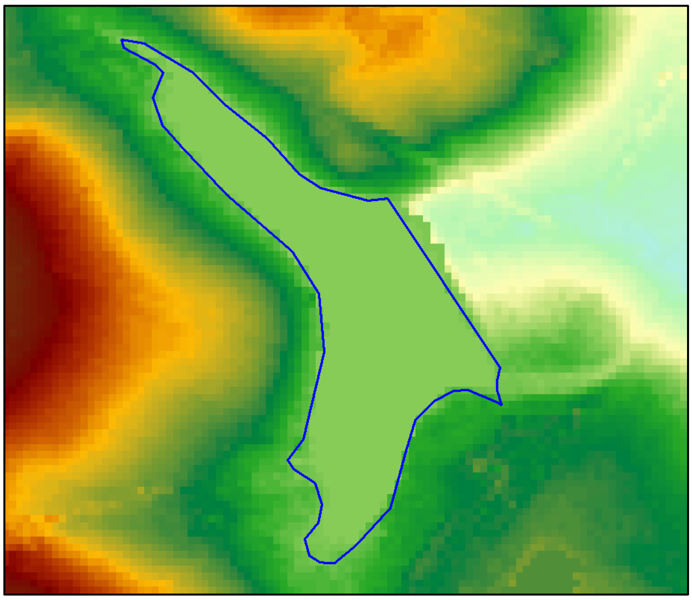

Figure 3: Showing the dam digitised dam outline an the flat area of satellite derived topography

In order to determine the dam bathymetry the flat water surface from the DEM is removed. Thereafter, a new DEM is created by using an appropriate interpolation technique, creating the assumed underlying bathymetry from the surrounding topography data. An example of this is shown figure 4 below.

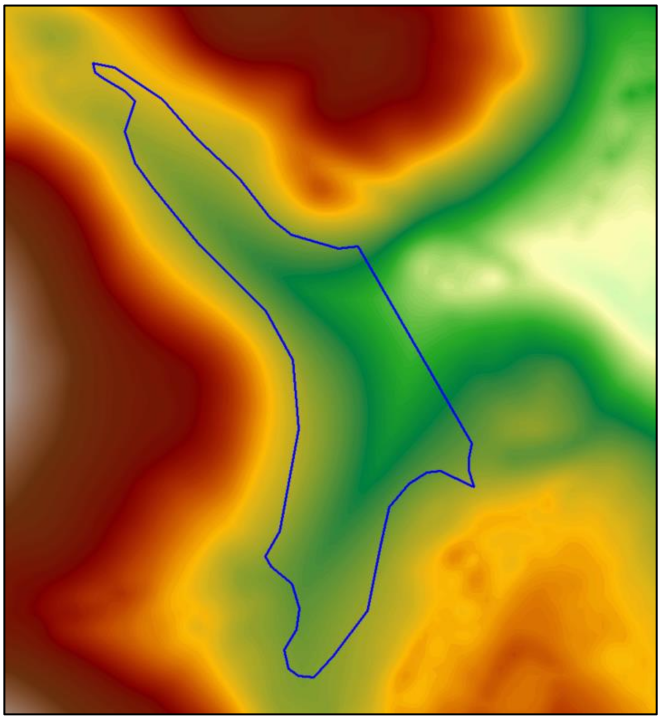

Figure 4. The interpolated surface is generated creating a representation of the dam bathymetry

This process results in the underlying dam bathymetry being represented. A process which is demonstrated in the figure 5 below is used in these cases the dam is essentially spliced into several layers and the volumes under each layer are calculated. The figure below shows a series of inundation areas at a consistent 4m elevation interval. The dams volumes are then generated using this technique, usually obtaining between 5 and 10 data points to fit a volume area relationship.

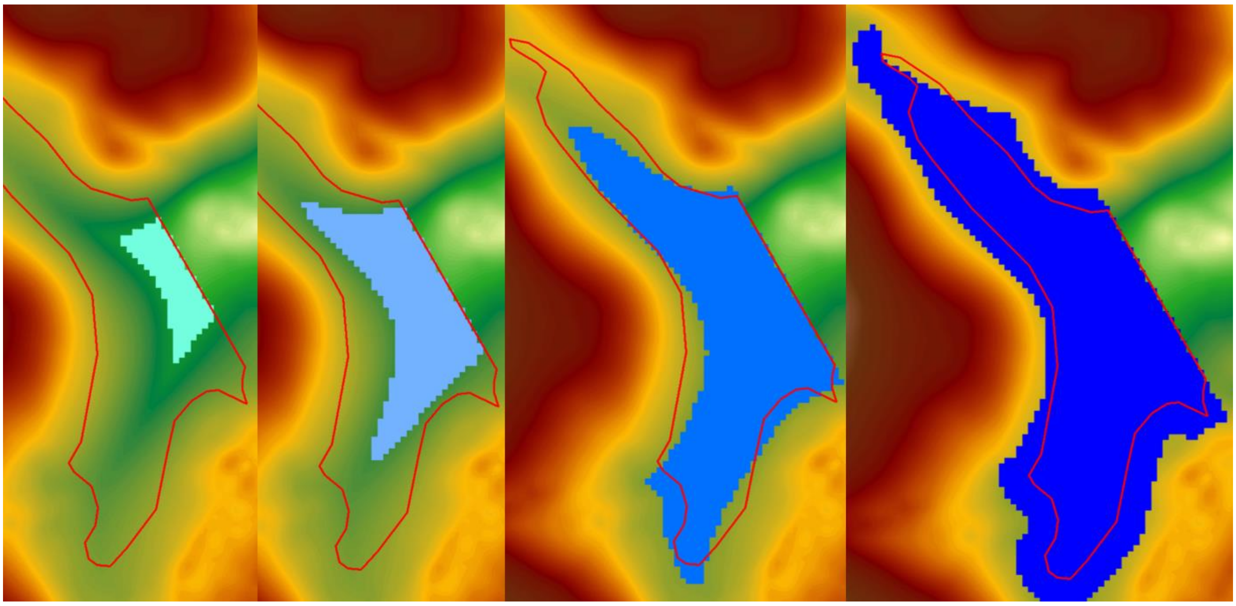

Figure 5: Graphical representation of process used to calculate the Area volume relationships

In the figure 6 below each of the intervals from the figure 5 are presented. It is thus possible to calculate an area and volume relationships.

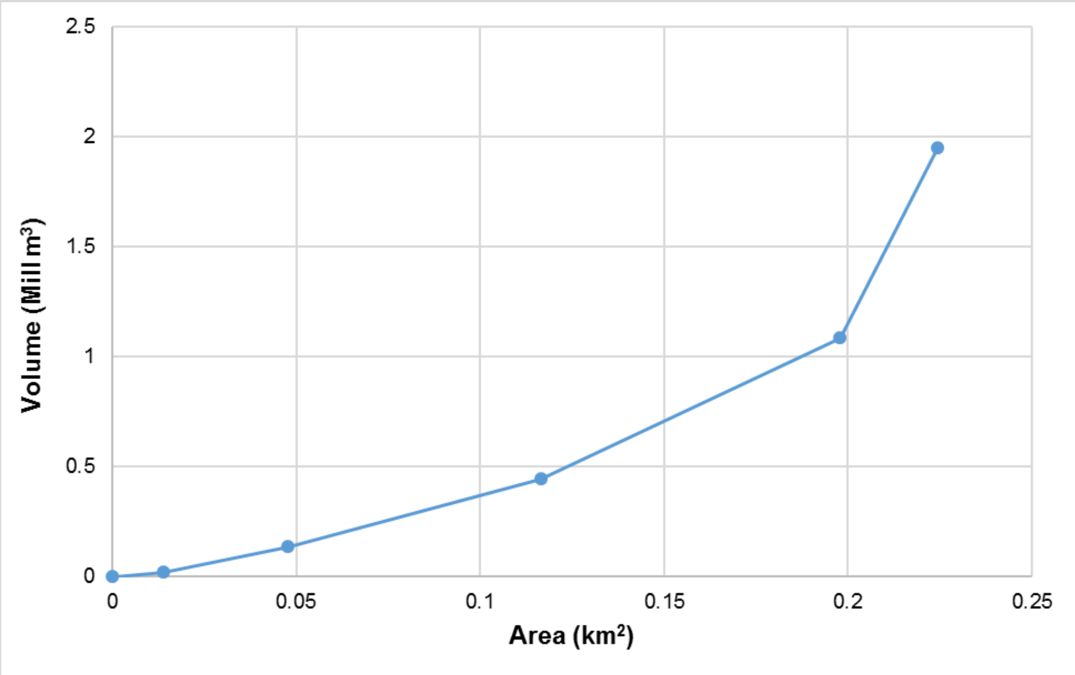

Figure 6: Representation of the individual calculated volume area relationship

It is hence possible to generate information which provides catchment based Volume Area relationships across South Africa and eventually across the globe using this technique. The next element to consider is the errors associated with these techniques. What level of accuracy is generated from the assessment techniques and calculations presented so far in this paper. There are 3 different types of errors that have been identified that could occur from this approach. It is possible to quantify these errors. The main errors that could occur are from:

- Misidentification of water bodies
- Errors in the calculation of surface areas
- Errors resulting from the calculation of dam volumes

Misidentification of water bodies

The misidentification of a dam or surface water feature is a function both of the spatial resolution of the Sentinel2 20m image format used, and the spectral modelling parameters. In terms of spatial resolution, water body detection is a function of what is referred to as (Strahlers) high or low-resolution modelling theory, which relates to how the image cell size (i.e. pixel) relates to the size of the landscape object. With a 20m pixel format, it is possible to ID the existence of a water body in a given area even if the size of the water body is smaller than 20x20m, if the spectral characteristics of the water body are significantly different from the surrounding landscape, and the area of the water feature typically exceeds 40% of the pixel area. In such cases, the water body will be potentially detectable, but the area representation of the water feature will be incorrect and over-estimated, since the entire pixel cell will be classified as water, even though the actual water feature is smaller. If the water feature is considerably larger than the 20m image pixels, and thus contains many image pixels with spectral water characteristics, the water feature will be identified, and well represented spatially. Simply put, the more 'whole' (as opposed to partial or edge) image pixels that represent a particular water body, the more accurate the expected detection and area estimation.

In terms of spectral modelling, the primary sources of commission error (i.e. too much water is "found" within the landscape) are typically within dark, low reflectance areas, such as within cloud and topographic shadows and wildfire burn scars, all of which exhibit similar low reflectance characteristics as water bodies. The water detection algorithms and image modelling principles that have been developed have been designed (and subsequently refined) to minimise all such commission errors. Anecdotal visual estimates of the accuracy of water body detection is around, we think 90% or better, in terms of visual, on-screen comparison to equivalent-date imagery representing the same time period as the image-modelling outputs. This includes both the ability to detect the existence of a water body as well as the area representation of the water feature in comparison to what a skilled image interpreter would delineate manually. Unfortunately, without same-date real reference data on water boundaries for a given monthly window, and actual statistical evaluation is a challenge.

Water identification omission errors, i.e. where an actual water feature has not been identified, other than due to small size, are typically the result of either 'impure' spectral characteristics, as a result for example of a heavy sediment load or very shallow conditions, where the reflectance characteristics are more aligned with a non-water bare ground surface. Or where the water body is "lost", even within multiple image observations as a result of persistent cloud shadow coverage or "deep" multi-seasonal terrain shadows, such as occurs in kloofs and gorges.

Errors related to calculating dam surface areas

Dam Surface Areas detection inaccuracies decrease as the size of the dams increase (Figure 7), this is due to the size of the dam area compared to the pixel resolution of 20 meters. The consequent commission and exclusion errors on the smaller dams thus have a higher impact on the accuracy. A graph concentrating on the initial part of the curve shows that the dams with volumes of 10 000 m2 present errors of less than 10% while smaller dams the errors can increase up to 80%. Errors associated with dams above 50 000 m2 present errors in the 2.5% range.

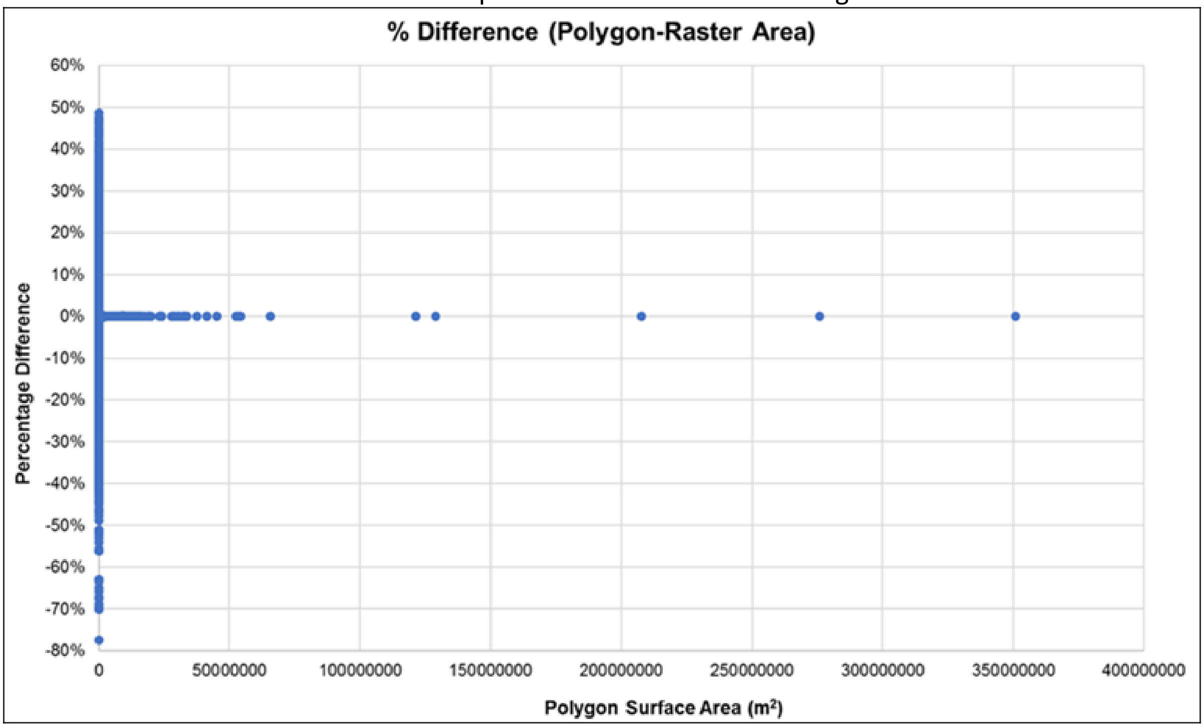

Figure 7. Errors for individual dams

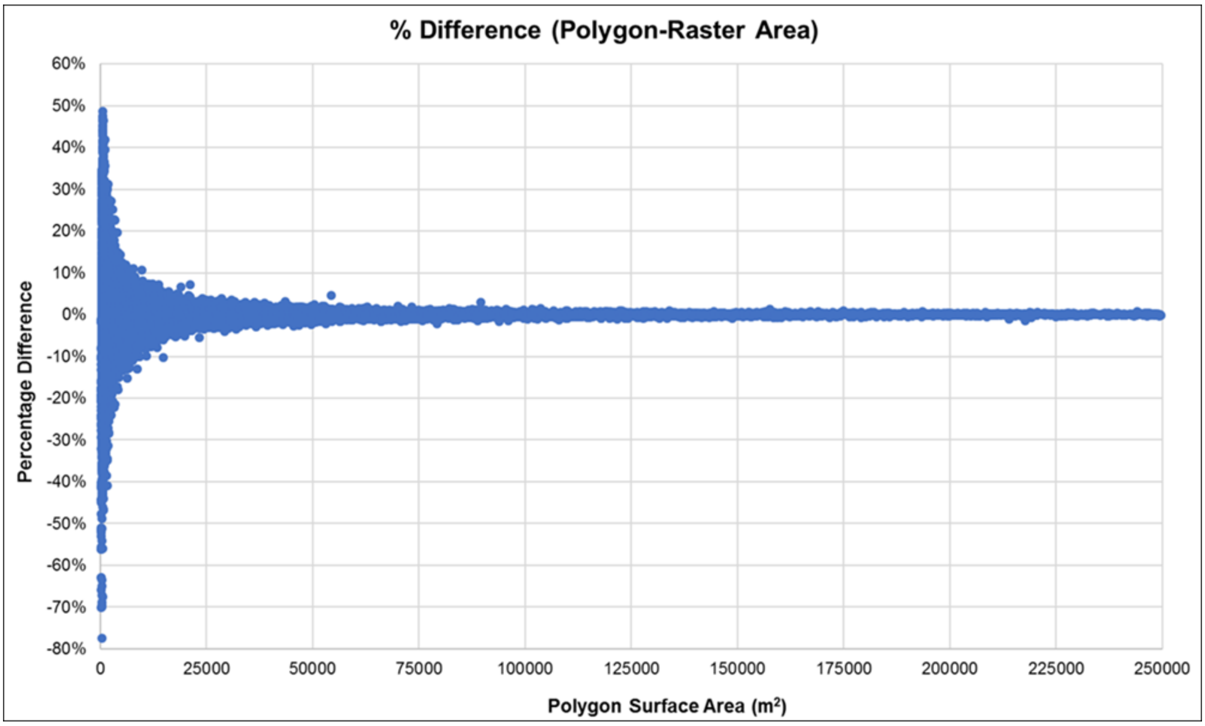

Figure 8. Errors for individual dams excluding large dams above 250000 m^2.

The statistics of errors, associated with the surface area commission errors is provided below. While it can be seen that the maximum errors associated with individual dams can be as high as 80% for the larger dams in general the mean error is about 10% for even the smaller dams (Figure 9). The lumping of smaller dams at a catchment scale produces an impact where there is a regression towards the mean and in the majority of cases the estimates of the smaller dam areas is less than 10%. A summary of these findings is presented in figures 10 to figure 12 below.

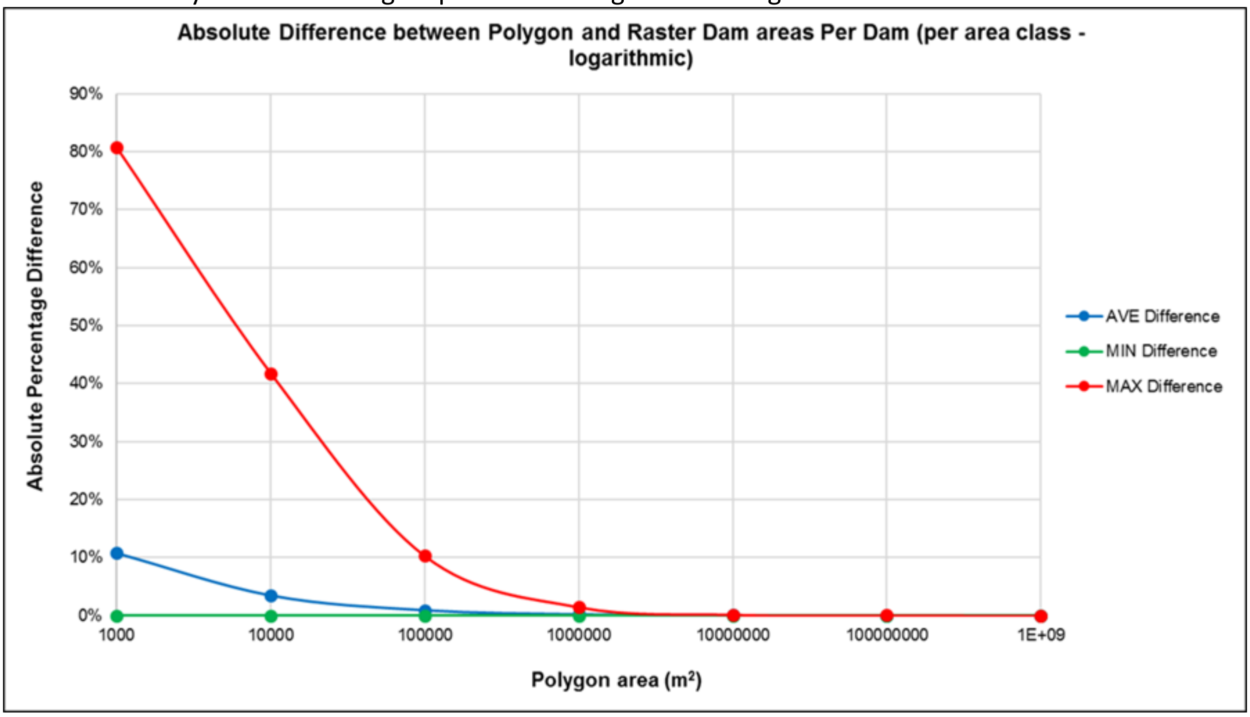

Figure 9: Summary of errors between polygons and raster estimates for area of individual dams.

Observations show that at the quaternary catchment level summary is the errors associated with the lumping of dams decreases the overall error to a range of less than 15% (Figure 10). Analysis reveals that that the maximum error is in the region of 15% per quaternary and it converges to less than 1%. (Figure 10 to 12)

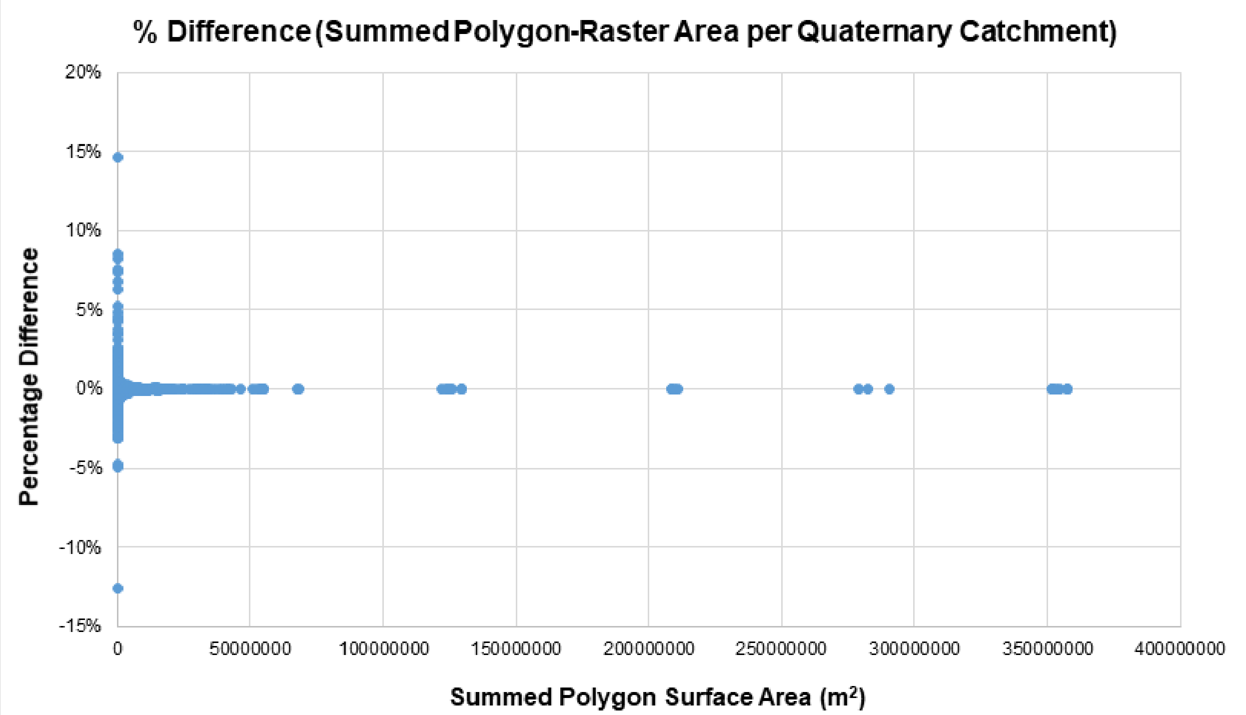

Figure 10: Difference between the raster volumes and polygon estimates.

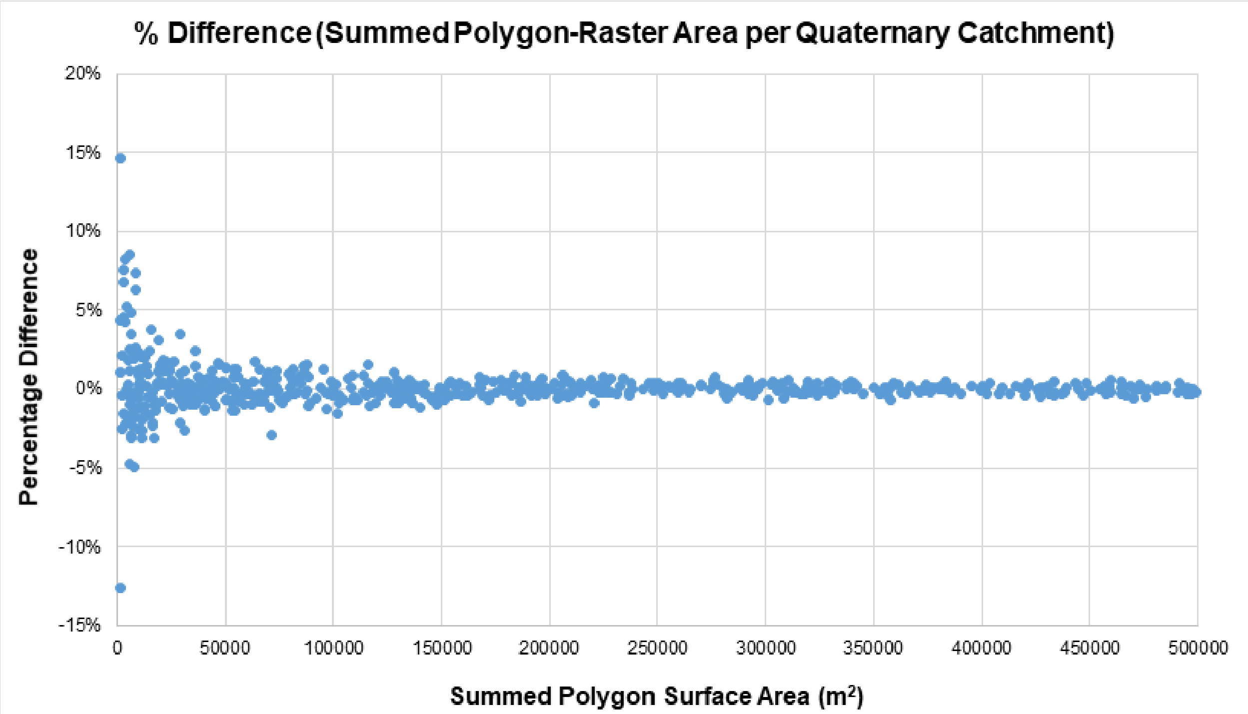

Figure 11: Zoomed in difference below 500 000 m^3.

The function giving the difference in error estimates on the average dam size is provided in the graphs below. We can see at a quaternary catchment level the errors are generally less than 10% and converge very quickly to less than 1% (Figure 12).

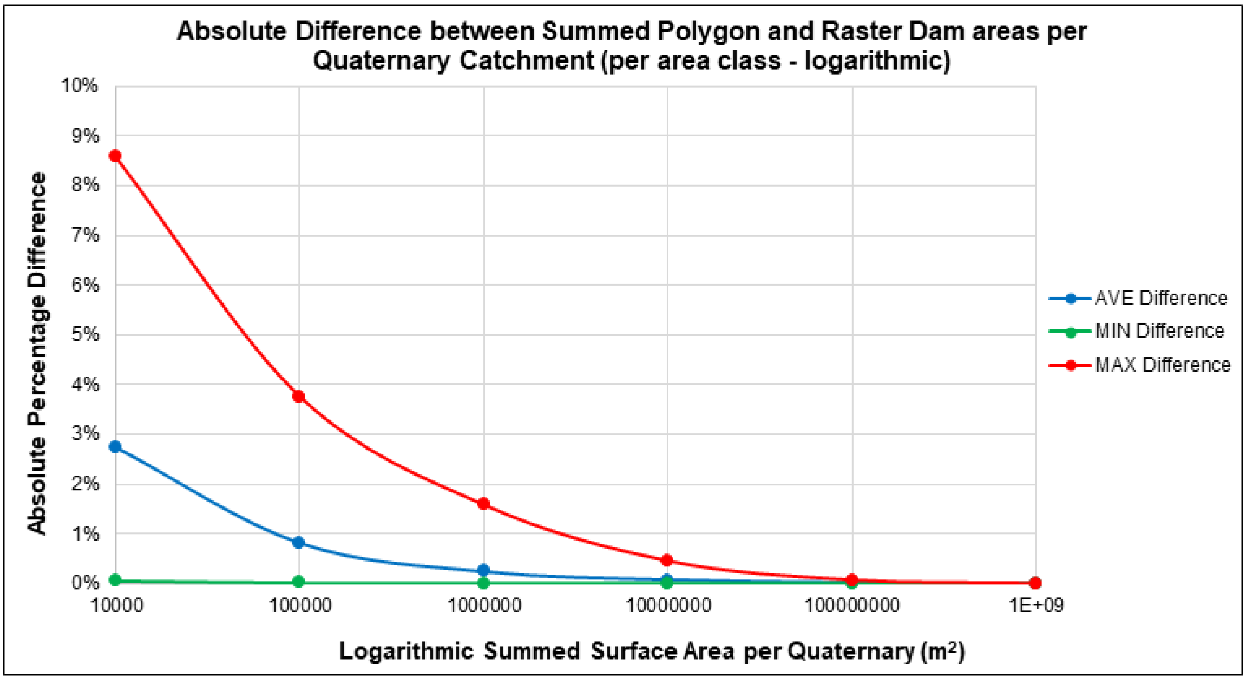

Figure 12: Difference as a mathematical functions.

Errors in calculating dam volumes

The above gives us an idea of the estimation errors we could expect in determining the overall error for calculating the surface area from the raster versus the digitised polygons of the dams from photography. The next component to investigate is the errors that could occur in the volume estimates. This varies from region to region and depends on the topography in a particular area. This estimate has been done in selected areas and processing is currently underway to do this at a country wide level. Initial estimates in pilot catchments suggest that we are able generate volume estimates from topographic data relatively effectively and that the errors can effectively be calculated. Error estimates increase as the size of the dam increases this is due to larger uncertainties that occur with the interpolation of the bathymetry as the dam sizes increase. It can be seen that at low volumes and smaller dams the errors are relatively low but as the dam volumes increase it appears that there is a decrease in the volume accuracy. These errors are generally in the region of 0 – 10% but can be as high as 20% in specific cases. The initial comparisons on estimate approximations are provided below in figures 13 and 14.

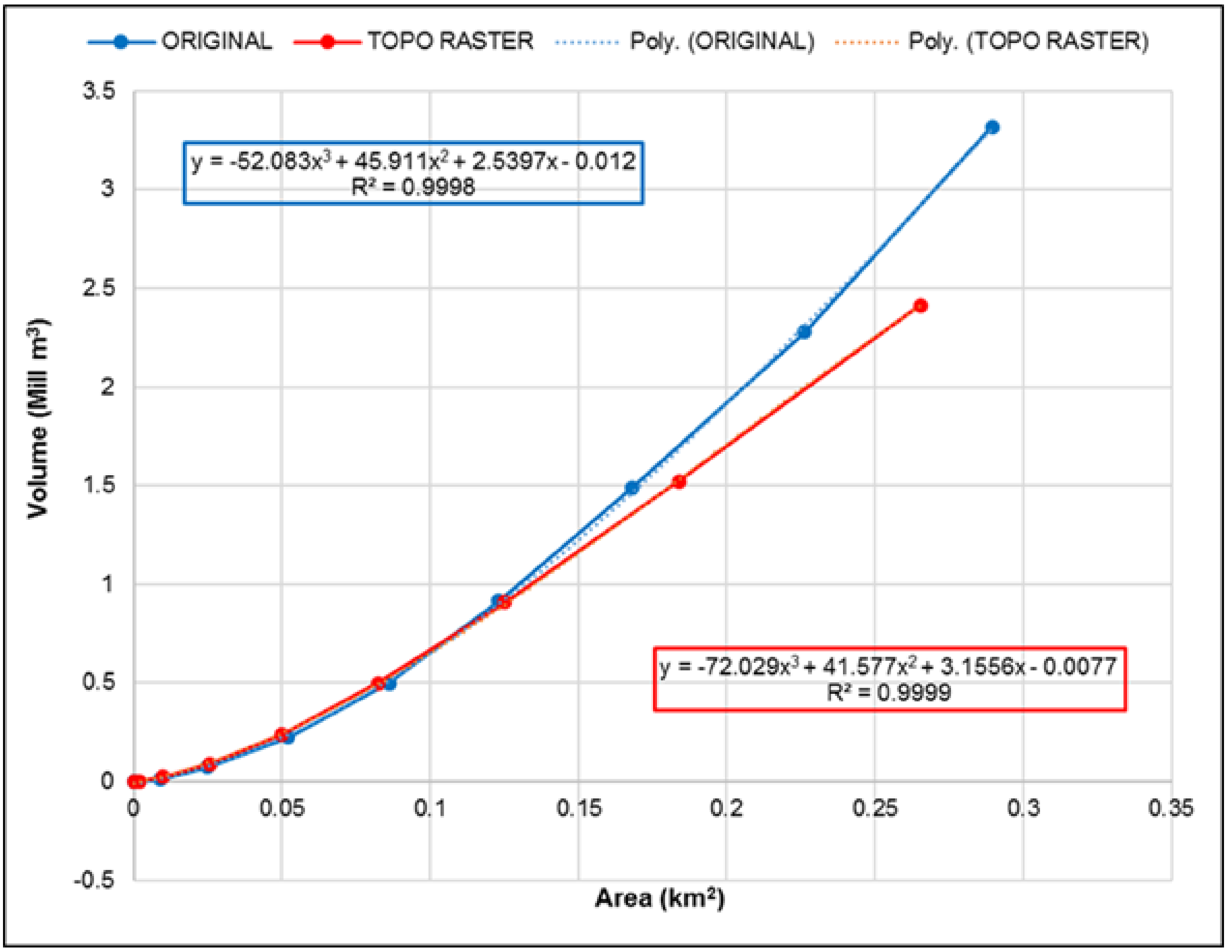

Figure 13: Comparison of actual dam volume compared to interpolation measurement method.

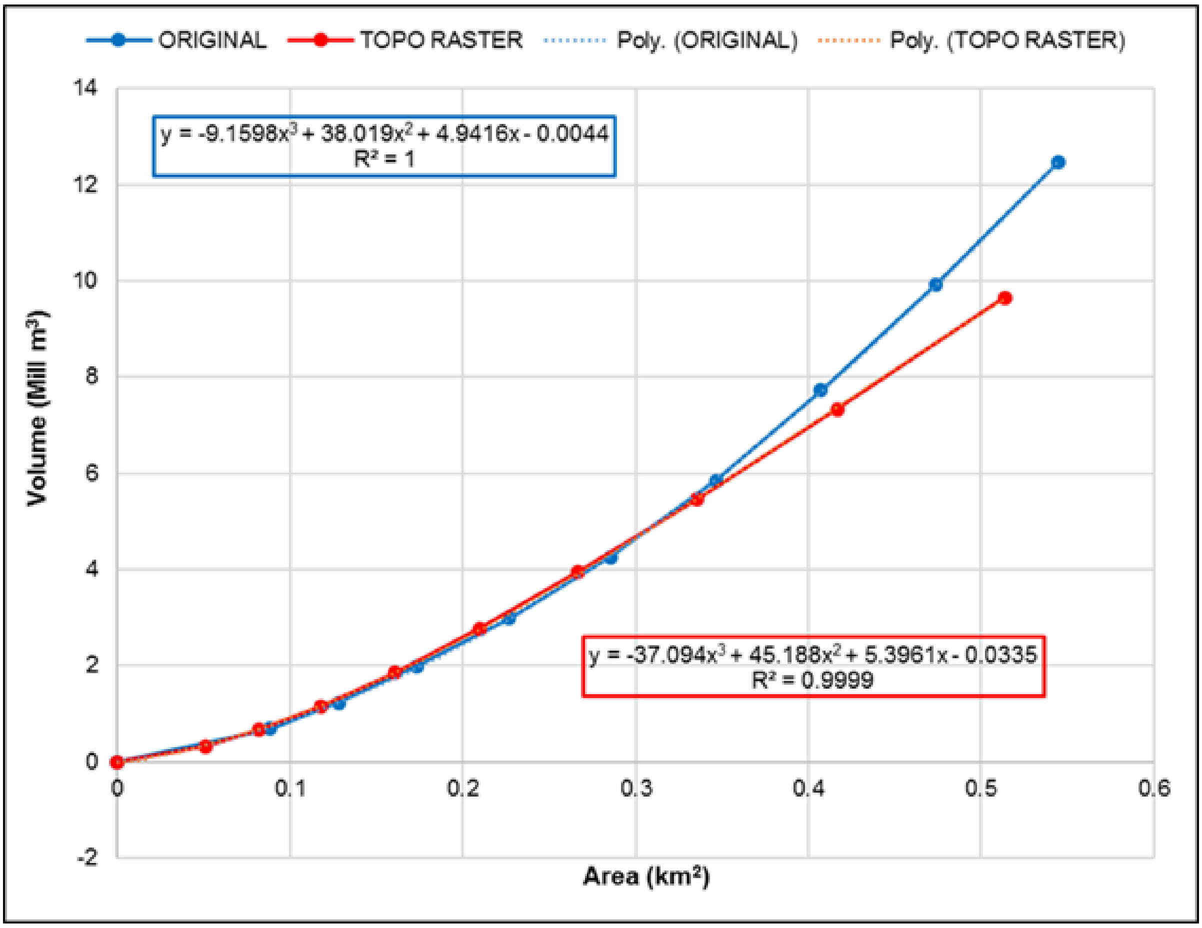

Figure 14: Comparison of same technique in different area.

These initial estimates are provide promising results and a family of curves giving an overall estimate of differences per quaternary catchment will be done using the topographic data. Again it is anticipate that there may be a regression towards the mean and it appears a definite bias in the technique occurs above a certain size and this can effectively be compensated for once more data is available reducing error estimates in specific areas.

Overall errors

An interesting inverse relationship occurs in the estimation of errors. While the errors associated with spatial area estimates decreases with bigger dams, the error associated with volume estimates increases as the dam size increases. The result is that the dam volume estimates from satellite generally present errors which are in less than 10% overall in the areas where the techniques have been tested. This suggests that an overall accuracy associated with this technique will generate estimates of better than 90% accuracy at a catchment level for all the unmonitored dams. This is improved if the larger dams have defined HVA tables in which case errors present are in the region of 2%.

Water surface area monitoring services

Mzanzi Amanzi, the monthly water monitoring service, is an operational, fully-automated procedure using cloud-based computing and data archive technologies. The cloud-based processing makes use of Google's Earth Engine infrastructure, which provides access to global image archives, scalable computing power and flexible, large-volume data storage options. This is the most efficient way to support the water monitoring web-based platform and ensure the provision of monthly national coverage information. Image archives on the Google servers include full global records of a range of ESA imagery from the agency's Copernicus programme, which includes Sentinel. Within the Google Earth Engine workflow, the GeoTerraImage developed water detection models are uploaded to the cloud-based system and applied to the relevant imagery in the cloud-based image archives. This approach has many advantages in comparison to conventional desktop procedures and workflows using proprietary GIS and image processing software: A cloud-based approach significantly improves data processing speeds, efficiency and levels or operational automation (using open-source programming languages). Most importantly, it removes the need to download and pre-process imagery. For example, a full year's database of Sentinel-2 imagery across South Africa, assuming all five-day overpasses generate cloud-free usable data, would be equivalent to roughly 7500 GBs in size, which would impose significant data access, storage and processing challenges to any desktop-based water monitoring process. At the start of every month the workflow procedures are activated. The cloud-based procedure loads all the Sentinel-2 satellite imagery over South Africa taken during the previous six weeks. The automated process initiates its two steps: first using a specific set of rules to identify and mask any cloud obscured imagery, after which a second set of rules is applied to the non-cloud obscured data in order to identify and classify all water areas at an individual pixel level. The derived water datasets are stored in GeoTerraImage's allocated Google storage and then synced into the web-based application, at which point it becomes publicly viewable. Anyone can visit and review the website.

Product

The public web-accessible information describes both visually and as tabulate summaries, the current months total surface water, area as well as the previous two months total water areas, the long-term maximum water extent and an index of long-term water occurrence levels.

All water statistics and spatial maps are updated in the 1st week of each month, according to the latest surface water areas detected in the preceding month, i.e. on 1st June the water for May is calculated and uploaded to the web-site.

The long-term data sets are derived from the total archive of Sentinel2 imagery, covering the period from late 2015 to present, and are again updated and revised and updated each month.

By clicking on one or more quaternary catchments, the website will display the statistics of the current, previous and long-term comparative monthly total water surface areas in the selected catchment(s).

Subscribers to the commercial paid-for-service can access the actual digital monthly water maps (which are provided in GIS compatible raster format, based on 20m cells), and comprehensive tabulated water data that contain monthly and long-term surface water area statistics per quaternary catchment, from October 2015 to present; all of which are updated monthly.

Conclusion

Over the second quarter of the year, the companies will be implementing further changes and improvements to the website, and users feedback and suggestions will be considered in these future development plans. A service now exists that provides a reliable solution to ensuring timeous and accurate information on catchment water resource status, that is based on non-traditional / non-conventional technologies, that are primarily independent of the need to in-situ equipment, field based collection, telemetry based and data measurements. The information presented will enable the continued quantification of water resources status allowing the monitoring of resources in remote unmonitored areas. The information provided will also give a better understanding of the overall water resource status in areas. Information provided will enable the better understanding of drought and wet conditions in areas and enable the better planning of water restrictions in areas. In all the system will provide a reliable estimates of water volumes across the SADC region and similar climate zones globally.

Future

The application is being developed further presently. In areas of high cloud cover the service which is based on visual spectrum is being enhanced to include cloud penetrating radar information from the latest Sentinel satellites. This upgrade will deliver a completely global method of determining dam surface area estimates in all climate zones. Further to this current negotiations with Airbus will enable the use of the global based high resolution DEM to generate dam volume estimates on a global scale and it is anticipated this will be in operation in mid 2019.

Initial analysis suggests that the service can be used to perform extended monitoring activities and can reveal interesting information for the following applications:

- Flood modelling applications where water extent can be detected effectively
- Monitoring of wetlands wetting and drying cycles in specific areas
- Establishing sedimentation rates of reservoirs where sediment bathymetry changes have been identified on the inflow areas of reservoirs under drought conditions
- Water resources estimates providing projected dam volumes up to a year in advance related to hydrological modelling and water use estimates

There is a lot of promise offered by this service which could form a good basis for the improved water management in many developing countries around the globe.