



science & innovation

Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA



SANSa Professional Serviced Contract.

PROVISION OF NATIONAL WATER QUANTITY MONITORING PRODUCT.

Data Report and MetaData

(version 008)

For

SOUTH AFRICAN NATIONAL SPACE AGENCY (SANSa)

By

GeoTerra Image SA (Pty) Ltd, South Africa

www.geoterraimage.com

In collaboration with **EKOSOURCE** Pty Ltd South Africa

Date : July 2021



LIABILITIES

The user / client has to verify prior to his commitment, the adequacy of GEOTERRAIMAGE / EKOSOURCE datasets and products to the specific data and analyses needs as set out by the client. GEOTERRAIMAGE / EKOSOURCE will in no event be liable for any direct, indirect, incidental or consequential damages including lost savings or profit, legal fees, lost data, lost sales, loss of commitments, business interruption, including loss of or other damage, resulting from its datasets and products, whether or not GEOTERRAIMAGE / EKOSOURCE has been advised of the possibility of such damage. This includes damages incurred by the client or any third party.

GEOTERRAIMAGE / EKOSOURCE datasets and products are processed from sources like topographic maps and satellite images. The data and data products are provided “as is”, without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose.

Whilst all possible care and attention will be taken in the production of the defined dataset and report product(s), neither GEOTERRAIMAGE / EKOSOURCE, nor any of their respective sub-organisations or employees, accept any liability whatsoever for any perceived inaccuracies or misrepresentations of the information presented within the digital data or accompanying report, as a result of the nature of land-cover mapping and modelling from coarse and medium resolution satellite imagery. Any image mapped boundaries should not be considered legal boundaries.

NATIONAL WATER QUANTITY DATA : INTELLECTUAL PROPERTY

The National Water Quantity (NWQ) information provided to SANSA is generated using pre-existing data modelling procedures developed and operationalized jointly by GEOTERRAIMAGE SA and EKOSOURCE. As such all Intellectual Property rights associated with these algorithms, models and methods remain at all times with GEOTERRAIMAGE SA and EKOSOURCE.

NATIONAL WATER QUANTITY DATA : USER LICENSE

The NWQ information is provided to SANSA on a monthly basis, covering the contractual period from January 2016 to June 2022 inclusive.

SANSA is granted for the contractual period a non-exclusive, **Multi-User License** to use and distribute the supplied monthly national surface water extent and associated dam volume information in both spatial and non-spatial formats to **all national government departments, provincial and municipal sub-departments, and any of their legally associated entities.**

Both SANSA and the indicated government departments (and associated legal entities) may use this information internally, or provide access to independent 3rd parties specifically undertaking contractual work on behalf of SANSA or any of the indicated government departments.

Neither SANSA, the indicated government departments or any contracted 3rd party may use the supplied information in order to develop a competitive, comparable national, monthly water and volume modelling capability and service, either for South Africa or elsewhere internationally.

At the end of the contractual period the supplied monthly surface water and volume data can be retained by SANSA and the government departments and associated legal entities. All 3rd, non-government parties provided with access to the monthly surface water and volume data must delete all copies of the data from their computer systems and data archives, once the contracted work has been completed. No 3rd party may retain nor use the SANSA or department supplied data for their own use or purpose. It is the responsibility of SANSA to ensure that these conditions are applied to by both departments and any departmentally contracted 3rd parties.

The multi-user license **does not allow public dissemination** of the supplied spatial and non-spatial data.

CONTENTS

LIABILITIES

NATIONAL WATER QUANTITY DATA : INTELLECTUAL PROPERTY

NATIONAL WATER QUANTITY DATA : USER LICENSE

NATIONAL WATER QUANTITY DATA : CITATION FOR PUBLICATION

1. INTRODUCTION

2. GEOGRAPHICAL COVERAGE

3. MONTHLY NATIONAL TOTAL SURFACE WATER EXTENT (AREA)

4. MONTHLY NATIONAL DAM WATER VOLUMES

CORE METADATA ELEMENTS (SANS1878)

ANNEXURE 1A. Monthly Surface Water Extent RASTER Datasets

ANNEXURE 1B. Monthly Surface Water Extent QUATERNARY SUMMARY Datasets

ANNEXURE 2A. Dam Categories Shapefile with Attributes

ANNEXURE 2B. Monthly Water Volumes Datasets for each Dam Category

ANNEXURE 2C. Quaternary Monthly Water Volumes Datasets for Dam Category 2,3 and 4

ANNEXURE 3 ADDITIONAL TECHNICAL INFORMATION TO MONTHLY SURFACE WATER AREA

ANNEXURE 4 ADDITIONAL TECHNICAL INFORMATION TO MONTHLY SURFACE WATER VOLUMES

The recommended citation to use when referencing the use of National Water Quantity Information is:

South African National Water Quantity Monthly Data (2016 – 2022). Monthly ('Mzansi Amanzi') datasets distributed under license by the South African National Space Agency. Monthly surface water and dam volumes data generated by GeoTerraImage SA Pty Ltd and EkoSource Insights Pty Ltd. 2020.
www.water-southafrica.co.za

1. INTRODUCTION

National Water Quantity (NWQ) information is provided to SANSA on a monthly basis, for a contractual period covering January 2016 to June 2022 inclusive. The NWQ information is generated monthly, and provides information on **surface water extent (area)** and **dam water volume**. The monthly information is provided in several GIS compatible spatial data formats and as associated tabulated (spreadsheet) information, namely:

1. spatial and numeric data on **total surface water location and extent** ,
2. spatial and numeric data on **individual *man-made dam volumes***,
3. spatial and numeric data on **summarised total dam volumes per quaternary catchment for the category 2,3 and 4 dams** (Category 1 which are large dams are reported individually as they often cross several quaternary catchments).

The national coverage, monthly **surface water extent (area)** and **dam water volume** information can also be viewed at the following two websites:

<https://www.water-southafrica.co.za>

<http://sbdvc.ekodata.co.za/>

This metadata report refers to the **version 2 data release (July 2021)** of the monthly water information, which contains a significant upgrade to the content of the each monthly dataset as a result of the integrated inclusion of Sentinel-1 radar imagery in the surface water classification methodology. The integrated use of both Sentinel-1 and Sentinel-2 imagery effectively eliminates any cloud-obscured areas in the monthly water mapping (which was a product limitation when using only Sentinel-2 imagery ~ version 1 data). The improved capability to detect and map full water areas in each month, regardless of cloud-cover conditions improves both the accuracy and representation of surface water areas as well as the dam volumes modelled from the observed surface water extents, since the full dam coverage is mapped rather than just cloud-free areas. The full history of the monthly water datasets from January 2016 to present has been re-processed to include this new radar-enhanced capability, and should ***replace all previously supplied version 1 monthly datasets.***

2. GEOGRAPHICAL COVERAGE

The monthly NWQ information and associated spatial data cover the full extent of South Africa within its international borders.

3. MONTHLY NATIONAL TOTAL SURFACE WATER EXTENT (AREA):

The monthly national coverage **total surface water** datasets represents the combined spatial extent of both natural and man-made water features. The monthly datasets cover the historical period 2016-01 to project inception date, and forward until June 2022. Each monthly national surface water extent dataset is delivered in digital, GIS compatible, raster format (i.e. GeoTIFF), based on 20m resolution cells, in Geographic Coordinates. A separate raster file is generated for each month. Water features are mapped from a combination of 20m resolution Sentinel 2 optical imagery and Sentinel 1 synthetic aperture radar (SAR) imagery. This image formats allows high detail identification and mapping of all surface water bodies across South Africa in a consistent, repeatable and accurate manner. All waterbodies typically > 0.25Ha will be included in the monthly datasets.

Each cell in the monthly surface water raster datasets contains only a single information class (1) which represents the surface water observed during the reporting month (based on all accumulated satellite overpass observations in that month). *Note that this differs from the version 1 (original) monthly datasets in that version 1 contained two information classes: (1) surface water observed during the reporting month (based on all accumulated satellite overpass observations in that month), and (2) cloud obscured locations (where cloud cover persisted in all satellite overpass observations in that month).*

The enhanced cloud-free content of version 2 monthly surface water detection compared to cloud-obscured version 1 data content is illustrated in Figure 1.

The integrated inclusion of Sentinel-1 SAR imagery in the monthly water detection modelling has resulted in several new surface water extent data characteristics that end-users should be aware of:

- The raster-based version 2 representation of surface water extent is typically much 'tighter' than the previous outputs associated with version 1 data products. This is because of less 'bleed-out' into very damp (rather than true water covered) areas. This especially noticeable in shallow water environments such as braided channels and/or wetland regions. The version 2 water extent is a more accurate representation of true open water extent. Note however that version 2 processes can result in the loss of some water feature edge pixels which are in reality transitional pixels between water and terrestrial surfaces. This can result in significant narrowing and even discontinuous representations of thin, linear river features; when comparing equivalent date content from version 1 and version 2 monthly water

datasets (see Figure 2). Considerable effort has been made to ensure version 2 contains less false positives than version 1, to improve the classification accuracy further.

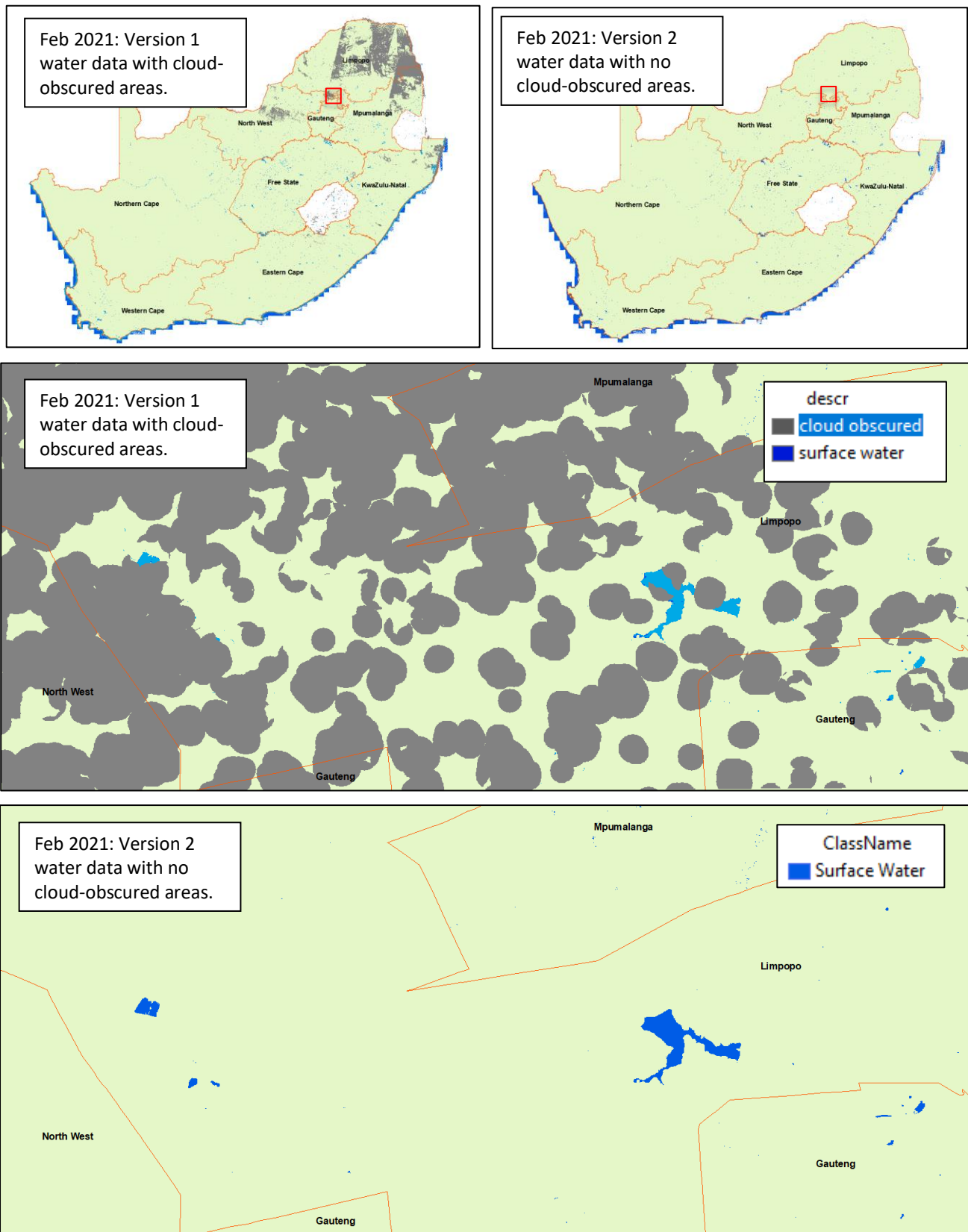


Figure 1. Comparison of integrated Sentinel-1 and Sentinel-2 enhanced monthly national surface water coverage raster data and class content, with no cloud obscured areas; with equivalent date, original Sentinel-2 only cloud obscured monthly surface water dataset (February 2021).

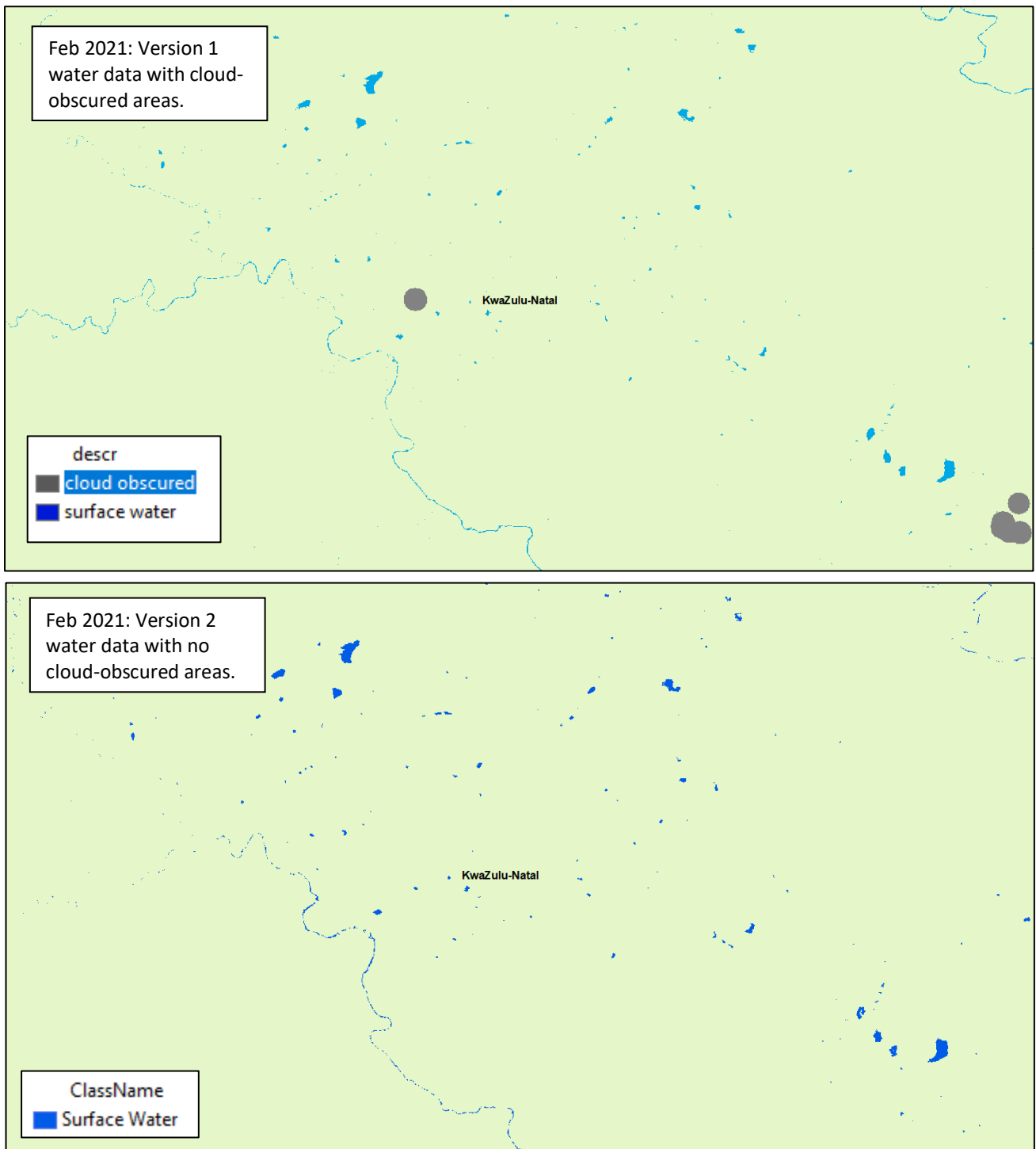


Figure 2. Comparison of narrow river feature representation mapped using integrated Sentinel-1 and Sentinel-2 enhanced modelling (version 2 product), with equivalent date, original Sentinel-2 only modelling (version 1 product), (February 2021).

- The raster-based version 2 representation of surface water extent is typically more accurate in classifying open water areas containing non-emergent aquatic vegetation just below the water surface, which was often mis-interpreted as emergent aquatic vegetation, and thus not part of the water surface

extent, with the Sentinel-2 only outputs associated with version 1 data products. The version 2 water extent is a more accurate representation of true open water extent. Note however that version 2 processes can result in the loss of water edge pixels, and thus narrower representations of linear water features such as rivers, when comparing equivalent date content from version 1 and version 2 monthly water datasets.

- The SAR raster-based classification component in version 2 representation of surface water extent, may result in some local surface water area detection losses as a result of significant surface wave conditions and associated backscatter characteristics, most likely in very large open water, or coastal and adjacent estuary areas. This localized effect is illustrated in Figure 3, which compares Richards Bay, Durban, Knysna and Cape Town harbors, estuaries and adjacent coastlines for the same month (Feb 2021).

Feb 2021: Version 1 water data with cloud-obscured areas.

Feb 2021: Version 2 water data with no cloud-obscured areas.

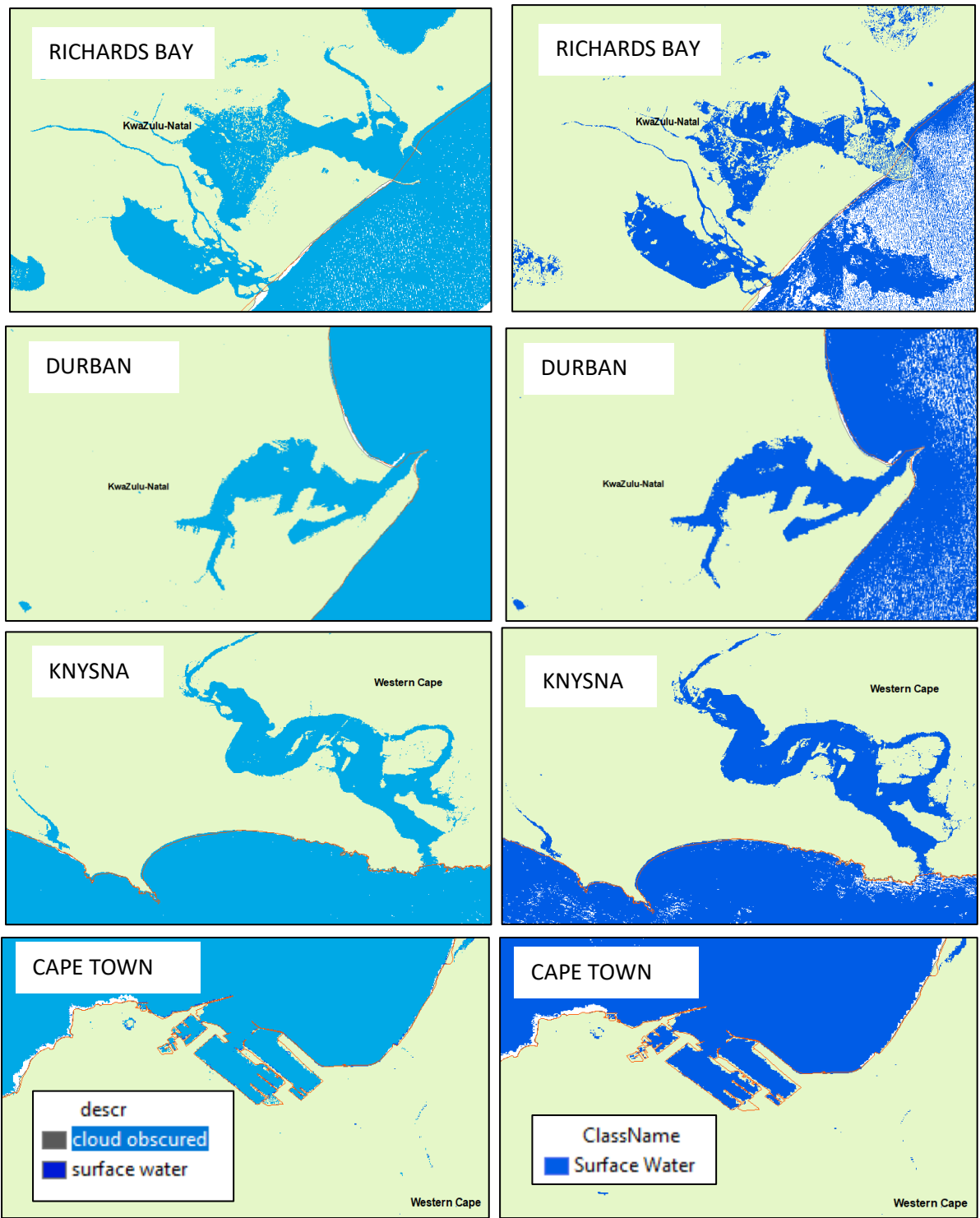


Figure 3. Comparison of coastal water representation mapped using integrated Sentinel-1 and Sentinel-2 enhanced modelling (version 2 product), with equivalent date, original Sentinel-2 only modelling (version 1 product), (February 2021); and localized wave impacts and mapping effects.

In addition to the monthly surface water raster (*tiff) coverages, a separate set of accompanying ESRI vector shapefiles is also provided, that report the summarized / aggregated total area of surface water within each Quaternary Catchment, for each month, since 2016-01. Each monthly shapefile has attributes to quantify the hectares of water area, cloud obscured area and the historical long-term maximum area. The example shapefile attribute shown below illustrates the following attribute content:

- a. QUATERNARY - The quaternary catchment identifier
- b. TERTIARY - The tertiary catchment identifier
- c. SECONDARY - The secondary catchment identifier
- d. QUAT_AREA - The quaternary catchment area size in hectares
- e. water_area - The area of water classified in the catchment for that particular month, in hectares
- f. cld_perc - The percentage of the quaternary catchment affected by cloud coverage within the particular month
- g. max_area - The maximum area of water classified in the quaternary catchment for all recorded months since 2016-01, in hectares

Table									
sansa_rsa_water_2020_06_quat_stats									
FID	Shape	QUATERNARY	TERTIARY	SECONDARY	QUAT_AREA	water_area	cld_perc	max_area	
0	Polygon	A71L	A71	A7	177165.58381	379.96	0	901.76	
1	Polygon	A63E	A63	A6	199695.873302	401.12	0	722.36	
2	Polygon	A71K	A71	A7	167338.369175	75.92	0	226.2	
3	Polygon	A92D	A92	A9	80878.370716	62.6	0	491.52	
4	Polygon	A80G	A80	A8	123318.216852	42.64	0	361.92	
5	Polygon	A80J	A80	A8	87437.59141	122.92	0	539.8	
6	Polygon	A91K	A91	A9	67308.032639	58.84	0	443.4	
7	Polygon	A63C	A63	A6	132946.483004	495.6	0	632.48	
8	Polygon	A91J	A91	A9	57182.918395	25.4	0	79.68	
9	Polygon	A63D	A63	A6	131941.689943	49.28	0	51.96	
10	Polygon	A92C	A92	A9	45593.431149	21.4	0	20.28	
11	Polygon	A72B	A72	A7	155546.993597	12.24	0	29.8	
12	Polygon	A92B	A92	A9	56653.99448	8.44	0	17.32	
13	Polygon	A63B	A63	A6	150467.191422	131.4	0	93.44	
14	Polygon	B90A	B90	B9	69478.225833	2.6	0	4.6	
15	Polygon	A71J	A71	A7	116295.254303	52.08	0	80.24	
16	Polygon	A80H	A80	A8	26599.210038	180.8	0	168.96	
17	Polygon	A91H	A91	A9	45126.808329	106.24	0	144.4	
18	Polygon	A80F	A80	A8	63061.618259	2.64	0	9.28	
19	Polygon	A50J	A50	A5	125814.959722	337.68	0	389.84	
20	Polygon	B90E	B90	B9	47516.835451	0.24	0	0.72	
21	Polygon	A80C	A80	A8	29414.670069	517.36	0	469.44	
22	Polygon	A92A	A92	A9	32960.931103	137.64	0	148.52	
23	Polygon	B90B	B90	B9	75597.15346	258.2	0	266.12	
24	Polygon	A91G	A91	A9	40676.473897	353.92	0	360.08	
25	Polygon	A80B	A80	A8	25173.389139	7.4	0	15.32	
26	Polygon	A80E	A80	A8	24732.713427	4.32	0	25.44	
27	Polygon	B90D	B90	B9	44813.944324	0.56	0	7.32	
28	Polygon	A80A	A80	A8	28791.316823	49.08	0	56.6	
29	Polygon	A91F	A91	A9	58124.199321	1544.84	0	1640.64	
30	Polygon	A80D	A80	A8	12786.346992	1.08	0	2.16	
31	Polygon	A71H	A71	A7	101346.676157	45.64	0	106.16	
32	Polygon	B90C	B90	B9	53595.868029	3.6	0	12	
33	Polygon	B90H	B90	B9	89184.196632	6.16	0	38.76	
34	Polygon	A72A	A72	A7	190848.472324	8.88	0	36.48	
35	Polygon	A63A	A63	A6	192741.272997	31.36	0	30.68	
36	Polygon	A50H	A50	A5	194721.237548	358.8	0	422.72	

Metadata for the national total surface water monthly raster data is contain in Annexure 1a. Metadata for the Quaternary Catchment surface water area monthly summaries dataset is contained in Annexure 1b.

Additional background information on the image data modelling procedures used to generate the monthly national surface water is contained in Annexure 2a.

4. MONTHLY NATIONAL DAM WATER VOLUMES

Deriving the information

The surface water area information generated by the process described in section 3 is used to generate an estimate of the monthly National Dam Water Volumes. This information is only generated for the man-made water bodies in a catchment. The reason for this as the main distinction is that the product is designed to provide an estimate of the usable water supply volume in a particular area, for this reason natural water bodies in South Africa have been excluded as they tend to be of a temporary nature and in most cases are not sustainably exploitable for water supply. In order to obtain this information the following elements have been undertaken.

Man-made versus natural water: A mask has been generated which encompasses all the water supply dams identified across South Africa. This mask distinguishes between natural and man made water bodies. Each water body has been identified and buffered according to a formula. This is to ensure if a dam is above the identified full supply level/area that we are still able to capture the volume. The mask has been separated into 4 different categories, namely.

- i. All large dams already monitored by DWS are Category 1 dams
- ii. All dams that are on the Dam Safety Database are classified as Category 2 dams
- iii. All dams remaining dams larger than 25 000 m³ are Category 3 dams
- iv. All the remaining dams that are considered man made are Category 4 dams

Equations: A set of equations have been developed to determine the surface water volume from the surface water areas identified on the satellite. An equation has been developed for each dam identified as man-made across the country. The method of deriving these equations is different for the different classes of dams. This process is described in detail in the paper written for this purpose (Thompson et al. 2018). The equations were determined using the following methodology for each of the different dam categories

- **Category 1 dams** already had known Height Volume Area (HVA) equations generally done from various topography surveys prior to construction or alternatively, through bathymetry surveys later. These were converted into equations and used for the satellite information. Several of these have been found to be slightly inaccurate and adjustments have been made to the tables.
- **Category 2 dams**, some of the larger category 2 dams had existing HVA tables from which equations were derived. Some of the dams have had their own Volume Area (VA) equations derived for them either using topographic data or other derived equations at a quaternary catchment level.
- **Category 3 dams**, these dams have equations that were generated at a quaternary catchment level and assigned to the individual dams. It has been noted that certain of these dams need to be added to the dams safety database from a size perspective. The catchment based VA equations have been generated for dams per quaternary at two different levels of accuracy. The first level is at an individual catchment level where an equation has been derived for dams in those catchments using topography. The second level of accuracy is for dams that have had their equations derived through a cluster analysis using various geomorphological attributes in an area.
- **Category 4 dams**, have had their equations determined in the same way the category 3 dams have had their VA equations derived.

Derived volume estimates: The derived volume estimates have been generated from the process where for each individual dam the equation is used along with the monthly area estimates to determine the monthly volume estimates, this information has also being used to derive the full supply volume for these particular dams. Thus FLC information plus the present supply volume as a percentage of the FLC is provided for each of the dams.

Estimates of accuracy and uncertainty: There are several different elements which can impact the accuracy of the volume estimates, these include:

- i. Errors of commission (Water surface identified but does not exist)
- ii. Errors of omission (Water surface identified as something else)
- iii. Fractal or pixilation errors resulting from the pixel size where water is identified
- iv. Bias in determining Volume Area equations
- v. Uncertainty in deriving single catchment equations
- vi. Errors in curve fitting VA equations

Catchment based water resources availability: the combination of the information provided in this particular product are thus able to provide us with an assessment of the overall water availability in a catchment. The idea is that we are able to calculate the current volume of water against the full supply capacity of water resources in an area and thus obtain an estimate of the overall supply in an area. These assessments have been done at a quaternary catchment level. There are two sets of information provided here, namely:

- i. Category 1 dams: the percentage full and actual volumes are supplied
- ii. Category 2,3 and 4 dams are all lumped together at a quaternary catchment level to provide an overview of the percentage of full supply capacity as well as the estimated volumes for those dams.

Information provided

A set of shapefiles has been provided with all the dams throughout the country identified as man-made water resources. In the shapefiles buffered polygons of all the dam features are provided. A shapefile is provided for each of the category 1, 2, 3 and 4 dams individually. In the shapefiles each dam has been given a unique ID and has a number of other attributes associated with it that are populated according to the dam safety database and are assigned to specific property boundaries. Each dam also has a Full Supply Capacity level and an error estimate is provided from a single dam entity and a catchment perspective. In the shapefile an estimate of the current monthly volume is provided. Additionally, there is further information which enables a history of the dam volumes to be calculated and attached to the dam's unique ID, enabling this to be provided for each dam since 2016.

An additional set of information is provided which summarises the above information into a set of shapefiles at a quaternary catchment level. This information is essentially a summation of the individual dam information found in categories 1, 2, and 3 dams.

The individual shapefiles contain the following information:

- i. Ekosource_Ddam_SHP_with_Attributes: Contains the shapefile with what we believe is all the manmade water bodies across South Africa, with a set of location and name associated attributes as defined in the bid submission. These waterbodies have been buffered.
- ii. EkoSource_Historical_Monthly_Volumes: contains all the dams in 4 different shape files, each with the historical volume attribute for each dam polygon (January 2016 – April 2020). Each attribute with “ve” suffix represents a monthly volume of the individual dam or a quaternary. The format of

the heading is (YYYY_MM_ve). The CL suffix provides an indication if there was cloud cover present during the month. This would make the estimate inaccurate as cloud cover prevents a good area estimate from being derived. This is illustrated below:

2016_01_cl	2016_02_ve	2016_02_cl	2016_03_ve	2016_03_cl	2016_04_ve	2016_04_cl	2016_05_ve	2016_05_cl
0	2.441971	1	19.039777	0	19.068594	0	18.531647	0
0	0	1	130.941909	0	121.803476	0	112.824074	0
0	31.263313	0	28.651838	0	27.923864	0	26.757858	0
0	1.832843	0	1.613308	0	1.507844	0	1.413315	0
0	2.820053	0	2.353135	0	1.872041	0	0.053418	0
0	0.317459	0	0.365806	0	0.330164	0	0.302355	0
0	3.438583	0	3.530632	0	3.836635	0	3.993699	0
0	69.687871	1	0	1	38.763262	0	41.246513	0
1	0	1	48.19875	0	48.286397	0	41.496017	1
1	264.162964	0	193.051126	1	231.198433	0	239.731269	0
0	31.638926	0	35.469013	0	44.202373	0	42.196803	0
0	4.172752	1	5.853554	0	5.852452	0	5.712054	0
0	2.860772	0	3.046511	0	2.322379	0	2.69046	0
0	12.222947	0	12.465413	0	14.274369	0	15.12199	0
0	14.659689	0	12.58138	0	13.487197	0	12.880542	0
0	191.533748	1	0	1	108.942145	0	103.386647	1
1	15.501515	1	98.482874	0	95.853512	0	91.364192	0
0	43.371664	0	33.29202	1	45.427384	0	44.983798	0
0	3.400253	0	3.358439	0	3.30121	0	0.825867	0
0	0.014652	1	1.995275	0	2.241422	0	2.082398	0
0	7.565889	0	8.49027	0	8.280603	0	8.174459	0
0	3.387514	0	3.136878	0	2.876611	0	3.013702	0
0	4.43014	1	4.625936	0	4.553279	0	3.210474	1
0	4.632275	0	4.475994	0	4.194924	0	3.950581	0
0	2.684881	0	2.209388	0	2.136706	0	2.659692	0
0	34.430253	0	0	1	31.053951	0	31.044478	0
0	37.363001	1	27.527686	0	16.50939	0	18.845738	0
0	0	1	0.388482	0	0.410507	0	0.391215	0
0	0	1	17.916442	0	16.237136	0	14.832894	0
0	0	1	0	1	0	1	7.594858	0
0	1.085851	0	1.405104	0	1.533609	0	1.237191	0
0	47.413671	0	0	1	75.5809	0	78.154636	0
0	4.263083	0	0.014952	1	3.503456	0	3.5641	0
0	296.266672	0	312.815815	0	310.941631	0	305.353374	0
0	1.390717	0	1.400106	0	1.41741	0	1.411104	0
1	0	1	5.826358	0	5.855798	0	4.455849	1

- iii. EkoSource_Historical_Quaternary_Summaries: is the same information but instead of for individual dams it is per quaternary catchment; and this is however only done for category 2,3 and 4 dams. The dataset covers the historical period January 2016 – April 2020. The reason category 1 dams are not included in this information is that they overlap and cover in many cases more than 1 quaternary catchment, and thus the summaries would be meaningless. Again the shapefile has the same information as provided above but for lumped Quaternary catchment information. An additional file that lumps all the category 2,3 and 4 dams together is also provided.
- iv. The two monthly folders contain individual monthly information from May 2020 going forward.

ANNEXURE 1A

Monthly Surface Water Extent RASTER Datasets

1(M) **Dataset title:** sansa_rsa_water_2021_06_mosaic_v2.tif

2(M) **Dataset reference date:** See date (format YYYY_MM) in the dataset title

3(O) **Dataset responsible party:** Prepared by GeoTerra Image (GTI) Pty Ltd (Mark Thompson, www.geoterraimage.com) for SANSA, South Africa.

4(C) **Geographic location of the dataset.**

WestBoundLongitude: 16.406471023 (Upper Left X)
EastBoundLongitude: 32.9424791101 (Lower Right X)
NorthBoundLongitude: -22.0787930531 (Upper Left Y)
SouthBoundLongitude: -34.8785282104 (Lower Right Y)

Projection coordinates based on Latitude & Longitude (Geographic), based on WGS84 Spheroid and Datum. Raster cell size is 0.00019766 degrees representing 20 x 20 meters.

5(M) **Dataset language :** "English" (eng)

6(C) **Dataset character set:** UTF8 (8-bit data)

7(M)**Dataset topic category:** 010 = Base Map earth coverage

8(O) **Scale of the dataset:** Surface water information has been modelled from 20m resolution Sentinel 2 imagery, therefore recommended for $\pm 1:75,000$ scale or coarse mapping & modeling applications.

9(M) **Abstract describing the dataset:** The monthly surface water datasets cover the period January 2016 to contract end date, and have been generated by GTI using internal, bespoke water detection algorithms developed by GTI specifically for Sentinel imagery. Each monthly surface water dataset is provided in digital, GIS compatible GeoTIFF raster format, based on Geographic Coordinates. Each raster cell represents 0.00017966 x 0.00017966 degrees, which is equivalent to the 20m resolution of the Sentinel 2 source imagery. The detected surface water class represents the modelled median surface water extent for the assessment month, derived from the analysis of all Sentinel 1 & 2 satellite image acquisitions in the assessment month.

10(O) **Dataset format name:** GeoTiff *.tif raster formats

11(O) **Dataset format version:** version 1

12(O) **Additional extent information for the dataset: (vertical and temporal)**

Vertical Extent:

Minimum Value: n/a
Maximum Value: n/a
Unit Of Measure: n/a

Vertical Datum: n/a

Temporal Extent: Datasets are prepared monthly and represent monthly surface water coverage for the period January 2016 to contract end date.

14(O) **Reference system:** Geographic Coordinates (Lat / Lon), WGS84

CRS:

Projection Used: Geographic Coordinates (Lat / Lon)
Spheroid used: WGS84
Datum used: WGS 84

Ellipsoid parameters:

Ellipsoid semimajor axis
axis units
denominator of flattening ratio

Projection Parameters:

Standard parallel: n/a

Longitude of central meridian: East n/a

Latitude of projection origin: East n/a

False easting: meters n/a

False northing: meters n/a

Scale factor at equator: n/a

Projection units: degrees (0.00017966 degree raster, representing 20 x 20 meter units)

15(O) **Lineage statement:** The monthly surface water datasets are generated from Sentinel 1 and 2 imagery, using imagery originally sourced ingested from scihub.copernicus.eu (ESA's cloud image archive repository) by GTI via an analytical cloud-platform using internal, bespoke water detection algorithms developed by GTI specifically for Sentinel imagery.

16(O) **On-line resource:** n/a

17(O) **Metadata file identifier:** n/a

18(O) **Metadata standard name:** SANS I878

19(O) **Metadata standard version:** version 01

20(C) **Metadata language:** English (eng)

21(C) **Metadata character set:** 021 (UsAscii)

22(M) **Metadata point of contact:**

Name: Ms Nomnikelo Bongoza or Ms Thando Oliphant
Organisation Name: South African National Space Agency (SANSA), Directorate: Earth Observation
Tel: +27 12 844 0500
Fax: +27 12 844 0396

Email: customers-ee@sansa.org.za

Physical Address:

Building	Enterprise Building
Street	Mark Shuttleworth Street
Suburb	Innovation Hub
City	Pretoria
Zip	0087
State	Gauteng
Country	South Africa

Postal Address :

PO Box	484
Suburb	Silverton
City	Pretoria
Zip	0127
State	Gauteng
Country	South Africa

GPS: S25 53 15.7 E27 42 43.2

23(M) **Metadata time stamp:** June 2021

ANNEXURE 1B

Monthly Surface Water Extent QUATERNARY SUMMARY Datasets

1(M) **Dataset title:** sansa_rsa_water_2021_06_quat_stats_v2.shp

2(M) **Dataset reference date:** See date (format YYYY_MM) in the dataset title

3(O) **Dataset responsible party:** Prepared by GeoTerra Image (GTI) Pty Ltd (Mark Thompson, www.geoterraimage.com) for SANSA, South Africa.

4(C) **Geographic location of the dataset.**

WestBoundLongitude: 16.456949 (Upper Left X)

EastBoundLongitude: 32.892281 (Lower Right X)

NorthBoundLongitude: -22.123983 (Upper Left Y)

SouthBoundLongitude: -34.833454 (Lower Right Y)

Projection coordinates based on Latitude & Longitude (Geographic), based on WGS84 Spheroid and Datum.

5(M) **Dataset language :** “English” (eng)

6(C) **Dataset character set:** UTF8 (8-bit data)

7(M) **Dataset topic category:** 010 = Base Map earth coverage

8(O) **Scale of the dataset:** Surface water information has been modelled from 20m resolution Sentinel 2 imagery, and then classified water areas have been aggregated to a quaternary catchment level. The scale of the dataset is at quaternary catchment level (1:500 000).

9(M) **Abstract describing the dataset:** The monthly surface water datasets cover the period January 2016 to contract end date, and have been generated by GTI using internal, bespoke water detection algorithms developed by GTI specifically for Sentinel imagery. Each monthly surface water dataset is provided in digital, GIS compatible GeoTIFF raster format, based on Geographic Coordinates. Each raster cell represents 0.00017966 x 0.00017966 degrees, which is equivalent to the 20m resolution of the Sentinel 2 source imagery. The detected surface water class represents the modelled median surface water extent for the assessment month, derived from the analysis of all Sentinel 1 and 2 satellite image acquisitions in the assessment month. Using aggregation spatial analysis technique the raster dataset is summarized and aggregated for each quaternary catchment area. The final shapefile dataset reports the “area of water in Hectare units”, the “area of cloud obscured areas impacting the catchment”, and the “maximum ever recorded water extent per catchment”.

10(O) **Dataset format name:** ESRI (*.shp) Shapefile formats

11(O) **Dataset format version:** version 1

12(O) **Additional extent information for the dataset: (vertical and temporal)**

Vertical Extent:

Minimum Value: n/a

Maximum Value: n/a

Unit Of Measure: n/a

Vertical Datum: n/a

Temporal Extent: Datasets are prepared monthly and represent monthly surface water coverage for the period January 2016 to contract end date, summarized to quaternary catchment level.

14(O) **Reference system:** Geographic Coordinates (Lat / Lon), WGS84

CRS:

Projection Used: Geographic Coordinates (Lat / Lon)

Spheroid used: WGS84

Datum used: WGS 84

Ellipsoid parameters:

Ellipsoid semimajor axis

axis units

denominator of flattening ratio

Projection Parameters:

Standard parallel: n/a

Longitude of central meridian: East n/a

Latitude of projection origin: East n/a

False easting: meters n/a

False northing: meters n/a

Scale factor at equator: n/a

Projection units: degrees

15(O) **Lineage statement:** The monthly surface water datasets are generated from Sentinel 1 and 2 imagery, using imagery originally sourced ingested from scihub.copernicus.eu (ESA's cloud image archive repository) by GTI via an analytical cloud-platform using internal, bespoke water detection algorithms developed by GTI specifically for Sentinel imagery.

16(O) **On-line resource:** n/a

17(O) **Metadata file identifier:** n/a

18(O) **Metadata standard name:** SANS I878

19(O) **Metadata standard version:** version 01

20(C) **Metadata language:** English (eng)

21(C) **Metadata character set:** 021 (UsAscii)

22(M) **Metadata point of contact:**

Name: Ms Nomnikelo Bongoza or Ms Thando Oliphant
Organisation Name: South African National Space Agency (SANSA), Directorate: Earth Observation
Tel: +27 12 844 0500
Fax: +27 12 844 0396
Email: customers-eo@sansa.org.za

Physical Address:

Building Enterprise Building
Street Mark Shuttleworth Street
Suburb Innovation Hub
City Pretoria
Zip 0087
State Gauteng
Country South Africa

Postal Address :

PO Box 484
Suburb Silverton
City Pretoria
Zip 0127
State Gauteng
Country South Africa

GPS: S25 53 15.7 E27 42 43.2

23(M) **Metadata time stamp:** June 2021

ANNEXURE 2A

Dam Categories Shapefile with attributes

1 **Dataset title:** Cat1_DWS_Dams_FINAL_dd_v2.shp, Cat2_Dams_FINAL_dd.shp_v2, Cat3_Dams_FINAL_dd_v2.shp, Cat4_Dams_FINAL_dd_v2.shp

2 **Dataset reference date:** 2020_06 (format YYYY_MM)

3 **Dataset responsible party:** Prepared by Ekosource Insight Pty Ltd (Jason Hallowes, <https://www.ekosource.co.za/>) for SANSA, South Africa.

4 **Geographic location of the dataset.**

WestBoundLongitude: 18.778095 (Upper Left X)
EastBoundLongitude: 32.181910 dd (Lower Right X)
NorthBoundLongitude: -22.632792 (Upper Left Y)
SouthBoundLongitude: -34.369669 (Lower Right Y)

Projection coordinates based on Latitude & Longitude (Geographic), based on WGS84 Spheroid and Datum.

5 **Dataset language:** "English" (eng)

6 **Dataset character set:** UTF8 (8-bit data)

7 **Dataset topic categories:**

8 **Scale of the dataset:** The category 1 dam shapefile has been derived from DWS dam shapefiles, category 2,3 and 4 from the water surface coverage that is derived from Sentinel 2 imagery. The scale of the dataset is at quaternary catchment level (1:500 000).

9(M) **Abstract describing the dataset:** This dataset represents the best estimate of the man-made water bodies across South Africa. Each water body has been recorded and a buffer has been added to the full supply spatial extent to ensure that flooded water flows are recorded. The information provided has similar information to that supplied by Department of Water and Sanitation dams safety dataset. Each element has been populated with as much data as possible. The dams have also been setup with a unique ID which is based on which quaternary catchment they are in and the spatial position.

10 **Dataset format name:** ESRI (*.shp) Shapefile formats

11 **Dataset format version:** version 1

12 **Additional extent information for the dataset: (vertical and temporal)**

Vertical Extent:

Minimum Value: n/a
Maximum Value: n/a
Unit Of Measure: n/a

Vertical Datum: n/a

Temporal Extent: No temporal extent for these datasets.

14(O) **Reference system:** Geographic Coordinates (Lat / Lon), WGS84

CRS:

Projection Used: Geographic Coordinates (Lat / Lon)

Spheroid used: WGS84

Datum used: WGS 84

Ellipsoid parameters:

Ellipsoid semimajor axis

axis units

denominator of flattening ratio

Projection Parameters:

Standard parallel: n/a

Longitude of central meridian: East n/a

Latitude of projection origin: East n/a

False easting: meters n/a

False northing: meters n/a

Scale factor at equator: n/a

Projection units: degrees

15(O) **Lineage statement:** The shapefile show that dam list and a number of attributes in the category 1,2,3 and 4.

16(O) **On-line resource:** n/a

17(O) **Metadata file identifier:** n/a

18(O) **Metadata standard name:** SANS I878

19(O) **Metadata standard version:** version 01

20(C) **Metadata language:** English (eng)

21(C) **Metadata character set:** 021 (UsAscii)

22(M) Metadata point of contact:

Name: Ms Nomnikelo Bongoza or Ms Thando Oliphant

Organisation Name: South African National Space Agency (SANSA), Directorate: Earth Observation

Tel: +27 12 844 0500

Fax: +27 12 844 0396

Email: customers-eo@sansa.org.za

Physical Address:

Building	Enterprise Building
Street	Mark Shuttleworth Street
Suburb	Innovation Hub
City	Pretoria
Zip	0087
State	Gauteng
Country	South Africa

Postal Address :

PO Box	484
Suburb	Silverton
City	Pretoria
Zip	0127
State	Gauteng
Country	South Africa

GPS: S25 53 15.7 E27 42 43.2

23(M) Metadata time stamp: June 2021

ANNEXURE 2B

Monthly Water Volumes Datasets for each Dam Category

1 **Dataset title:** Cat1_DWS_Dams_May2020_Volumes_v2.shp, Cat2_Dams_May2020_Volumes_v2.shp, Cat3_Dams_May2020_Volumes_v2.shp, Cat4_Dams_May2020_Volumes_v2.shp

2 **Dataset reference date:** See date (format Month_YYYY) in the dataset title

3 **Dataset responsible party:** Prepared by Ekosource Insight Pty Ltd (Jason Hallowes, <https://www.ekosource.co.za/>) for SANSA, South Africa.

4 **Geographic location of the dataset.**

WestBoundLongitude: 18.778095 (Upper Left X)
EastBoundLongitude: 32.181910 dd (Lower Right X)
NorthBoundLongitude: -22.632792 (Upper Left Y)
SouthBoundLongitude: -34.369669 (Lower Right Y)

Projection coordinates based on Latitude & Longitude (Geographic), based on WGS84 Spheroid and Datum.

5 **Dataset language:** "English" (eng)

6 **Dataset character set:** UTF8 (8-bit data)

7 **Dataset topic categories:**

8 **Scale of the dataset:** The monthly water volume for category 1 dam calculated using the equation developed from corrected DWS HVA relationship and the category 2,3 and 4 monthly volumes are calculated using the equation developed using Sentinel 2 imagery. The scale of the dataset is at quaternary catchment level (1:500 000).

9(M) **Abstract describing the dataset:** The monthly surface water volume datasets cover the period January 2016 to contract end date, and have been generated by GTI and EkoSource using internal, bespoke water detection algorithms developed by GTI specifically for Sentinel imagery and a set of catchment specific volume/area estimate equations calculated using catchment specific topography. Each monthly surface water dataset is provided in digital, GIS compatible shapefile format, based on Geographic Coordinates. Each polygon represents a buffered dam that has been derived from numerous sources including imagery, maps, DWS geodata and more. The detected surface water class represents the modelled median surface water volumes for the assessment month, derived from the analysis of all Sentinel 2 satellite image acquisitions in the assessment month. The final shapefile dataset reports the "volume of water", the "whether it is cloud obscured", "the expected errors" and the "volume recorded" per dam polygon.

10(O) **Dataset format name:** ESRI (*.shp) Shapefile formats

11(O) **Dataset format version:** version 1

12(O) **Additional extent information for the dataset: (vertical and temporal)**

Vertical Extent:

Minimum Value: n/a

Maximum Value: n/a
Unit Of Measure: n/a

Vertical Datum: n/a

Temporal Extent: Monthly

14(O) **Reference system:** Geographic Coordinates (Lat / Lon), WGS84

CRS:

Projection Used: Geographic Coordinates (Lat / Lon)

Spheroid used: WGS84

Datum used: WGS 84

Ellipsoid parameters:

Ellipsoid semimajor axis

axis units

denominator of flattening ratio

Projection Parameters:

Standard parallel: n/a

Longitude of central meridian: East n/a

Latitude of projection origin: East n/a

False easting: meters n/a

False northing: meters n/a

Scale factor at equator: n/a

Projection units: degrees

15(O) **Lineage statement:** The monthly dam water volume datasets are calculated based on the equations developed that show the relationship between volume and area. The surface water areas are generated from Sentinel 2 imagery, using accessed from Google Earth Engine by GTI using internal, bespoke water detection algorithms developed by GTI specifically for Sentinel imagery. Statistics for the water volume are reported at individual dam level.

16(O) **On-line resource:** n/a

17(O) **Metadata file identifier:** n/a

18(O) **Metadata standard name:** SANS I878

19(O) **Metadata standard version:** version 01

20(C) **Metadata language:** English (eng)

21(C) **Metadata character set:** 021 (UsAscii)

22(M) **Metadata point of contact:**

Name: Ms Nomnikelo Bongoza or Ms Thando Oliphant
Organisation Name: South African National Space Agency (SANSA), Directorate: Earth Observation
Tel: +27 12 844 0500
Fax: +27 12 844 0396
Email: customers-eo@sansa.org.za

Physical Address:

Building Enterprise Building
Street Mark Shuttleworth Street
Suburb Innovation Hub
City Pretoria
Zip 0087
State Gauteng
Country South Africa

Postal Address :

PO Box 484
Suburb Silverton
City Pretoria
Zip 0127
State Gauteng
Country South Africa

GPS: S25 53 15.7 E27 42 43.2

23(M) **Metadata time stamp:** June 2021

ANNEXURE 2C

Quaternary Monthly Water Volumes Datasets for Dam Category 2,3 and 4

1 **Dataset title:** Monthly_Quaternary_Volume_Summary_May2020_v2.shp, Cat2_Quaternary_Monthly_Volume_Summary_May2020_v2.shp, Cat3_Quaternary_Monthly_Volume_Summary_May2020_v2.shp, Cat4_Quaternary_Monthly_Volume_Summary_May2020_v2.shp

2 **Dataset reference date:** See date (format Month_YYYY) in the dataset title

3 **Dataset responsible party:** Prepared by Ekosource Insight Pty Ltd (Jason Hallowes, <https://www.ekosource.co.za/>) for SANSA, South Africa.

4 **Geographic location of the dataset.**

WestBoundLongitude: 18.778095 (Upper Left X)
EastBoundLongitude: 32.181910 dd (Lower Right X)
NorthBoundLongitude: -22.632792 (Upper Left Y)
SouthBoundLongitude: -34.369669 (Lower Right Y)

Projection coordinates based on Latitude & Longitude (Geographic), based on WGS84 Spheroid and Datum.

5 **Dataset language:** "English" (eng)

6 **Dataset character set:** UTF8 (8-bit data)

7 **Dataset topic categories:**

8 **Scale of the dataset:** The monthly summarized Quaternary water volume for dam category 2,3, and 4 are summarized from the monthly water volume of the categories. The scale of the dataset is at quaternary catchment level (1:500 000).

9(M) **Abstract describing the dataset:** The information presented in this particular shapefile is the summation of all the dam volume data associated with individual dams in appendix 2.a to a quaternary catchment level. The final shapefile dataset reports the "volume of water in m³ per catchment", the "the percentage of water compared to full supply capacity", "the error estimates at a quaternary catchment level", and the "the full supply capacity".

10(O) **Dataset format name:** ESRI (*.shp) Shapefile formats

11(O) **Dataset format version:** version 1

12(O) **Additional extent information for the dataset: (vertical and temporal)**

Vertical Extent:

Minimum Value: n/a
Maximum Value: n/a
Unit Of Measure: n/a

Vertical Datum: n/a

Temporal Extent: Monthly

14(O) **Reference system:** Geographic Coordinates (Lat / Lon), WGS84

CRS:

Projection Used: Geographic Coordinates (Lat / Lon)

Spheroid used: WGS84

Datum used: WGS 84

Ellipsoid parameters:

Ellipsoid semimajor axis

axis units

denominator of flattening ratio

Projection Parameters:

Standard parallel: n/a

Longitude of central meridian: East n/a

Latitude of projection origin: East n/a

False easting: meters n/a

False northing: meters n/a

Scale factor at equator: n/a

Projection units: degrees

15(O) **Lineage statement:** The summarized monthly water volumes datasets for category 2,3 and 4 are summation of dam volumes of the category 2,3 and 4. Statistics for the water volume are reported at quaternary catchment level.

16(O) **On-line resource:** n/a

17(O) **Metadata file identifier:** n/a

18(O) **Metadata standard name:** SANS I878

19(O) **Metadata standard version:** version 01

20(C) **Metadata language:** English (eng)

21(C) **Metadata character set:** 021 (UsAscii)

22(M) Metadata point of contact:

Name: Ms Nomnikelo Bongoza or Ms Thando Oliphant

Organisation Name: South African National Space Agency (SANSA), Directorate: Earth Observation

Tel: +27 12 844 0500

Fax: +27 12 844 0396

Email: customers-eo@sansa.org.za

Physical Address:

Building	Enterprise Building
Street	Mark Shuttleworth Street
Suburb	Innovation Hub
City	Pretoria
Zip	0087
State	Gauteng
Country	South Africa

Postal Address :

PO Box	484
Suburb	Silverton
City	Pretoria
Zip	0127
State	Gauteng
Country	South Africa

GPS: S25 53 15.7 E27 42 43.2

23(M) Metadata time stamp: June 2021

ANNEXURE 3

ADDITIONAL TECHNICAL INFORMATION TO MONTHLY SURFACE WATER AREA

Water is a critical and scarce resource in regions such as 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, which means that it is difficult to perform assessments effectively. Satellite data can now be considered a reliable and robust source of regular and repeatable data that can provide water resource information at a level of accuracy that can support many water resource assessments in an operational manner. The advantage of such a system is that it provides extremely cost-efficient, desk-top based access to reliable monthly data for all surface water bodies across geographically large target areas, without reliance on expensive in-situ based monitoring equipment; with the associated risks and costs of instrumentation installation, maintenance, and possible loss.

GeoTerraImage (GTI), in collaboration with EkoSource, has developed a unique, operational water resource monitoring service, that provides, on a monthly basis, detailed, wide area coverage inventory of all surface water features. This capability has been developed using cloud-based technologies, such as cloud-based (satellite) image archives and cloud-based computing facilities. These support the necessary big data analyses and rapid output generation required for such a water monitoring service. The service is based on an advanced machine learning and deep learning classification approach that uses a combination of European Space Agency (ESA) Sentinel-2 optical imagery and Sentinel-1 synthetic aperture radar (SAR) imagery. The service collectively facilitates a all-weather water feature detection, regardless of cloud cover. The broad approach is to use Sentinel-2 imagery as the primary water detection dataset, with the all-weather capabilities of Sentinel-1 SAR imagery being used to “fill-in” cloud obscured water surfaces. This approach ensures that the high spatial detail and accuracy of optically detected surface water areas from the 20m resolution Sentinel-2 imagery is retained as the primary controller of output content wherever possible. These surface water detection capabilities support the identification of water features typically > 0.25 ha to be individually identified and mapped on a repeatable and accurate manner.

This monthly surface water monitoring capability is currently used to provide a national, wall-to-wall monthly information service on the state of all surface water resources across South Africa, which is accessible as a web-based tool (www.water-southafrica.co.za). The South African-focussed 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, each month, in support of activities such as compliance monitoring, water security, water licence applications and water resource inventory.

The monthly surface water datasets represent the *median surface water extent for that month*, rather than the average or (absolute) maximum extent, as a result of the multi-date image acquisition date compositing approach used to modelling water features. This approach is strongly influenced by the timing and intensity of rainfall within a given month. For example, the effect of single, significant rainfall event occurring at the very end of a given month is only likely to be reflected in the median surface water extent reported for the subsequent month, and not the month within which the rainfall event occurred.

Surface Water Detection Modelling Principles.

The ‘target’ month’s total surface water represents the median 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 that integrates all available imagery for the target ‘month’.

A ‘target month’ is defined as a full calendar month, plus the two preceding weeks. This 6-week window allows the potential acquisition of up to 10 Sentinel-2 images, as a result of the 5-day revisit cycle of the Sentinel-2A and 2B satellite constellation, which is the primary dataset for water detection. The same data acquisition window format is also used for Sentinel-1 data.

The modelling and detection of surface water extents is based on a multi-step process:

1. Generation of a Sentinel-2 only, multi-date image composition representing maximum cloud free content from all image acquisition dates in the ‘target’ month window. The median value for each image pixel (per spectral band), from all image acquisition dates within the six-week window¹ is constructed,

¹ The current Sentinel 2 programme consists of a constellation of two satellites, namely Sentinel 2A and 2B. Working in tandem, Sentinel 2A and 2B are able provide image updates every 5-days (i.e. 4 – 5 acquisitions per 4 week monthly period), anywhere across the globe. This high frequency update cycle supports the capability to generate cloud-free image composites from which the monthly surface water datasets, which are not impacted by loss of information in cloud-obscured areas. Prior to the launch of Sentinel 2B in late 2017, when only Sentinel 2A was in operation, the update cycle was every 10-days. This unfortunately means that the possibility of cloud-obscured areas in each monthly image composite is higher for the period start 2016 – late 2017; than compared to post late-2017 when both Sentinel 2A and 2B were operational. Since the actual image acquisition window used to generate each cloud-free image composite is not based on individual calendar months but rather weekly windows, it is possible to acquire up to ten image acquisitions in the period used to represent each ‘month’; where each ‘month’ consists of the actual ‘calendar’ month plus the preceding 2 weeks. This slightly extended period is used to ensure as far as possible, at least one cloud-free observations is possible for each image pixels. Unfortunately even with this approach, some locations will still be cloud-obscured, due to locally persistent cloud-cover on all acquisition dates; even in southern Africa.

and then used as the final dataset on which the presence or absence of water is modelled from the Sentinel-2 imagery in that month. The median (rather than average) pixel value is used in order to 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 affected² over several acquisition dates so that it is not possible to extract a cloud-free pixel value for that month, then that pixel is classified as a “cloud-obscured” pixel (regardless of whether or not it covers a surface water feature).

The construction of the ‘target’ month Sentinel-2 only, multi-acquisition date image composite is governed by the following rules:

- If 3 or more cloud-free image acquisition dates exist, then only these image dates are used to generate the multi-date, median pixel value monthly composite, otherwise ...
- If a total of 3 or more cloud-free *and/or* cloudy (defined as $\leq 30\%$ total cloud cover) exist, then only these image dates are used to generate the multi-date, median pixel value monthly composite, otherwise ...
- Use *all* available cloud-free, cloudy and cloud-covered (defined as $> 30\%$ cloud cover) to generate the multi-date, median pixel value monthly composite.

Once the Sentinel-2 only monthly image composite has been constructed, modelling rules are applied to all composite image pixels that have not been classified as “cloud-obscured”, to identify open surface water extents. The pre-defined suite of Sentinel-2 modelling rules consists of two rulesets, which are used / applied to the Sentinel-2 only monthly composite according to the characteristics of the individual acquisition date images used to construct the monthly composite, namely:

- If 3 or more cloud-free image acquisition dates exist for the target month, then the ruleset representing all available water modelling rules is applied (‘liberal’ ruleset application), otherwise ...
- If a total of 3 or more cloud-free *and/or* cloudy (defined as $\leq 30\%$ total cloud cover) exist for target month, then the same ‘liberal’ ruleset is applied, otherwise ...
- For all other image acquisition content profiles, the more water restrictive ‘conservative’ ruleset is applied (which minimises any potential misclassification of water in dark cloud shadow areas), missed by the prior cloud cover exclusion classification).

² As part of the generation of the monthly cloud-free image Sentinel-2 composite, cloud-top and cloud-shadow exclusion masks are generated for each image acquisition date. These masks effectively identify and mask out $\pm 95\%$ of cloud affected areas. This is then combined with the median pixel value approach in order to provide near 100% exclusion of cloud-affected pixels in the surface water modelling process. The significant 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.

2. All composite image pixels classified as 'cloud-obscured' are then assessed using Sentinel-1 SAR imagery and the associated surface water detection modelling rules. A Sentinel-1 only, multi-date image composition representing median values from all Sentinel-1 image acquisition dates within the six-week window is constructed, and used as the final dataset on which the presence or absence of water is modelled from the Sentinel-1 imagery in that month; but only in the 'cloud-obscured' pixels as determined from the Sentinel-2 water modelling. The Sentinel-1 water modelling rules are designed so that a positive water identification can only be an outcome if the target 'cloud-obscured' pixel was previously determined to contain water in any of the preceding 12 months, using either Sentinel-1 or Sentinel-2 based modelling. This rule ensures that any Sentinel-1 based water detection has minimal / zero commission error outcomes (i.e. false positive water detection), since detection is limited to previously determined highly likely water extent areas.

Data Modelling and Automation (Sentinel-2)

The core procedural objective in the surface water monitoring service 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; which has removed the need for downloading, preparation and conventional modelling and analysis of large volumes of image data using office-based proprietary software. This has resulted in significant enhancements in operational procedural efficiencies that requires minimal office-based support infrastructure.

The surface water detection algorithms have been developed by GTI 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 1). As such decision tree classifiers are ideal tools for deriving standardised, threshold-based rules for image classification, since 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 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. These rulesets have been successfully tested across a range of variable landscapes (with similar climatic conditions), across southern Africa, Australia and South America, proving the versatility and transportability of the GTI water detection and monitoring service.

The reference dataset consists of $\pm 60,000$ sample points that represent a wide range of seasonal and geographical variations in both water (i.e. turbidity, depth) and non-water surface conditions with potentially similar spectral characteristics, such as burn scars, terrain shadows, and dark non-vegetated surfaces from both natural and man-made environments. Collectively these points ensure full representation of all spectral characteristics required in the water detection modelling process.

A range of spectral values were extracted from the Sentinel imagery for each sample point, representing 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.

Potentially suitable spectral indices for water detection were sourced from various publications, with the final selection based primarily on proven usefulness with Sentinel-2 imagery, and to a lesser degree, on the ability to replicate the same water detection modelling with Landsat 8 spectral information content, should this ever be necessary. 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, which provides the necessary optimal rule generation capability.

The WEKA software includes the open source Java J48 version of the C4.5 algorithm. This algorithm, is considered one of the top performing data mining algorithms, and is ideally suitable 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 Figure 2 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.

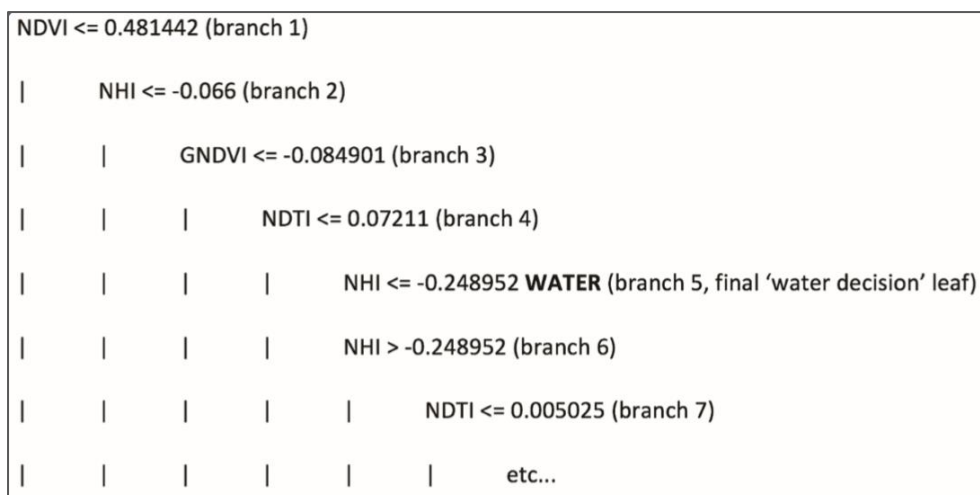


Figure 2. Decision tree hierarchical rule structure.

Data Modelling and Automation (Sentinel-1)

The Sentinel-1 SAR modelling rules have been developed using a different approach. The full history of Sentinel-2 only identified surface water extents (2016 – 2020) has been used to develop accurate reference data, from which definitive ‘water’ and ‘non-water’ temporally defined reference points have been extracted. Deep Learning (DL) machine learning procedures have then been used to define and generate the Sentinel-1 SAR based surface water detection rules, and ultimately the final surface water classification.

How Accurate are the Modelled Surface Water Extents?

Large waterbodies and/or water bodies with unique spectral characteristics, especially if they are different to the surrounding non-water landscape are the easiest to detect, identify and delineate. For example, it is possible to ‘detect’ the presence of a water feature that is smaller than the 20m Sentinel 2 image pixel if the spectral characteristics of the water feature are significantly different from the surrounding landscape, and the area of the water feature exceeds $\pm 40\%$ of the pixel area. In such cases, the water feature will be potentially ‘detectable’, but the surface area representation of the water feature will be incorrect and over-estimated, since the entire pixel cell will be coded as water, even though the actual water feature is smaller.

If the water feature is considerably larger than a single 20m image pixel, and contains *many* image pixels with spectral water characteristics, the water feature will be both 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 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, as far as possible, to minimise all such commission errors. Anecdotal visual estimates of the accuracy of water body detection appears to be 85 – 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), are, other than due to small water feature size, typically the result of either 'impure' spectral characteristics, such as a heavy sediment load or very shallow conditions; where the reflectance characteristics are more aligned with a non-water bare ground surface. A water feature can also be potentially "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.

Statistical Errors Related to Calculating Water Surface Areas

Whilst it is challenging to provide statistical measures of surface water extent related to linear features such as rivers and channels, it is possible to compare the image-derived surface water extents to independently defined man-made impoundment boundaries (i.e. dams), when suitable reference data exists for statistical comparison. The results presented below are based on comparison of image-generated and map digitised dam areas in test areas in South Africa.

Dam surface areas detection inaccuracies decrease, as would be expected, as dam size increases due to the size of the dam area compared to the pixel resolution of 20 meters. The combination of both commission and exclusion errors on smaller dams having a greater negative impact on image mapping accuracy. Figure 3 shows that for surface water feature, such as dams, larger approximately 10,000 m², the difference between the image-modelled surface extent and the map-digitised extent is less than 10%; although for very small dams the errors can increase to 80%. Errors associated with dams above 50,000 m² typically present errors in the 2.5% range.

European Space Agency (ESA) WorldWater Round Robin Evaluation

The 'Mzansi Amanzi' surface water detection capability was selected for evaluation as part ESA's WorldWater Project Round Robin evaluation (<https://worldwater.earth/elementor-4282/>). The objective of the WorldWater project's Round Robin exercise was an inter-organisational comparison of Earth

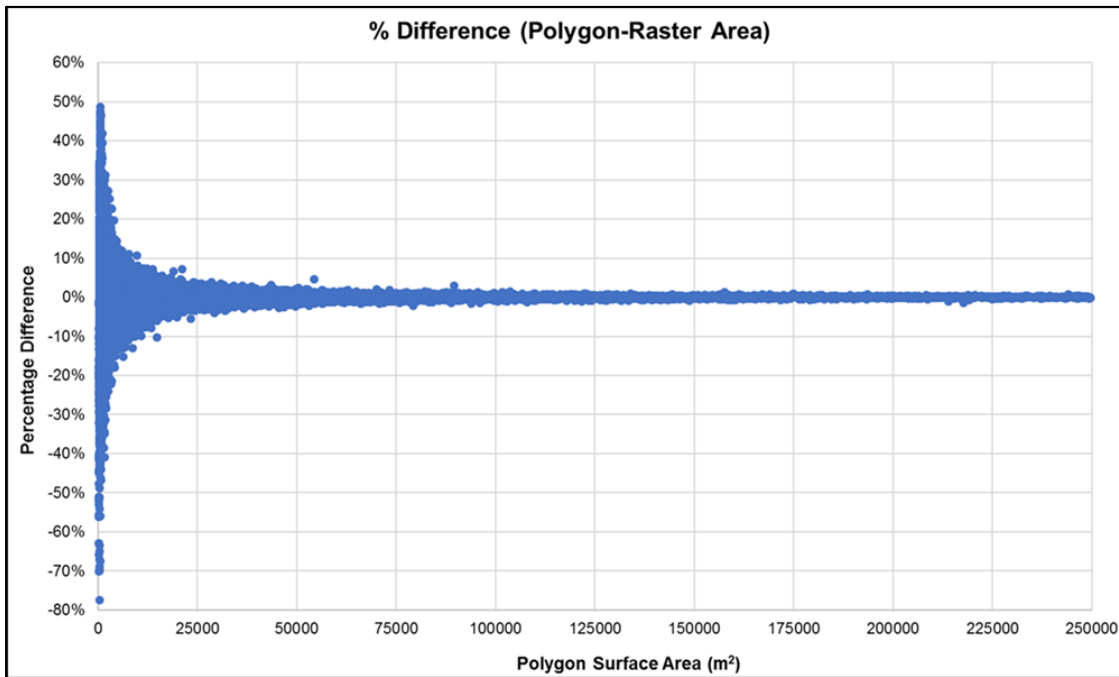


Figure 3. Dam surface area errors for individual dams up to 250,000 m² (image-modelled vs map digitised boundaries).

Observation (EO) algorithms for surface water detection, using the latest generation of free and open satellite data from Sentinel-1, Sentinel-2 and Landsat 8. The Round Robin is organised under the auspice of the European Space Agency and being supported by a number of international organisations and initiatives including CNES, NASA, the European Association of Remote Sensing Companies (EARSC), the CEOS Ad hoc team on Sustainable Development Goals (CEOS SDG AHT) as well as GEO and their Earth Observations for the Sustainable Development Goals (EO4SDG) initiative.

Preliminary results of the Round Robin (RR) confirm that the 'Mzansi Amanzi' enhanced surface water detection capability (based on integrated Sentinel-1 and Sentinel-2 processing), is comparable in accuracy to the majority of other international water mapping capabilities evaluated in the RR exercise³. Statistical map accuracies were determined independently by RR coordinators over four globally distributed water test sites in Colombia, Ghana, Mexico and Zambia, which included a significant range of coastal, terrestrial, deep, shallow, permanent and highly seasonal water feature characteristics. In each case, monthly surface water datasets were created for the two years 2019 and 2020. Mzansi Amanzi outcomes for these four sites were an average Overall Accuracy of 92% (group average 92%), Producer Accuracy of 84% (group average 81%), and User Accuracy of 87% (group average 90%)⁴.

³ 15 international organisations were selected to participate in the RR

⁴ Based on ESA verified but unpublished results distributed to RR participants 07-2021

ANNEXURE 4

ADDITIONAL TECHNICAL INFORMATION TO MONTHLY SURFACE WATER VOLUMES

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.

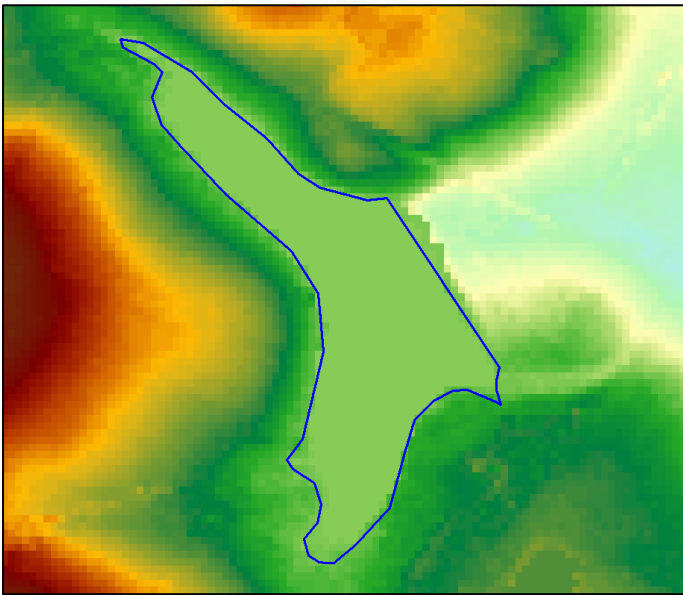


Figure 3: Showing the dam digitised dam outline on the flat area of satellite derived topography

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.

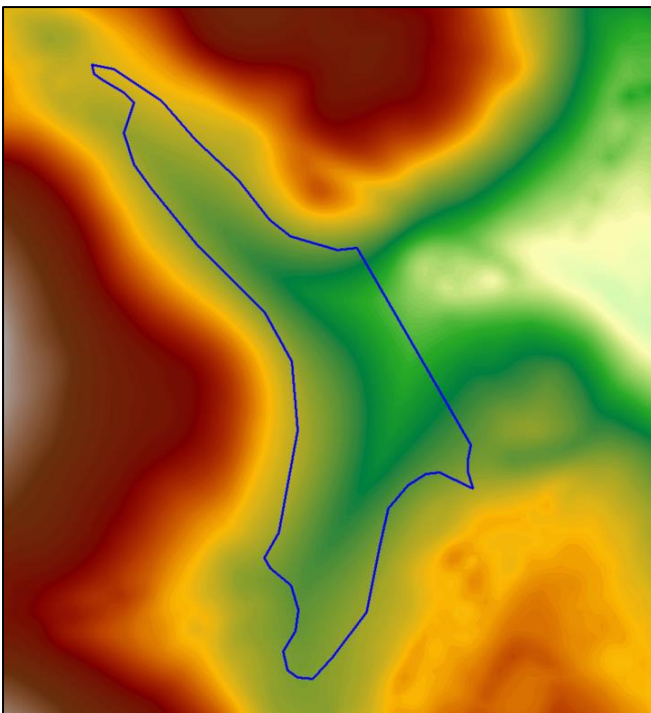


Figure 4. The interpolated surface is generated creating a representation of the dam bathymetry

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.

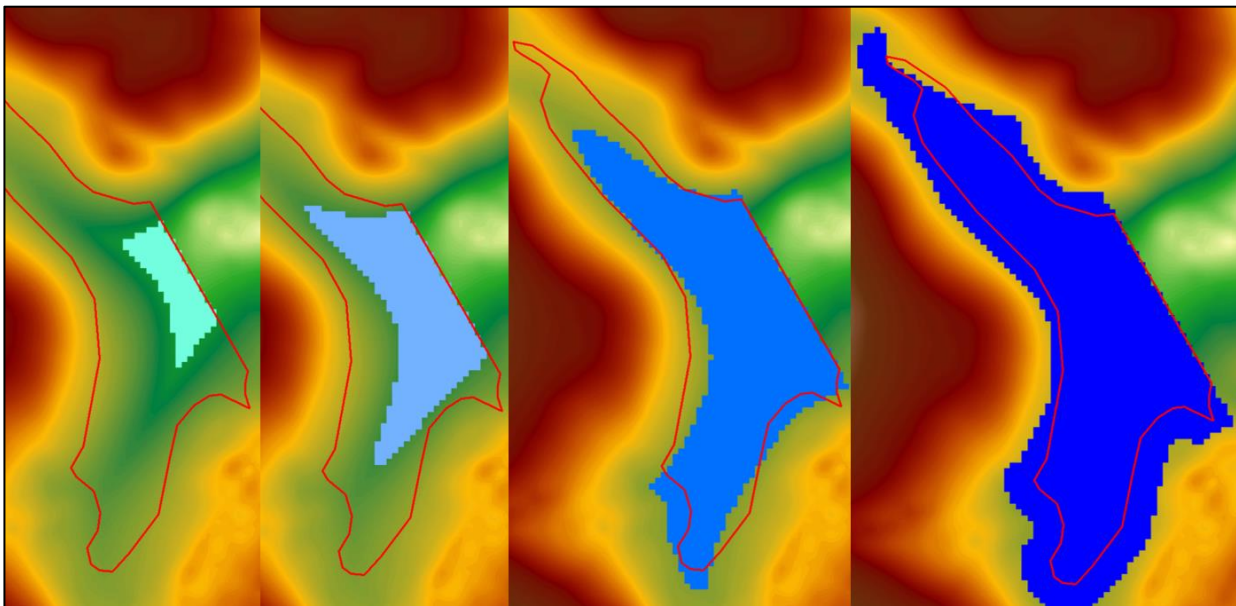


Figure 5: Graphical representation of process used to calculate the Area volume relationships

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.

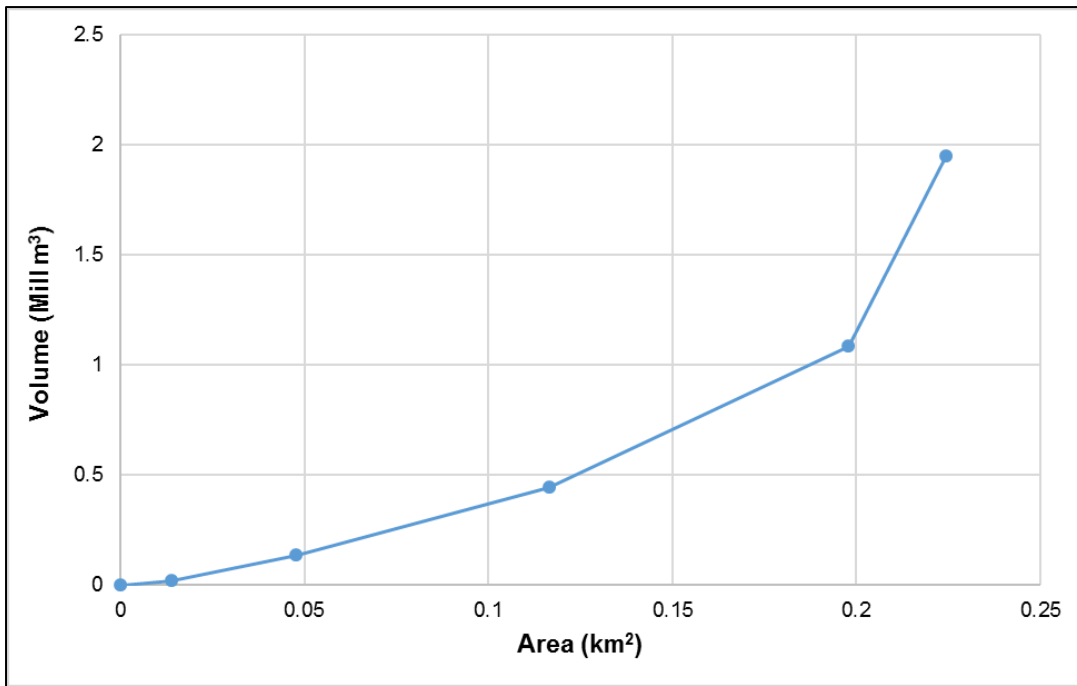


Figure 6: Representation of the individual calculated volume area relationship

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.

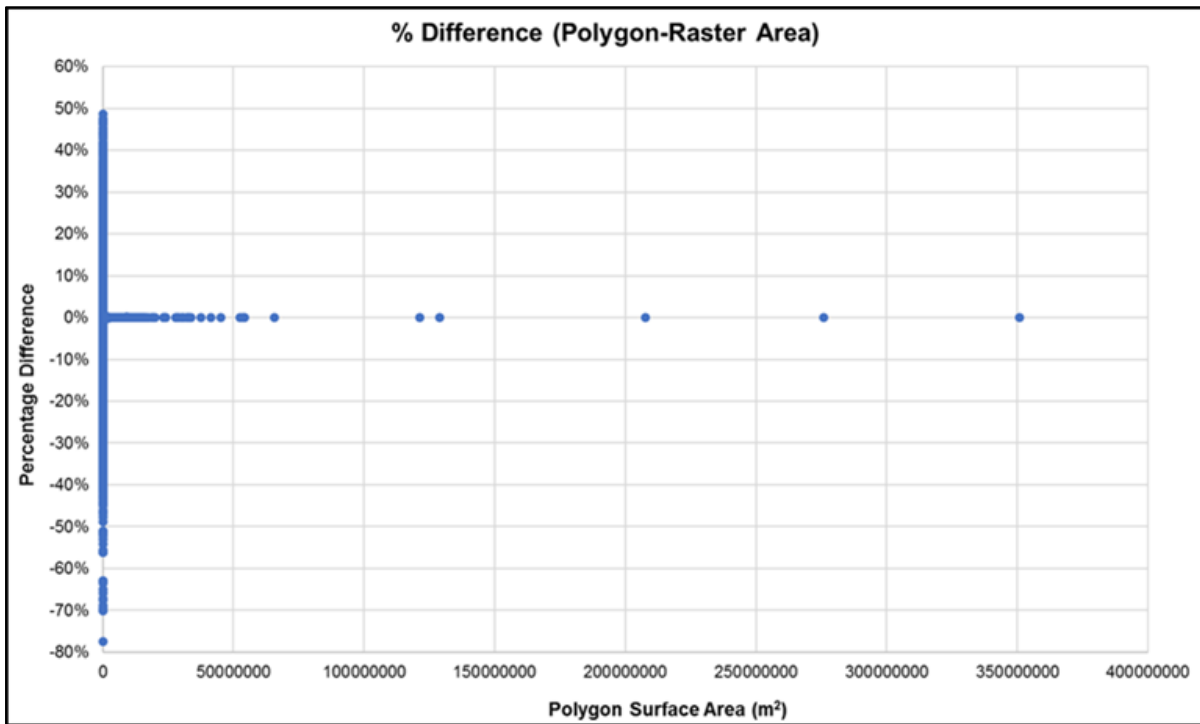


Figure 7. Errors for individual dams

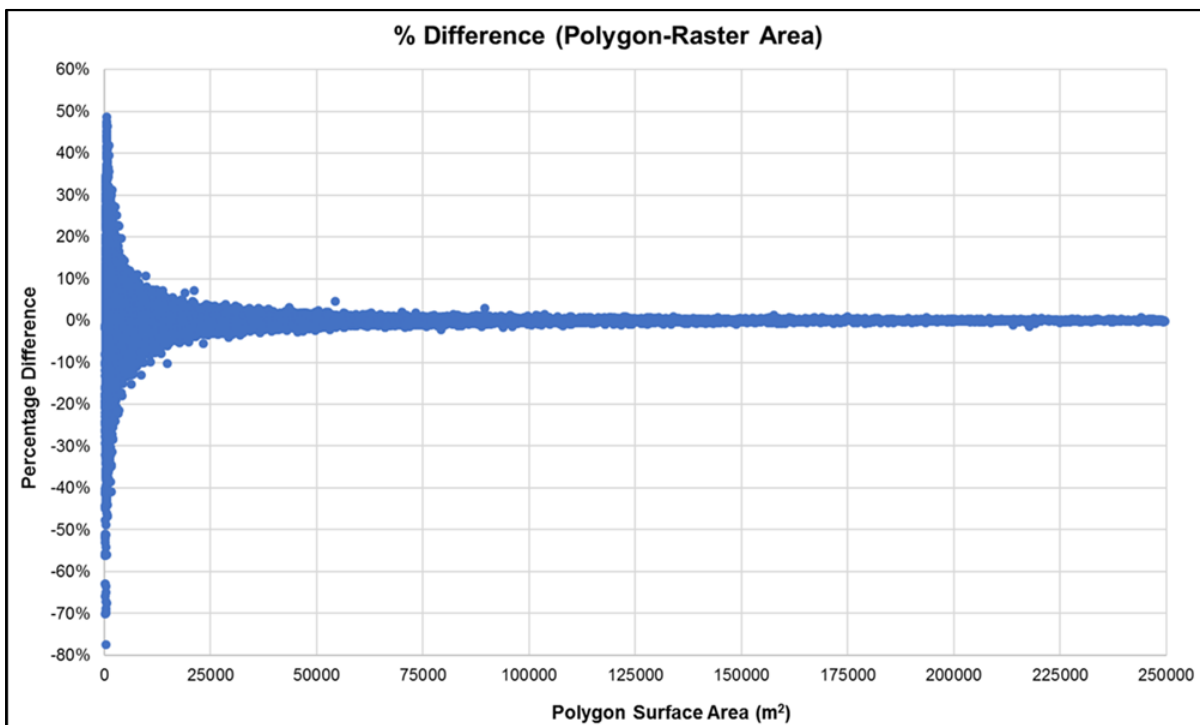


Figure 8. Errors for individual dams excluding large dams above 250000 m².

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.

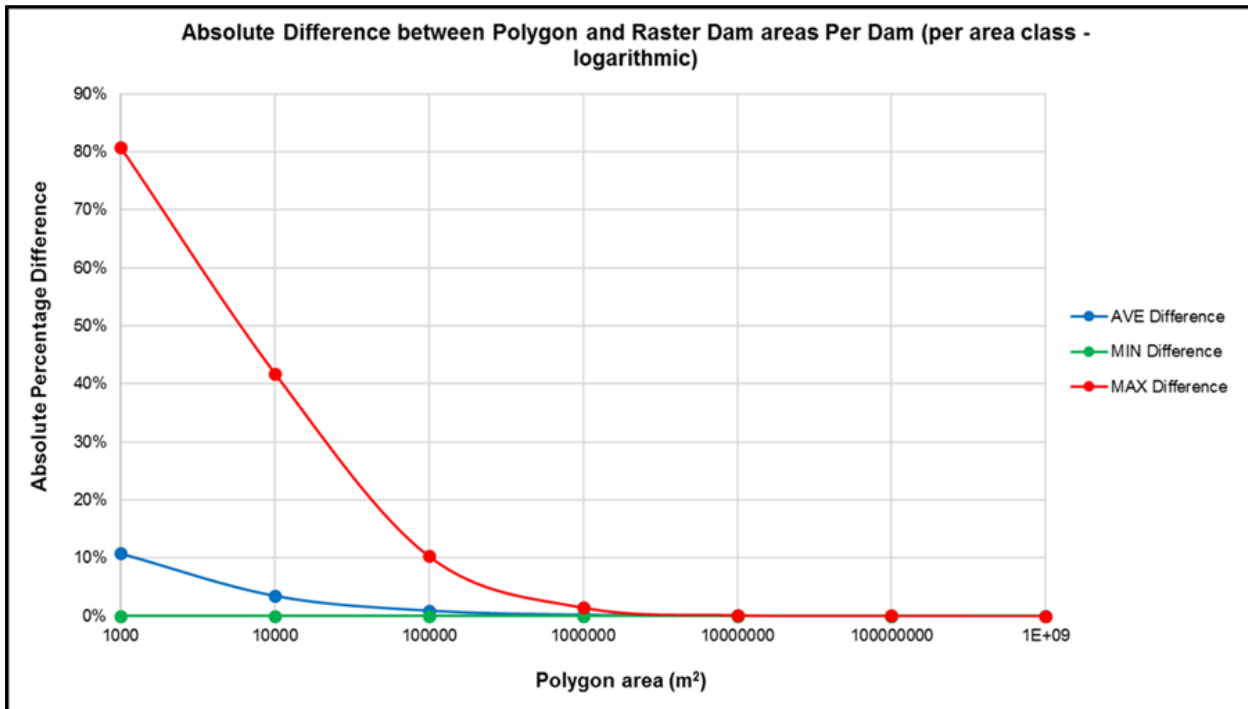


Figure 9: Summary of errors between polygons and raster estimates for area of individual dams.

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)

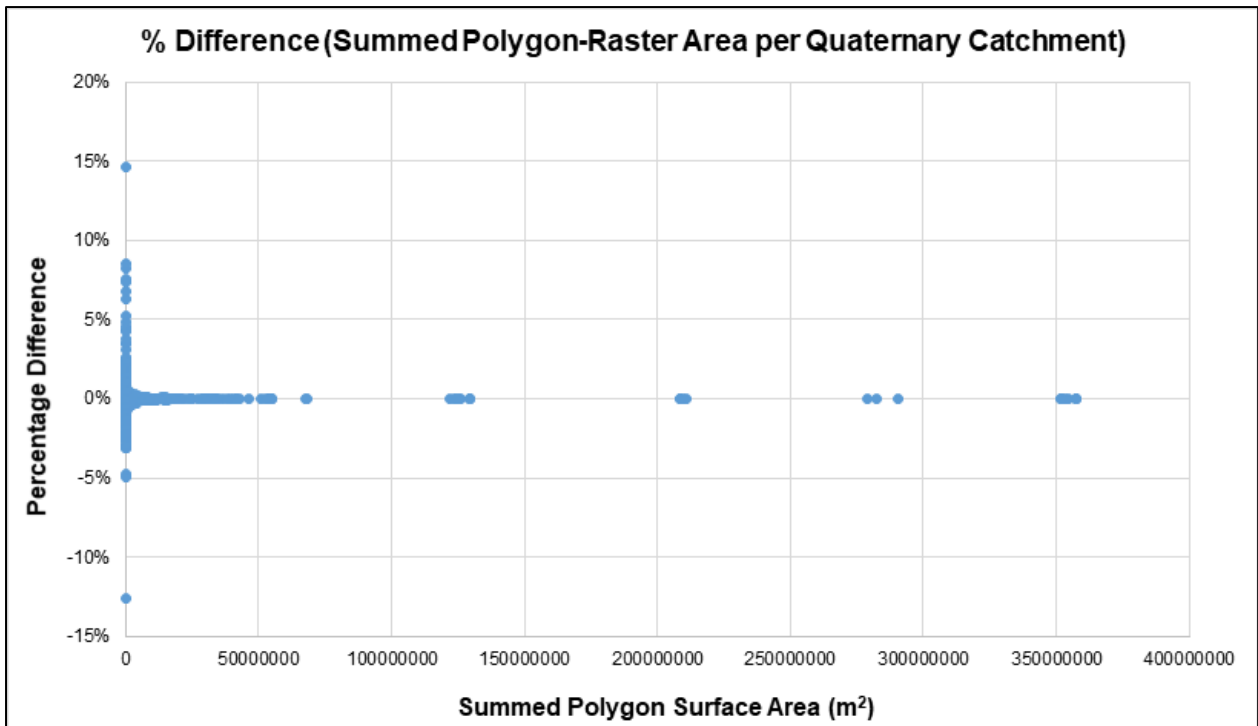


Figure 10: Difference between the raster volumes and polygon estimates.

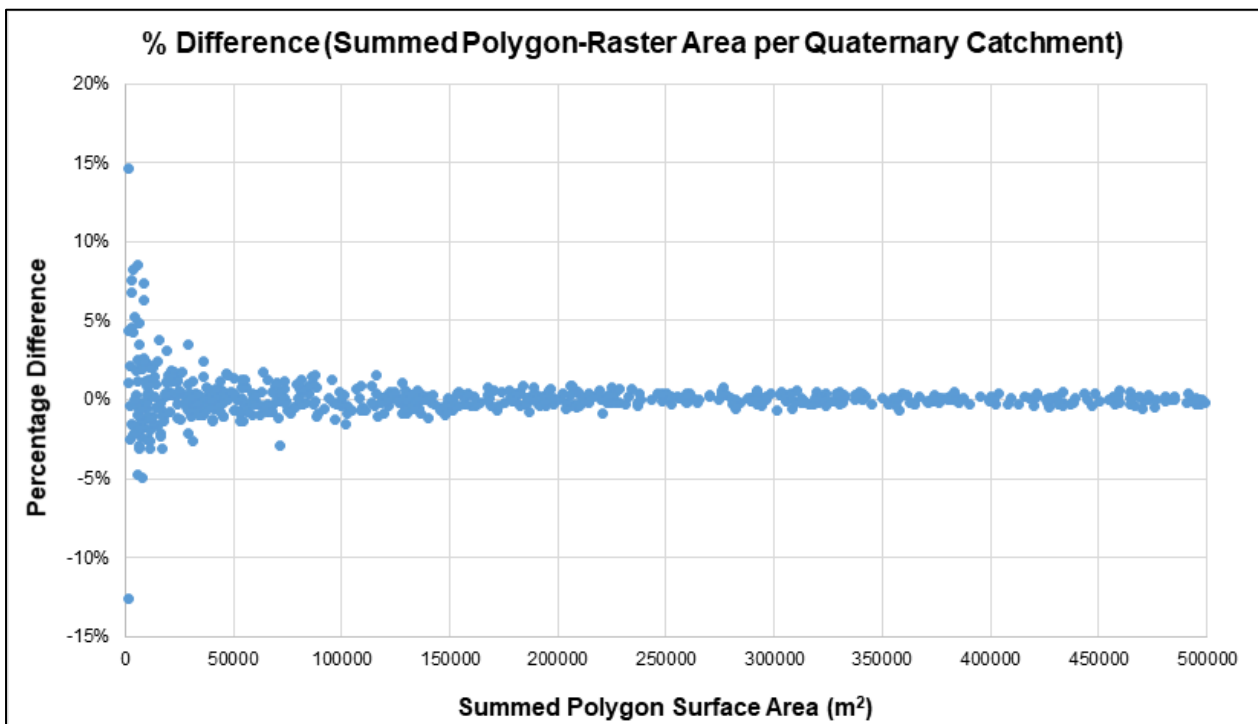


Figure 11: Zoomed in difference below 500 000 m³.

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).

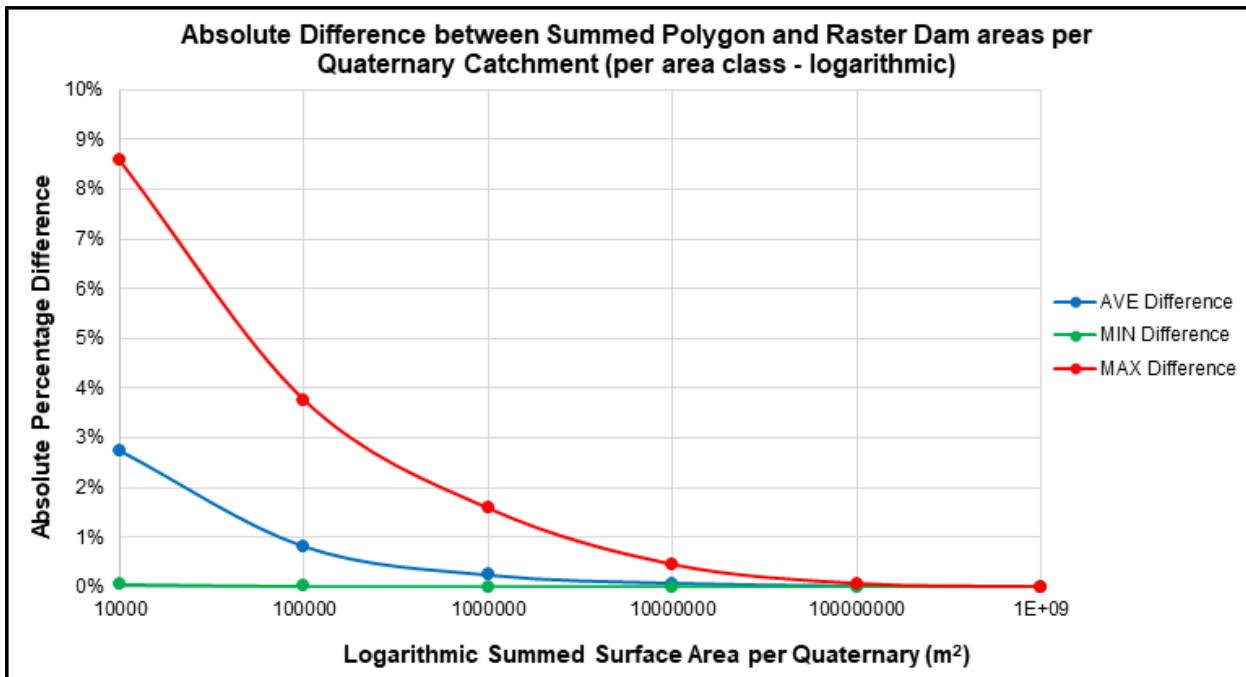


Figure 12: Difference as a mathematical functions.

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.

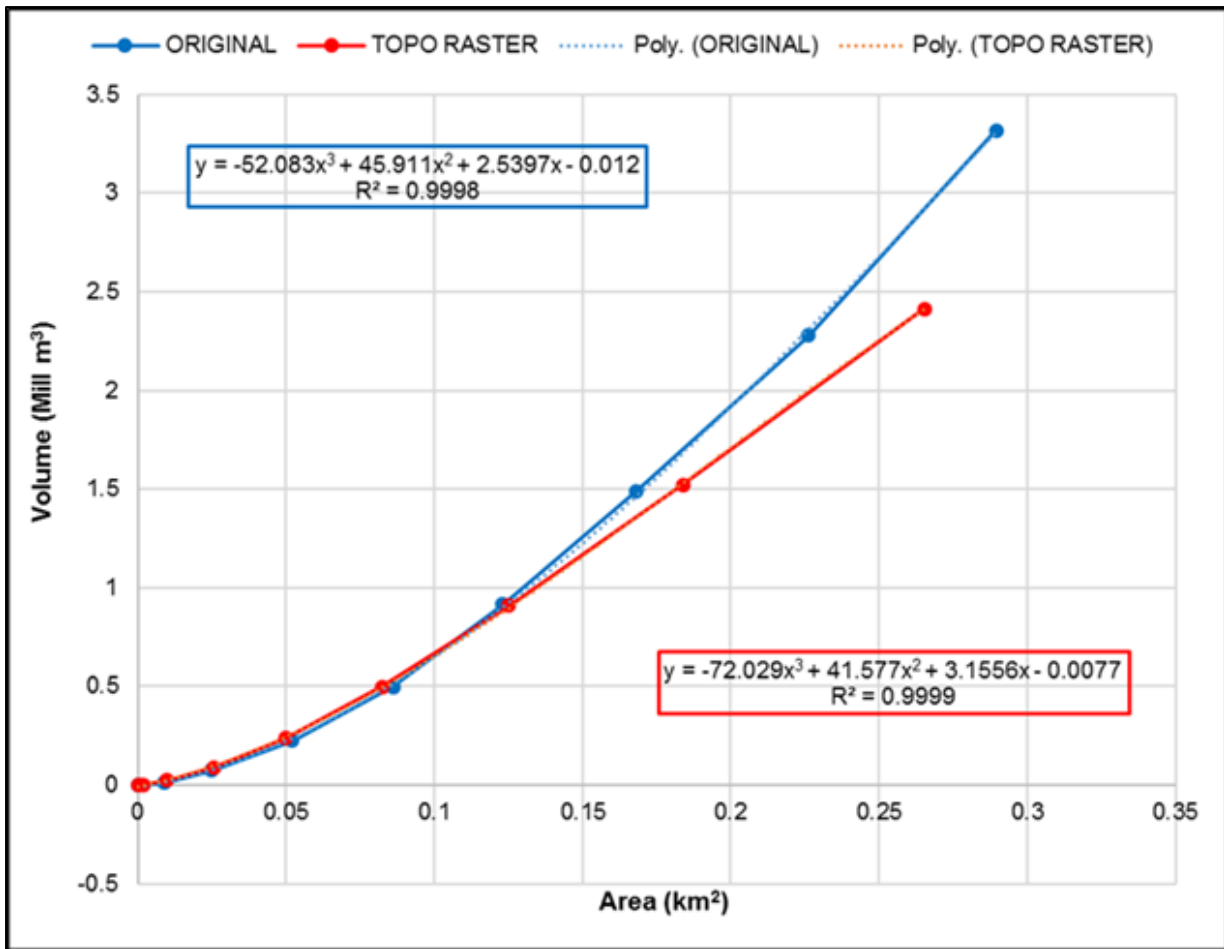


Figure 13: Comparison of actual dam volume compared to interpolation measurement method.

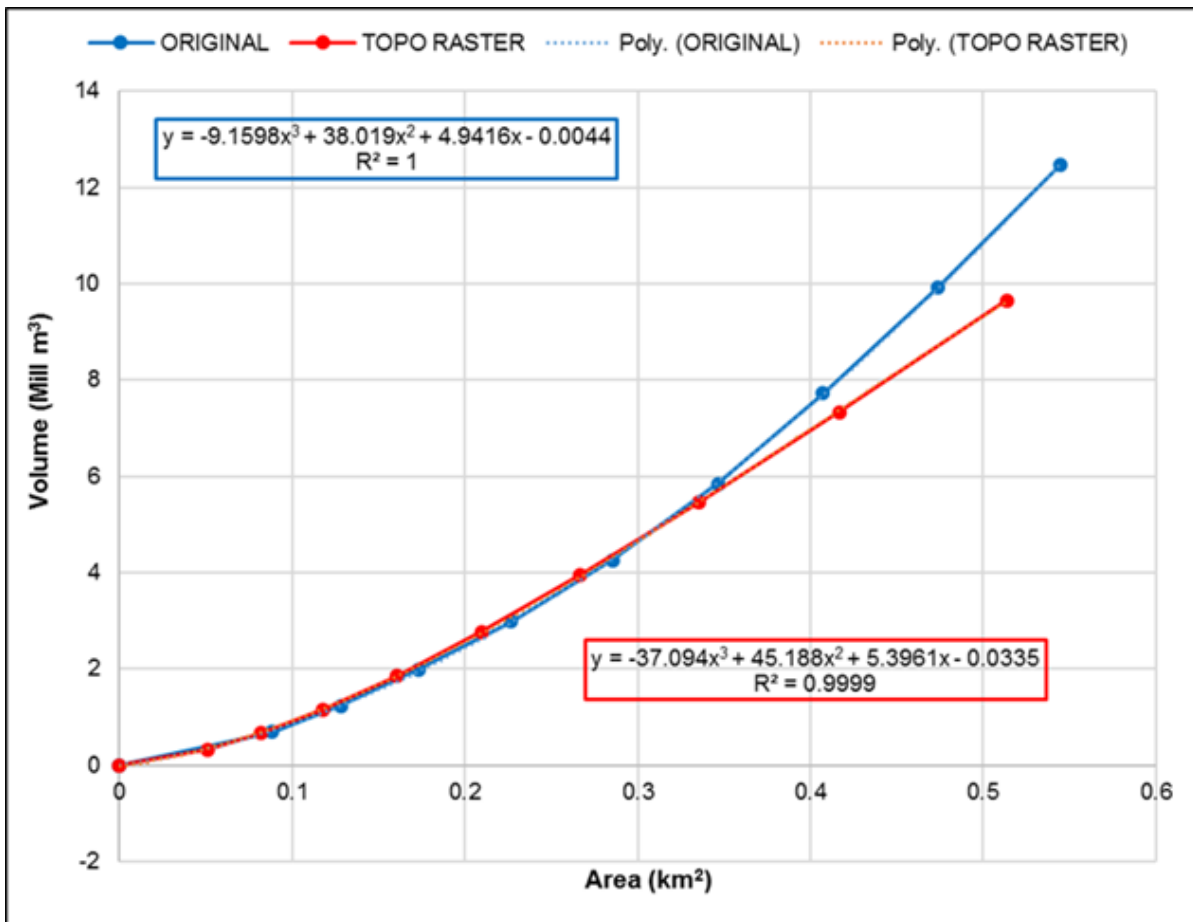


Figure 14: Comparison of same technique in different area.

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%.