EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Groundwater flow modelling in the Zambezi river basin to investigate the interaction with surface water bodies

FINAL REPORT

Prepared by:

Dr. Kawawa Banda
Water Resources Consultant
University of Zambia
School of Mines
P.O. Box 32379
Lusaka
Zambia

Tel: +260 977 533 041
E-mail: Kawawabanda@yahoo.co.uk
Kawawa.Banda@unza.zm

Prepared for:

European Union – Joint Research Commission
Client contact:
César Carmona Moreno

DG – Joint Research Centre
Directorate D – Sustainable Resources
Unit D.02 Water and Marine Resources
Via E. Fermi, T.P. 440, I-21027 Ispra (VA) Italy
Building 100 01/1318

Tel: +39 0332 789654
E-mail: cesar.carmona-moreno@ec.europa.eu

This report is to be referred to in bibliographies as:

European Union - Joint Research Centre, Zambia. 2020. *Final Report*. Prepared by Kawawa Banda as part of Groundwater flow modelling in the Zambezi river basin to investigate the interaction with surface water bodies (Second Phase activities).

Abbreviations

ACE WATER	African Networks of Centers of Excellence on Water Sciences
FEFLOW	Finite Element subsurface FLOW system
SANWATCE	Southern Africa Network of WATer Centres of Excellence
SPATSIM	SPAtial and Time Series Information Modeling
WEFE	Water-Energy-Food-Ecosystem
ZAMCOM	Zambezi Watercourse Commission
ZAMWIS	Zambezi Water Information System
ZEMA	Zambia Environmental Agency
ZRB	Zambezi River Basin

Acknowledgments

This work was funded in the framework of the European Commission support to the African Networks of Centres of Excellence on Water Sciences PHASE II (ACE WATER 2). I would like to acknowledge Ezio Crestaz from the Joint Research Centre for the support in this assignment and the many hours of revision and guidance in R and FEFLOW – I am highly indebted. Fabio Farinosi and Denis Hughes for discussions and oversight during this assignment. Finally, to my family and friends for the support – God bless you all.

Blank Page

TABLE OF CONTENTS

List Of Figures

List of Tables

1. Introduction

1.1 Background

This work is defined within the project, “The African Networks of Centres of Excellence on Water Sciences PHASE II (ACE WATER 2)” focusing on an overall WEFE (Water-Energy-Food-Ecosystem) nexus assessment over the Zambezi River Basin (ZRB). The project, whose geographical scope extends over most of the sub-Saharan Africa focusing on three distinct and complementary networks in Western, Central-Eastern and Southern Africa, is run under the coordination of the Joint Research Centre of the European Commission (JRC/EC), responsible for the scientific research, in collaboration with UNESCO in charge of the HCD (Human Capacity Development) component. Leading scientific experts from ten Southern Africa Centers of Excellence (CoEs, geographically spanning through South Africa, Botswana, Namibia, Zimbabwe, Zambia, Malawi and Mozambique), as well as from the Un. of Rhodes and the Un. of Florida, collaborated at different extent to such a challenge joint effort, addressing different topics as; the climate variability and climate change; the surface hydrology; the groundwater hydrology; the hydropower and the agriculture current status and future developments under the various constraining factors, as water availability reduction and increasing pressure due to population growth and activities development. Key regional and basin management Institutions, as the SADC (Southern Africa Development Community), SADC-GMI (SADC-Groundwater Management Institute) and ZAMCOM (ZAMbezi water course COMmission) supported the activities, providing guidance with respect to key policies and access to relevant datasets.

As for the groundwater hydrology, the University of Western Cape (Mengistu, 2018) compiled an updated geological and hydrogeological map complementing the work from the University of Zambia (Banda, 2018) and the NUST of Zimbabwe (Chinyama and Makaya, 2018), who focused on the compilation of detailed databases and the analysis of the hydrogeological and hydro-chemical status at respective country scale (Zambia and Zimbabwe respectively). The analysis implemented in the framework of the ACEWATER2 project contributed to identify: (i) areas relevant to groundwater use, as inferred from wells spatial distribution, characteristics (e.g. yield, hydrogeological properties estimate after pumping tests, water sampling and water quality analysis) and any further evidence of groundwater withdrawal, for human supply (e.g. from population density), irrigation in agriculture (e.g. pivoting systems), cooling of industrial plants and water use in mining activities; (2) groundwater accessibility, as related to wells and water table depth (from the ground), and future potential, relevant to expected socio-economic development (e.g. growing population, expanding irrigated agriculture); (2) few groundwater vulnerability issues, as due to contamination from surficial or deep origin sources (e.g. fertilizers and pesticides in agriculture, fecal choliforms from untreated water, leakage from landfills, upconing of salinized water trapped in deep aquifers, as in western Zambezi, or salt water intrusion along coastal areas).

Very few quantitative studies and datasets at the basin scale exist, among which include: the SADC hydrogeological map and atlas, SADC HGM (Pietersen *et al*., 2010), the groundwater hydrology and hydrochemistry database from the SADC-GMI, continent quantitative hydrogeological maps of Africa from the BGS (MacDonald *et al*., 2012) and global coverage of potential recharge estimates calculated using WaterGap Model from BGR (Doell and Fiedler 2008). This work focuses on a further specific analysis that was undertaken to investigate the potentials and the bottlenecks in the application of groundwater flow modelling at the basin scale, with the following objectives, (i). to investigate the reliability of hydrogeological parameters estimates and the groundwater resources availability; (ii). to support the assessment of interlinks between the surface water bodies and the aquifer systems. Given the areal extent, the Zambezi river basin being the fourth largest one in Africa after Nile, Niger and Congo, and its geological/tectonic complexity, the OS (Open Source) state-of-the-art USGS codes MODFLOW and MODPATH (Pollock, 2012) were considered not the most suitable platform, due to the limited discretization flexibility of the finite difference scheme. Instead, the DHI-WASY finite element code and modelling environment (Diersch, 2009) was adopted; thanks to the finite element numerical formulation, the high flexibility of triangular meshing makes possible to capture the relevant features (e.g. drainage network, geological and tectonic limits), while adopting a rough resolution over more remote and unknown areas.

1.2 Study area

The Zambezi river basin (ZRB), Figure 1, the largest African river system flowing into the Indian Ocean (Balek, 1977), straddles across eight Southern African countries. The basin is home to about 40 million people predominantly living in Malawi, Zimbabwe, and Zambia (Wirkus and Boge, 2006). The larger part of the basin is occupied by the Central African Plateau which lies more than 900 m above sea level, rising in places near the rim to over 2500 m (Leenaers, 1991). Climatically, the north regions of the Zambezi basin have mean annual rainfall of 1100 to 1400 mm which declines towards the south, reaching about half that figure in the south-west (Beilfuss and Santos, 2001). The rain falls in a 4-to-6-month summer rainy season when the Inter-Tropical Convergence Zone moves over the basin from the north between October and March. Evaporation rates are high (1600 mm-2300 mm) and much water is lost this way in swamps and floodplains, especially in the south-west of the basin (Beilfuss and Santos, 2001).

Figure 1: The Zambezi river basin and its 13 sub basins (Source: https:// www.grida.no/resources/5207)

Based on hydrology and geophysical characteristics, the basin can be divided into the upper, middle and lower ZRB (Chandiwana and Snellen, 1994):

- The Upper Zambezi reach defined as the area above the Victoria Falls. The source of the Zambezi river is a marshy bog near Kalene Hills in Zambia ranging from 1500 - 1600 m above sea level. The river flows north for about 30 km, then west and southwards through Angola for about 280 km and re-enters Zambia just north of the Chavuma Falls. After the falls the river begins to meander through broad and marshy plains. The last of these plains is the Barotse floodplain. At an altitude of 1000 m the river enters a 100 km stretch of rapids. After the Katima rapids, the Zambezi flows through a sandy plain where it meanders widely and its floodwaters join with those of the Chobe river from Angola, thereby creating another permanent swamp. When the river approaches Victoria Falls it has a mean width of 1350 m.

- The Middle Zambezi reach which stretches between Victoria Falls and Cahora Bassa dam. After the Zambezi river plunges some 100 m down the Victoria Falls, the river flows eastwards for almost 1000 km through gorges and the man-made lakes Kariba and Cahora Bassa. In the middle Basin, the river is joined by its two largest tributary the Kafue and the Luangwa.

- The Lower Zambezi reach, from downstream Cahora Bassa dams until the outfall to the Indian Ocean. At Cahora Bassa, the Zambezi river begins its descent from the Central African Plateau to the coastal plain. The Shire river, which drains Lake Malawi, joins the Zambezi river near Caia. On the Mozambique plain, the river occupies a broad valley with a width up to 7 km. The river may flow in several channels in the dry season, which merge again into a single river in the wet season. The delta of the Zambezi is wide, marshy and obstructed by sandbars.

The riparian countries of the Zambezi basin have different levels of wealth, population, literacy, and access to clean water and sanitation. Furthermore, they are all plague with by the lack of information about available water resources and their uses. The increasing demand for water is a concern because of population growth, exceedingly rapid urban and increasing irrigation to increase food production, climate variability and change.

1.3 Geology and Hydrogeology

Generally, ZRB is endorsed with various rocks and layers dating over a million years (Precambrian era) to recent times. The rock formations consist of igneous, sedimentary and metamorphic rocks. The basin is poorly characterised hydrogeologically because of limited borehole drilling records with pump test results. The hydrogeological description of the ZRB has typically be inferred from the lithological description (Figure 2). Mengistu, 2018, highlighted nine aquifers namely:

- Quaternary Alluvial Aquifer
- Kalahari Group Sediment Aquifers
- Mesozoic Marine Sediment Aquifers
- Karoo Volcanic Aquifer

- Karoo Sedimentary Aquifers
- Palaeozoic Marine Sediment Aquifers
- Katanga Carbonate (dolomite) aquifers
- Precambrian metasediment and metavolcanic Aquifers
- Basement Complex Aquifers

Quaternary alluvial aquifers are alluvial aquifers mainly composed of unconsolidated coarse gravel and minor fine sand, situated on either side of major river flood plains and interestingly scattered around the river basin. The yields range from 5 to 20 l/s (MacDonald *et al.*, 2012). These unconsolidated clastic sediments have a representative transmissivity range of 30 – 150 m^2/d (Yachiyo Engineering Co. Ltd, 1995, Baumle et al, 2019). The Kalahari Group Sediment Aquifer, mainly outcropping in the western portion, covers 26% of the basin by surface area amounting to about 52 x 10^6 hectares (Mengistu, 2018). This formation is Pleistocene in age and consists of sands, siltstone, sandstones, orthoquartzites and duricrusts. Transmissivities range from 0.2 to 200 m^2/d (Yachiyo Engineering Co. Ltd). Mesozoic marine sediment aquifers are limited to central east portion of the basin in close proximity of the downstream portion of Zambezi River Basin (mainly in Mozambique). There are reported to be relatively high yield boreholes, likely tapping the sandy portion of the aquifer of more than 20 l/s. However, relatively low borehole yield have also been observed likely due to the shale portion of the aquifer drastically minimize the transmissivity and storage capacity (Mengistu, 2018).

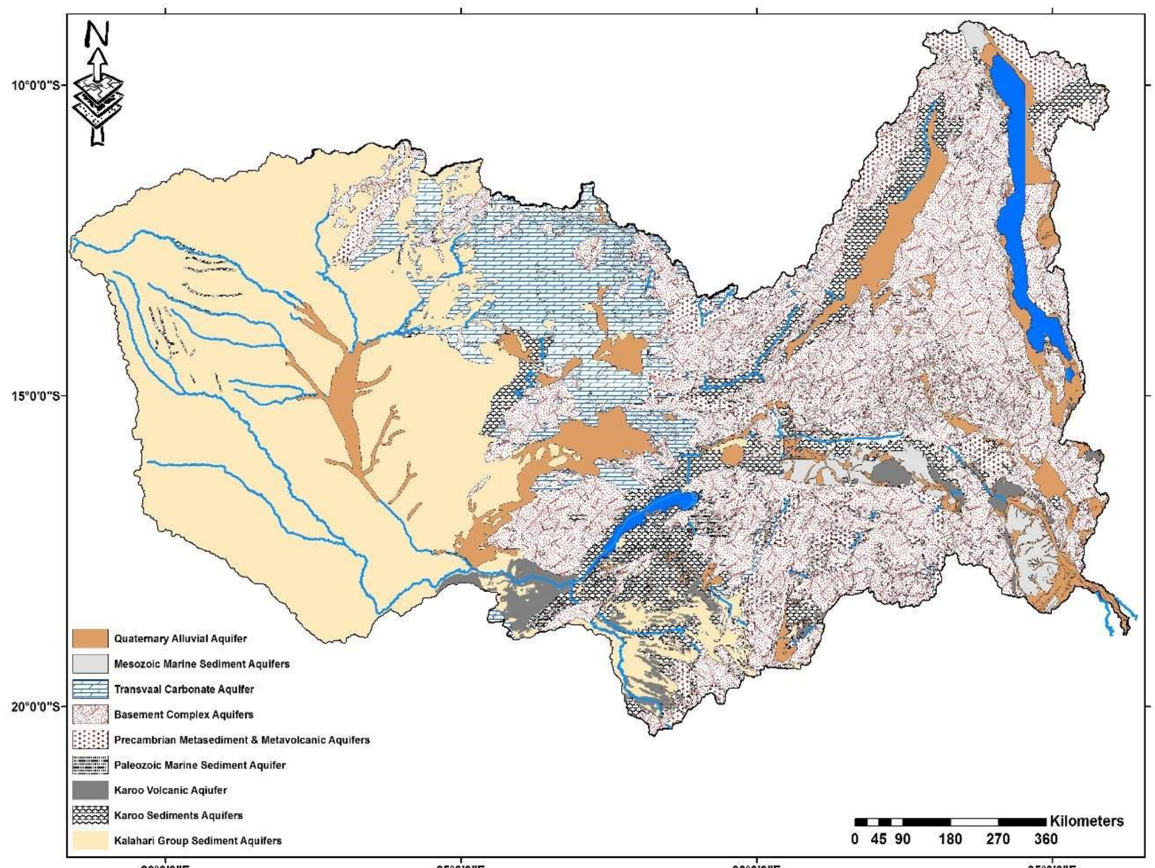


Figure 2: Lithological description of the Zambezi River Basin (Source: Mengistu, 2018)

Karoo volcanic aquifers occur due to fracturing from dolerite dykes and are relatively low yielding and range from 2 to 5 l/s. The Karoo sedimentary aquifer is Carboniferous to early Jurassic in age. Thick successions of coarse clastic sediments, which can be linked to fault-controlled basin margins, occur in the Cahora-Bassa and Mid-Zambezi basins in Zimbabwe and the Waterberg Basin in Namibia. Transmissivities in this aquifer range from 10 – 500 m^2/d (Lekula et al, 2018). Palaeozoic marine sediment aquifers comprise mainly mudstone and shale silts and marl covering only 0.1% of the Zambezi Basin. These typical have low yield ranging from 0.1 - 0.5 m^2/d. These aquifers show dual aquifer classes (intergranular and fracture) because of the presence of units with primary and secondary porosities (Mengistu, 2018).

Katanga Carbonate (dolomite) aquifers are highly productive with yield more than 20 l/s where karst structures occur. Transmissivity have been reported of more than 300 m^2/d (Yachiyo Engineering Co. Ltd, 1995). Precambrian metasediment and metavolcanic aquifers are exposed in the far northeast margin of the basin but also it is evident that the central margin and few areas in Zimbabwe and in the downstream areas of the Zambezi River in Mozambique. The aquifer is categorized as being generally controlled by network of fractures and faults that create secondary porosities. Transmissivity ranging from 0.2 to 10 m^2/d have been observed (Yachiyo Engineering Co. Ltd, 1995). These are low yielding typically of < 0.5 l/s but with estimated total storage capacity is roughly 34 x 10^6 m^3, which is considered a significant volume taking into account the limited occurrence compared to the basement complex aquifers (Mengistu, 2018). Basement Complex Aquifers comprise mainly archaean age granites, gneisses as well as other subordinate units. The aquifers cover roughly 35% of the basin outcropped largely in the eastern half of the basin. Intense weathering and alteration could justify the low yield aquifer of < 0.1 l/s. Basement rocks dominate the eastern margins of the ZRB. Transmissivities in this aquifer range from 0.1 – 5 m^2/d (Yachiyo Engineering Co. Ltd, 1995).

Table 1 below, compares the transmissivity estimates from literature and the work by the British Geological Survey (MacDonald et al, 2012). The Quaternary Alluvial Aquifers transmissivity range from 30 – 150 m^2/d compared to 500 – 1000 m^2/d from the BGS which is an over estimation. However, the rest of the aquifer systems seem well represented by the BGS datasets as the variability in Transmissivity is captured. The BGS dataset therefore captures the variability in the hydraulic properties of the major aquifers. It is therefore an appropriate dataset to represent the spatial variability in the groundwater modelling framework at Catchment level. However, it is likely that some of the local variability is lost in generalisation of representative transmissivity estimates. Variability is likely complex in parts of the basin and is lost in this reductionist approach.

Table 1: Aquifer units in the Zambezi River Basin compared with available transmissivity estimates and the British Geological Survey (BGS) estimates.

Aquifer Unit	Predominate lithology	Transmissivity Range	BGS Transmissivity
Quaternary Alluvial Aquifer	unconsolidated coarse gravel and minor fine sand	30 – 150 m^2/d	500 – 1000 m^2/d
Kalahari Group Sediment Aquifers	sands, siltstone, sandstones, orthoquartzites and duricrusts	0.2 to 200 m^2/d	50 – 500 m^2/d
Mesozoic Marine Sediment Aquifers	Mostly Sand but with occurrence of shale	-	5 – 10 m^2/d
Karoo Volcanic Aquifer	Dolerite dykes	0.1 - 0.5 m^2/d	10 – 50 m^2/d
Karoo Sedimentary Aquifers	Clastic sediments e.g. sandstone, conglomerate	10 – 500 m^2/d	10 – 50 m^2/d
Palaeozoic Marine Sediment Aquifers	Clastic sediments such as silt	-	10 – 50 m^2/d
Katanga Carbonate (dolomite) aquifers	Dolomites	> 300 m^2/d	500 – 1000 m^2/d
Precambrian metasediment and metavolcanic Aquifers	Volcanic Tuffs	0.2 to 10 m^2/d	1 – 5 m^2/d
Basement Complex Aquifers	Gneisses and Schists	0.1 – 5 m^2/d	1 – 5 m^2/d

2.0 Groundwater flow model

2.1 Model design rationale

The surface hydrology modelling benefits from many different freely available datasets, continuous over large regions being derived from remote sensed products (e.g. SRTM, CHIRPS). These datasets contribute to the characterization of the main hydrological drivers (climate, land use, topography), that govern the main components of the hydrological cycle (recharge, evapotranspiration, surface runoff, sub-surface and deep percolation). Ground based data (e.g. meteo-climate stations, river discharge time series measured at gauging stations) complements the overall framework, contributing to model refinement and calibration. On the other hand, the groundwater hydrology analysis generally suffers from the lack of continuous and detailed datasets. The airborne geophysical surveys (e.g. magnetometry, gravimetry), that could support an interpretation of the main geological and tectonic features, are rarely available over large regions, particularly in Africa; global datasets such as GRACE (Gravity Recovery and Climate Experiment) are very promising to detect aquifers replenishment or depletion trends (Doell et al, 2014) but lack the detail that most groundwater modelling would demand for; datasets derived from geophysical prospecting for oil and gas exploration are strictly confidential and hence not accessible; ground based geo-electrical and electromagnetic surveys generally cover areas of limited extent, being relevant mainly to local scale analysis.

The ground data, as is the case for direct observations derived by drilling or sampling (e.g. stratigraphy, parameters estimate after pumping tests, chemical analysis), contributes, together with the knowledge of the geological and structural framework, to build the hydrogeological conceptual model. Still, the ground data are often spatially scattered over the study areas, unevenly distributed, limited in depth. The resulting conceptual model is often affected by a high degree of uncertainty. Also with all the above challenges and limitations in mind, a large effort to conduct an extensive bibliographic review of former hydrogeological studies at the African continental scale and to build quantitative hydrogeological maps was completed by the BGS (MacDonald et al., 2012). The BGS maps account for the estimation of the range of main aquifers properties: aquifers' yield, porosity, thickness and effective recharge. Other datasets can be used to characterize the surface water bodies, that act, from the groundwater hydrology perspective, as upper bounding recharge or draining conditions:

- DTM (Digital Terrain Model) e.g. SRTM (Koch, A. and Heipke, C., 2001), ASTER (Abd-Elmotaa et al, 2017) and the hydrologically corrected DTM, relevant to deriving the drainage network (Hydrosheds (Lehner et al, 2006)
- Perennial rivers and most relevant lakes (Hydrosheds);
- Global wetlands (Gumbricht, 2012, 2015),
- Surface water bodies temporal dynamics, as derived from the analysis of the Landsat time series (JRC global surface water) (Tulbure, M.G. and Broich, M., 2013)

Given the framework above, the decision was taken to design and implement a groundwater flow model at the Zambezi river basin scale, that would account for both freely available datasets, the BGS hydrogeological maps being the most prominent ones, and the ground data, as compiled in the mentioned databases from SADC-GMI and Banda, 2018. The compiled databases both contribute to the initial setup of the model and to the subsequent calibration and validation tasks.

2.2 Groundwater flow model setup

A first major objective was identified in relation to the willingness and need itself of critically assessing the viability of a deterministic modelling approach for such a large groundwater system. The hydrogeological system underlying the Zambezi River Basin characterizes for a huge complexity, as detailed in previous paragraphs, and particularly for a major difference in between the western and the eastern regions of the basin, resulting in highly different transmissivities. The idea was to build a model that, combining recharge and transmissivity estimates with the surface water draining bodies (rivers, lakes and humid areas), would allow a sensitivity analysis of the underlying assumptions. In perspective, hopefully based on the calibrated model, other potential challenge objectives were identified, including the assessment of:

- the spatial distribution of groundwater depth over the region. This has strong implication on groundwater accessibility and can be validated in most developed areas based on the indirect evidence of irrigated agriculture with groundwater (e.g. presence of central pivoting schemes, where no other direct information is available);

- the groundwater flow field, with all the related implications in terms of estimation of flow directions, groundwater residence times, and contaminants fate & transport. The latter is of course particularly challenge, still it can be initially addressed in a WHAT-IF scenarios perspective;

- the groundwater return flow (entity and spatial distribution) to the surface water drainage system, relevant, among others, to the conceptualization of surface water hydrological models that generally adopt a very simplified schematization of the groundwater component.

- the groundwater system behavior in the long run, under changing climate. Differently from the relatively rapid responses of surface hydrology to climate drivers, the groundwater system characterizes for a much higher latency, resulting effects being potentially detected after years. This is typically the case for salt water intrusion along coastal areas, as an effect of groundwater overexploitation or reduction of aquifers recharge. In the context of the Zambezi river basin, there is space to investigate WHAT-IF scenarios impact on groundwater availability, aquifers depletion trends, any reduction and redistribution of groundwater return flow to the drainage network.

- last but not least, the groundwater transient behavior as related to major modifications of the surface water bodies extent and level. While this is evidently an interesting opportunity in case of naturally shrinking or expanding large water bodies at regional scale (e.g. Chad lake, Aral sea), the application turns to be relevant also face to the construction of human made infrastructures. Particularly dams have a relatively immediate and easy to assess impact on surface water extent, as based on digital terrain model, and consequently on the immediate social impacts as related to loss of terrain and people displacement. On the other hand, groundwater again undergoes much slower transient modifications, that merit an in-depth analysis of their medium-long term implications.

The groundwater flow model was implemented using the finite element code and modelling/simulation environment Feflow 7.0 from the DHI (Danish Hydraulic Institute). Despite its proprietary nature, the code has been selected for the higher flexibility of the numerical finite element solution scheme compared to the well-known finite difference one, as implemented in the Open Source USGS flow and transport codes MODFLOW and MT3D. The Zambezi River Basin is the fourth wider in Africa and the largest one flowing towards the Indian ocean. With its 1,390,000 km2, the willingness to attempt at simulating the water table along many flood plains and hence the need to follow the details of a highly complex drainage network, a finite element mesh was the only viable option.

2.3 Initial model setup

The model was initially setup based on following assumptions:

- groundwater system extent was the same as the surface drainage basin; although this assumption does not necessarily hold true everywhere, both the extent of the study area and the lack of detailed aquifer geometry framework justifies this choice (see later point for the connection with the Kalahari aquifer system);

- groundwater model simulated as confined over the entire domain; again this is a simplification, that finds justification in the complexity of the groundwater system, the available information and the need to limit instability of numerical solution process;

- no flow with the external world; occasional inflow that could occur from the Kalahari aquifer system along the southern boundary, south to Barotse plain and upstream of Victoria falls, was neglected;

- fine discretization mesh (triangular elements up to 500m, later derefined to 1500 - 2000 m for computational efficiency reasons) constrained to the geometry of surface drainage network and lakes, as derived from SRTM 90m, HydroLakes and JRC global surface water (

- Figure 3);

- model nodes along major perennial rivers, lakes and humid areas (e.g. Barotse and Kafue plains) were setup as 1st Dirichlet constant head Boundary Conditions (BCs), with elevation inferred from the SRTM 90m or, where available, from attributes of the reference dataset (e.g. for HydroLakes). Dirichlet BCs resulted in potential artifacts of local relevance, as unrealistic in and out flow at local scale due to water bodies relatively close but at sensibly different elevations (Figure 4); in these cases the impossibility to capture the local hydrogeological complexity (e.g. transmissivity not capturing local nuances, presence of perched aquifers and multi-aquifer systems) may cause these anomalous flows whose relevance remain local, still being clearly reflected in the final water balance. For all the reasons above, inflow from nodes at 1st type BCs was finally inhibited (max inflow constrain set to 0 m3/d), converting them to potential (in case of higher piezometric level) drainage conditions only.

- mean transmissivity T (Figure 5) and mean effective recharge (Figure 6) of the aquifer systems was setup at mesh finite elements based on continental scale BGS estimates; it was spotted that T measurement units were erroneously reported as m2/s in BGS documentation, actually being reasonably m2/d, such an order of magnitude being consistent with ground evidences as from pumping tests. Most of the groundwater modelling sensitivity analysis was later focused on the assessment of the reliability of these parameters as a first estimate in the setup of the groundwater flow model at the river basin scale. The use of other data sources, as the hydrogeological map and the databases available for the basin, was the basis for finalizing the model sensitivity analysis and reasoning on BGS datasets challenges and opportunities when attempting a quantitative assessment.

- for the moment being, withdrawal from the aquifer system was not accounted for; estimate for spatsim should be splitted between surface water and groundwater and properly spatialized, e.g. based on pumping evidences from the presence of pivot irrigation systems for irrigated agriculture

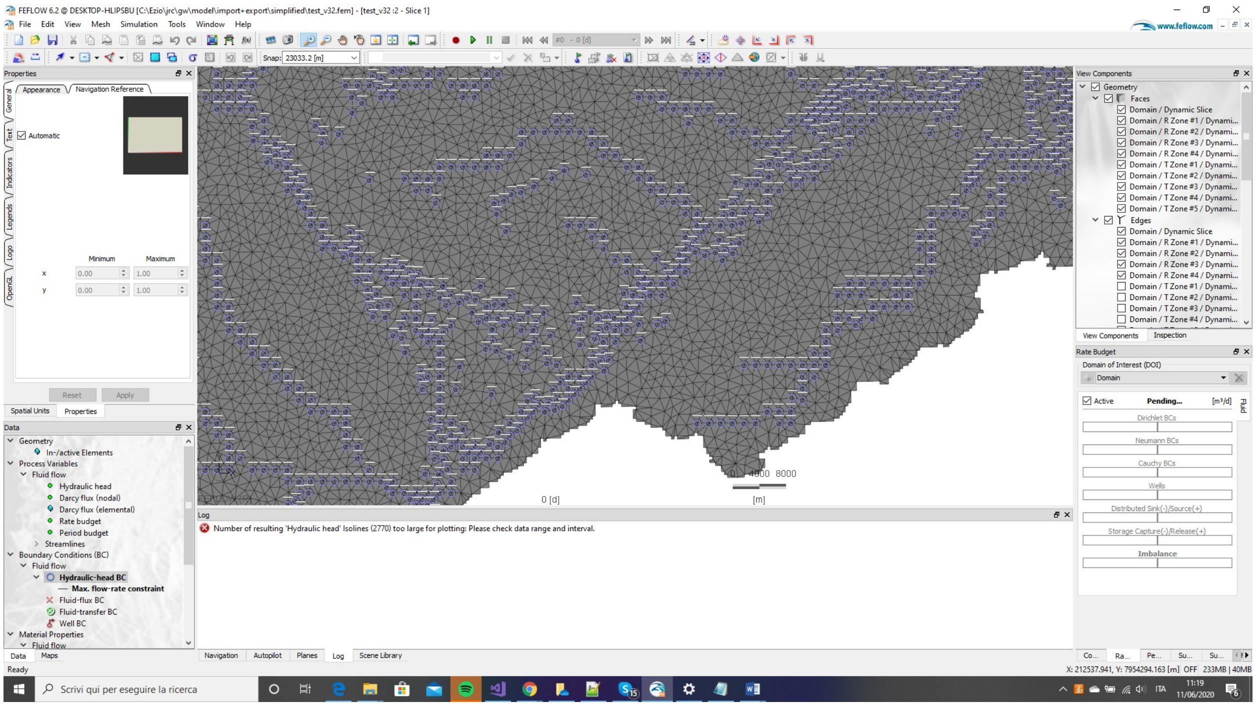


Figure 3 - Groundwater flow model: detail (upstream of Victoria falls) of discretization mesh and setup of inner BCs along the drainage river network

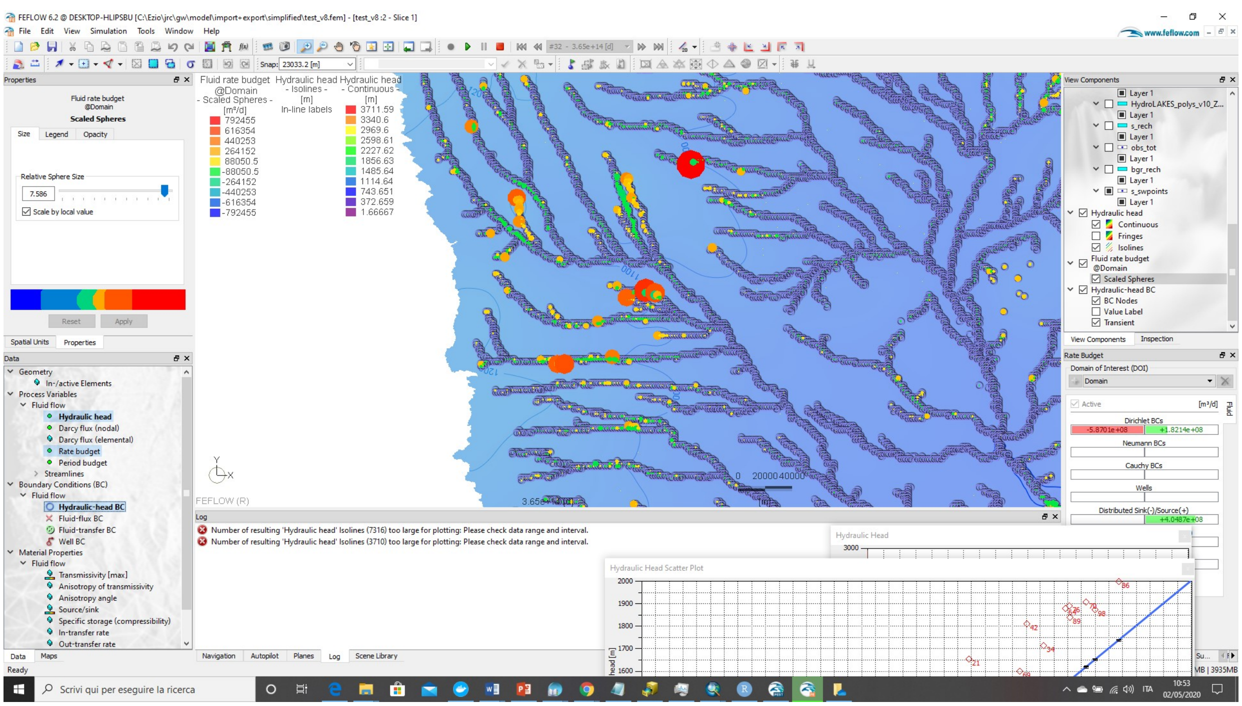
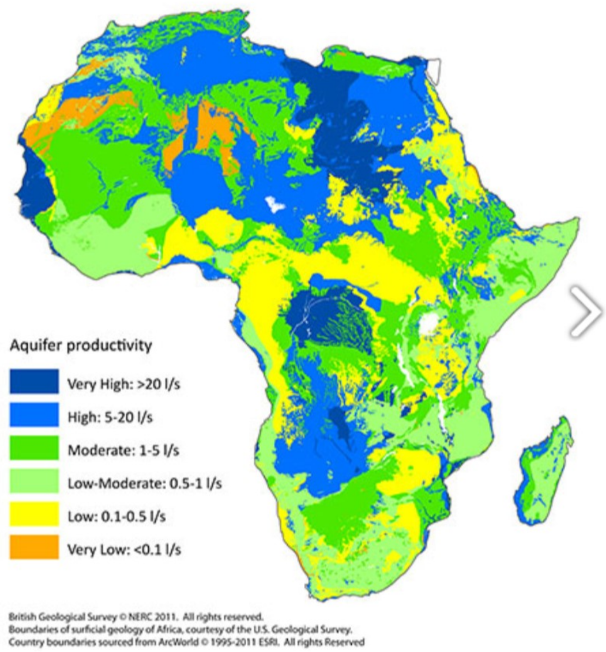


Figure 4 - Fictitious in and out flows at unconstrained 1st type BCs due to poor aquifers characterization at local scale



Code	GW productivity (l/s)	Tmin (m^2/d)	Tmax (m^2/d)
VH	>20	500	1000
H	5-20	50	500
M	1-5	10	50
LM	0.5-1	5	10
L	0.1-0.5	1	5
VL	<0.1	0.1	0.5

Figure 5 – Model initial setup: aquifer productivity and transmissivity (BGS)

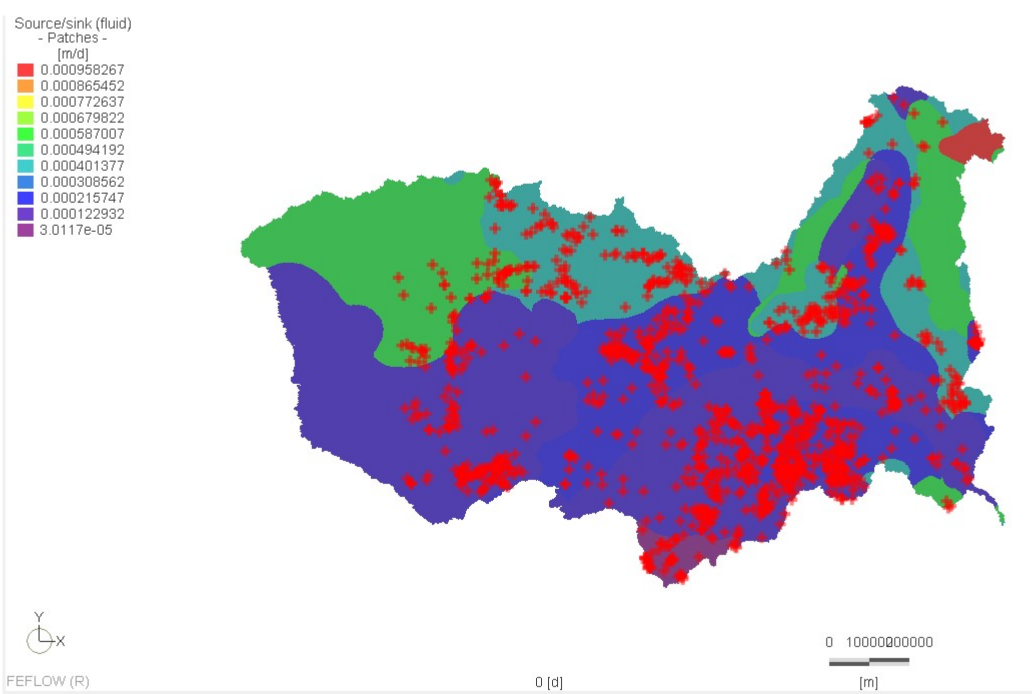


Figure 6 – Model initial setup: estimated recharge as from BGS

The observed piezometric head, as from the SADC-GMI database, locally complemented by the Zambia hydrogeological database (Banda, 2018 in few uncovered areas (e.g. Barotse plain), was used to setup observation points for model calibration purposes (Figure 7).

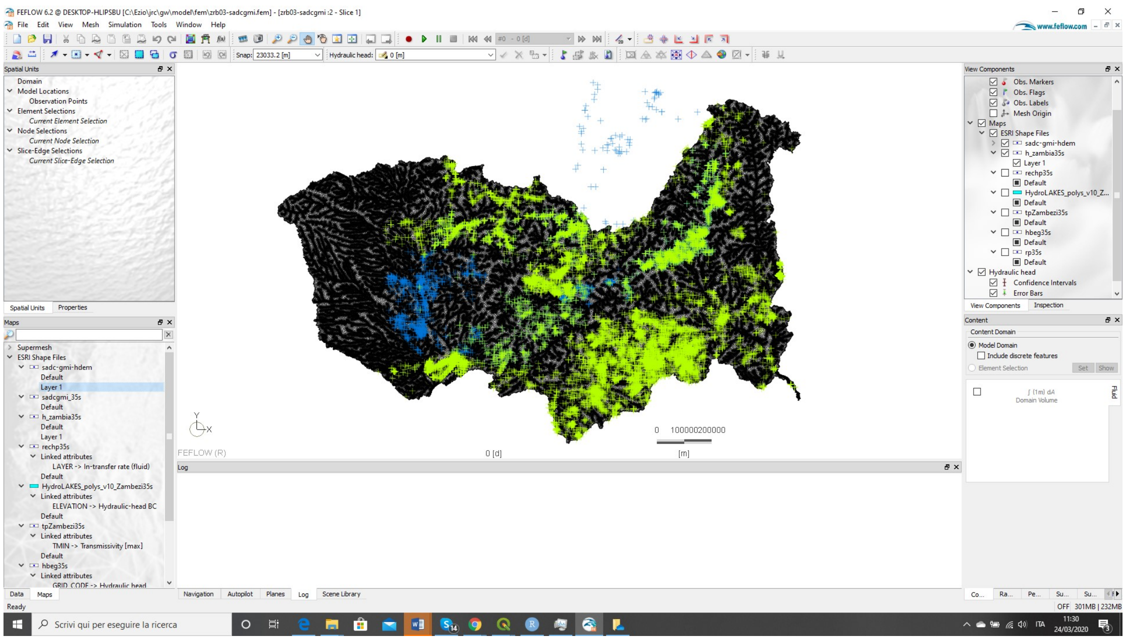


Figure 7 - Groundwater model domain, initial mesh setup (later derefined) and observed data points (yellow: SADCGMI db; blue: Banda db)

The model was run in transient conditions over a long time (1e[11] yrs) to produce a reference steady state condition. Initial piezometric head within the modelling domain was setup based on interpolation of elevation at river nodes; this choice was relevant only to the efficiency of the numerical convergence process towards the final steady state conditions, but not relevant to the final outcome.

2.4 Model Sensitivity and Calibration

The model sensitivity analysis was conducted, based on the initial setup accounting for transmissivity (with correct units m2/d) and recharge as estimated from BGS. Surface water drainage network and bodies were overall reliable and, being inflow at nodes inhibited, activated only for groundwater drainage; of course, this condition occurs only where piezometric level in the aquifer exceeds nodes elevation at surface water bodies. Any evidence of the presence of permanent water bodies and its comparison with (activated) draining nodes can be used for further validation; photographic documentation, as from Google Earth (Figure 8), confirmed that even at the most western part of the basin, active water courses occur, consistently with the extension of the drainage network in the model.

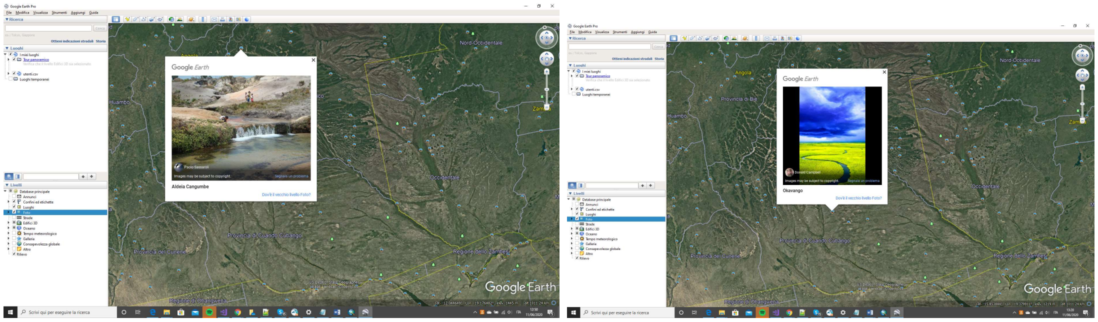


Figure 8 - Evidence of surface water bodies in the far semiarid western region of the Zambezi river basin

Manual trial and error calibration attempt highlighted that, given the initial assumptions for hydrogeological parameters as from the BGS (T, effective recharge), were not consistent with piezometric observations. Being the model confined, the computed piezometric heads were not constrained by any seepage at the topographic surface, locally rising far beyond any realistic value in the eastern (lower T) part of the basin.

The calibration process led to the conclusion that:

- T was strongly underestimated, particularly in the eastern part of the modelling domain; or
- recharge was overestimated; or
- somehow a combination of both of the above conditions occurred.

The problem is ill-posed. In absence or in the impossibility of collecting observed variables, other than the piezometric heads reported at wells, and in the light of the highly simplified hydrogeological system conceptualization, the only option was to investigate the model sensitivity to different parameters value and spatial arrangements. Expert judgement, any ground evidences or outcomes from other scientific research would help to discriminate in between realistic ranges of model parameters. Examples include: i. the T estimate deriving from pumping tests at wells, still being aware about their often limited depth compared to aquifer thickness; ii. effective recharge estimate and uncertainty assessment. In it's current form, we accept the level of uncertainty given the data scarcity.

In view of conducting further modelling assessment of medium or long term impact of key climate drivers on groundwater hydrology, two options remained opened:

- Run the simulation scenarios, based on the different outcomes of the calibration process;
- Select the most conservative conditions, as dictated by the problem statement (e.g. the calibration option resulting in lower recharge at current conditions).

In light of the inherent uncertainties of the calibration process, the error and trial manual analysis was complemented by an automated calibration process in FePEST, the Feflow interface implementation to optimization code PEST.

The analysis involved both:

- T, constrained at the lower limit of the range variability as from BGS estimate. Taking into account the evidence that T could be strongly underestimated in the eastern part of the basin, the upper limit of the T range was left quite open, up to even two order of magnitude higher than the top BGS estimate. Large fractured formations outcrop in the eastern part of the basin and groundwater circulation could effectively be controlled locally by high conductivity fractures;

- Effective recharge, generally referring to BGS estimate as the maximum value in the variability range. Effective recharge is highly uncertain, provided that spatsim hydrological model assumed far lower values than BGS estimates.

Given the framework above, one, among the many possible realizations, was setup based on estimated recharge of spatsim hydrological model at calibration (Figure 9). Such recharge was sensibly lower than BGS estimate and hence consistent with the evidences emerged from the model sensitivity analysis. The T spatial arrangement (Figure 10), as resulting from FePEST automatic inversion, captured the strong difference in between eastern and western part of the hydrogeological basin, in line with the main geological units, as from the revised SADC-GMI hydrogeological map (Mengistu, 2018;

Figure 11). The order of magnitude of T was fully aligned with BGS estimate in the western part of the basin, but it turned out to be much higher in the eastern part, always in line with evidences from the sensitivity analysis.

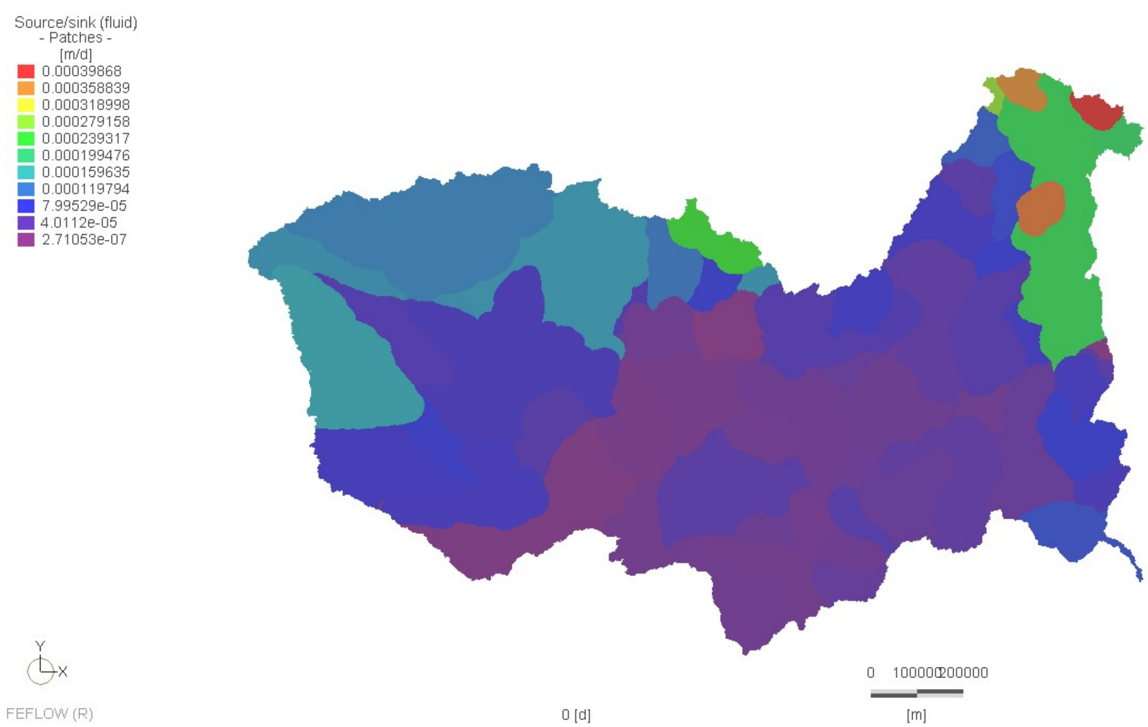


Figure 9 - Recharge over the modelling domain as from spatsim calibration

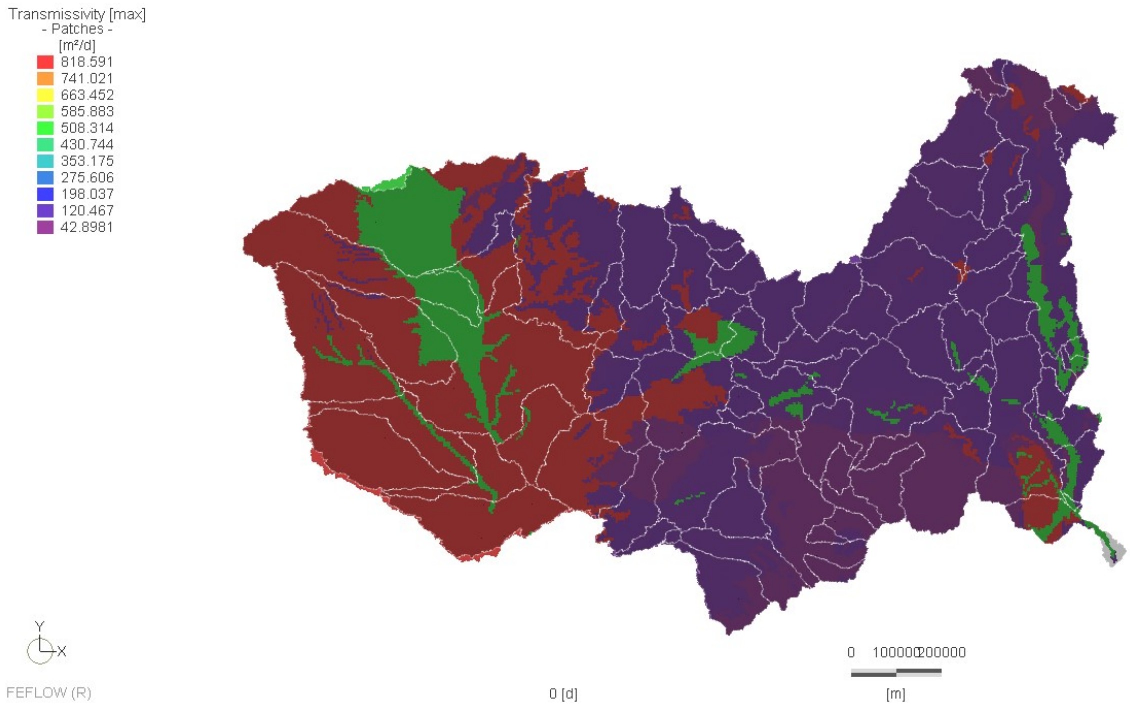


Figure 10 - Transmissivity distribution at calibration, sub-basins boundary as from spatsim setup

The outcome of the model is captured as below:

- the two maps on difference between computed and observed piezometric head, respectively in selected areas of the western (Figure 12) and the eastern (Figure 13) sectors of the river basin. The former map illustrates an overall good model fitting in the western sector; this is an expected outcome, given the thickness and continuity of the sedimentary deposits and the strong control on piezometric trends exercised by the low topographic gradient of the drainage network. The second map reports a more controversial picture, with a good fitting in main plain but complex piezometric patterns, controlled by tectonic features and clearly distinct within the different geological formations.
- the multi-facet scatter plot of computed vs. observed piezometric head, organized by different lithological units. The scatter plot clearly shows that piezometric trends are generally well captured and also spreading is relatively limited. A clear exception occurs for the basement, largely outcropping over the eastern sector, where the complex piezometric patterns has already been commented above. These rocks are deformed crystalline basement units overlain by quartzites and pelites with sparse volcanic tuffs (De Waele, 2006).

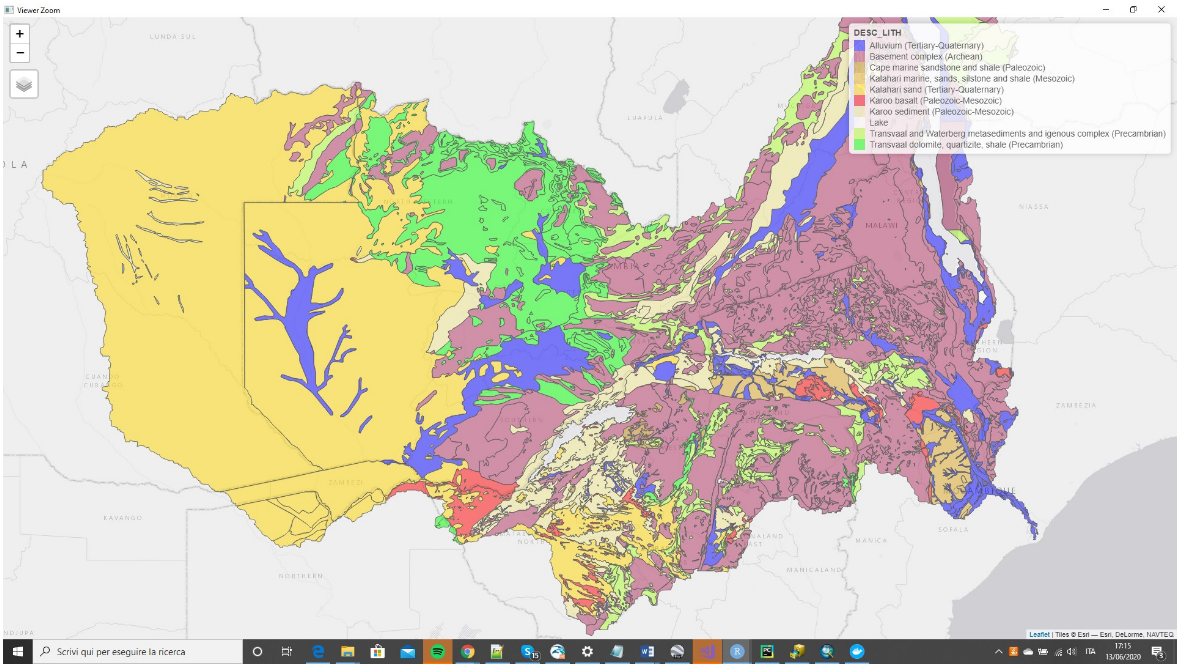


Figure 11 – Hydrogeological map (from Mengistu, 2018)

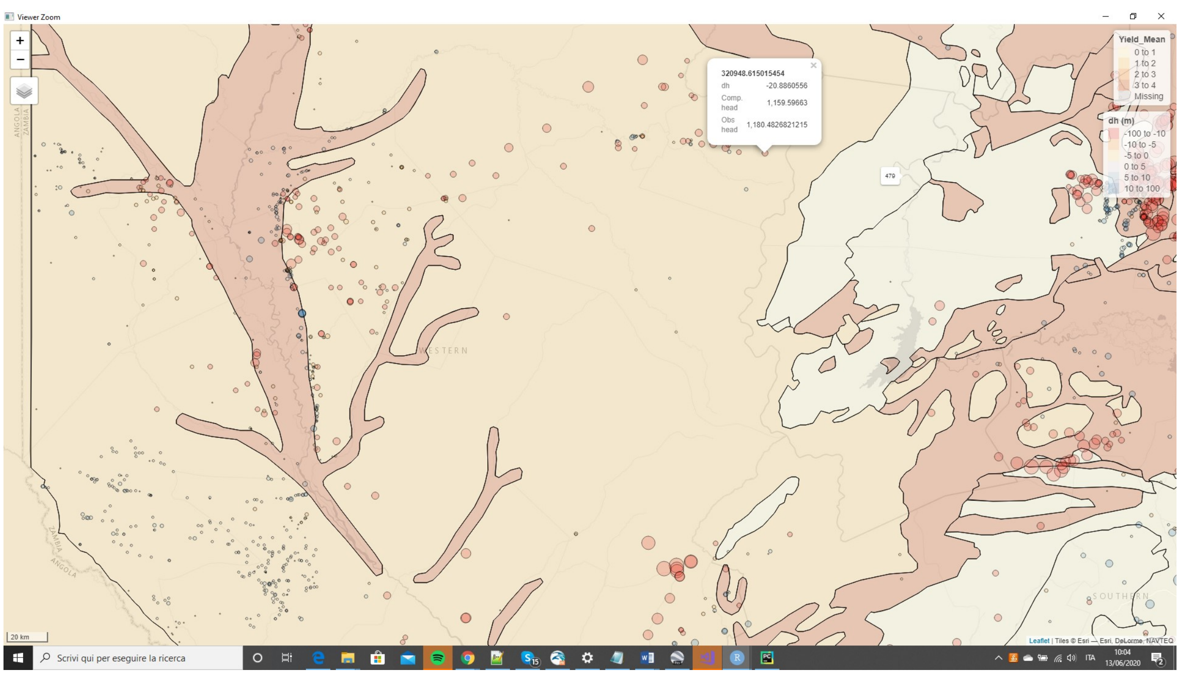


Figure 12 - Difference between computed and observed piezometric heads in the western sector of the basin

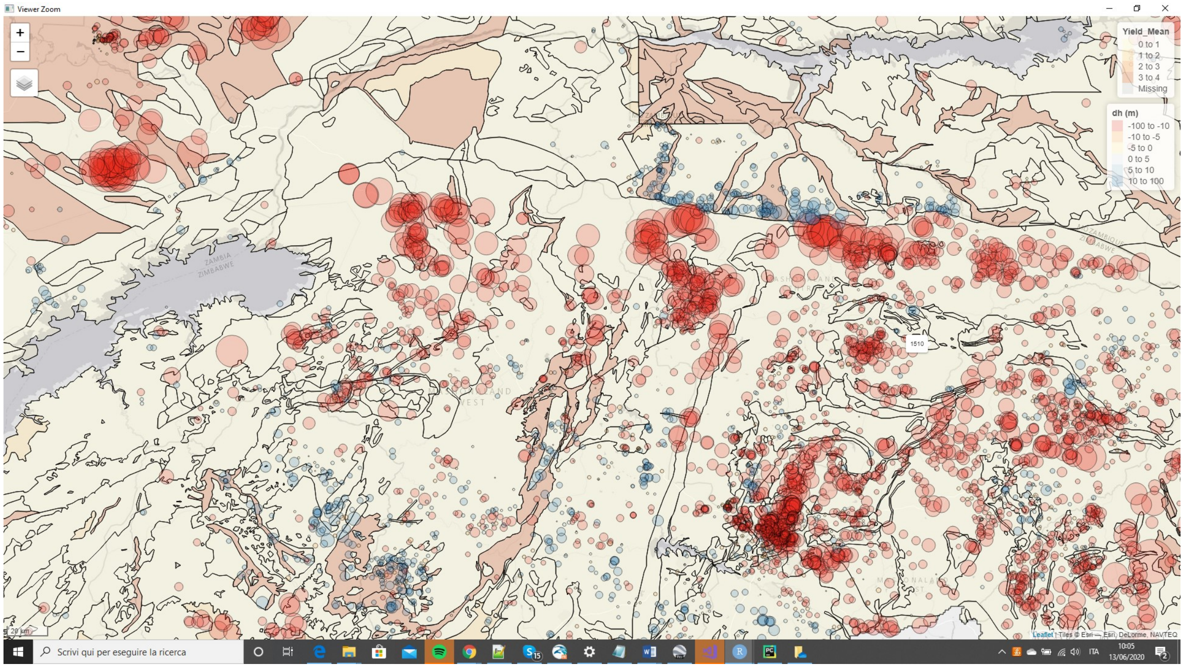


Figure 13 - Difference between computed and observed piezometric heads in the western sector of the basin

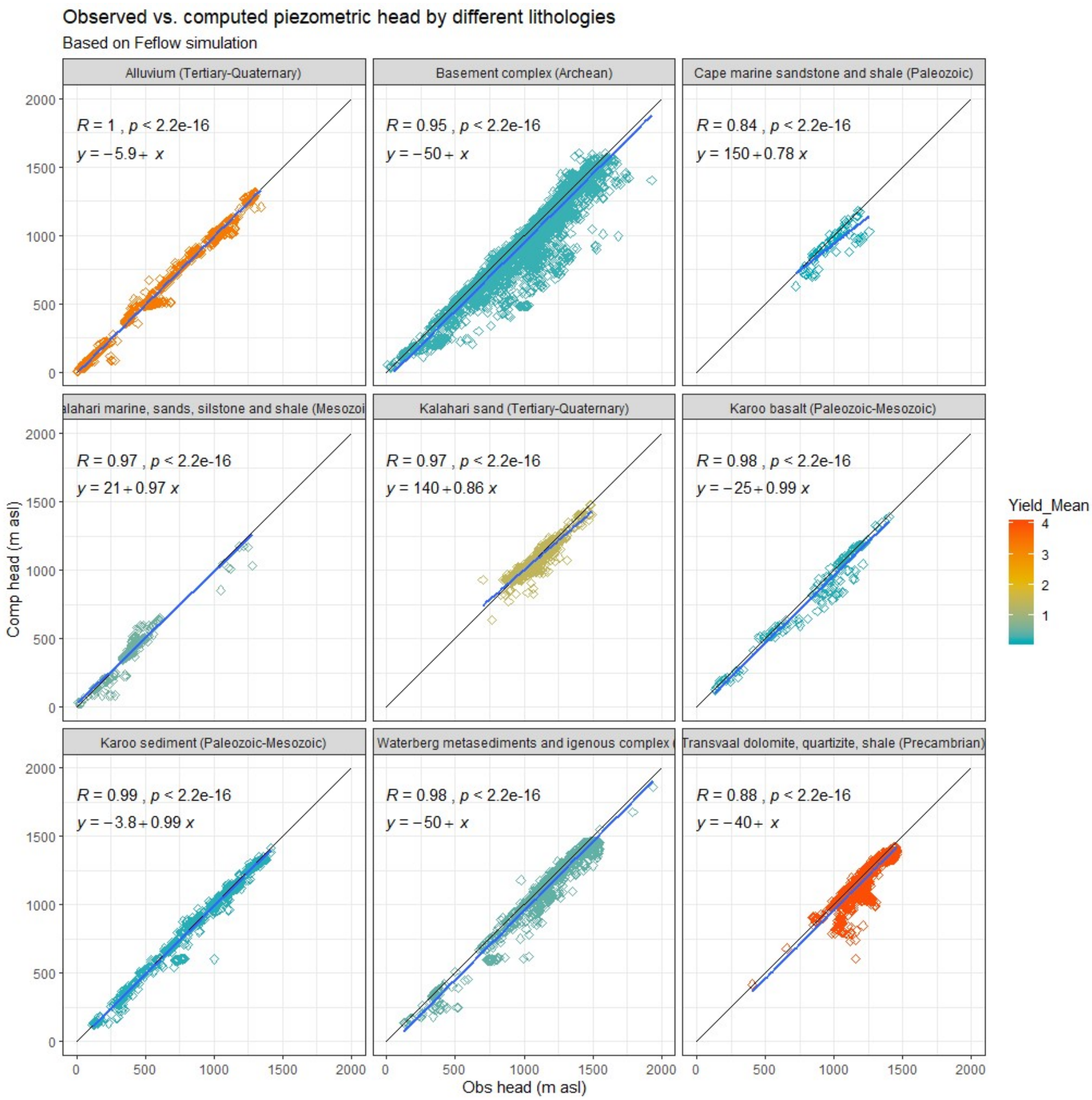


Figure 14 – Observed vs. computed piezometric head at calibration

3. Conclusions and recommendations

The following are key conclusions of this work:

- Freely available global or continental hydrogeological databases can be used to setup groundwater flow models at regional scale in Africa, as the current case study for the transboundary Zambezi river basin clearly demonstrates. These resources include, among others, the quantitative hydrological and hydrogeological maps, as from the BgR (recharge) and the BGS (British Geological Survey) (recharge, transmissivity, specific yield, aquifer thickness), as well as topographic information, as from SRTM DTM and Hydroshed.

- Both finite difference codes, as the USGS Modflow and Modpath, and finite element codes, as the DHI-WASY Feflow, can be used to setup this kind of models. Still, face to complex large scale geological, hydrogeological and tectonic settings, the finite element formulation grants a higher discretization flexibility, face to the need to closely follow the drainage network and the relevant features at hydrogeological discontinuities (e.g. faults, fractures, contact between formations). On the other hand, the finite element formulation guarantees a rough mesh resolution in the most remote or less known regions, hence optimizing the overall number of elements and improving the effectiveness of the numerical solution.

- The numerical flow modelling initial attempt based on the large scale datasets clearly reveals major shortcomings. The large scale hydrological and hydrogeological datasets capture the key differences or trends; this is the case, in the Zambezi river basin, for the reducing rainfall rates (and aquifers' recharge) from the N to the S, or the clear distinction between the high transmissivity thick porous media alluvial deposits in the western basin vs. the large outcrops of fractured basement rocks with much lower transmissivity in the eastern basin. Still, the highly simplified hydrogeological conceptualization implies a high uncertainty in hydrological and hydrogeological parameters. In the Zambezi river basin case study, overestimation of effective recharge and/or underestimation of transmissivity in the eastern sector is highlighted, based on initial setup simulations.

- The availability of ground based data, as from existing hydrogeological databases, is key to improve groundwater model conceptualization and to critically review the parameters as initially inferred from the large scale datasets. In the Zambezi river basin case study, the SADC-GMI hydrogeological database geographically extended over most of the Southern Africa, the hydrogeological map, compiled databases and analysis conducted in the framework of the ACEWATER2 project (Mengistu, 2018; Banda, 2018; Chinyama and Makaya, 2018) provided a key support to model sensitivity analysis and calibration; among others, the observed piezometric heads have been used to support model calibration. Literature review of ground based hydrogeological parameters estimate (e.g. transmissivity deriving from pumping tests) contributed to investigate a much more scattered picture than the BGS estimate would do at large scale and definitely to quantitatively support the assumptions emerged at model calibration stage.

- The numerical problem is an ill-posed one, calibration mainly relying upon observed piezometric heads only and expert judgement as for the hydrogeological parameters distribution. As such it leads to a not unique solution. While the drainage network is generally well captured, both in terms of location and water elevation, all other parameters (e.g. recharge, transmissivity) are highly uncertain and potentially variable, even highly variable over short distances. Hence automatic inversion attempts potentially lead to an ensemble of quite different solution, all equally likely. This is for sure a major limitation, but it should not prevent from investigating the system behaviour based on a set of different model realizations, to distil general lessons learned, or, alternatively, make use of the most conservative outcome given a specific problem at hand (e.g. lower groundwater flow). Any other independent information, as deriving from geophysical prospecting, groundwater dating and hydrochemical characterization (the SADC-GMI also reports quality parameters, as salinity) potentially leads to a better hydrogeological conceptualization, helping in constraining parameters to more realistic variability ranges.

- With all the limitations above in mind, a groundwater flow model still can support many different and valuable tasks: i. spatial exploratory analysis, to investigate hydrogeological parameters ranges, based on different equally likely distinct assumptions of groundwater system conceptualization; although not necessarily leading to a unique solution, the model can be used to critically investigate and sort out alternative solutions as related to the specific task at hand (e.g. from a cautelative perspective) or distil relevant information through multiple solutions analysis; ii. to analyse interactions between surface water and groundwater, including return flow entity and spatial distribution over the drainage network; iii. to support surface hydrology model parameterization, as it is the case for the spatsim hydrological model in the Zambezi river basin case study; iv. to investigate medium and long term implications of climate change scenarios on groundwater hydrology.

- Overall the groundwater flow model implemented over the Zambezi river basin well captures the observed piezometric trends in the western part of the basin, that is not surprising given the aquifer nature, thickness and strong control exercised by the surface water drainage network. The picture is more scattered and controversial in the eastern part of the basin, due to higher geological and tectonic complexity. Still, despite higher spreading of computed vs. observed heads, the gradient trend is overall captured. The lower regions of the Barotse Flood Plains, around the Kariba, Luangwa and delta seem to have a significant contribution of groundwater. We postulate, these sites are groundwater dependant ecosystems

The recommendations of this study are the following:

- further work is required to validate water balance relationship between the FEFLOW model and the hydrological model SPATISM

- further analysis with FEPEST to further explore the model sensitivity to both recharge and transmissivity and overall impact on model calibration.

- further validation of the proposed groundwater dependant ecosystems with insitu data such as environmental isotopes, time series data or baseflow assessment from literature records where available.

4. References

Abd-Elmotaa, H.A., Makhloof, A., Abd-Elbaky, M. and Ashry, M., 2017. The African 3 ″× 3 ″DTM and its validation. In *International Symposium on Gravity, Geoid and Height Systems 2016*, pp. 79-86. Springer, Cham.

Banda, K., 2018. Country scale groundwater hydrology for Zambia. Technical report for the ACEWATER II Project, Supported by EU JRC.

Balek j, 1977. Hydrology and water resources in tropical Africa. First edition. Elsevier Science, Amsterdam.

Bäumle, R., Himmelsbach, T. and Noell, U., 2019. Hydrogeology and geochemistry of a tectonically controlled, deep-seated and semi-fossil aquifer in the Zambezi Region (Namibia). *Hydrogeology Journal*, *27*(3), pp.885-914.

Beilfuss, R. and Santos, D.D., 2001. Patterns of hydrological change in the Zambezi Delta, Mozambique. *Program for the sustainable management of Cahora Bassa Dam and the Lower Zambezi Valley*.

Chandiwana SK and Snellen W.B., 1994. Incorporating a human health component into the integrated development and management of the Zambezi river basin. Report of a Panel of Experts on Environmental Management for Vector Control mission to Zimbabwe, Zambia and Mozambique. World Health Organization, Geneva, Switzerland.

Chinyama and Makaya, 2018. Country scale groundwater hydrology for Zimbabwe. Technical report for the ACEWATER II Project, Supported by EU JRC.

De Waele, B., Kampunzu, A.B., Mapani, B.S.E. and Tembo, F., 2006. The mesoproterozoic Irumide belt of Zambia. *Journal of African Earth Sciences*, *46*(1-2), pp.36-70.

Diersch, H.J.G., 2009. DHI-WASY Software FEFLOW Finite Element Subsurface Flow and Transport Simulation System. *Reference Manual: DHI-WASY GmbH. Berlin*.

Doell P., Fiedler K., 2008. Global-scale modeling of groundwater recharge. Hydrology and Earth System Sciences Discussions, European Geosciences Union,12 (3), pp.863-885.

Gumbricht, T., 2012. Mapping global tropical wetlands from earth observing satellite imagery. Working Paper 103. CIFOR, Bogor, Indonesia.60 pp.

Gumbricht, T., 2015. Hybrid mapping of pantropical wetlands from optical satellite images, hydrology, and geomorphology. *Remote sensing of wetlands: applications and advances*, p.435.

Koch, A. and Heipke, C., 2001. Quality assessment of digital surface models derived from the Shuttle Radar Topography Mission (SRTM). In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)* (Vol. 6, pp. 2863-2865). IEEE.

Leenaers H., 1991. Estimating the Impact of Land Use Change on Soil Erosion Hazard in the Zambezi River Basin. International Institute for Applied Systems Analysis-IIASA Working Paper 90-024. Austria.

Lehner, B., Verdin, K. and Jarvis, A., 2006. HydroSHEDS technical documentation, version 1.0. *World Wildlife Fund US, Washington, DC*, pp.1-27.

Lekula, M., Lubczynski, M.W. and Shemang, E.M., 2018. Hydrogeological conceptual model of large and complex sedimentary aquifer systems–Central Kalahari Basin. *Physics and Chemistry of the Earth, Parts A/B/C*, *106*, pp.47-62

MacDonald, A. M., Bonsor, H. C., Dochartaigh, B. É. Ó., & Taylor, R. G., 2012. Quantitative maps of groundwater resources in Africa. *Environmental Research Letters*, *7*(2), 024009.

Mengistu, 2018. Groundwater hydrology for the Zambezi River Basin. Technical report for the ACEWATER II Project, Supported by EU JRC.

Pietersen K., N. Kellgren, O. Katai and Roos M, 2010. The SADC Hydrogeological Map & Atlas: Towards an improved understanding of groundwater regimes in Southern Africa. *International Conference "Transboundary Aquifers: Challenges and New Directions"(ISARM2010). 6-8 December.*

Pollock, D.W., 2012. User guide for MODPATH version 6: a particle tracking model for MODFLOW. P. 58. US Department of the Interior, US Geological Survey.

Tulbure, M.G. and Broich, M., 2013. Spatiotemporal dynamic of surface water bodies using Landsat time-series data from 1999 to 2011. *ISPRS Journal of Photogrammetry and Remote Sensing*, *79*, pp.44-52.

Wirkus, L. and Böge, V., 2006. Transboundary water management on Africa's international rivers and lakes: current state and experiences. *Transboundary water management in Africa*, p.11.

Yachiyo Engineering Co. Ltd.,1995 The study on the national water resources master plan in the Republic of Zambia, Final Report -Supporting Report [B] Hydrogeology. Japan International Cooperation Agency & Republic of Zambia, Ministry of Energy and Water Development, Lusaka, pp 1–82

ANNEX 1: R-Script used for the database curation and analysis of results

```
# Program: calib.r
# Author:  E. Crestaz & K.Banda
# Date:    2020-03-05
# Scope:   Support Feflow groundwater flow model calibration assessment
#          over the Zambezi river basin.
#          In order to facilitate the use of the program, snippets of code are
#          provided to introduce and document key processes
# References:
#
# Resources:
# https://ryanpeek.org/mapping-in-R-workshop/vig_spatial_joins.html
# https://gis.stackexchange.com/questions/312413/plotting-spatial-points-data-in-r
# https://www.r-spatial.org/r/2018/10/25/ggplot2-sf.html
# https://rgeomatic.hypotheses.org/1205
# https://rdrr.io/cran/fasterize/man/fasterize.html
#
library(rgdal)
library(gdalUtils)    # To use GDAL functions call
library(dplyr)
library(tidyr)
library(rgeos)
library(raster)       # To support raster management and crs
library(sf)           # Support to OGC vector simple features
library(shiny)        # Library to develop web applications (as needed for tmaptools)
library(shinyjs)      # Same as above...
library(units)
library(tmap)         # Note installation requires (install.package ('tmaps', type = 'binary')
library(tmaptools)    # Tool to obtain colors' palette
library("ggplot2")    # Advanced plotting
library("ggspatial")  # Advanced geospatial data management extending ggplot2
library("rnaturalearth")
library("rnaturalearthdata")
library("cartography")
library("ggpubr")
library("fasterize")  # Fast rasterization of sf vector data
library(leaflet)
library(widgetframe)
library(RColorBrewer)

# Check the version of a specific library. Cartography gt 2 supports sf library
packageVersion("cartography")

############################
# Load Feflow model elements
############################

# Feflow Dirichlet (constant heads) internal boundary conditions have been set by
# integrating the drainage network inferred after the hydroshed raster dataset and
# the existing lakes
```

```
# River points, derive from the SRTM raster and resampled at about 400-500m distance
setwd("D:\\JRC_EU\\GW_model\\Calibration")
hriv <- st_read("rp35s.shp")
# Not needed as the input shape file above is already projected in WGS84 UTM35S
#p4s <- "+proj=utm +zone=35 +south +datum=WGS84 +units=m +no_defs"
#st_crs(hriv) <- p4s
head(hriv)
#plot(hriv, pch=3, cex=0.2)  # It takes long time (not to be run each time)

# Lake polygons from HydroLAKES used to spatially integrate the rivers drainage network
#setwd("D:\\JRC_EU\\GW_model\\Calibration")
hlake <- st_read("HydroLAKES_polys_v10_Zambezi35s.shp")
plot(hlake)

##################################################################
# Load observation points and Feflow results using the sf package
##################################################################

# Observation points as from Kawawa Banda hydrogeological database
# over the Zambia section of the river basin (cleaned after the SADC-GMI and expanded)
# Alternatively the original SADC-GMI database that has been integrated with the geological
# information (see later for how the full database has been compiled)

setwd("D:\\JRC_EU\\GW_model")

#obs <- st_read("results/obs.shp")          # db Kawawa for Zambia
obs <- st_read("Refined/obs_Feflow.shp")    # combined db for the entire SADC region + Kawawa db.
attr(obs, "sf_column")
print(obs[3:6], n = 3)
methods(class = "sf")
class(obs)
plot(obs)

# Check for and assign the proper projection system (WGS84 UTM 35s)
 # The original file was not assigned a projection reference system

st_crs(obs) <- 32735

st_crs(obs)  # After assigning the reference system

# Quick thematic mapping of observed data, based on tmap library
tmap_mode(mode = "view")   #"plot"  # Set dynamic view map or static map
qtm(obs) +
  tm_legend(show = FALSE)



# Zambezi river basin

zrb <- st_read("zambezi35s.shp")
```

```
# Oservation points from SADC-GMI
# This is a comprehensive database extending over the entire Zambezi river basin
# It has not been cleaned, but, further to the geographical scope, it has a detailed set of variables
# NOTE: this dataset was not used as calibration control points in Feflow

# Options below must be executed before loading the XLConnect. Still the file is very huge and
# even 16Gb is not enough. Hence a csv file is created and read in
#options(java.parameters = c("-XX:+UseConcMarkSweepGC", "-Xmx8192m"))
#library(XLConnect)  # Attention: Java and R version must be for the same platform (e.g. x64bits)
#setwd("C:/Ezio/jrc/gw/model/import+export/sadc-gmi")
#wb <- loadWorkbook("gip_BHdata_mozambique.xlsx")  #

# Note: in the original MsExcel file, the ';' have been substituted with '|' before exporting to csv
# The csv file itself has the ';' as the fields separator
# Piezometric head (m asl) must be computed based on ground elevation and gw depth
# Relevant attributes:
#   - Elev_m_dem (estimated from dem?)
#   - Elev_m_sub (ground based measurement?)
#   - Bh_depth_m   Borehole depth
#   - Swl_m   Groundwater level from surface (in m)
#   - Swl_date   Date when groundwater level has been measured
#   - Yield_l_s
#   - ec_ms_m  Electrical conductivity

obs_gmi <- read.csv(file = 'h_gmi.csv', sep=';') # entire SADC GMI database
head(obs_gmi)

prj4 <- "+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
obs_gmi <- st_as_sf(x = obs_gmi, coords = c("long_wgs84", "lat_wgs84"), crs = prj4)
class(obs_gmi)
plot(obs_gmi)
str(obs_gmi)
# Modify attributes' types
obs_gmi$country <- as.character(obs_gmi$country)
obs_gmi$compltdate <- as.Date(obs_gmi$compltdate)
obs_gmi$swl_date <- as.Date(obs_gmi$swl_date)
obs_gmi$yieldtype <- as.character(obs_gmi$yieldtype)
obs_gmi$lith_subm <- as.character(obs_gmi$lith_subm)
str(obs_gmi) # check the character type update

# Transform coordinates from geographic WGS84 to projected UTM 35S
obs_gmi <- st_transform(obs_gmi,32735)

# Add three attributes: 1. piezometric heads h (m asl) based on dtm;
#                       2. piezometric heads h (m asl) based on sub attribute (ground data?)
#                       3. difference between the two estimated h in order to assess their accuracy
# Data reported as -9999 are not available, hence the derived attribute is set to NA
obs_gmi <- obs_gmi %>%
  mutate(hdem = ifelse(elev_m_dem == -9999 | swl_m == -9999, NA, elev_m_dem-swl_m))%>%
  mutate(hsub = ifelse(elev_m_sub == -9999 | swl_m == -9999, NA, elev_m_sub-swl_m)) %>%
  mutate(hdelta = ifelse(is.na(hdem) | is.na(hsub), NA, hdem-hsub))
```

```
# Show only few attributes. Note library naming conflict is solved by including library reference
obs_gmi %>% dplyr::select(elev_m_dem,swl_m,hdem,hsub,hdelta)
head(obs_gmi %>% dplyr::select(elev_m_dem,swl_m,hdem,hsub,hdelta))

# Select only points falling within the Zambezi river basin, before saving
# https://stackoverflow.com/questions/50144222/how-to-mark-points-by-whether-or-not-they-are-within-a-polygon
# Note that the operation below is computationally demanding (at least 5' on a i7/16GB RAM). Once
# the output is saved, no need to rerun again
obs_gmi <- obs_gmi %>% mutate(inZRB = lengths(st_within(obs_gmi, zrb)))
str(obs_gmi)
obs_gmi$inZRB <- as.logical(obs_gmi$inZRB)  # Convert numeric to boolean


# Compute frequency histogram of difference between the two alternative estimated piezometric heads
# (after the dtm and the sub attribute)
tmp <- obs_gmi$hdem-obs_gmi$hsub
tmp <- tmp[!is.na(tmp)]
length(tmp)
hist(tmp,breaks=seq(from = -5000, to = 5000, by = 1),xlim=range(-50,50),
    main='Histogram of differences between estimated well elevation (dtm vs. sub)')


##################################################################################################################
# The SADC-GMI database can be updated with the one from Kawawa Banda (Zambia)
# For example areas as the Barotse plain lack of data in the SADC-GMI db, while they are
# available in the Kawawa db. Records are appended and source origin kept trace of

obs_kawawa <- st_read("h_zambia35s.shp")

# Create obs points code (not available in the original database), naming being consistent
# with SADC-GMI database convention, and information source reference
obs_kawawa$FID <- paste0('kawawa',seq(1, dim(obs_kawawa)[1]))
obs_kawawa$FID <- as.factor(obs_kawawa$FID)
obs_kawawa$source <- 'Banda db'
length(obs_kawawa)

obs_kawawa$X <- NULL # remove unneccessary columns
obs_kawawa$Y <- NULL
obs_kawawa$SLICE <- NULL



colnames(obs_kawawa) <- c('swl_m','swl_date','elev_m_dem','hdem','geometry','FID','source')

obs_kawawa <- obs_kawawa %>% mutate(inZRB = lengths(st_within(obs_kawawa, zrb)))
str(obs_kawawa)
obs_kawawa$inZRB <- as.logical(obs_kawawa$inZRB)  # Convert numeric to boolean
sum(obs_kawawa$inZRB == FALSE)  # Count how many points in the Kawawa db falls outside the basin
```

```
# Create a raster of distances from points in SADC-GMI db to select points from the
# database of Kawawa that falls far away enough. The objective is to keep on board data
# in the Barotse plain that are not in the original SADC-GMI db while preserving from
# creating duplicates.
r <- raster(zrb, res = 1000)
r <- fasterize(zrb, r, field = "BASIN_ID", fun="sum")
plot(r)
values(r) <- NA

# Raster distance computation is time-consuming. You may need to use the already saved file
tmp <- as_Spatial(obs_gmi)  # Create a temporary spatial point data frame to be used below
rtmp1 <- distanceFromPoints(r, tmp)

plot(rtmp1)
plot(as_Spatial(zrb), add=TRUE)
plot(as_Spatial(obs_gmi), add=TRUE, pch = 1, size = 0.02, col="blue")
plot(as_Spatial(obs_kawawa), add=TRUE, pch = 2, size = 0.02, col="red")

outdir <- "D:\\JRC_EU\\GW_model\\Calibration"
outfile <- "obs_gmi_dist.tif"
rf <- writeRaster(rtmp1, filename=file.path(outdir, outfile), format="GTiff", overwrite=TRUE)

# Based on the raster of distances from points in the SADC-GMI database, query the distance
# layer at the points in the Kawawa db, assign the distance to the points and then filter out
# all points that are too close to current points in SADC-GMI database
# https://gis.stackexchange.com/questions/271268/assigning-raster-values-to-spatial-point-using-r
dist <- extract(rtmp1,as_Spatial(obs_kawawa))
sort(dist, decreasing=TRUE)
obs_kawawa$dist <- dist
obs_kawawa <- obs_kawawa %>% filter(dist>2000)

# Append records of Kawawa db to SADC-GMI db. Recall that only points in Kawawa db being
# enough away from those in SADC-GMI db have been retained. Given that the SADC-GMI db
# is richer in information content, common attributes are attached and the other ones set to
# NULL
obs_gmi$source <- 'SAD-GMI db'

# Binding sf objects with different number columns can be challenging. Actually common columns
# have been named consistently. Dataframes (without) the geometry are bind together by row,
# not existing attributed in the second dataframe being set to NA. The geometry is extracted
# and added back as a new attribute
#https://gis.stackexchange.com/questions/242395/is-there-an-equivalent-of-dplyrbind-rows-or-the-old-
plyrrbind-fill-for-spa?rq=1

df1 <- data.frame(obs_gmi)
geo <- df1$geometry
df1$geometry <- NULL

df2 <- data.frame(obs_kawawa)
geo <- c(geo,df2$geometry)
df2$geometry <- NULL
df2$swl_date <- NULL
```

```
df <- bind_rows(df1, df2)
df$geometry <- geo

obs_tot <- st_as_sf(df)

obs_tot$source <- as.factor(obs_tot$source)

class(obs_tot$source)
levels(obs_tot$source)

# Build map with watershed boundary and point locations integrated after the 2 dbs
plot(st_geometry(zrb))

plot(as_Spatial(obs_tot),cex = 0.35,col=c('red','blue')[as.numeric(obs_tot$source)], add=TRUE)



##############################################################################
############
# Create a subset of randomly selected observation points to check for obs points in Feflow
# Actually the test was not relevant at all, as the obs. vs. compt. difference bars show up
# anyway, indipendently from the number of points or the length of their label

#tmp <- obs_tot
#tmp <- tmp[2001:2050,]
#tmp$id <- seq(1:dim(tmp)[1])
#plot(st_geometry(zrb))
#plot(st_geometry(tmp), add=TRUE)
#
#tmp <- as(tmp, 'Spatial')
#dsn <- 'C:/Ezio/jrc/gw/model/import+export'
#outfile <- 'tmp'  # .shp extension must be omitted
#writeOGR(obj=tmp, dsn=dsn, layer= outfile, driver="ESRI Shapefile", overwrite_layer=TRUE)
##############################################################################
############
# The attempt to write the sf using st_write results in an error! Hence the sf is converted to
# a spatial point data frame and saved by using the writeOGR function
# Issue to be further investigated.
# Hence the new file sadc-gmi-h.shp contains computed piezometric heads in m asl, ready for being
# used as a reference calibration layer in Feflow

# Note: only observation points falling within the Zambezi river basin boundaries are selected
# otherwise loading the points to Feflow would take too long

obs_gmi_hdem <- obs_gmi %>% filter(!is.na(hdem) & inZRB)
tmp <- as(obs_gmi_hdem, 'Spatial')
dsn <- 'D:\\JRC_EU\\GW_model\\Calibration'
outfile <- 'sadc-gmi-hdem2'  # .shp extension must be omitted
writeOGR(obj=tmp, dsn=dsn, layer= outfile, driver="ESRI Shapefile", overwrite_layer=TRUE)
```

```
# Really not too interesting, hsub being not clear what it is!
#obs_gmi_hsub <- obs_gmi %>% filter(!is.na(hsub) & inZRB)
#tmp <- as(obs_gmi_hsub, 'Spatial')
#dsn <- 'C:/Ezio/jrc/gw/model/import+export'
#outfile <- 'sadc-gmi-hsub'  # .shp extension must be omitted
#writeOGR(obj=tmp, dsn=dsn, layer= outfile, driver="ESRI Shapefile", overwrite_layer=TRUE)

# Save the entire observation points db (SADC-GMI + part of Kawawa db)
# Note: only observation points falling within the Zambezi river basin boundaries are selected
# otherwise loading the points to Feflow would take too long
obs_tot <- obs_tot %>% filter(!is.na(hdem) & inZRB)
plot(st_geometry(zrb))
plot(st_geometry(obs_tot),col='red',add=TRUE)
plot(st_geometry(obs_tot),col=obs_tot$source,add=TRUE,legend=TRUE)



# Save the complete database, ready for being imported as observation points to Feflow
tmp <- as(obs_tot, 'Spatial')
dsn <- 'D:\\JRC_EU\\GW_model\\Calibration'
outfile <- 'obs_tot'  # .shp extension must be omitted
writeOGR(obj=tmp, dsn=dsn, layer= outfile, driver="ESRI Shapefile", overwrite_layer=TRUE)

obs_tot$FID <- as.integer(obs_tot$FID) # set the FID as an integer
str(obs_tot)


##############################################################################
#######################

# TO BE MOVED TO A FUNCTION TO ASSIGN PROJECTION SYSTEM... (FEFLOW ISSUE WHEN
EXPORTING)
# Assign projection reference system to a shape file in output from Feflow, as it does not
# save it!, then save the file back to the original
#setwd("C:/Ezio/jrc/gw/model")
#tmp <- st_read("results/obs-tot.shp")
#p4s <- "+proj=utm +zone=35 +south +datum=WGS84 +units=m +no_defs"
#st_crs(tmp) <- p4s
#tmp <- as(tmp, 'Spatial')
#dsn <- 'C:/Ezio/jrc/gw/model/results'
#outfile <- 'obs_tot'  # .shp extension must be omitted
#writeOGR(obj=tmp, dsn=dsn, layer= outfile, driver="ESRI Shapefile", overwrite_layer=TRUE)

# Building obs_gmi or obs_tot above can be challenging from the computational point of view,
# hence the spatial data can be read back from the shape file, stored to disk
# Just in case you also need to modify attributes types, as in the next
#setwd('C:/Ezio/jrc/gw/model/import+export')
#obs_gmi_hdem <- st_read("sadc-gmi-hdem.shp")
# Modify attributes' types
#obs_gmi_hdem$country <- as.character(obs_gmi_hdem$country)
#obs_gmi_hdem$compltdate <- as.Date(obs_gmi_hdem$compltdate)
#obs_gmi_hdem$swl_date <- as.Date(obs_gmi_hdem$swl_date)
```

```
#obs_gmi_hdem$yieldtype <- as.character(obs_gmi_hdem$yieldtype)
#obs_gmi_hdem$lith_subm <- as.character(obs_gmi_hdem$lith_subm)


# Load observation points exported from Feflow. This is the database of SADC-GMI complemented
# with data from the kawawa db to fill the gaps (not all data)
# ATTENTION: not clear, but probably when more points fall within the same discretization
# element, only one of them is retained. The observation points created in Feflow and that
# will match the file of output (results) do not necessarily contain the same number of
# records of the input file. Suggestion: have a look at the files in Feflow
#setwd('C:/Ezio/jrc/gw/model/results')

obs <- st_read("results/Obs_V12.shp") # observation from Feflow
dim(obs)


# If the spatial reference system has not been set, no need to update the shape file
# Assign crs based on basin boundary coverage, knowing that the spatial ref systems are the same
obs <- obs %>% st_set_crs(st_crs(zrb))
obs$LABEL <- as.numeric(levels(obs$LABEL))[obs$LABEL] # convert label to numeric
str(obs)
obs <- obs %>% dplyr:: rename(FID = LABEL) # rename


# Attach the complete database to the Feflow observation points, removing records with empty
# geometries
# NEED TO CHECK THAT THE FILES ARE ALIGNED

tmp <- obs_tot
tmp$FID <- as.numeric(tmp$FID)
str(tmp)
st_geometry(tmp) <- NULL
class(tmp)
#tmp <- tmp %>% dplyr::rename(LABEL = FID)

obs <- obs %>% full_join(tmp, by = "FID")

b <- st_is_empty(obs)
obs <- obs %>% filter(!b)


##################################################################################
################

# Upload Feflow results

##################################################################################
#################
# It is a pity, but the Feflow-saved shape file reports obs vs. comp heads as they
# would appear in a scatter plot. Hence, the geometry has no meaning at all, except
# capturing the coordinates in the scatter plot itself. Also a clear identifier to
# join the output with the observation points location is lacking. The good point is
```

```
# that records are in the same sequence, hence the columns can be simply attached to
# the observation points (cbind in following statement)
#setwd('C:/Ezio/jrc/gw/model')
#results <- st_read("zrb03_results.shp")         # Results to be compared with Kawawa db
#results <- st_read("zrb03-sadcgmi_results.shp")  # Results to be compared with SADC-GMI db

results2 <- st_read("results\\GW_V12_results.shp")
#results2 <- st_read("zrb03-sadcgmi_results.shp")

head(results2)
plot(results2)  # Obs vs. computed scatter plot (not really of interest)
# Bind Feflow results to the observation points location (note: you can check that
# the join is consistent having a look at the attributes)
obs <- cbind(obs,results2)
head(obs)
str(obs)

# Rename attributes and clean the table by removing not relevant attributes
# ATTENTION: note that FREF_RANGE is not always defined in the Feflow output file
#colnames(obs) <- c("X","Y","ELE","LABEL","REF_PARID","REF_RANGE",
#                  "ID","hobs","hcomp","geometry","geometry1")  #??
#colnames(obs) <-
c("X","Y","ELE","LABEL","REF_PARID","REF_VALUE","REF_RANGE","SADCID","NATID","BH
Type","country","Datum")



obs$geometry.1 <- NULL
obs$ELE <- NULL
obs$REF_PARID <- NULL
obs$REF_VALUE <- NULL

obs <- obs %>% dplyr::rename(hobs = X,hcomp = Y.1)

# Add columns reporting (i). the diff of computed and observed heads;
# (ii). if the computed value is higher than the observed one
# This example uses the dplyr syntax
# https://datacarpentry.org/R-ecology-lesson/03-dplyr.html
obs <- obs %>% mutate(dh=hcomp-hobs,dhabs=abs(hcomp-hobs),above=FALSE) %>%
  mutate(above = replace(above, dh >= 0, TRUE))

# Build a map of observed points and countries' boundaries
# Note that, being sf objects, the geometry must be passed in calling the st_geometry function
world <- ne_countries(scale = "medium", returnclass = "sf")
world35s <- st_transform(world, 32735)

plot(st_geometry(obs), pch=3, cex=0.5, col="blue")
plot(st_geometry(world35s), border="orange", add=TRUE)
plot(st_geometry(zrb), border="gray", add=TRUE)
#plot(st_geometry(hriv), pch=1, cex=0.2, col="grey", add=TRUE)   # Just half a million points!

# Load geology from the Un. of Western Cape
```

```
geo <- st_read("SADC_Geologyrevised.shp")
geo35s <- st_transform(geo,32735)
plot(st_geometry(geo35s), border="gray")
plot(obs,add=TRUE)
str(geo35s)

# Join properties of geological formations to the observation points
obs <- st_join(obs,geo35s)
head(obs)
str(obs)

# Plot dynamically linked multifaced syncronized maps of different attributes spatial distribution
# of the geological map
tmap_mode("view")
#tmap_mode("plot")  # This is for static map plotting

# Two maps of aquifer type and mean yield, with observation points location
tm_shape(geo35s) +
  tm_polygons(c("AQU_TYPE", "Yield_Mean")) +
  tm_facets(sync = TRUE, ncol = 2) +
  tm_shape(obs) +
  tm_dots(c("hcomp"))

# A unique map of mean yield (based on geological map) and observation points
# Points popup window report piezometric heads (computed, observed and difference)
tm_shape(geo35s) +
  tm_polygons(c("Yield_Mean"), alpha=0.10) +
  tm_facets(sync = TRUE, ncol = 1) +
  tm_shape(obs) +
  tm_dots(c("hobs"),
          popup.vars=c("Computed head"="hcomp","Observed head"="hobs",
                       "Delta head"="dh","Mean yield"="Yield_Mean"))


# A final map of mean yield (based on geological map) and observation points
# Points popup window report piezometric heads (computed, observed and difference)
# Points size is proportional to the dh, but both size and color scheme would need
# to be further refined

# Note that using the OSM will result in a misalignment with the shape files
# Suggestion is to use EPSG 3857 (WGS 84 / Psuedo Mercator aka "Web Mercator")
# https://gis.stackexchange.com/questions/234658/qgis-misaligned-shape-with-google-maps-osm

# Run the web application below to investigate the different color palettes and their names
#tmaptools::palette_explorer()

# Example of function call to obtain a colors' vector given a palette
# This is the perfect diverging scale to be used around the zero
#get_brewer_pal("RdYlGn", n = 6, plot=TRUE)
bubble_palette <- get_brewer_pal("RdYlGn", n = 6, plot=FALSE)

# Build a custom palette, aimed at differentiating trends of positive vs. negative dh
```

```
pal_red <- get_brewer_pal("-YlOrRd", n = 3, plot=FALSE)
pal_blue <- get_brewer_pal("Blues", n = 3, plot=FALSE)
bubble_palette <- c(pal_red,pal_blue)

tm_shape(geo35s) +
  tm_polygons(c("Yield_Mean"), alpha=0.25, legend.hist = TRUE) +
  tm_layout(legend.outside = TRUE) +
  tm_facets(sync = TRUE, ncol = 1) +
  tm_shape(obs) +
  tm_bubbles(col = "dh",              # Color mapped to piezometric head difference
           size = "dhabs",          # Size proportionality to piezometric difference
           scale = 1/2,             # Size symbol rescale
           border.col = "black",
           border.alpha = .5,
           style="fixed",
           breaks=c(-100,-10,-5,0,5,10,100),  # Note one break more than colors in palette
           #palette="-RdYlGn", contrast=1,  # The hyphen - inverts the colorimetric scale
           palette = bubble_palette,
           midpoint = TRUE,
           title.size="Metro population",
           title.col="dh (m)",
           alpha=0.2,
           popup.vars=c("dh"="dh","Comp. head"="hcomp", "Obs head"="hobs"),
           legend.hist = TRUE) +
  tm_scale_bar(position=c("left", "bottom")) +
  tm_layout(legend.outside = TRUE)

# https://stackoverflow.com/questions/41940403/popup-on-a-shape-using-tmap
# How to implement finer control on popup windows

# Multi-facet scatter plots hobs vs. hcomp by geological units
# https://www.datanovia.com/en/lessons/ggplot-scatter-plot/
# https://rpkgs.datanovia.com/ggpubr/reference/stat_regline_equation.html

str(obs)
tmp <- obs %>% drop_na(REG_LITH)   # Remove NAs for the attribute used in facet_wrap

b <- ggplot(tmp, aes(x = hobs, y = hcomp), na.rm = TRUE) +
  #geom_point(color = "#00AFBB", size = "Yield_Mean", shape = 23) +
  #geom_point(color = "#00AFBB", size = 2, shape = 23, alpha=0.5) +
  geom_point(aes(color = Yield_Mean), size = 2, shape = 23) +   #, alpha=0.5) +
  scale_color_gradientn(colors = c("#00AFBB", "#E7B800", "#FC4E07")) +
  theme(legend.position = "right") +
  # Set minimum and maximum coordinates (m asl), fix ratio and add convergence line
  xlim(0, 2000) +
  ylim(0, 2000) +
  coord_fixed() +
  geom_segment(aes(x = 0, y = 0, xend = 2000, yend = 2000)) +
  #scale_color_manual(values = c("#00AFBB", "#E7B800", "#FC4E07")) +
  #scale_size(range = c(0.5, 12))  # Adjust the range of points size
  geom_smooth(method = lm, se = FALSE) +
  #geom_smooth(method="auto", se=TRUE, fullrange=FALSE, level=0.95) +
```

```
  labs(title="Observed vs. computed piezometric head by different Aquifer Types", subtitle="Based on
Feflow simulation",
       y="Comp head (m asl)", x="Obs head (m asl)") +
  facet_wrap(~hgm_aqtype) +    #REG_LITH hgm_aqtype
  #facet_wrap(~ Yield_Mean) +
  stat_cor(label.y = 1800.0) +
  stat_regline_equation(label.y = 1600.0)

# Print multi-faced diagram
b

#...............................................................................
# Groundwater model error metrics

library(hydroGOF)
library(zoo)

gof(sim= obs$hcomp, obs= obs$hobs)

Sta = gof(sim= obs$hcomp, obs= obs$hobs)
write.csv(Sta,'modelstatistics.csv')

#ggof(sim= tmp2$hcomp, obs= tmp2$hobs, ftype="dm", gofs = c("NSE", "rNSE", "ME", "MSE",  "d",
"RMSE", "PBIAS"), FUN=mean)
```