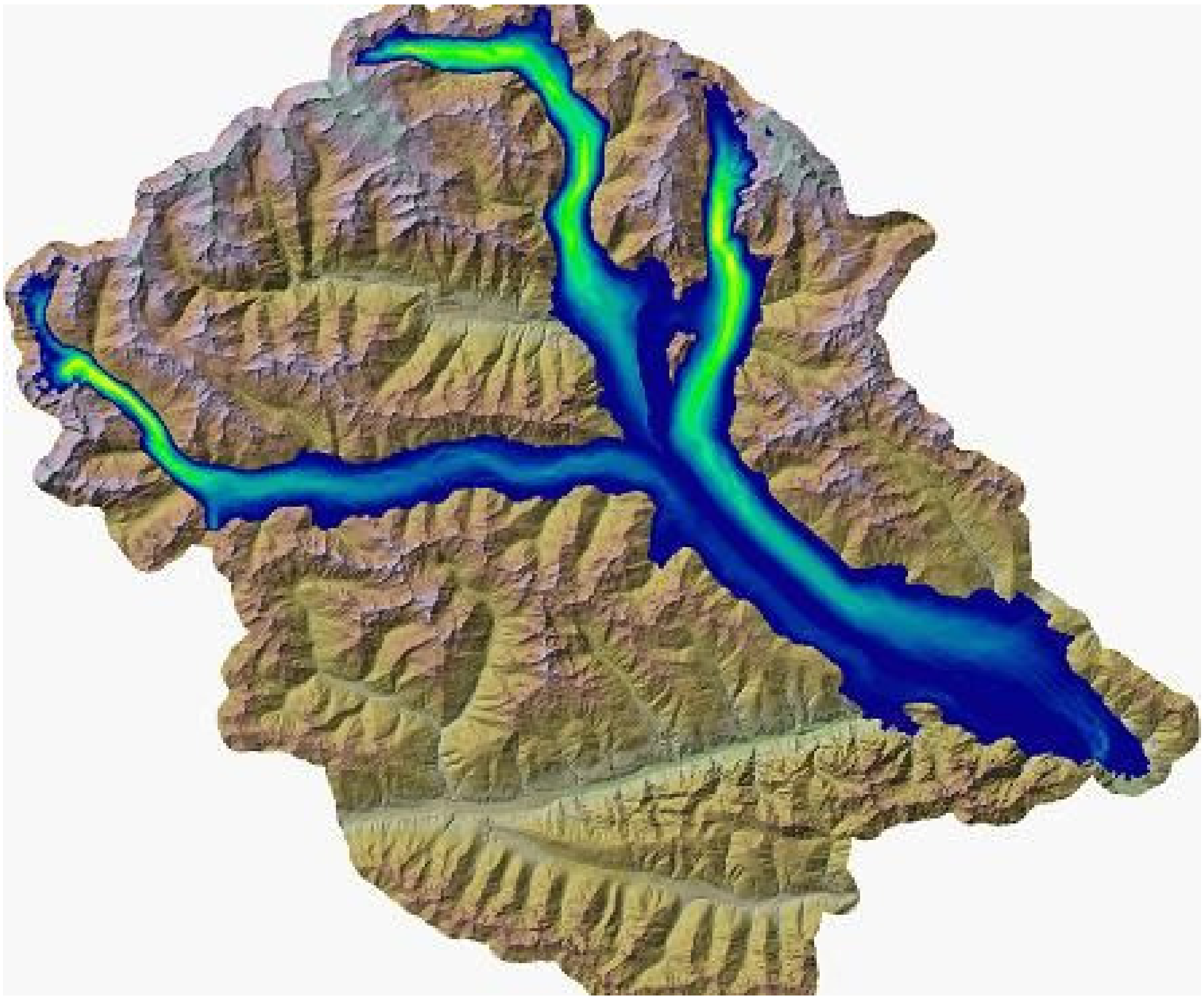
EMMC case study:

TerraMath - software solutions for geological and environmental applications

Interview with Dr Robert Faber, TerraMath

Writers: Alexandra Simperler and Gerhard Goldbeck



About Robert Faber

Dr Robert Faber (https://www.terramath.com/) wrote his doctoral thesis about software development for the synoptic analysis of geological and geophysical at the University of Vienna. In 2001 when Dr Faber worked on an FWF[1] project on the Vienna Basin, numerical simulation of erosion and sedimentation was very limited, available software being restricted to 2D sections.

[1] The Austrian Science Fund (FWF) is Austria's central funding organization for basic research.

Thus, Dr Faber developed the methods he needed by himself and started to work on the 3D visualisation and analysis software of digital elevation, satellite and other geologically relevant datasets using C/C++ and OpenGL. In 2003, still in Austria, he founded TerraMath, a company specialized in software and (later) sensor solutions for geological and environmental applications. In 2010 he moved his company to Indonesia. Since then, his focus has extended to field data recording solutions (hard- and software) for research projects, which included projects for gathering seismic data, data from volcanic sites and even wild/domestic animal monitoring. Besides his career as software owner, Dr Faber is engaged in providing English and STEM education to Indonesian children. As he is discontent with often incomplete information in the media about how to behave during earthquakes, Dr Faber offers education and seminars to aid people to react in the right way in dangerous situations.

Essentials about data and modelling in Geology

Similar to materials modelling, geology works in four dimensions, 3D space and time. However, the geologist's laboratory is the whole Earth and geological formations of interest can cover 1000s of km^2 and cannot be translated (or just in a very simplified way) into small lab-size units.

For geologist, modelling is particularly important because getting field data is expensive and not always possible (missing permissions, difficult terrain). Even not very deep drilling (100m) might exceed the funds available in a research project. Drilling several kilometres costs easily millions of €. Sometimes the oil or mining industry can aid with sharing data if these are not business critical. However, drilling happens at particular spots or where wells are needed which may not be the area of interest of a research geologist. Seismic tomography is another technique for imaging the subsurface of the Earth and can provide more data; however, although cheaper than drilling – in comparison to the area / volume covered in one campaign - it is also expensive and mainly operated by the oil and gas industry and costs cannot be easily covered by a government looking into contamination patterns of their water supply. Road building uncovers certain layers of ground, but geologists often tend not to be involved at building sites to collect data. Also, in the latter case, the geologist has no influence on where the data can be taken. There are of course more methods available but, in all cases, they are too expensive and work intensive to get a complete image of the subsurface – therefore the need for interpolation.

A peculiarity of geology is that relevant timeframes can span over millions of years. Thus, geological data once taken is valid for a long time. It is, however, important to learn about their quality. Methods and knowledge improved continuously and it is not uncommon that data from different times even contradict each other and sometimes this happens even with interpretations of the same age done by different workgroups. Drilling cores may give information about grain sizes, mineral composition, deposition environment and chemistry of different layers and tell the relative age of layers with respect to each other. Exact dating would require measuring of radioactive isotopes like ^{14}C (only for very young sediments up to 50,000 years) and $^{40}K/^{40}Ar$ or palaeontologists who can analyse microorganisms and index fossils. So, if data from core drilling is used for model calibration, the modeller has to be aware that the age given of different layers by reference data may not be a very reliable parameter; as sediment sequences are not always complete, may change laterally and are easy to be mixed up especially if the environment under which the sediments were deposited was not changing much (and therefore look very similar).

To get hold of data, geologists will try to get project funding to acquire the necessary data (satellite data, field surveys, core drilling data, geophysical measurements) and work on joint projects with the oil or mining industry. There are also databases, provided by organisations such as the British Geological Survey[2], the U.S. Geological Survey[3], which is a primary source of geographic information system (GIS) data, and regional ones such as the Geologische Bundesanstalt in Austria[4]. They provide elevation data, geological maps, etc. Publications provide geochemical and geophysical data such as hardness and other descriptors. This data is not always digitised.

Geology data are complex and data of other disciplines may be needed to accomplish work. For example, the parameters that can influence the path of drinking water from precipitation [1] to tap, cover everything from geological layers, to behaviour of rivers at different sections and seasons and sources of pollution and much more. Of importance are also meteorological data as they provide climate tables and thus, information about weathering, erosion and vegetation.

Essentials about TerraMath Software

If geologists want to do modelling, they are confronted with discrete data they have to interpolate to continuous data. It is tempting to use standard mathematical or statistical interpolation methods between data points, and indeed, this is what most often is done. Data points comprise the spatial position and associated property values, and a prohibitive amount of data would be needed to get realistic results. Figure 1 shows a time layer of a digital twin of a river, when only a low density of data points is used and standard interpolation techniques. Having in mind that drillings and other available methods are most often just point data and will most often not pass through the actual river channel – we will not be able to see a continuous river channel in a standard interpolation.

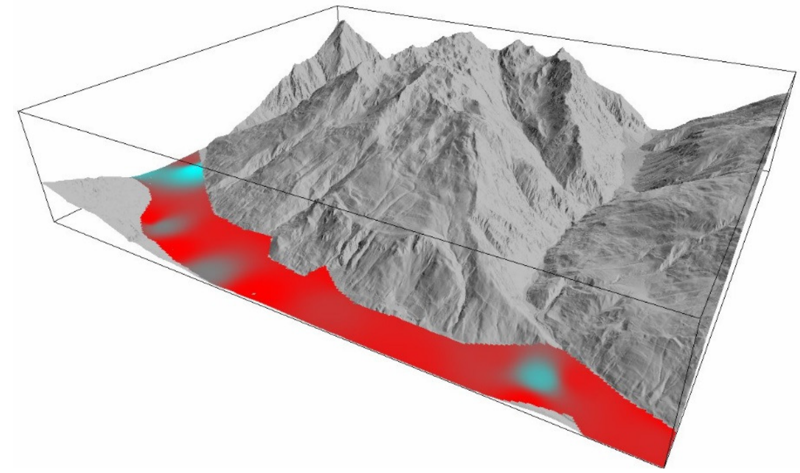


Figure 1. Negative effects of low density input data: Digital twin of a river, created with low density data sampling and standard interpolation techniques. Blue areas mark paleo river channel sediments and red areas other sediments.

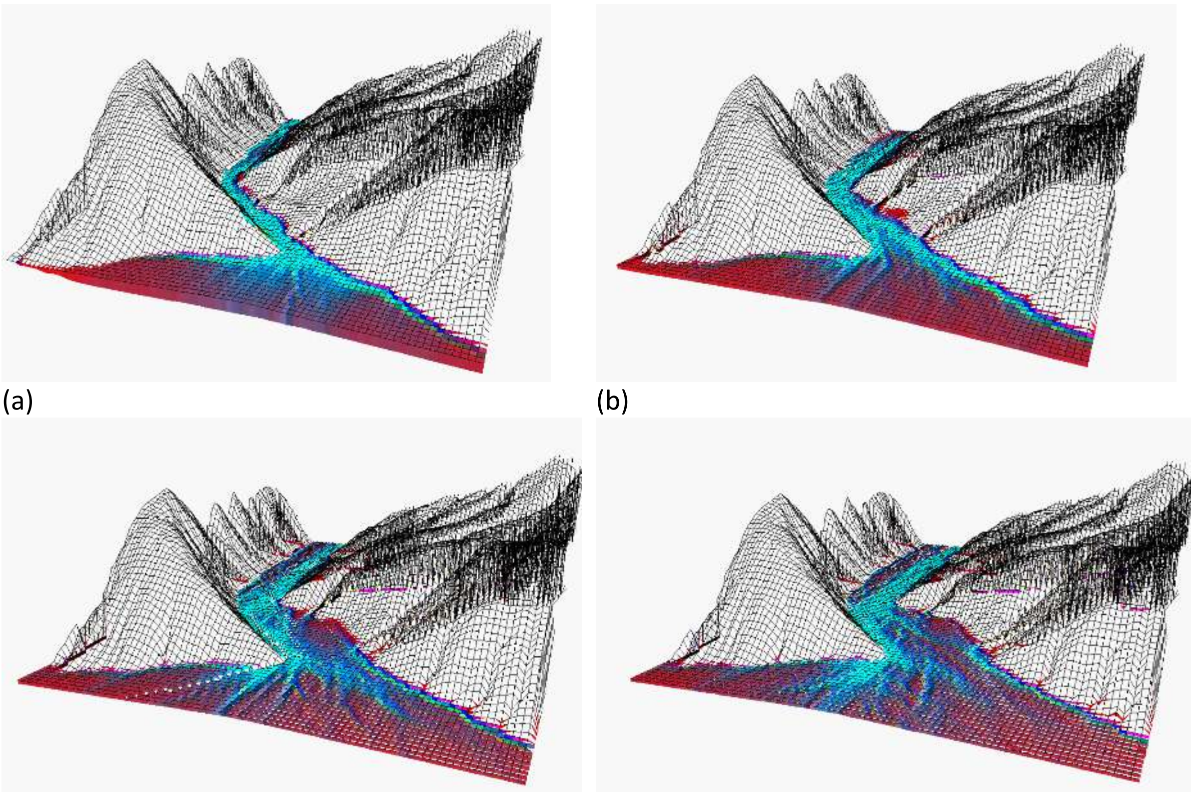
Therefore the geological modeller will look at the strata[5] and set it up as homogeneous layers with average permeability. However, in reality the materials have a granularity and are very complex so there is an attempt to use software as WinGeol/SedTec that can aid to provide a digital twin which is more realistic and can go beyond a simplified layer one (Fig. 2 a-d).

Dr Faber's WinGeol is a tool for visualisation and analysis of geological or geology related datasets and is written in C/C++. Generally, Dr Faber prefers standard languages such as C, C++, C#, Java, and

[2] https://www.bgs.ac.uk/
[3] https://www.usgs.gov/
[4] https://www.geologie.ac.at/
[5] A stratum (plural: strata) is a layer of sedimentary rock or soil, or igneous rock that were formed at the Earth's surface.

Python with as little 3rd party function libraries and tools as possible. WinGeol overlaps with standard geographic information system (GIS) software as it comprises layer bases data organisation, permits usage of vector, raster and geology database data, offers tools for data manipulation, and aids the digitisation of vector and tabular data. WinGeol exceeds standard software (see Figures 2a-d) by offering FaultTrace, a tool for determining azimuth and dip of geological strata and faults from digital elevation models[6] and optional satellite/aerial photography data – in this way it allows us to extend the structures visible at the surface to the subsurface. SubsurfaceModeller, is a tool to create a mesh; i.e. the volume of a geological structure is divided into cells (the mesh). This can be used for subsequent numerical groundwater modelling.

A major module within WinGeol is SedTec, which can simulate erosion and deposition in dependency on topography, tectonic movements, lithological[7] properties and sea-level.



(a) (b) (c) (d)

Figure 2. Complex digital twin of a river over time; red areas show bad/low and blue ones good/high flow properties/permeability

Dr Faber's goal is to provide functions to create as realistic digital twins of geological structures as possible. He uses VOXEL based approaches which are extended to store all relevant data for each location and not only at the volume surface. Further, calibration against real data is very important to make data more reliable. The granularity a geological model can reach, depends on the existence and accuracy of calibration data.

Data imported to the software are ASCII tables and formats like well logs, strike/dip, 2D sections (seismic or drawings) or lithological surface data and also data derived from AutoCAD 3D solid can

[6] A digital elevation model (DEM) is a 3D computer graphics representation of a terrain's surface.

[7] Lithology: the study of the general physical characteristics (colour, texture, grain size, and composition) of rocks.

be imported. For the actual modelling data are required to be converted into binary as they are easier to handle for computation. The algorithms used in the software are finite difference methods.[2]

About the Case Study

The case is based on workflows similar to

"Modelling of topography and Sedimentation along syn-sedimentary faults: WinGeol/SedTec" Faber,R.; Wagreich, M.; Austrian Journal of Earth Sciences 97(2005)60-66.

"Climate as main factor controlling the sequence development of two Pleistocene alluvial fans in the Vienna Basin (eastern Austria) – a numerical modelling approach" Salcher, B.C.; Faber, R.; Wagreich, M.; Geomorphology 115 (2010) 215-227. DOI: 10.1016/j.geomorph.2009.06.030

For this particular case, which were your objectives?

We want to model sedimentary processes and use digital twins to test geological hypotheses and simulate real world cases. For example, a hydrogeologist has interest in us providing a geological 3D mesh of an area of interest. They would like to know if water contamination happens, how it will distribute spatially over time. They also would like to be able to pose "what if scenarios" to these digital twins, e.g. assigning a variation of permeability values or changing erosion and sedimentation parameters to see possible variations. Other scientists are interested in the geomorphological / landform development - what happens over several years or many thousands or millions of years, when layers erode under different climate conditions.

For this Case study did you create the input structure?

We would use data from meteorological institutes about the climate and geological data from national institutes such as U.S. GIS, NASA, and from the EU and Japan, dependent on the geographical area we look at. The colour codes of geological maps are linked to look up tables containing data of the respective types of rock formation. Based on the provided information we generate a 3D input structure of the geology which is an extended VOXEL model (3D pixel) which is able to store additional information in each cell like: quality of interpolation, strike/dip, lithology, etc., depending on the focus and needs of the actual project. The concept of a VOXEL model storing orientation data can be seen in Figure 3a, as well as a mesh we can build using them in Figure 3b.

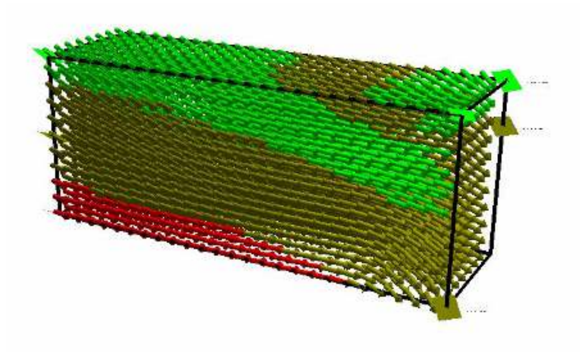


Figure 3a. A small volume interpolated using WinGeol based on just a few input points (represented by the square elements) – every Voxel contains orientation and rock type information (therefore the VOXEL is shown as a coloured arrow; level of confidence is calculated as well but not displayed here). This is used as an input for a further erosion / deposition simulation using SedTec. But can be already a valuable input by itself (e.g. for groundwater flow simulations).

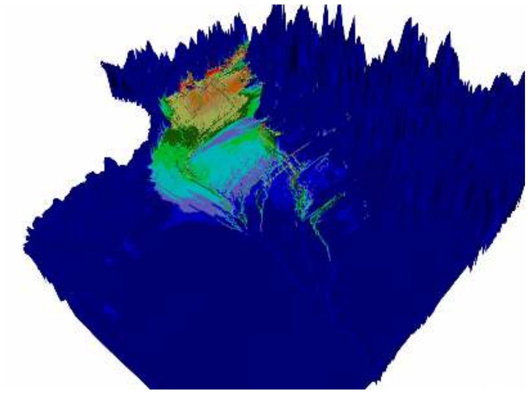


Figure 3b. Higher level of detail (like material and grain size distribution) added by SedTec for a sediment fan from the Alps to the southern Viennese Basin at the end of the last ice age.

We use existing data from literature for sediment transport parameters, digital elevation models as topographic input and fault geometry to obtain a realistic structure. Other properties can be added such as water content, porosity and temperature and also more model "behaviours" such as weathering and different algorithms for sediment transport can be used or added to the simulation.

How did data play a key role in problem solving?

Geological data are key to generate a digital twin of a geological structure at all, and the context to field data is vital. The data are also used to determine and improve parameters in my numerical models (i.e. in the materials relations) and to validate them. We have very concrete questions about geographical areas and thus, must make the models as realistic as possible.

What methodologies have been applied?

We used WinGeol/SedTec for the cases discussed here, which uses the following computational representation of a continuum mechanics model. The actual geology digital twin consists of cells with equally spaced x,y-dimensions and a variable z dimension. Every cell gets assigned state variables like local flow, cumulative erosion, … and material content (material and grain/particle size). Sedimentation modelling is based on physics such as mass transport etc. and solved by cellular automata [3,4][8]. Sediment can be transported by elevation or concentration differences between neighbouring cells. The main algorithm represents the mass transport function based on the neighbourhood matrix, and extensions of this algorithm describe what happens to the transported sediment (amount, grain size reduction, etc.). With each time step we can compute a new topography, mass balance (erosion vs sedimentation for each grid point), amount of erosion for each cell, amount of sediment per cell, grainsize distribution for each lithologies and we can match this against control points where we know the sediment thickness. The latter can be used to stop the

[8] See RoMM, Chapter 5.1.; https://publications.europa.eu/en/publication-detail/-/publication/ec1455c3-d7ca-11e6-ad7c-01aa75ed71a1

cycle. Once the final data are processed one can “drill” a virtual well, look at single simulation layers and the complete digital twin as a multilayer grid dataset. Real world well log profiles can be imported and compared to the simulation.

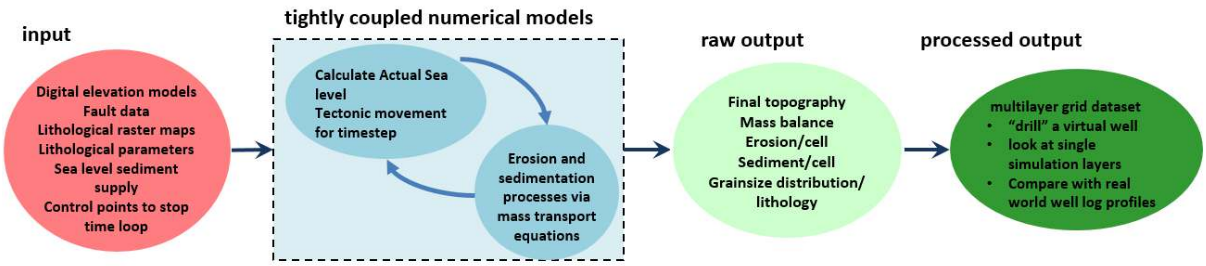


Figure3: MODA flow chart for a generic SedTec workflow

What were the expected improvements by adding data to your modelling?

Adding data or adding more algorithms can improve a geological numerical model, however the model must not become incalculable. It is always better, to be very selective on what to add makes sense.

For this particular case, did you have to invest a lot of work to make the data usable?

Geological databases and services are providing digitised data and aim for accessibility to a wide variety of stakeholders. When confronted with analogue data it may become more time intensive to make data usable. For example, older maps do not match with contemporary elevation or thematic information, the annotations are erroneous, etc. We need to scan maps or sections and the correct them. We also needed to learn how data acquisitions and storing changed over time to estimate their quality and correctness. Also, natural phenomena such erosion can make earlier recorded data unusable.

Figures 4a-e represent a storyboard of difficulties we encounter:

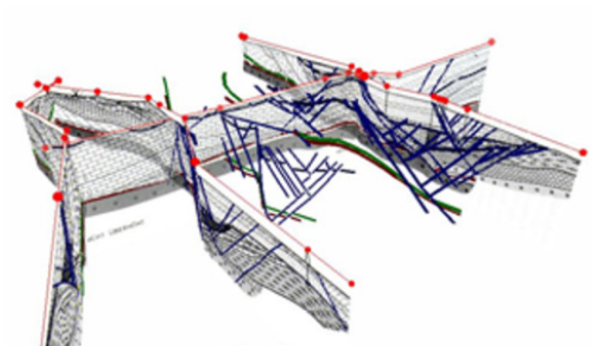


Fig.4a

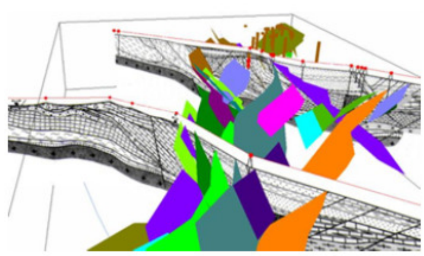


Fig.4b

Especially difficult is the processing and interpretation of fault lines from heterogeneous datasets. Figure 4a shows lines representing faults digitized based on section images and Figure 4b displays some of these lines connected to fault planes. Faults are extremely important elements as they might seal volumes from each other or in other cases might do the exact opposite (paths for fluid flow).

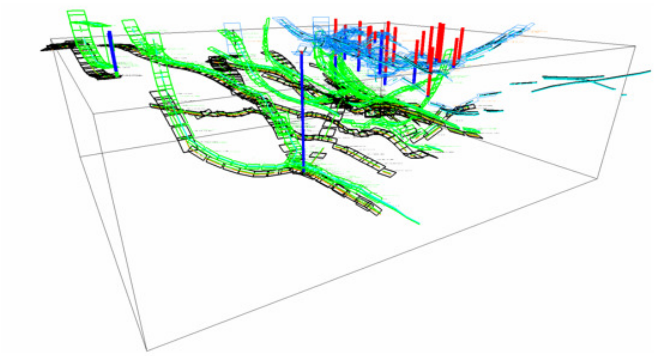


Fig.4c

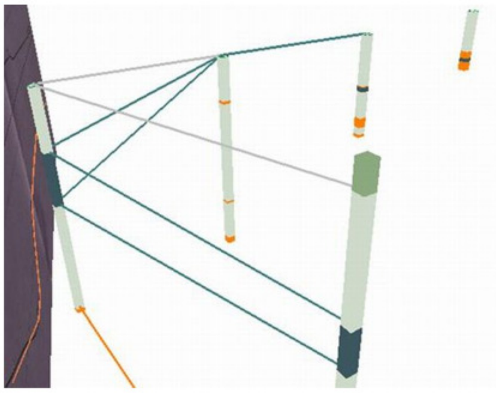


Fig.4d

Also, very difficult is the combination of bore logs (Figure 4c) to represent 3D bodies. Figure 4d shows an interpreted detail - the lines connecting the various bore logs are interpretation. What is already very visible here - although with just 5 logs – all of them show quite different sequences and the geologist has to decide, if there is a fault cutting through (and causing offsets) or if it is a typical variation by the sedimentation environment such as channels and river planes, for example.

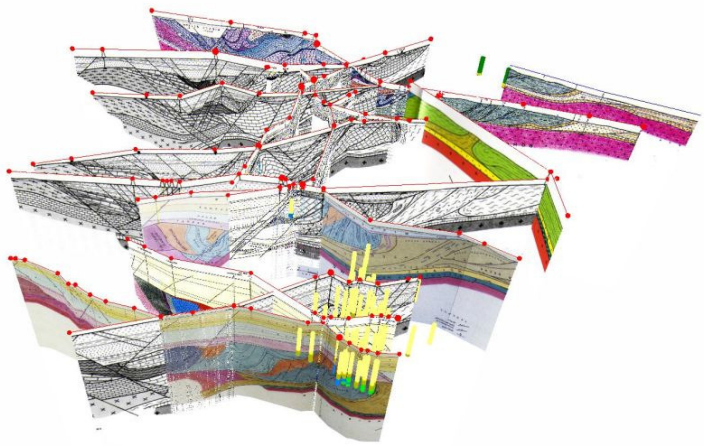


Fig.4e

Figure 4e shows what we can obtain by combination of all available input data: A digital twin of geological sections created over more than 30 years and put together (plus additional data from bore logs) using WinGeol.

The before mentioned modules like FaultTrace, SedTec and Subsurface Modeller are then used to convert this data into a continuous model.

For this particular case, what did you do with the data w.r.t. data-science?

We extract data from data bases and reduce them to make sure the once we use are well validated and reliable.

For this particular case, what did you do with the data w.r.t. geology applications?

It is important to calibrate the data to reflect the properties of the geological layers. It can be also important to add situation specific knowledge such as fluvial transport processes, tectonic movements, volcanic layers, vegetation, and changes in environmental conditions over time, etc. to describe the case at hand best.

What is the quantitative value of combining data with geology modelling?

Geology modelling enables us to look into processes at all, e.g. we can study different formulations of mineral or lithological compositions and physical properties such as grain sizes on a large area, which is not possible in a real-life scenario. The interaction of various processes, although one by one not difficult to understand, will result in very high complexity which is more or less impossible to imagine without the aid of the software. Such "what if" scenario may give us new (research) ideas.

Satellite data are very valuable as they can help us to get a better overview of geological formations and fault lines within a large area and we can use this in our modelling. Their value is to be able to cover large areas in short time and can also relates to geologist's safety. We can have projects in areas of conflict and it is not possible to send a researcher there for fieldwork. Also, some terrains are physically not accessible.

What investments were made during the project?

We were applying for grants and looked into collaboration to finance the actual software development.

What sort of obstacles or barriers (if any) did you have to overcome to use data driven geology modelling?

From a data perspective, we have to reduce and evaluate data. Geology does not only vary with different geographical locations/terrains but also geological time frames. There are different domains and the experts may communicate differently.

From a modelling perspective, as we need validation with field data, we have to understand how geologists work in the field, what conditions they face and how they inter-/extrapolate their observations. This may differ from what our simulation shows and we have to understand the possible uncertainties of the field data used for calibration and we have to communicate what and what not the numerical simulation can do.

Did using data improve your competitiveness/innovation power?

Certainly, using data the right way and working with models as realistic as possible did improve our competitiveness. Our software development happens with close links to our network of researchers which contributes to our innovation power.

What would you need the community to provide to enable data-driven geology modelling?

We want our peers to become more exact and reproducible when it comes to data acquisition – this should include information of the reliability of field observations. Sometimes to be aware of limitations means to be more exact. We as geologists might have the largest lab – the earth – but sometimes it is hard to find the right things or to see them clearly in “the mess” in front of us. Therefore, to compensate the very heterogeneous nature of our data plus the huge gaps in it, we have to be open to apply many different methods of very different fields, scales and so on. Like in other sciences the tendency is that researchers become more and more specialized which is not always helpful. The specialists tend to stay on their secure ground but when we make our simulations and modelling, we have to build bridges over the gaps – spatial and thematic ones.

References

[1] “Karst morphology and groundwater vulnerability of high alpine karst plateaus.” Plan, l.; Decker, K.; Faber, R.; Wagreich, M.; Grasemann, B.; Environmental Geology 58 (2009) 285-297. DOI:10.1007/s00254-008-1605-5

[2] “Climate as main factor controlling the sequence development of two Pleistocene alluvial fans in the Vienna Basin (eastern Austria) – a numerical modelling approach” Salcher, B.C.; Faber, R.; Wagreich, M.; Geomorphology 115 (2010) 215-227. DOI: 10.1016/j.geomorph.2009.06.030

[3] “Interpreting the Stratigraphy of the Baltimore Canyon Section, Offshore New Jersey With PHIl, a Stratigraphic Simulator.” Bowman, S.A.; Vail P.R.; SEPM Special Publication 62 (1999) 117-138. DOI: https://doi.org/10.2110/pec.99.62.0117

[4] “Numerical modelling of drainage basin evolution and three-dimensional alluvial fan stratigraphy” Clevis, Q.; de Boer, P.; Wachter, M.; Sedimentary Geology 163 (2003) 85-110. DOI: https://doi.org/10.1016/S0037-0738(03)00174-X