




# GemGIS - Spatial Data Processing for Geomodeling

Alexander Jüstel<sup>1,2</sup>, Arthur Endlein Correia<sup>3</sup>, Marius Pischke<sup>1</sup>, Miguel de la Varga<sup>4</sup>, and Florian Wellmann<sup>1</sup>

<sup>1</sup> RWTH Aachen University, Computational Geoscience and Reservoir Engineering, Wüllnerstraße 2, 52062 Aachen, Germany <sup>2</sup> Fraunhofer IEG, Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems, Am Hochschulcampus 1, 44801 Bochum, Germany <sup>3</sup> Geovariances LATAM, Av. Do Contorno, 7218 Sala 1201, 30110-048 Belo Horizonte, Brazil <sup>4</sup> Terranigma Solutions GmbH, Laurentiusstraße 59, 52072 Aachen

DOI: [10.xxxxxx/draft](https://doi.org/10.26434/chemrxiv-2021-08-01)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Chris Vernon](#) 

## Reviewers:

- [@omshinde](#)
- [@kanishkan91](#)

Submitted: 29 August 2021

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

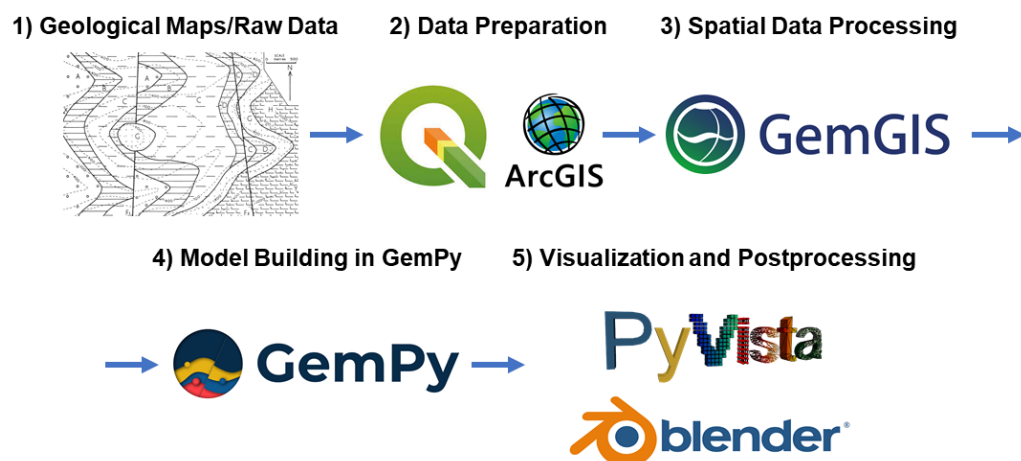
**GemGIS** is an open-source Python package for processing spatial data for geological modeling. **GemGIS** wraps and extends the functionality of packages known to the geo-community such as GeoPandas, Rasterio, OWSLib, Shapely, PyGEOS, PyVista, Pandas, NumPy, the geomodelling package GemPy and others. The aim of **GemGIS**, as indicated by the name, is to become a bridge between conventional geoinformation systems (GIS) such as ArcGIS and QGIS, and currently open-source geomodelling tools such as GemPy, enabling simpler and more automated workflows from one software environment to the other. Next to the Github-hosted open-source repository, the package concludes a dedicated documentation page containing detailed installation instructions, tutorials for beginners and more advanced users and sample models to walk users through the entire workflow from creating the data to processing the data and to construct a structural geological model. These examples include the well-known Bennison maps for geological mapping. Functionality for visualizing 3D models is available through the PyVista package and through Blender in the near future.

## Statement of need

**GemGIS** is intended for students, lecturers, researchers and anyone else working with spatial data with the aim of constructing 3D structural geological models both from example data or real world data for teaching purposes or real world projects. These real world projects may include models for the exploration of deep geothermal energy, carbon capture and storage, nuclear waste disposal or mining activities. The aim of **GemGIS** is hereby to act as a connecting bridge for users being familiar with conventional Geographic Information Systems (GIS) such as ArcGIS or QGIS and currently open-source geomodeling software such as GemPy ([Varga et al., 2019](#)). Additionally, **GemGIS** builds upon the common data classes such as (Geo-)Pandas (Geo-)DataFrames, NumPy arrays or PyVista meshes. Data collected in the field or geological data obtained from state or country authorities, respectively, can be visualized and preprocessed within a known GIS environment whereas the processing of the available data for geomodeling can be done with **GemGIS** instead of handling multiple versions of CSV files or vector/raster data files such as shape files within a GIS environment. While similar projects such as Map2Loop ([Jessell et al., 2021](#)) offer fully-automated workflows based on open-source outcrop and subsurface data, easily accessible in countries like Australia, **GemGIS** aims to offer the user maximum control by automating purely the technical challenges of data integration. Hence, we encourage every user to also work with and understand the data for each model.

The great advantage of using **GemGIS** in open-source web applications such as Jupyter

Notebooks is that processing steps are always reproducible without the need of managing different versions of input data files. This is of extreme importance when dealing with multiple data sources and large volumes of data in general. Traditional GIS systems with graphical user interface are capable of managing this data whereas **GemGIS** will perform the data processing and conversion of input data into formats ready for the construction of structural geological models. However, in comparison to GIS systems, all intermediary results are stored in memory when working within the Python environment. It remains the choice of the user to store intermediary results or only the final results.



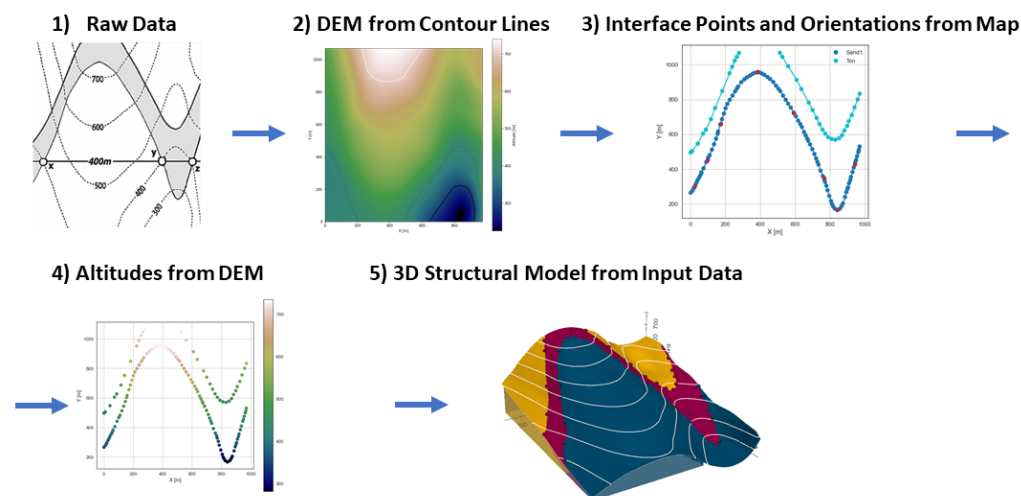
**Figure 1:** Graphical Workflow from raw geological data via data preparation in a GIS environment to GemGIS, model building in GemPy and visualization in PyVista and/or Blender.

## GemGIS Functionality

**GemGIS** is capable of working with different types of vector and raster data sets as well as XML-based formats and mesh formats. It is achieved through either parsers that were created for special data formats or through parsers already available in existing packages. And this is the strength of **GemGIS**: Rather than reinventing the wheel, **GemGIS** builds upon well-known packages within the geo-community such as GeoPandas (Jordahl et al., 2021), Rasterio (Gillies et al., 2013--), OWSLib (Gillies et al., 2006--), Shapely (Gillies et al., 2007--), PyGEOS (Wel et al., 2021), PyVista (Sullivan & Kaszynski, 2019), Pandas (team, 2021), NumPy (Harris et al., 2020), the geomodelling package GemPy (Varga et al., 2019) and others. **GemGIS** wraps, combines and extends the functionalities of these different packages in order to allow for a more automated processing of spatial data for geomodeling and visualization of input and output data.

The functionality of **GemGIS** includes:

- Editing vector geometries and raster data
- Extracting interface point positions from maps, digital elevation models, cross sections and already existing meshes
- Calculating orientations and respective positions from maps and cross sections
- Obtaining data stored on web servers through OWSLib
- Visualizing vector data, raster data (maps and cross sections), meshes and boreholes with PyVista
- Post process structural geological models created with GemPy



**Figure 2:** Workflow from raw data to creating a digital elevation model, extracting interface points and orientations for geomodeling, extracting altitudes for geomodeling and building the final model.

## GemGIS Outlook

**GemGIS** is designed to enable a further extension of functionalities in the future. This includes the capability of working with even more data formats through self-written parsers or parsers from already existing packages, working with more geophysically related data such as seismic data, data obtained from borehole geophysics, magnetic and gravitational field measurements and many more.

A more direct link to Google Earth and a QGIS Plugin are planned. In addition, the creation of data, which is usually done in conventional GIS systems, could also be transferred to a web application. A plugin for the Blender Software to visualize GemPy models is currently under development at the Department for Computational Geoscience and Reservoir Engineering at RWTH Aachen University.

## GemGIS Resources

Various tutorials including supplementary data and an elaborative documentation including introductions to various topics related to **GemGIS** and an API Reference are available:

- GemGIS Repository - <https://github.com/cgre-aachen/gemgis>
- GemGIS Documentation - <https://gemgis.readthedocs.io/en/latest/>
- GemGIS Installation Instructions - [https://gemgis.readthedocs.io/en/latest/getting\\_started/installation.html](https://gemgis.readthedocs.io/en/latest/getting_started/installation.html)
- GemGIS Tutorials - [https://gemgis.readthedocs.io/en/latest/getting\\_started/tutorial/index.html](https://gemgis.readthedocs.io/en/latest/getting_started/tutorial/index.html)
- GemGIS Tutorial Notebooks - [https://rwth-aachen.sciebo.de/s/AfXRzZywYD-bUF34/download?path=%2F&files=tutorials01\\_53.zip](https://rwth-aachen.sciebo.de/s/AfXRzZywYD-bUF34/download?path=%2F&files=tutorials01_53.zip)

## Acknowledgements

All authors would like to thank the Software Underground (<https://softwareunderground.org/>) for providing a platform to interact with users and to organize hackathons (Transform 2020/2021) to further develop the open-source package. We also acknowledge contributions

from, and thank all our users for reporting bugs, raising issues and suggesting improvements to the API of **GemGIS**.

## References

- Gillies, S.others. (2006--). *OWSLib*. Mapbox. <https://github.com/geopython/OWSLib>
- Gillies, S.others. (2007--). *Shapely: Manipulation and analysis of geometric objects*. toblerity.org. <https://github.com/Toblerity/Shapely>
- Gillies, S.others. (2013--). *Rasterio: Geospatial raster i/o for Python programmers*. Mapbox. <https://github.com/mapbox/rasterio>
- Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., R'io, J. F. del, Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Jessell, M., Ogarko, V., Rose, Y. de, Lindsay, M., Joshi, R., Piechocka, A., Grose, L., Varga, M. de la, Ailleres, L., & Pirot, G. (2021). Automated geological map deconstruction for 3D model construction using *map2loop* 1.0 and *map2model* 1.0. *Geoscientific Model Development*, 14(8), 5063–5092. <https://doi.org/10.5194/gmd-14-5063-2021>
- Jordahl, K., Bossche, J. V. den, Fleischmann, M., & et al. (2021). *Geopandas/geopandas: v0.9.0* (Version v0.9.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.4569086>
- Sullivan, C. B., & Kaszynski, A. (2019). PyVista: 3D plotting and mesh analysis through a streamlined interface for the visualization toolkit (VTK). *Journal of Open Source Software*, 4(37), 1450. <https://doi.org/10.21105/joss.01450>
- team, T. pandas development. (2021). *Pandas-dev/pandas: pandas* (Version v1.2.3) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.4572994>
- Varga, M. de la, Schaaf, A., & Wellmann, F. (2019). GemPy 1.0: Open-source stochastic geological modeling and inversion. *Geoscientific Model Development*, 1–32.
- Wel, C. van der, Bossche, J. V. den, Ward, B., Chaitanya, K., Taves, M., Fleischmann, M., Lackner, M., & Clancy, T. (2021). *Pygeos/pygeos: v0.9.0* (Version v0.9.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.4459137>