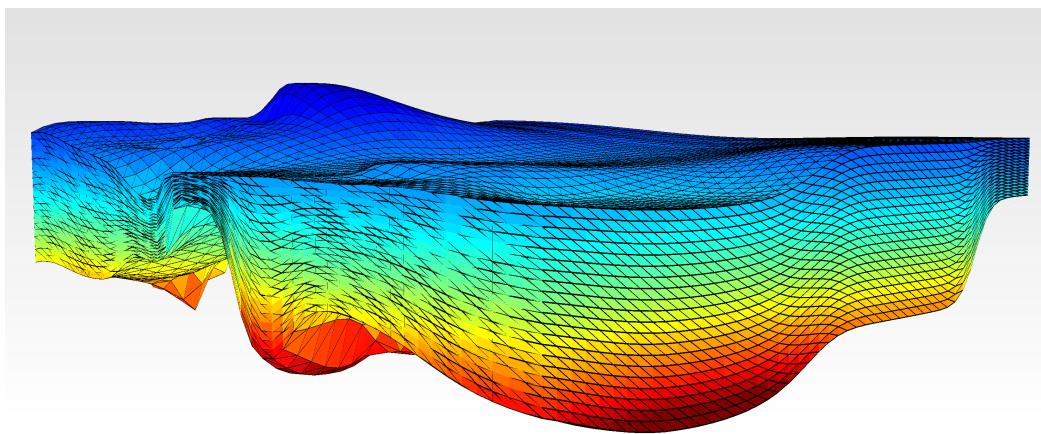


# User Manual and Tutorial for GeoReVi

---

## Geological Reservoir Virtualization

Date: November 6, 2019



---

## Foreword

This manual is under preparation and subject to future changes.



The most recent version of the manual can be found on [GitHub](#)

## Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Built with . . . . .	7
1.2	Authors and Contributors . . . . .	8
1.3	Hard- and software requirements . . . . .	8
1.4	Installation . . . . .	8
<b>2</b>	<b>Quick start</b>	<b>9</b>
2.1	Importing a local data set . . . . .	9
2.2	Univariate statistics . . . . .	10
2.3	Empirical distribution . . . . .	11
2.4	Creating a line and bubble chart . . . . .	11
2.5	Box-Whisker plot . . . . .	13
2.6	Variography . . . . .	13
2.7	Handling meshes . . . . .	14
2.7.1	Generating custom meshes . . . . .	17
2.7.2	Copying, removing and storing meshes . . . . .	19
2.7.3	3D meshes . . . . .	21
2.7.4	Complex geometries . . . . .	22
2.8	Performing Predictions . . . . .	22
2.9	2-D predictions . . . . .	22
2.9.1	Checking the interpolation for consistency . . . . .	24
2.9.2	3-D predictions . . . . .	26
2.9.3	Rescaling . . . . .	27
2.9.4	Error integration . . . . .	27
<b>3</b>	<b>Getting started in depth</b>	<b>28</b>
3.1	Introduction . . . . .	28
3.1.1	Starting the application . . . . .	29
3.1.2	Registration and Sign in . . . . .	31
3.2	Obtaining a Bing Maps Key . . . . .	32
3.3	Sample data . . . . .	32
3.4	Data management . . . . .	32
3.4.1	Supported data types . . . . .	33
3.4.2	Provided data set . . . . .	36
3.4.3	Project Management . . . . .	36
3.4.4	Objects of investigation . . . . .	36
3.4.5	Lithologies . . . . .	37
3.4.6	Rock samples . . . . .	38
3.4.7	Measurements . . . . .	38

## CONTENTS

---

3.4.8	Import to the database . . . . .	39
3.5	Data analysis and visualization . . . . .	41
3.5.1	Data transformation . . . . .	41
3.5.2	Univariate data analysis . . . . .	42
3.5.3	Mesh generation . . . . .	46
3.5.4	Mesh statistics . . . . .	48
3.5.5	Spatial interpolation and simulation . . . . .	49
3.5.6	Conditional Simulation . . . . .	53
3.5.7	Measurement Error Integration . . . . .	53
3.5.8	Cross-validation . . . . .	54
3.6	Multivariate statistics . . . . .	55
3.6.1	Dimensionality reduction . . . . .	55
3.6.2	Classification . . . . .	57
3.6.3	Correlation analysis . . . . .	57
3.7	Data visualization . . . . .	58
3.7.1	3-D chart . . . . .	58
<b>4</b>	<b>Developing Plug-ins</b>	<b>59</b>
<b>5</b>	<b>Personal data security</b>	<b>60</b>
<b>6</b>	<b>Related works</b>	<b>61</b>

## License

### GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc. 51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

#### Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software—to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced

## CONTENTS

---

by others will not reflect on the original authors' reputations. Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

## Reporting issues

Unfortunately projects of GeoReVi's size are seldom bug free. If you find bugs, we would be grateful if you report them either directly to the author under contact@georevi.com or on our GitHub repository

<https://github.com/ApirsAL/GeoReVi/issues>

# 1 Introduction

GeoReVi is a software developed for geoscientific information management and knowledge discovery. GeoReVi can deal with many types of subsurface characterization yet being specialized on geothermal reservoir characterization. The software is developed under the MIT license making it free to use and open source.

The software is split up into two components. One component is the client application, the user interacts with, and the other component is the so called back end located either on an accessible server or on local machines. The front end uses a communication framework to access the database located in the back end.

## 1.1 Built with

- [Entity Framework 6](#) - Database access
- [HelixToolkit.WPF](#) - Providing the 3D components
- [Managed Extensibility Framework](#) - Providing the modular structure of GeoReVi
- [Accord.NET](#) - Linear Algebra and Machine Learning framework
- [Windows Presentation Foundation](#) - Framework for creating Windows Desktop Applications
- [Caliburn.Micro](#) - Framework for MVVM development in XAML platforms
- [FontAwesome.WPF](#) - Providing nice icons
- [DotSpatial](#) - Spatial algorithms
- [GeoAPI](#) - Coordinate Conversion
- [ProjNET](#) - Coordinate Projection
- [LiteDB](#) - Embedded NoSQL database
- [MoreLinq](#) - Query helper
- [Extended WPF Toolkit<sup>TM</sup>](#) - Providing nice UI controls

### 1.2 Authors and Contributors

Adrian Linsel, Technische Universität Darmstadt, [contact@georevi.com](mailto:contact@georevi.com)

### 1.3 Hard- and software requirements

GeoReVi is built with the .NET Framework. This makes the software executable on every machine with Windows XP and newer OS. Since GeoReVi is a desktop database application, a valid connection to the database server has to be provided. When GeoReVi is used internally, please contact your admin for further assistance or use the software in local mode.

The software is tested under the following hard- and software configurations:

Hard-/Software	Database	User interface
OS	Windows Server 2012 R2	Windows 10
CPU	Intel Xeon CPU E3-1276 v3 @ 3.6 GHz	Intel Core i5-7200U 2.5 GHz
RAM	24 GB DDR4	16 GB DDR4

### 1.4 Installation

Download the GeoReVi.exe file from GitHub (<https://github.com/ApirsAL/GeoReVi>) and perform the installation.

To initiate a client-server architecture, a more advanced installation has to be performed. Researchers interested to establish a multi-user environment at their institute should feel free to contact us under [contact@georevi.com](mailto:contact@georevi.com) via Email.

## 2 Quick start

This section will not cover any theory or background information on how any functionality works in detail. The reader will be guided with the help of the tutorial data set that can be found under

*<https://github.com/ApirsAL/GeoReVi/tree/master/docs/Tutorial>*

completely from making basic statistical analysis to the final reservoir property model. First download the complete **Tutorial** folder from GitHub and save it anywhere on your hard drive. Start GeoReVi and enter the **Local mode** in the login screen.

### 2.1 Importing a local data set

Under **Data → Measurements** we enter the measurements screen. Now we navigate to the **Data analysis** tab. There, under **Univariate** click on **Mesh action → Import mesh**. Select **.gmsh** as file extension and navigate to the downloaded **Tutorial** folder and there to **FieldParameters**. Select the *ThermalConductivity.gmsh* file and click open. This file contains thermal conductivity measurements from samples taken from six drill cores penetrating the Rotliegend formation in southwestern Germany. Alternatively, individual data sets from common spreadsheet files can be imported. Therefore, we import another mesh in the same folder called *ThermalConductivity.csv*. When importing spreadsheet files, a dialog will pop up where the user must define, which column should be mapped to which entity. We will produce following mappings between the **Imported headers** and the **Database headers**

1. Thermal conductivity → Value
2. x [m] → X
3. y [m] → Y
4. elevation [m] → Z
5. Date → DateTime
6. Name → Name

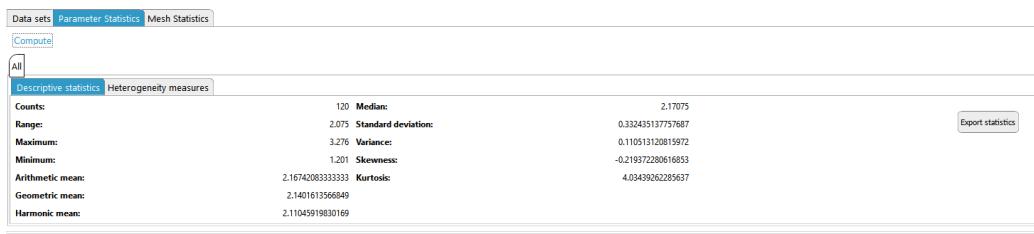


Figure 1: Basic statistics result.

When clicking **Import** the data set will be added to the **Data sets** with the name *New data set*. When we want to remove a data set, we can perform **Mesh action → Remove dataset**. After deploying the command the selected mesh will be removed.

The measurement results can be seen under **Univariate** and **Data set**. The first column shows the actual value of the measurement. Thermal conductivity is provided in the SI unit  $W/m \cdot K$ . The second column shows the x coordinate or longitude respectively. The third and fourth column display y/latitude and z/elevation respectively. The uploaded values are provided as metric x,y and z coordinates in the UTM coordinate system. The fifth column shows the data and the sixth column the name of the individual sample.

The name *All* of the imported mesh is meaningless. So we will change the name to something more meaningful. Therefore, open the expander left from the data table by clicking on it. Open the topmost expander called **Display**. Here, the selected mesh name can be changed. The selected mesh is always the tab selected in the **Data sets** view. Type in "Thermal conductivity Rotliegend" into the textbox and click **Save** left beside it.

## 2.2 Univariate statistics

To get a first overview over the statistics of the measurement, got to **Univariate** and there to **Parameter statistics**. When clicking on **Compute** the basic parameters of the distribution will be calculated. The result should look like the results shown in fig. 1.

Results from the basic statistics analysis can be exported as .CSV file. Results presented are the statistics of all sample values of the particular data set. The provided measures are the total count of the samples, the *range*, *maximum*, *minimum*, *arithmetic mean*, *geometric mean*, *harmonic mean*, *median*, *standard deviation*, *variance*, *skewness*, *kurtosis*, *coefficient of variation* and *Dykstra-Parson coefficient*.

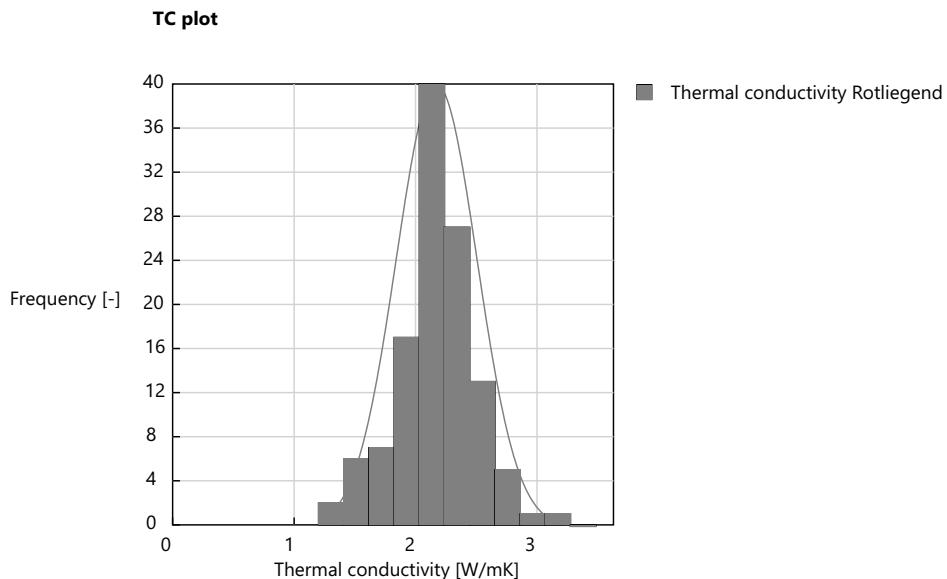


Figure 2: Empirical and theoretical distribution of the thermal conductivity measurement.

### 2.3 Empirical distribution

Empirical distributions can be visualized in histograms. Therefore, return to the **Data sets** and **right-click** on the mesh *Thermal conductivity Rotliegend* and press **Add data set to bar chart**. Navigate to **Charts → Histogram** and refresh the chart with the **Refresh** button in the top-left corner. Under **Data series → Theoretical distribution** make a tick in the field **Show** and refresh . The resulting distribution chart should look like fig. 2.



By **right-click → Export chart** a chart can be exported as .png, .bmp, .pdf or .emf file. Latter one is a vector format that can be opened with popular vector graphic programs like CorelDRAW or Inkscape. When you export the chart from the **Empirical distribution** section it will look like fig. 2)

### 2.4 Creating a line and bubble chart

To create a line chart, we will delete all values from the **Thermal conductivity Rotliegend** table that do not belong to the drill core *B8*. Therefore, navigate back to the data sets and sort the table by name by clicking on the header of the name column. Now delete all samples that do not belong to the drill core *B8*. Rows can be selected by clicking in the leftmost column like shown in figure 3. When pressing the **Delete** key, the selected rows will

Data sets					
Parameter Statistics					
Mesh Statistics					
All					
Item1	Item2	Item3	Item4	Item5	Item6
2.101	3483830.30522338	5531903.92101447	162.7	12/10/2016 12:00:00 AM	BB-01
2.213	3483830.30522327	5531903.92097749	159.1	12/10/2016 12:00:00 AM	BB-02
2.113	3483830.30522311	5531903.92092716	154.2	12/10/2016 12:00:00 AM	BB-03
2.00	3483830.30522302	5531903.92089634	151.2	12/10/2016 12:00:00 AM	BB-04
2.105	3483830.30522287	5531903.92084653	146.35	12/10/2016 12:00:00 AM	BB-05
1.745	3483830.30522278	5531903.9208152	143.3	12/10/2016 12:00:00 AM	BB-06
1.906	3483830.30522259	5531903.92075203	137.15	12/10/2016 12:00:00 AM	BB-07
2.445	3483830.30522244	5531903.92070324	132.4	12/10/2016 12:00:00 AM	BB-08
2.335	3483830.3052224	5531903.92069143	131.25	12/10/2016 12:00:00 AM	BB-09
2.173	3483830.3052221	5531903.92062518	124.8	12/10/2016 12:00:00 AM	BB-10
1.997	3483830.30522332	5531903.92069485	160.79	12/10/2016 12:00:00 AM	BB-14
2.197	3483830.30522318	5531903.92094914	156.34	12/10/2016 12:00:00 AM	BB-15
2.111	3483830.30522309	5531903.92091946	153.45	12/10/2016 12:00:00 AM	BB-16
2.114	3483830.30522292	5531903.92086399	148.05	12/10/2016 12:00:00 AM	BB-17

Figure 3: Selecting rows in the Data sets view.

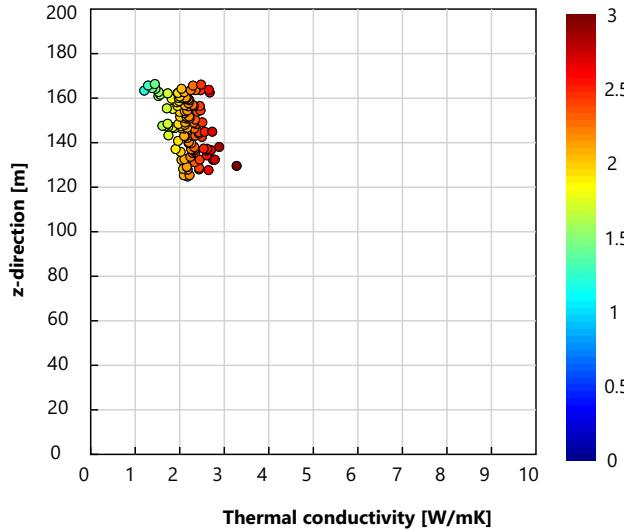


Figure 4: Bubble chart of the z-direction thermal conductivity measurements.

be removed from the data set.

Now, again we must **right-click** on the mesh *Thermal conductivity Rotliegend* and press **Add data set to line chart**. When we navigate to **Charts → Log chart** we can see under **Editing** that our data set was added. Now, we must specify the direction that we want to visualize. You can find the option above the chart. In our case, we want to display the *Z-direction*. Again, adapt the chart axes to 200 (Y max) and 0 (Y min) and under **Chart style → Legend → Colorbar** make a tick in **Is color map**. After refreshing, the chart should look like like fig. 6

## 2.5 Box-Whisker plot

Box-Whisker plots visualize a set of basic statistic measures of a sample. These measures compose of minimum, maximum, upper and lower quartile, median and outliers as illustrated in fig. 5.

When we want to create a box-whisker chart of our data set, we need to **right-click** on the mesh *Thermal conductivity Rotliegend* and press **Add data set to box-whisker chart**. For each data set added to the chart, a new box-whisker will be created. Outliers are shown as circles when the **Remove outliers** checkbox is activated.

## 2.6 Variography

To analyze the spatial correlation of the data set we can conduct a variographic analysis. Therefore, return to the **Data sets** and **right-click** on

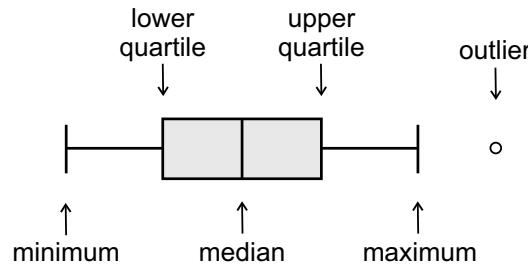


Figure 5: Schematic of a box-whisker plot.

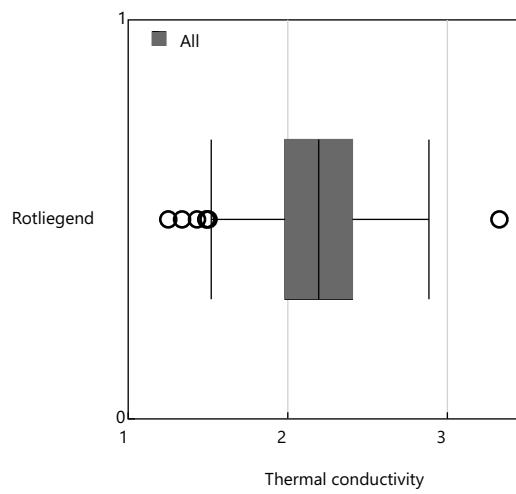


Figure 6: Box-Whisker chart of the thermal conductivity measurements with removed outliers.

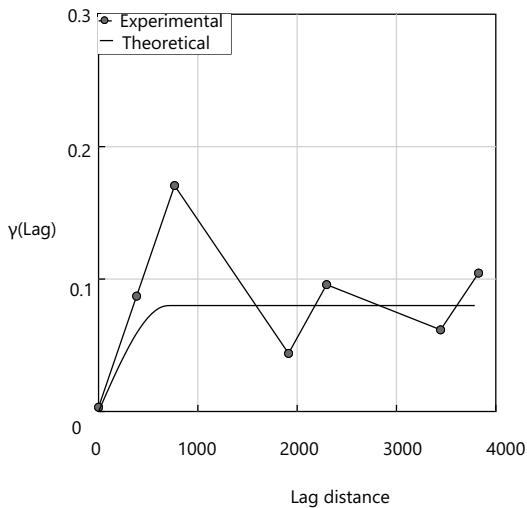


Figure 7: Semivariogram of the thermal conductivity measurement.

the mesh *Thermal conductivity Rotliegend* and press **Add data set to variogram chart**. Now we navigate to the **Charts** tab and then to the **Variogram**. Here, navigate to **Semi-variogram** and open the expander on the left side. In the **Editor → Editing** expander menu you will find our data set that was added to the chart in the last step. When you press the button in the left upper side of the chart that shows the tooltip **Refresh** the chart will be created. The x axis shows the lag vector distance and the y axis shows the cumulative semivariance. We want to consider the semivariance in three-dimensional space. Accordingly, **Model → Parameters → Direction** we have to switch the direction from *Z-direction* to *XYZ-direction*, we insert 0.08 in the **sill** and 700 in the **range** textbox and we will increase the number of **Bins** to 15 . Under **Chart style → Axes → X axis** insert 4000 in the **X max**, 1000 in the **X tick**, 0.5 in the **Y max**, 0.1 in the **Y tick** textboxes and refresh the chart.

## 2.7 Handling meshes

Each data set in GeoReVi is treated as a mesh. A mesh consists of nodes, faces and cells. A one-dimensional mesh consists of nodes only, while a two-dimensional mesh consists of nodes and faces and a three-dimensional mesh of nodes, faces and cells. Each node is indexed in a mesh and therefore is aware of its neighboring nodes and if it is a boundary node of the domain. Two sample surfaces are provided in the **Tutorial → Surfaces** folder. First, we will import the surface *BaseExample.gmsh* with the same procedure as we imported the thermal conductivity measurements before. The surface is a

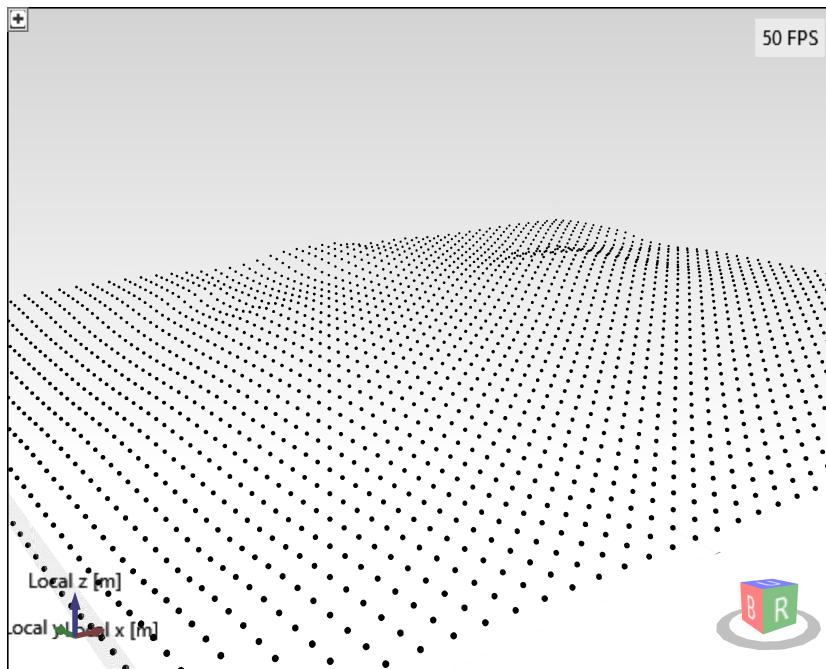


Figure 8: Point cloud visualization of a surface in GeoReVi

hiatal surface called "post-Variscan non-conformity". After import, we want to get an overview about the mesh statistics. As we go to **Mesh statistics** and press **Compute** the statistics will be provided. The imported 2-D mesh consists of 2601 nodes and 2500 faces.

We can display the mesh in 3D when we **right-click** when the mesh *Base Rotliegend Interpolated* is selected and press **Add data set to 3D chart**. When we navigate to **Charts → 3D** we will find a 3D viewport where all types of spatial data can be visualized. Since the imported mesh provides a very big extend, we will configure the axes to be displayed accordingly. Therefore, under **Chart style → Axes** we insert the value *1000* in the **X Tick** and **Y Tick** boxes and *200* in the **Z Tick** box. Once we click **Refresh** in the dropdown button in the left upper corner, the mesh will be visualized as scatterplot which is the default for all imported meshes. Under **Data series → Symbols** we will select a **Symbol size** of *5 m*. All values in the 3D port are referenced by a metric system. When we refresh, the port should look like fig. 8.

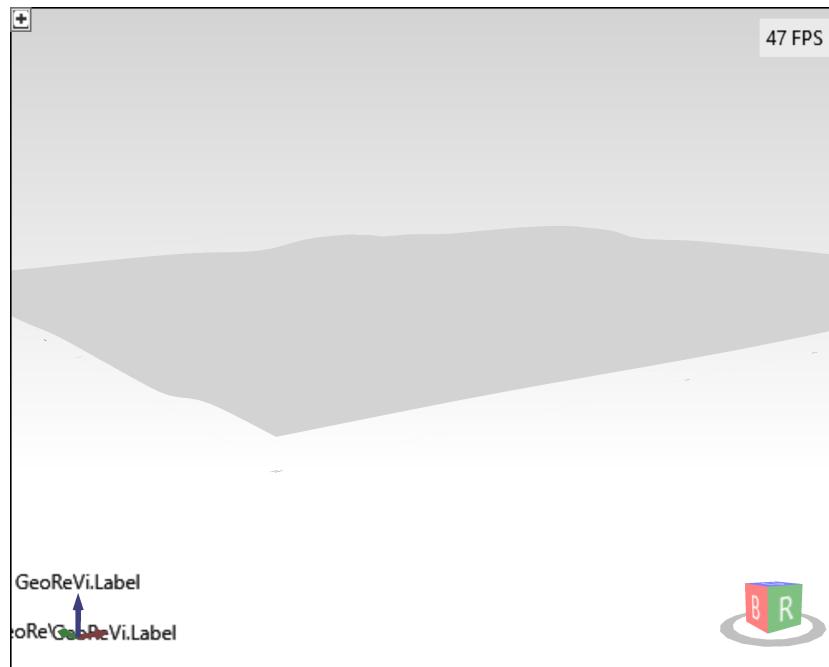


Figure 9: Faces visualization of a surface in GeoReVi



If the 3D viewport starts flickering, we need to make sure using an appropriate scaling of the axes. This problem occurs when **X min**, **X max**, **Y min** and **Y max** are all defined as  $0$ . This problem can be fixed by clicking **Auto subdivision** in the topmost part of the **Axes** menu. After refreshing, the problem should not come up again.

Another way to visualize meshes is with faces. To display the faces of the 2-D mesh select **Surface chart** as **Display type** and change the **Symbol → Fill color** to a type of *Gray*. After refreshing we will get a result that looks like fig. 9.

Another way to display meshes is to add the wireframe of the triangles to the model. Therefore, select **Wireframe and faces** as **Mesh visualization** and define a **Wireframe thickness** of *2*. The result should look like fig. 10. As of now there are no labels visible and the surface is very even. To display the labels appropriately we need to define the label size accordingly. Therefore, we go to **Chart style → Labels** and determine the **Label font size** to be *200*. When we want to exaggerate the surface there are two ways to do so.

1. We go to **Chart style → General** and define a vertical exaggeration

of  $5x$ . This option will exaggerate each object in the viewport by a factor of  $x$ .

2. To only rescale a single object we go to **Data series → Transformation → Scaling** and define a **Scale** of  $5$ . **IMPORTANT:** This option will transform ONLY the vertical scale of the SELECTED SERIES. Vertical scales will not be georeferenced anymore.

We will try both of the aforementioned methods. For better visualization we deactivate the **XZ grid** and **YZ grid** under **Chart style → Gridlines**. Now we will see a visualization according to fig. 11.

### 2.7.1 Generating custom meshes

For custom mesh generation the **Discretization** section of the **Data operations** menu can be used. The **Source data sets** serve as input meshes for newly created ones. Exemplary, we will create a 2-D mesh in the bounding box of our base surface. Firstly, we will select *Base Rotliegend interpolated* as a **Source data set**. Then we will define *Hexahedral regular grid* as **Discretization method**, *3D Bounding box* as **Boundary type** and define *0* as **Bins z direction** to create a two-dimensional data set. We will leave the default values in the other boxes.

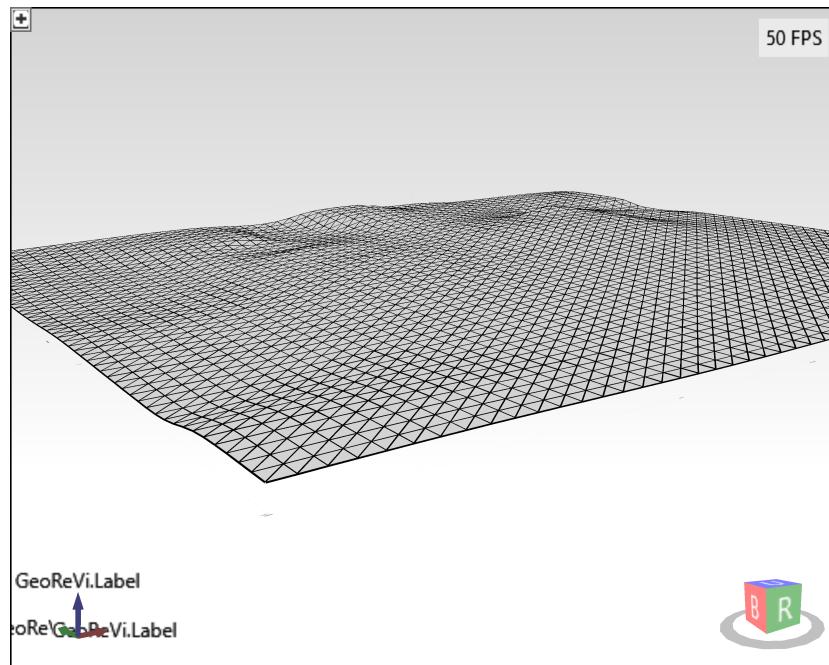


Figure 10: Faces visualization of a surface in GeoReVi



Users can define custom bounding areas using the **Start point x direction** and **End point x direction** textboxes. For y, and z direction the same applies. However, retaining the default value *0* in both **Start** and **End** box the minimum and maximum value of the **Source data sets** in the particular direction serve as boundaries.

When pressing **Compute discretization** a new mesh will be created and added to the data sets with the name *Discretized mesh*. When we give out the **Mesh statistics** we can see that a 2-D mesh with 441 vertices and 400 faces was created. When we add the mesh to the **3D chart** it will look like fig. 12.

Please, rename the mesh to *Base Model*. Now we want to interpolate the custom mesh through the already provided surface. Therefore, we will select the *Base Rotliegend Interpolated* as **Source data set** and *Base Model* as a **Target data set**. Now we will perform an interpolation, thus we need to open the **Interpolation** expander. As **Interpolation feature** we select *Elevation* and as **Interpolation method** we define *Inverse Distance Weighting*. Other options will be left as the default values. When we press

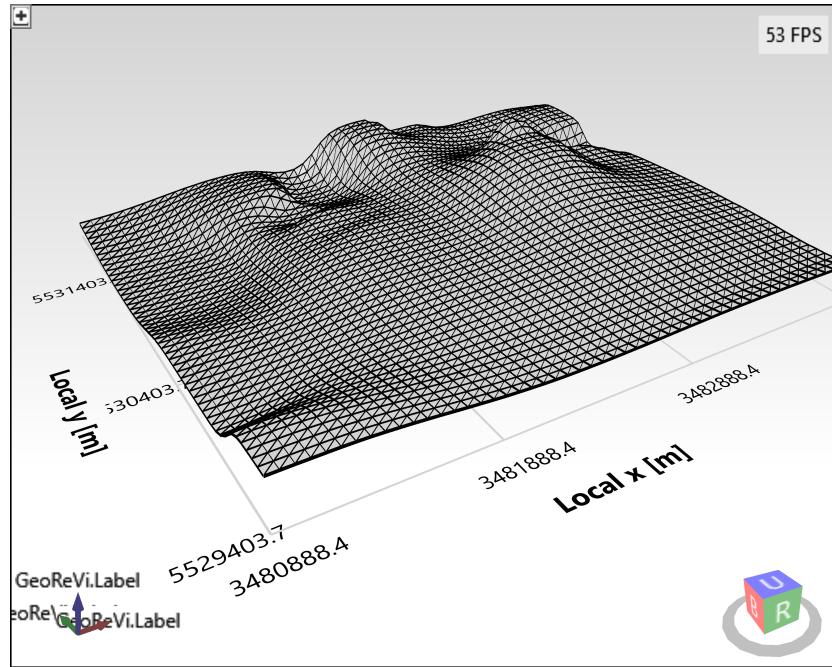


Figure 11: Exaggerated faces visualization of an IDW interpolated surface in GeoReVi

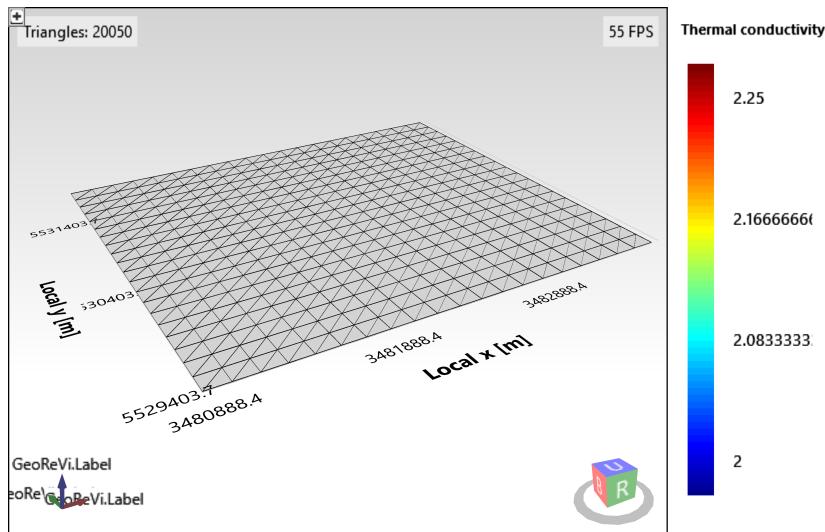


Figure 12: Custom surface.

**Compute interpolation** the new surface will be interpolated through the points provided by *Base Rotliegend Interpolated*. This can take a few seconds since the target mesh consists of 2600 nodes that will be included in the calculation. When the interpolation has finished, a new data set called *Interpolated mesh* will be added to the data sets. We rename it to *Interpolated surface*. When we add all three meshes, namely *Base Model*, *Base Rotliegend Interpolated* and *Interpolated surface* to the **3D chart** and display them as wireframe surfaces it will look like fig. 13.

### 2.7.2 Copying, removing and storing meshes

Since we do not want our newly created mesh to be lost once the application gets closed we can store it in a custom-developed, human-readable XML format called .gmsh. To export the mesh we click **Mesh action** and **Export mesh**. There we select our *Tutorial/Surfaces* folder and save *Interpolated surface* as *InterpolatedSurface.gmsh*. Now we navigate to the newly created file and see that a file of ca. 1.3 MB was exported. When we open it with a conventional text editor we can find the basic structure of a GeoReVi mesh (fig. 14).

We really want to make sure that we do not lose our mesh. So, we create a copy of it by again clicking on **Mesh action** and **Copy mesh**. A copy of the recently selected mesh will be created and added to the data sets. We will rename this mesh to *Interpolated surface - Copy*. Now we are sure to remove our created mesh from the data sets and click **Mesh action** and **Remove**

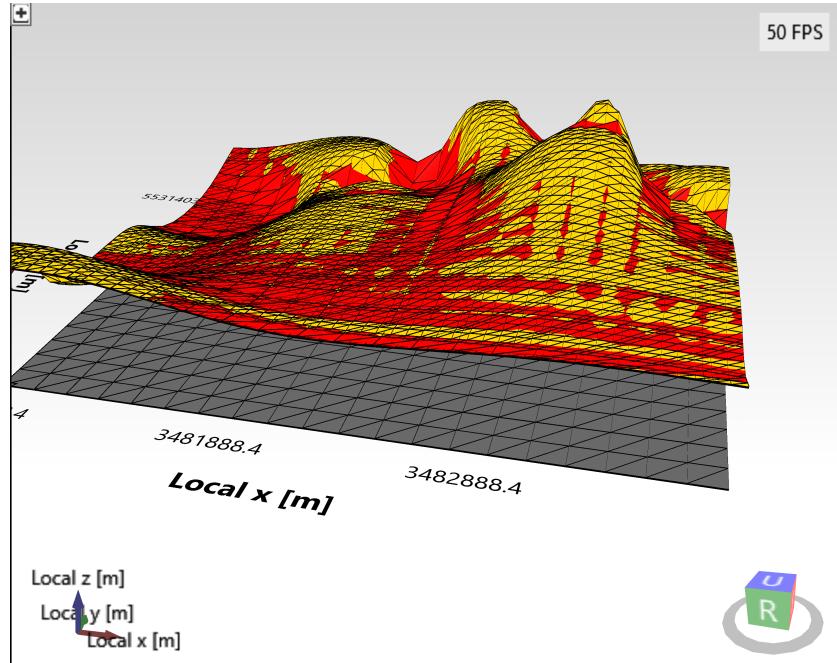


Figure 13: Comparison of the *Model base* (dark gray), *Base Rotliegend Interpolated* (yellow) and *Interpolated surface* (red) meshes.

```
[root@centos-1 ~]# curl -s -o test.xsd "http://www.w3.org/2001/XMLSchema-instance"
<?xml version="1.0" encoding="utf-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:i=http://www.w3.org/2001/XMLSchema-instance">
  <xsd:element name="test">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="id" type="xsd:string"/>
        <xsd:element name="name" type="xsd:string"/>
        <xsd:element name="age" type="xsd:int"/>
        <xsd:element name="discretions">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element name="id" type="xsd:string"/>
              <xsd:element name="name" type="xsd:string"/>
              <xsd:element name="age" type="xsd:int"/>
              <xsd:element name="isactive" type="xsd:boolean"/>
              <xsd:element name="isdiscretion" type="xsd:boolean"/>
            </xsd:sequence>
            <xsd:choice>
              <xsd:when test="false">
                <xsd:element name="location" type="xsd:string"/>
                <xsd:element name="value" type="xsd:string"/>
                <xsd:element name="geograph" type="xsd:string"/>
                <xsd:element name="isdiscretion" type="xsd:boolean"/>
              </xsd:when>
              <xsd:otherwise>
                <xsd:element name="location" type="xsd:string"/>
                <xsd:element name="value" type="xsd:string"/>
                <xsd:element name="geograph" type="xsd:string"/>
                <xsd:element name="isdiscretion" type="xsd:boolean"/>
              </xsd:otherwise>
            </xsd:choice>
          </xsd:complexType>
        </xsd:element>
        <xsd:element name="details">
          <xsd:complexType>
            <xsd:sequence>
              <xsd:element name="id" type="xsd:string"/>
              <xsd:element name="name" type="xsd:string"/>
              <xsd:element name="age" type="xsd:int"/>
              <xsd:element name="isactive" type="xsd:boolean"/>
              <xsd:element name="isdiscretion" type="xsd:boolean"/>
            </xsd:sequence>
            <xsd:choice>
              <xsd:when test="true">
                <xsd:element name="location" type="xsd:string"/>
                <xsd:element name="value" type="xsd:string"/>
                <xsd:element name="geograph" type="xsd:string"/>
                <xsd:element name="isdiscretion" type="xsd:boolean"/>
              </xsd:when>
              <xsd:otherwise>
                <xsd:element name="location" type="xsd:string"/>
                <xsd:element name="value" type="xsd:string"/>
                <xsd:element name="geograph" type="xsd:string"/>
                <xsd:element name="isdiscretion" type="xsd:boolean"/>
              </xsd:otherwise>
            </xsd:choice>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

Figure 14: File structure of the custom file format .GMSH to store GeoReVi meshes.

**mesh** after selecting the *Interpolated surface* mesh. Now, it will be removed from the data sets.

When we now import again the previously exported file *InterpolatedSurface.gmsh* we can see that our mesh was completely stored.

### 2.7.3 3D meshes

#### Regular meshes

Now, we want to go from 2D to 3D. In order to produce a 3D mesh of the bounding area of the input data set *Thermal conductivity* we go to **Data operations** → **Discretization**. Previously, we define our *Thermal conductivity* mesh as **Source data set**. As **Discretization method** we select *Hexahedral regular grid*, **Boundary type** is *3D Bounding box*, and bin counts should be 40 (x), 40 (y) and 20 (z). When we click **Compute discretization** a regular 3D mesh in the bounding domain of the thermal conductivity mesh will be generated like shown in fig. 15. We name that mesh *3D bounding box Rotliegend*



When adding three-dimensional meshes to the **3D chart**, it is best to select the **Display type** to be *Volumetric*. However, if you want to display all points of a 3-D mesh you can select *Scatter* as a display mode.

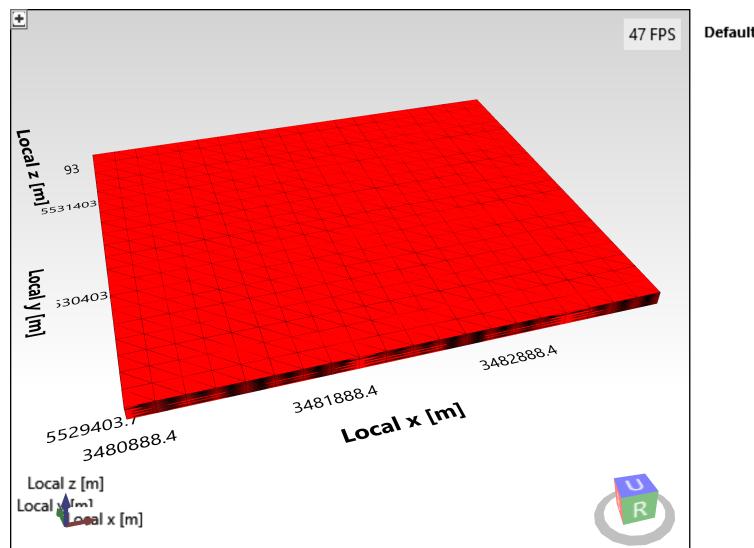


Figure 15: Three-dimensional regular, hexahedral mesh.



When displaying big meshes and selecting the *Wireframe and faces* or *Wireframe* option can lead to significant performance drawbacks since each wire on a wireframe is composed of ten triangles.

### Meshing two surfaces

Two surfaces can be meshed when they provide identical dimensionality. The functionality will be demonstrated by generation of a 3-D mesh from our surfaces *Base Rotliegend Interpolated* and *Top Rotliegend Interpolated*. Since both base on the same mesh, they also provide identical dimensionality). If the data sets are not present in your data sets anymore, load them again from the files *BaseExample.gmsh* and *TopExample.gmsh*. Go to the **Discretization** menu and select those two meshes as **Source data sets**. Then select the option *Two bounding surfaces* in the **Boundary type** drop-down box. We will define 10 bins in z direction and click **Compute discretization**. The result will be added to the data sets. Please, name this mesh *sGRID Rotliegend*.



If the meshes do not provide identical dimensionalities an empty mesh will be created. If that occurs feel free to generate new surfaces in the bounding area of your input surfaces with identical dimensionality and apply an IDW interpolation of the elevation to produce correct input meshes. The error can be assessed in the **Results** section that provides measures quantifying the quality of the interpolation.

The result should look like fig. 16.

#### 2.7.4 Complex geometries

Following section contains information on how to generate more complex geometries than the example before. Therefore, six surfaces are provided in the *Tutorial/Surfaces*

## 2.8 Performing Predictions

### 2.9 2-D predictions

Using our surface we will perform our first interpolation of the values of the thermal conductivity. Therefore, we make sure that we have both *BaseRotliegend-Interpolated.gmsh* as well as the thermal conductivity measurements loaded as **Data sets**. After that we will open the expander left from the data tables and select our *Thermal conductivity* mesh as a **Source data set**. Now we

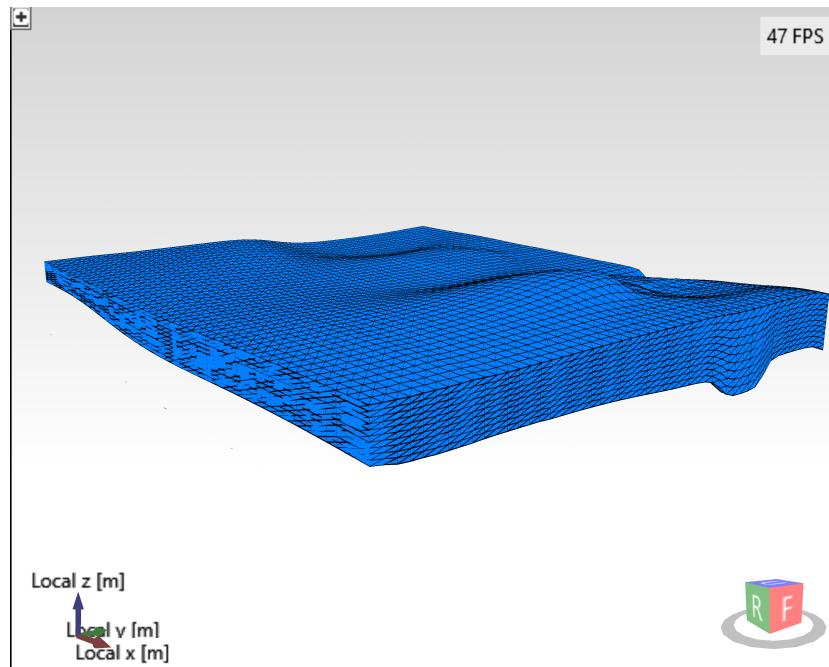


Figure 16: Stratigraphic grid of the Rotliegend group on the Sprendlinger Horst.

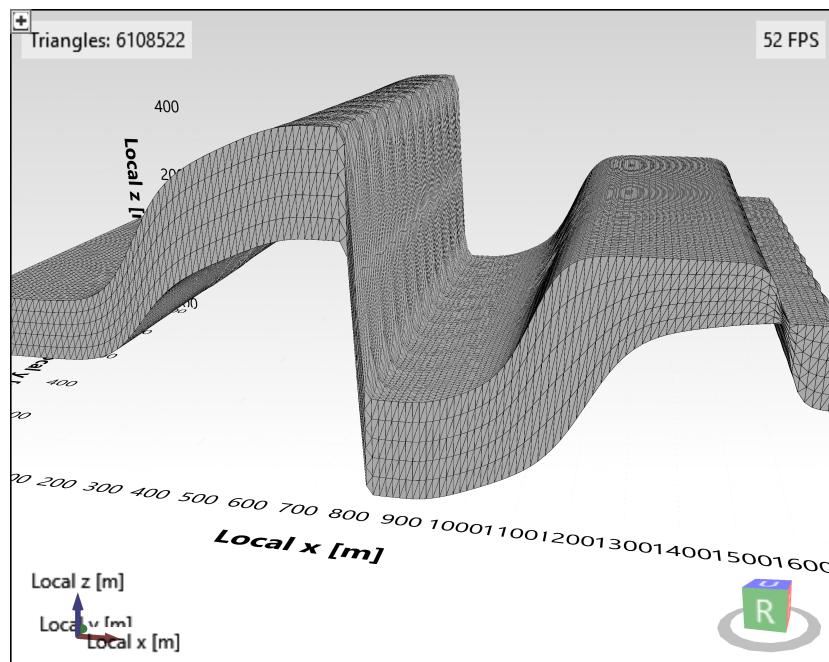


Figure 17: Complexly folded geological domain mesh.

will open the **Interpolation** expander and select *Base Rotliegend Interpolated* as **Target data set**. Our aim is to interpolate the unknown values of the surface using the thermal conductivity measurements provided in the source data set. When performing interpolations, a new data set called *Interpolated mesh* will be created. This data set contains all points from the **Target data set** with the interpolated values from the **Source data sets**. At first, we will perform an inverse distance weighting interpolation. Therefore, we define *Value* as interpolation feature, *Inverse Distance Weighting* as **Interpolation method** and a neighborhood of 1000 (x), 1000 (y) and 50 m (z) and we will use the maximum number of neighbors (9999). Rest of the option will remain the default values.

When pressing *Compute interpolation* the algorithm will be deployed and the results of the interpolation will be added as a new mesh. Each interpolation is automatically cross-validated using the root-mean-squared-error (RMSE) and the mean-absolute-error (MAE) that give quantitative insights into quality of the interpolation. In our example we could produce an RMSE of 0.32 and an MAE of 0.20.

We will add the newly generated mesh to the **3D chart** now and apply the settings from fig. 10 with exception of the **Is color map** option in the **Data series** menu that we will tick now and a **Wireframe thickness** of 1. Under **Chart style → Legend → Colorbar** we will define a max value of 2.25 and a Min value of 2. Additionally we will name our interpolated parameter *Thermal conductivity* in the **Title** option. Fig. 18 shows the result. High thermal conductivity values in the reddish areas indicate another type of basement (granodiorite) than in the low-thermal conductivity areas indicated in blue (monzodiorite).

### 2.9.1 Checking the interpolation for consistency

One way to assess quality of an interpolation is to evaluate the measures resulting from cross-validation. Another way is to check whether the original distribution is preserved over the interpolation. Therefor, we add both, the original data set and the interpolated mesh to the **Box-Whisker chart**. We will exclude outliers for better assessment. Now, we see that the interpolation does not represent the entire distribution of the original measurements (fig. 19).

This behaviour is due to the interpolated surface that only represents a small subset of the total domain that the original measurements are included in.

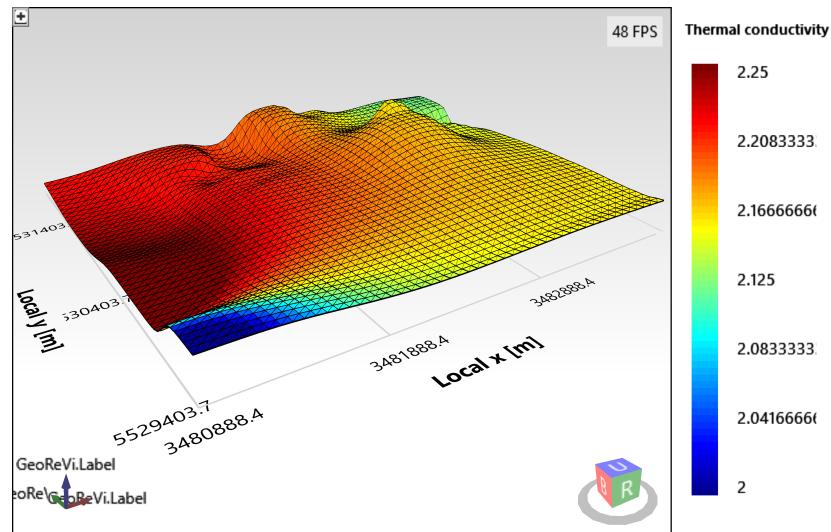


Figure 18: Exaggerated faces visualization of a surface in GeoReVi

#### Comparison interpolation/original values

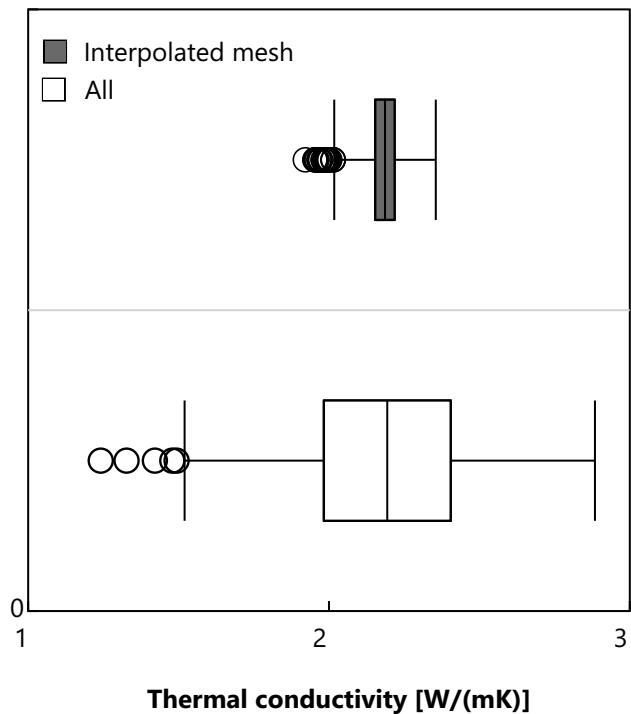


Figure 19: Comparison between the original distribution and the interpolated values.

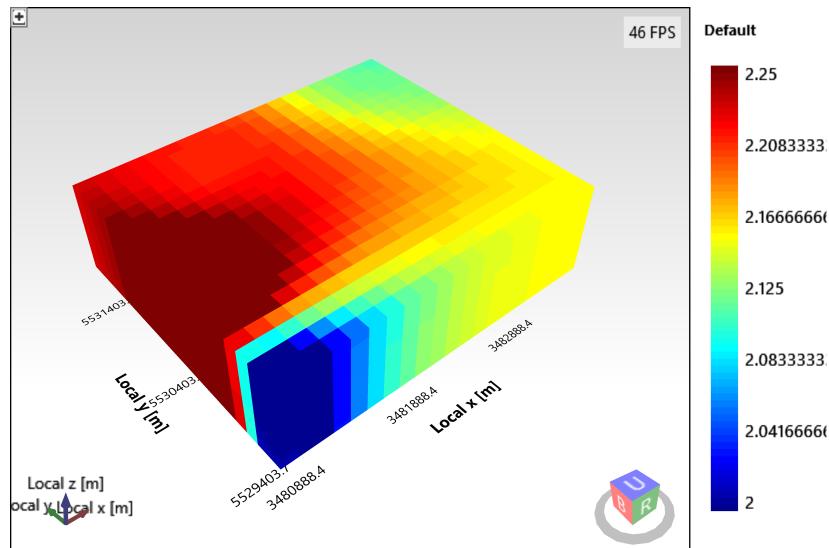


Figure 20: Three-dimensional property model of the thermal conductivity in the study area.

### 2.9.2 3-D predictions

#### Deterministic interpolation

Now, we want to create a 3-D prediction. Before we do that we remove all meshes except for *Thermal conductivity* and *SGrid Rotliegend*. In order to achieve this we select our *Thermal conductivity* data set as a **Source data set** and the 3D model as **Target data set**. Now, we select *Invese Distance Weighting* as **Interpolation method** and simply leave all parameters as default. When we press **Compute interpolation** the parametrized 3D model will be added and we name it *3D thermal conductivity Rotliegend*. When adding the result to the **3D chart** and exaggerating it by a factor of 5 it should look like fig. 21.

#### Stochastic interpolation

#### Simulation

Again, go to the semivariogram and define a range of 700, a nugget of 0 and a sill of 0.08. Go back to **Data sets** and select *Thermal conductivity* as **Source data set**. As **Target data set**, we again select our *SGrid Rotliegend*. We will again interpolate the **Value** with the **Euclidean distance** but this time with the **Sequential Gaussian Simulation** (SGS) method. This method simulates a gaussian random function at each interpolated point based on the previously derived simple kriging variance. When you press **Compute interpolation** a result of the SGS might look like fig. ??.

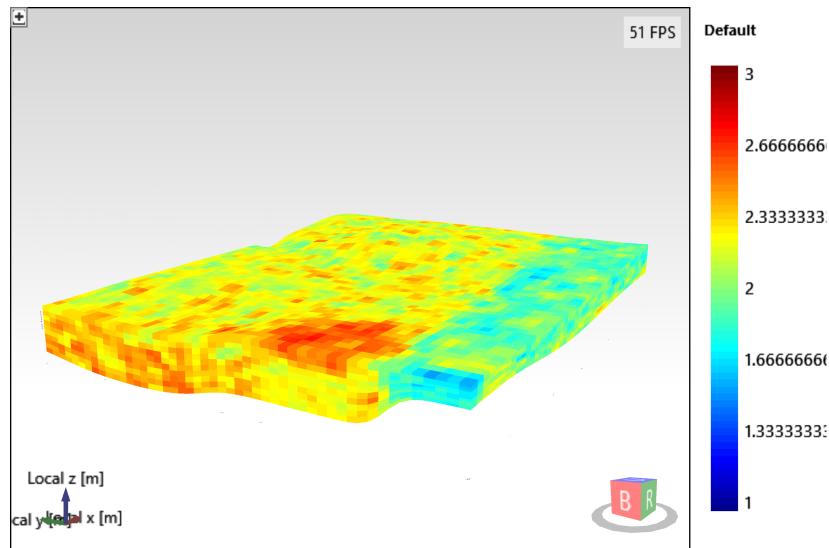


Figure 21: Three-dimensional property model generated by SGS of the thermal conductivity in the study area.

### 2.9.3 Rescaling

When interpolations do not reproduce the original distribution it can be useful to rescale the properties to the original distribution after interpolation. Therefore, a scaling is applied on the interpolated mesh values according to

$$z_s(x) = \frac{z(x) - z_{min}}{z_{max} - z_{min}} \cdot (t_{max} - t_{min}) + t_{min} \in [t_{min}, t_{max}], \quad (1)$$

where  $z(x)$  is the calculated value,  $z_{min}$  is the minimum value of the total sample  $\mathbf{z}$ ,  $z_{max}$  is the maximum value and  $t_{min}$  and  $t_{max}$  are the minimum respectively maximum value of the target range.

A rescaling can be performed when the **Transform to original distribution range** check box is selected in the **Interpolation** menu.

### 2.9.4 Error integration

Measurements often provide a strong susceptibility to errors. Therefore, errors can be considered by prediction algorithms. To integrate a measurement error, the **Error variance** text box in the **Interpolation** menu can be filled with a variance value. See more background in equation 23.

## 3 Getting started in depth

Following section contains comprehensive descriptions on the data, algorithms and the software structure of GeoReVi.

### 3.1 Introduction

Reservoir characterization is a crucial procedure for subsurface reservoir resource assessment and production performance prediction. [11] defines reservoir characterization as '*the process of preparing a quantitative representation of a reservoir using data from a variety of sources and disciplines*'. Accordingly, as numerous field parameters contribute to a reservoirs potential, this problem can be considered being high-dimensional with multiple contributing domains. Regardless of whether considering hydrocarbon, geothermal or groundwater systems, carbon sequestration or final disposal of radioactive material, a comprehensive knowledge on the underground architecture and field parameter distribution is substantial for making profitable business-critical and sustainable decisions [1].

In particular the spatio-temporal distribution of reservoir and fluid field parameters is critical for economic and sustainable exploitation of the resource hosted by the reservoir [36, 15]. Hence, we need quantitative information about a reservoirs physical and geological characteristics which is often provided as a tremendous number of heterogeneous, high-dimensional, cross-disciplinary data sets compiled into disparate data integration platforms. Additionally, due to modern sensor systems and low-cost data storage, the amount of data produced during exploration and exploitation is rapidly increasing in the last years. Thus, data sets produced in reservoir studies regularly extend the degree of complexity to be efficiently analyzed by domain experts. Consequently, modern automatized technologies to explore and model reservoir-related data are required to increase underground predictability and eventually optimize exploitation efficiency.

Reservoir-related data is mostly compiled from numerous domains with highly diverging ontologies [14]. This data is produced in well log measurements, during reservoir production, in outcrop analogue studies or in geophysical surveys and can be both static (e.g. depth of lithological bodies) or dynamic (e.g. production rates). Figure 22 conceptualizes an exemplary exploitation strategy for hydrothermal power production. Information about the lithological architecture and field parameter variability can be derived from both, the reservoir itself as well as outcrop analogues [12, 19, 3] that can be easily accessed. Paramount field parameters diverge in their mathematical nature. Permeability, stress, thermal conductivity/diffusivity for instance

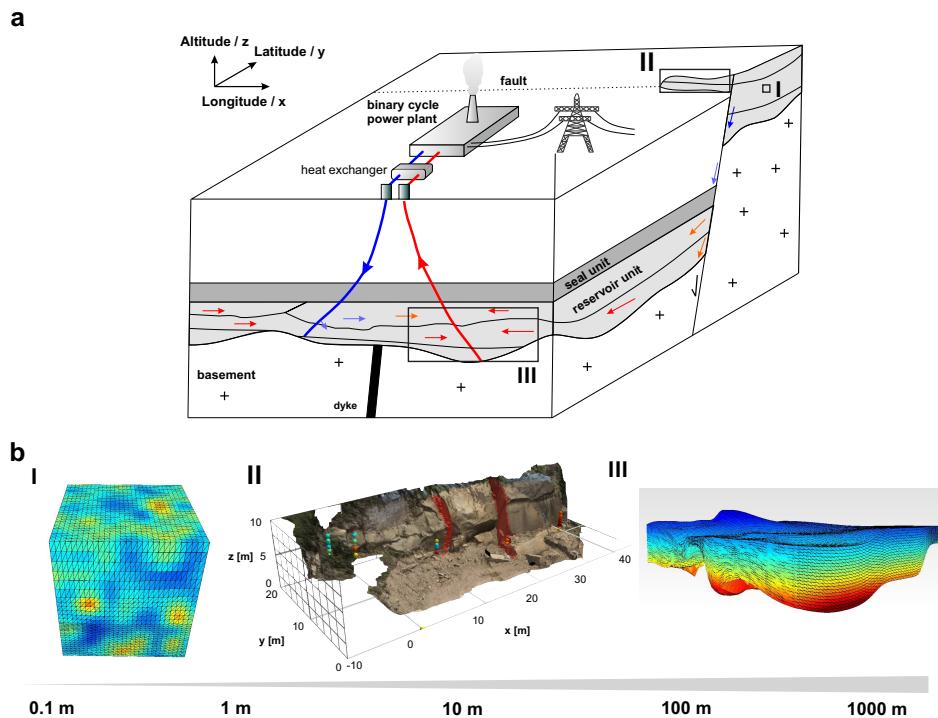


Figure 22: Concept of a multiscale outcrop analogue study.

are direction-dependent parameters and should be considered as tensors [27, 7, 10] whereas porosity or grain density are parameters of scalar type. This diversity in mathematical formats increases the effort to normalize reservoir-related data models and hampers data analytics and modeling of the field parameters often leading to inaccurate simplifications of reservoir simulations.

#### 3.1.1 Starting the application

First, the user will see the *Login* screen.

If a connection to the server database is established, the text block in the lower left corner is displayed in green color. Otherwise, the text block indicates a missing connection through showing a red color and associated text. If you are not registered in your institutional database yet go to the "Register and Sign In" chapter. Alternatively, you can work on a local database by clicking on "Local mode". When you work on your local database make sure that you update your database by clicking **Start → Update LocalDB** to get the domain data for chronostratigraphy, petrography and facies analysis. After login, the user finds the main navigation menu in the top part of the main screen.

### 3 GETTING STARTED IN DEPTH

---

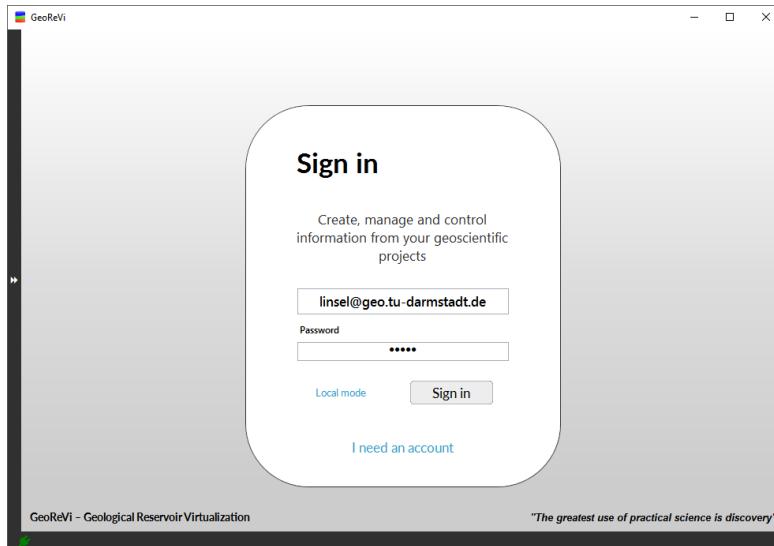


Figure 23: Login view.

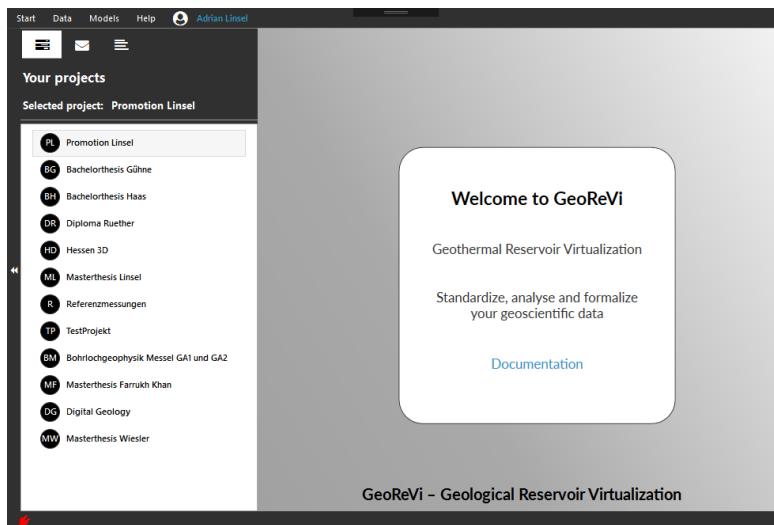


Figure 24: Home view with projects sidebar.

The general structure is

Start	Data	Map	Help
Home	Geological objects	Global Map	Documentation
Projects	Lithology		About GeoReVi
	Rock Samples		
	Measurements		

#### 3.1.2 Registration and Sign in

After installation of the server database, the user can create an individual account to access data from GeoReVi. Therefore, the user has to create an account by providing information. Other users will see the email address and affiliation provided in the registration when inviting users to their projects. The information will be encrypted and stored in the database. The account can be deleted or user information can be changed under *My Account*.

**Please, be aware, that all information you provided in GeoReVi won't be accessible anymore when you delete your account!**

Your projects and its related data will exist further on and subscribed users will have access to the projects related data unless those are transferred to another project or deleted, too.

After successful registration, the user can sign in by the provided user name and password. At the beginning, the user doesn't have access to any projects but he can see objects of investigation for planning studies and field campaigns.

Projects, the user participates at, will be loaded after sign in and can be selected in the drop-down box next to the main navigation menu. After selection, the user will work on the selected projects and has access to data related to that. Most times, many people will work on one specific project. Therefore, only the uploader and the project creator will have writing-access to the data sets to avoid unwanted data manipulation.

## 3.2 Obtaining a Bing Maps Key

In order to use the full functionality of GeoReVi the user has to obtain a Bing Maps Key under <https://docs.microsoft.com/en-us/bingmaps/getting-started/bing-maps-dev-center-help/getting-a-bing-maps-key>. The key has to be inserted in the K.csv file located in the installation folder under /Media/Data. Alternatively navigate to **Maps** → **Global map** and press the **Insert Bing Maps Key** in the left upper corner. This will open the previously mentioned table where you can insert your key. When opening the **Global map** again, there should not be a warning of an invalid key anymore.

## 3.3 Sample data

We provided a sample data set for the first use of GeoReVi. In the [GitHub directory](#) many samples can be found that serve as input data for GeoReVi. A comprehensive data set is provided in the Disibodenberg folder. Here, we produced a comprehensive case study to validate the software. The rock samples and measurements are already documented in the local database when GeoReVi is installed.

Additionally, we provide 3-D models of rock cubes and the entire outcrop that were generated with GeoReVi. Those files can be imported and directly visualized. In the following sections, we will often come back to the Disibodenberg sample data set and exemplary perform workflows with this data set.

## 3.4 Data management

Data produced in subsurface studies include various geoscientific domains like (petro)physics, sedimentology, hydrogeology, instrumental analytics or geo-physics. GeoReVi facilitates the user to manage data sets from all of these entities. Entities like chronostratigraphy or petrography are supplied according to international standards like chronostratigraphic units from the International Chronostratigraphic Chart 2017 (ICC, [33]) and the petrographic terms mostly by definitions from the British Geological Survey (BGS) or American Geological Institute (AGI).

Entities that are not standardized internationally were tried to be reduced on the most normalized level. Lithostratigraphy for instance provides a more or less internationally applied standard to subdivide units hierarchically like following:

1. Group

2. Subgroup
3. Formation
4. Subformation/Member

This schema was transferred into GeoReVi. Nevertheless, in contrast to the chronostratigraphic units, a user also has to be able to adapt information of a lithostratigraphic unit. Generally, data sets can be inserted, updated and deleted manually in accordance to common **CRUD** standards (*Create, Retrieve, Update, Delete*). A user always works on one of the projects where subscription is provided. Every data set is equipped with the ID of the user who generated the data set. Only this user is allowed to update or delete associated data sets. Users with project subscription may retrieve data sets.

#### 3.4.1 Supported data types

Data from various geoscientific disciplines is supported in the GeoReVi databases. Many types of information necessary for geothermal reservoir rock characterization are accomplished and rationalized in a multidisciplinary data model. The basis for GeoReVi's data model builds the data collection of [5] that was remodeled into a relational data model according to the relational theory after [8]. Additionally, selected fragments of the rock mechanical data model from [21] and the global geochemical database of [16] were used to extend GeoReVi's data model . However, most data models are developed within a specific frame, wherefore each data model had to be adapted to provide compatibility with the core scheme of GeoReVi.

Following information can be processed by the databases:

1. **Lithology**
  - (a) lithofacies
  - (b) architectural elements
  - (c) genetic units
  - (d) lithological logs
2. **Geological objects of investigation:**
  - (a) outcrops
  - (b) drillings/wells
  - (c) geophysical transects

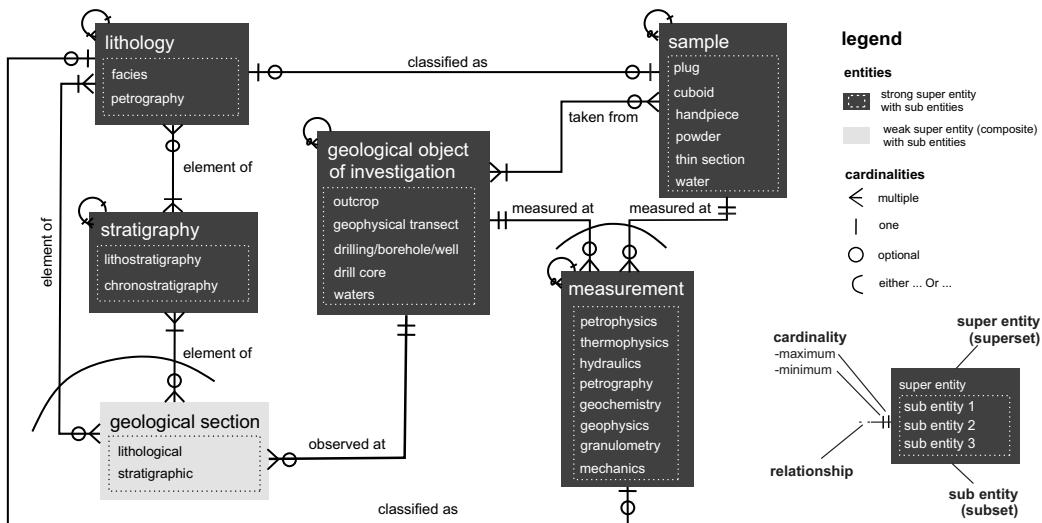


Figure 25: Core Entity-Relationship model of the GeoReVi database. To provide a better overview, attributes were excluded from the model.

(d) hydro(geo)logical objects

#### 3. Field/drill core samples

- (a) cylinders/plugs
- (b) cuboids
- (c) handpieces
- (d) thin sections
- (e) powders
- (f) soils
- (g) sediments

#### 4. Analytical instruments

#### 5. Laboratory measurements

- (a) grain density
- (b) bulk density
- (c) effective/total porosity
- (d) apparent/intrinsic permeability
- (e) (saturated) thermal conductivity

- (f) thermal diffusivity
- (g) (saturated) p- and s- wave velocity,
- (h) bulk geochemistry (oxides and trace elements)
- (i) isotopes
- (j) electrical resistivity
- (k) rock strength (uni- and triaxial)
- (l) grain size

#### 6. Field measurements

- (a) spectral  $\gamma$ -ray
- (b) total  $\gamma$ -ray
- (c) magnetic susceptibility
- (d) palaeo flow
- (e) bedding
- (f) lineaments
- (g) joints
- (h) temperature
- (i) sonic log
- (j) rock quality designation index
- (k) hydraulic head
- (l) bounding surface locations

#### 7. Stratigraphy

- (a) lithostratigraphic units
- (b) basins

All those entities include meta-information like measurement conditions, subtype definitions or detailed descriptions. However, the visualization and analysis functionality of GeoReVi can be used with local data of any kind as well. Users are encouraged to send us suggestions for data, they would like to have available in GeoReVi too. Please address your requests to [contact@georevi.com](mailto:contact@georevi.com). We will try to answer your requests as soon as possible.

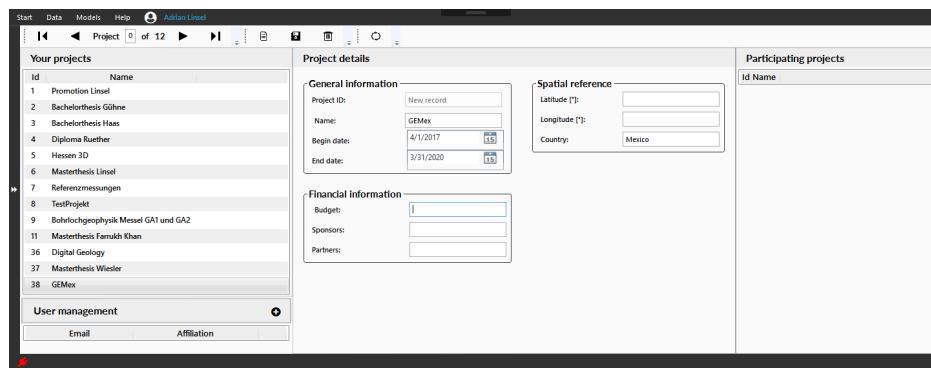


Figure 26: Projects view with a newly added project.

#### 3.4.2 Provided data set

In the local application we provided a data set from one of our publications (Linsel et al. 2019). The data set is located in the "Disibodenberg" project. Here you find 41 rock samples (plugs) taken from the outcrop "Sandstone Quarry Obersulzbach". Additionally, you find

#### 3.4.3 Project Management

Projects can be managed and created in the *Projects* menu. To improve the overview, duplicate project names are forbidden. Duplicate project names will be changed during registration and the user will be informed about it. After creating a project, the creator can add or remove users. By clicking on '+' in the user management frame a user can be added via dialog service. '-' will unsubscribe the participating user. A user can also unsubscribe from the project by clicking the '-' button in the 'Participating projects' frame.

#### 3.4.4 Objects of investigation

Objects of investigation are all object that can be investigated in a geoscientific context. You can get to the objects form via **Data → Geological objects**. In GeoReVi you can differentiate between "Outcrop", "Drilling", "Transect" and "Water body".

Objects of investigation are visible to all users working on the same database. Therefore, an object of investigation has to provide an unique name what is tested when entering it. If the name is already given to another object, you will have to change it accordingly.

### 3 GETTING STARTED IN DEPTH

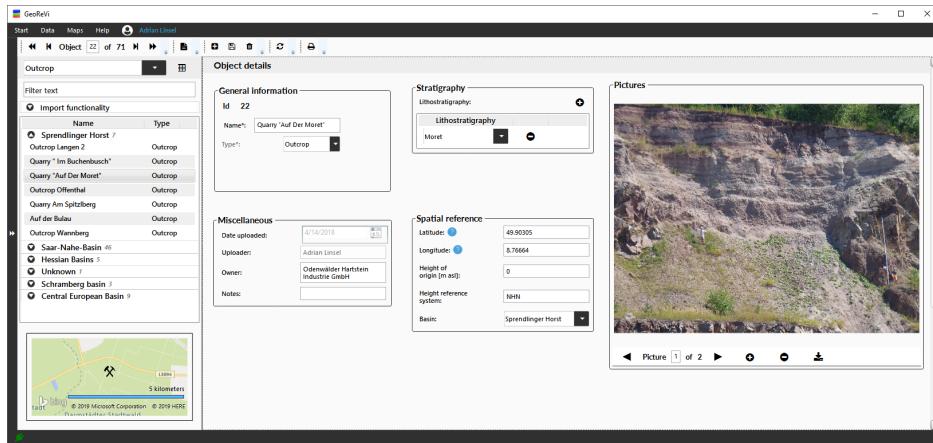


Figure 27: Objects of investigation form.

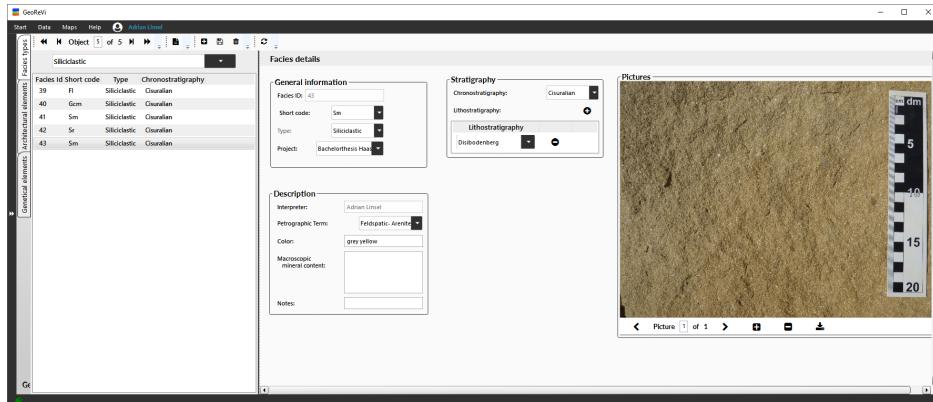


Figure 28: Lithologies form.

#### 3.4.5 Lithologies

Lithologies can be stored according to the lithofacies concept after Miall (1985). Here, geological units are subdivided into three classes namely lithofacies types, architectural elements and genetical units. Here applies

$$\text{lithofaciestype} \subset \text{architecturalelement} \subset \text{geneticalunit} \quad (2)$$

Lithologies can be updated in the "Lithologies" menu. When entering this menu you will find a tab-item with the names "Facies", "Architectural elements" and "Genetical units". General information can be documented in the main forms. To store details the user can open the "Details" window.

#### 3.4.6 Rock samples

Rock samples taken from the field can be recorded in the "Rock samples" form. Again the name of a rock sample must be unique across the entire database. An algorithm ensures this while inserting the name of the rock sample and provides an alternative if a sample name is already in use. To avoid this, we recommend to report all samples in the System for Earth Sample Registration (<http://www.geosamples.org/>). This will provide each sample an individual International Geosample Number (IGSN) that can be used globally.

Samples can be either recorded in the provided data form one by one or loaded in bulk using the import section. Therefore, a .CSV or .XLSX file containing the individual rock samples must be dropped in the "Import" field.

Details can again be documented in the "Details" window.

#### 3.4.7 Measurements

When you navigate to the 'Measurements' view, you will find two types of measurements in two separate views. Laboratory measurements are measurements related to a rock sample. When you open that view you can find the documented objects of investigations in the list on the left side. There you can select the rock sample where you want to archive a new measurement or from which you want to analyze the data. When selecting the rock sample, all types of measurements produced on that rock sample will be loaded and displayed in the list in the central part of the form.

Field measurements are measurements directly generated in or on an object of investigation. You can archive measurements such as well logs or structural measurements here. In the list on the left side you will find again all objects of investigation you documented beforehand. When selecting an object, all measurement related to that object are loaded in the list in the central part. To add measurements to a sample or to an object you can click the '+' button above the list view or you right-click on the list and select 'Add measurement'. A measurement of undefined type will be added to the list. To archive further information you will have to define the particular type of measurement under 'Type'. After saving the changes you can enter further information in the tab item 'Parameter'. If you want to import a set of measurements please refer to the chapter 3.4.8. To delete a set of measurement you can select multiple measurements in the list, **right-click** and select **Delete selected measurements**. Confirm the command to delete the selected measurement and all of those will be deleted from the database.

### 3.4.8 Import to the database

Data can be imported from spreadsheet files. For import .XLSX and .CSV files with one header row are accepted. We recommend .CSV for import. Most data forms provide an import field in the left upper corner. This field can be expanded and the preferred import objects can be selected. Please ensure, that the .CSV cells provide the correct format (number, string, date, etc.). For file-import, the file containing the data has to be dropped into the border of the field. A file-import dialog will show up immediately, where the imported data columns have to be mapped to the existing database headers. Therefore, single headers from the import file and from the database can be selected and mapped together with the map button. Headers can also be mapped automatically by using the mapping wizard. This algorithm checks for similar headers and maps them automatically.

Some database tables require redundancies for import. For instance, a plug must contain the sample name twice, since two tables are connected via this name to each other. Hence, two columns containing the sample name have to be included into the import spreadsheet. Also, following entities contain **obligatory** fields, where a value has to be provided:

1. Rock samples
2. Objects of investigation
3. Measurements

Please check beforehand, that the provided cells in the selected file have the adequate data format for the imported entity.

#### Exemplary proceeding

1. Open the rock sample form under **Data → Rock samples**
2. Expand the import field by clicking on the **Import functionality** expander
3. Select the type of object you wish to import, in our case, this is the **Plug** sample type
4. Drag & Drop your .XLSX or .CSV files with the rock sample information on the import field
5. In the import form, corresponding headers have to be mapped. This can be done one after another by selecting the headers in the select and by clicking the right-arrow. The mapping will show up in the right

### **3 GETTING STARTED IN DEPTH**

---

mapping-list. An automatic mapping can be done by clicking the "Auto connect" button. This will map headers with identical names together and adds the connection to the mappings list. It is recommended to use the .CSV lists from the data exports since those provide the same headers as required in the import form.

### 3 GETTING STARTED IN DEPTH

Sample ID	Label	Type	Lithology	Petrography	Lithofacies	Chirography	Object of investigation	Sample type	Architectural element	Depositional environment	Local x [m]	Longitude [°]	Latitude
75	OSB1_1	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		3			4,7	7,647276	49,523635
76	OSB1_2	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			4,7	7,647276	49,523635
77	OSB4_1	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		1			4,7	7,647276	49,523635
78	OSB4_2	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			4,7	7,647276	49,523635
79	OSB6	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			10	7,647276	49,523635
80	OSB6	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			10	7,647276	49,523635
81	OSB7	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		1			4,7	7,647276	49,523635
82	OSB8	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			4,7	7,647276	49,523635
84	OSB9	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			10	7,647276	49,523635
85	OSB10	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			10	7,647276	49,523635
86	OSB11	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			10	7,647276	49,523635
88	OSB12	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0	7,647276	49,523635
89	OSB13	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			0	7,647276	49,523635
90	OSB14	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0,5	7,647276	49,523635
91	OSB14	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0,5	7,647276	49,523635
92	OSB15	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			0,5	7,647276	49,523635
93	OSB16	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0,5	7,647276	49,523635
95	OSB17	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			0,5	7,647276	49,523635
96	OSB18	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0	7,647276	49,523635
97	OSB19	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			0	7,647276	49,523635
98	OSB20	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0	7,647276	49,523635
99	OSB21	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		9			0,5	7,647276	49,523635
100	OSB22	Plug	Coralian		Sandstone Quarry Oberaulbach	Plug		8			0,5	7,647276	49,523635

Figure 29: Measurements form with an import area where .CSV and .XLSX files can be dropped for import..

## 3.5 Data analysis and visualization

Data analysis and visualization takes place in the 'Measurements' view. The form is built up by two components namely "Measurements" and "Data analysis". First one contains measurements according to section 3.4.7. The "Data analysis" component is built up by data sheets, statistic forms and charts where you can analyze and visualize your data.

### 3.5.1 Data transformation

Prior to analysis it can be necessary to transform data according to their type of distribution. Accordingly, we provide following types of data transformation procedures:

*logarithmic* :  $y_i = \log(x_i)$

*exponential* :  $y_i = 10^{x_i}$

$$z-score : y_i = \frac{x_i - \bar{x}}{\sigma}$$

$$rescaling : y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

$$mean-rescaling : y_i = \frac{x_i - \bar{x}}{x_{max} - x_{min}}$$

$$subtract-mean : y_i = x_i - \bar{x}$$

### 3.5.2 Univariate data analysis

To load a set of univariate, spatially referenced measurements click the black 'Load univariate data set' button above the "Univariate" tab. This will download all measurements from the selected type from all samples where the selected sample belongs to. For example when you have the 'Porosity' measurement of a sample called 'OSB1' selected and you click the button, all data of the selected combo box property from all rock samples belonging to the associated outcrop will be loaded in the univariate data sheet.

In the tab "Parameter statistics" the user can retrieve basic descriptive statistic measures. Basic statistical parameters including

$n = \text{count of the samples}$

$\text{maximum} = \text{maximum value of the samples}$

$\text{minimum} = \text{minimum value of the samples}$

$\text{range} = \text{maximum} - \text{minimum}$

$$\text{arithmetic mean} = \bar{x} = \frac{\sum_{i=0}^n x_i}{n}$$

$$\text{geometric mean} = \sqrt[n]{\prod_{i=0}^n x_i}$$

$$\text{harmonic mean} = \frac{n}{\sum_{i=0}^n \frac{1}{x_i}}$$

$$\text{sample variance} = \sigma^2 = \frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2$$

$$\text{standard deviation} = \sqrt{\sigma^2}$$

$$\text{skewness} = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=0}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2}\right)^3}$$

$$\text{kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=0}^n (x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2}\right)^4} - 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)}$$

Physical and lithological heterogeneity has a strong influence on both the predictability of reservoir properties in a target formation and the lifetime and recovery rate of any type of geological reservoirs [1, 9]. [13] present a set of quantitative heterogeneity measures of a univariate field parameter data set. These measures, that indicate the dispersion of a univariate

data set  $\mathbf{x}$ , include the coefficient of variation  $c_v$ , which can be calculated as  $c_v = \sigma \cdot \bar{x}^{-1}$ , with  $\sigma$  as the empirical standard deviation and  $\bar{x}$  as the arithmetic mean and the Dykstra-Parsons coefficient  $c_{dp}$ , which can be calculated as  $c_{dp} = (p_{84} - p_{50}) \cdot p_{50}^{-1}$ , with  $p_n$  as the  $n$ th percentile.

### Variography

Originally defined by [24], the semivariogram describes the spatial dependence of a random field variable in space. Variability of a regionalized field parameter  $z$  is computed at different scales by calculating the dissimilarity between pairs of data values  $z(x_\alpha)$  and  $z(x_\beta)$  at the locations  $x_\alpha$  and  $x_\beta$  in the spatial domain. As a measure for dissimilarity  $\gamma_{\alpha\beta}$  is calculated as  $\gamma_{\alpha\beta} = 0.5 \cdot (z(x_\alpha) - z(x_\beta))^2$ . Since  $x_\alpha$  and  $x_\beta$  can be expressed as points separated by a lag vector  $\mathbf{h}$ , dissimilarity can also be formulated as:

$$\gamma_{\alpha\beta}(\mathbf{h}) = \frac{1}{2} \cdot (z(x_\alpha + \mathbf{h}) - z(x_\alpha))^2 \in [0, \infty] \quad (3)$$

Semivariogram calculation results in a set of points representing the cumulative dissimilarity  $\gamma$  of point-pairs with the Euclidean distance of  $|\mathbf{h}|$  in the domain. The experimental semivariogram, however, represents the cumulative dissimilarity of a discrete set of point-pairs  $x$  with  $n_c$  representing the count of point-pairs within the distance classes  $\mathbf{h}_k$  of identical distance increments (eq. 4).

$$\gamma(\mathbf{h}) = \frac{1}{2n_c} \sum_{\alpha=1}^{n_c} (z(x_\alpha + \mathbf{h}) - z(x_\alpha))^2 \in [0, \infty] \quad (4)$$

The continuous counterpart, represented by the theoretical semivariogram, is an approximation to the experimental semivariogram assuming  $z(\mathbf{x})$  to be a stationary random field [39]. A theoretical variogram  $\gamma_{theo}$  is represented by a covariance function  $C$  with the relationship  $\gamma_{theo}(\mathbf{h}) = C(0) - C(\mathbf{h})$ , where  $C$  is a positive definite, even function. If a nugget effect is observable, the theoretical semivariogram is translated by the magnitude of the nugget ( $n$ ) on the ordinate. Six covariance models are mostly used to fit the experimental semivariogram namely the spherical, gaussian, exponential, power, kardinal sine and linear model [4, 31] like:

#### spherical

$$C_{sph}(\mathbf{h}) = \begin{cases} b \cdot \left(1 - \frac{3|\mathbf{h}|}{2a} + \frac{|\mathbf{h}|^3}{2a^3}\right) & \text{for } 0 \leq |\mathbf{h}| \leq a \\ 0 & \text{for } |\mathbf{h}| \geq a. \end{cases} \quad (5)$$

#### gaussian

$$C_{gau}(\mathbf{h}) = \begin{cases} b \cdot \exp\left(-\frac{|\mathbf{h}|^2}{a^2}\right) & \text{for } 0 \leq |\mathbf{h}| \leq a \\ 0 & \text{for } |\mathbf{h}| \geq a. \end{cases} \quad (6)$$

**exponential**

$$C_{exp}(\mathbf{h}) = b \cdot \exp\left(-\frac{|\mathbf{h}|}{a}\right) \quad \text{with } a, b > 0 \quad (7)$$

**power**

$$C_{pow}(\mathbf{h}) = b \cdot h^a \quad \text{with } a, b > 0 \quad (8)$$

**cardinal sine**

$$C_{lin}(\mathbf{h}) = \sin\left(\frac{|\mathbf{h}|}{a}\right) \cdot \left(\frac{|\mathbf{h}|}{a}\right)^{-1} \quad (9)$$

**linear**

$$C_{lin}(\mathbf{h}) = \begin{cases} b & \text{for } |\mathbf{h}| > 0 \\ 0 & \text{for } |\mathbf{h}| = 0 \end{cases} \quad (10)$$

In order to calculate the covariance function, the variables nugget (n), range (a) and sill (b) must be determined.

Semivariograms can be used to quantify the spatial or time correlation of a random field parameter [31, 17, 29]. Further on, the differences in range and sill in dissimilar directional semivariograms can quantify the geometric anisotropy of a field parameter [31]. The resulting semivariogram and covariance functions are input variables for kriging analyses.

To produce a semivariogram analysis you have to add a mesh to the semivariogram chart with **Right click → Add to semivariogram chart**. There the user has to update the chart with the button in the top left corner. Sometimes, the distance (x axis) and variance (y axis) are exceeding the pre-defined range of 10 and 10 respectively. So, the user has to adapt the axes. Under "Model" the user can adapt the properties of the experimental and theoretical semivariogram.

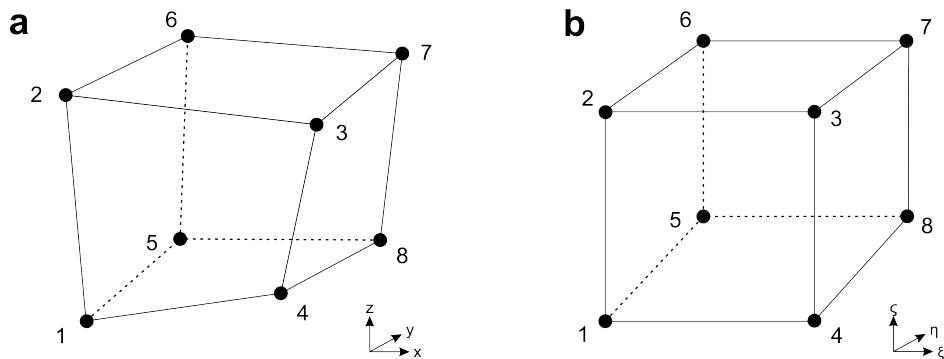


Figure 30: Conceptual representation of a hexahedral grid cell in GeoReVi as irregular shape (a) and with a local coordinate system (b).

### 3.5.3 Mesh generation

All data that is loaded in the univariate data sheet will be treated as a 'Mesh'. A mesh is a structure with vertices or nodes, faces and cells. When you load a data set it is a set of scattered, unconnected vertices. You can add the selected data sets to chart objects by right-clicking on the sheet and selecting the wanted option. If you want to create for instance a depth log of a property you have to right-click and select 'Add to line chart'. Then navigate to the line chart by 'Charts → Line chart' and select 'Z-direction'. When clicking the 'Refresh' button in the upper right corner, the graph will be drawn.

Each data set handled in the measurements view is a mesh. Meshes in GeoReVi consist of nodes (vertices), faces and cells. Additionally, a datatable is associated with each mesh that holds the measurement values. Data sets loaded from the database or from a .CSV, .XLSX or .XLS file are imported as discrete points into the data table without nodes, cells or faces. You can edit or remove single values in the data table displayed.

#### Regular mesh generation

Each node in a mesh created in GeoReVi is indexed according to a regular mesh. To create a simple hexahedral mesh in the bounding domain of one or more meshes the user has to expand the menu on the left side of the data set view. Under *Source data sets* you can select the meshes that should serve as source data sets for the meshing process. When you open the *Discretization* expander you can find define the type of mesh that should be created, the boundaries as x, y and z coordinates and the step width in each direction. You can create one-, two- and three-dimensional meshes. The default setting creates a three-dimensional mesh of 20x20x20 cells in the bounding domain

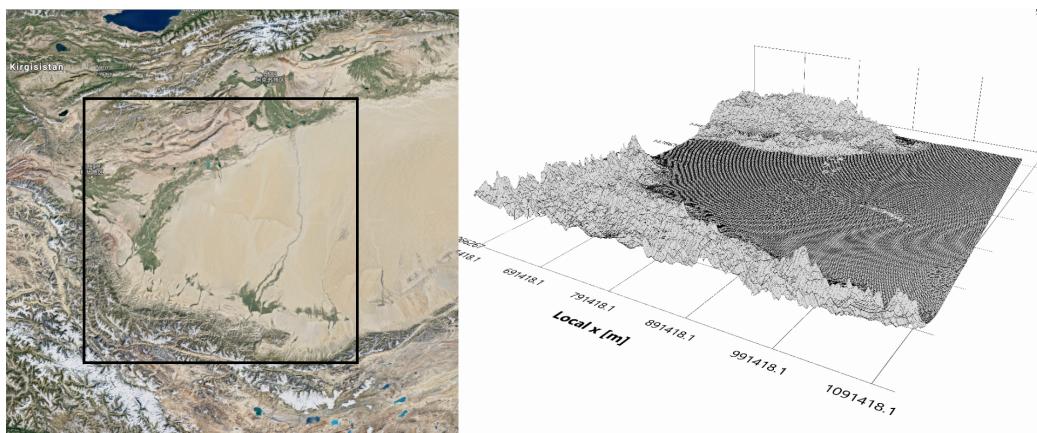


Figure 31: Digital elevation model of the area around the western part of the Xinjiang basin in China (15x exaggerated).

of the selected meshes. The dimension of a mesh is controlled by the step width. By selecting a step width of 0 the according dimension will be removed. If, for instance, step width z is reduced to 0, and x and y is kept as 20, a two-dimensional grid with 20 faces in x and 20 faces in y direction will be created. Similar, if two dimensions are reduced to 0, a line grid will be created.

**Digital elevation models** Based on the API provided by **Bing Maps** a two-dimensional data set can be assigned elevation values. Therefore, create a two-dimensional mesh in the area you want the elevation data from by providing start x, end x, start y and end y coordinates in WGS84 format.

### Gridding two surfaces

When you have two two-dimensional data sets you can create a three-dimensional grid out of those. However, both surfaces must provide identical dimensionality in x, y and z direction and one dimensionality less than the target mesh.

For 3-D mesh creating select the option "From two surfaces" in the "Boundary type" combo box and select those two surfaces that should be used in the "Source data set". When you press "Compute discretization" a 3-D grid will be created. Therefore, the two source data sets are projected from a physical in a local coordinate system (fig. 32. Nodes from the two source data sets serve as boundary nodes for the new mesh. After projection, the points in

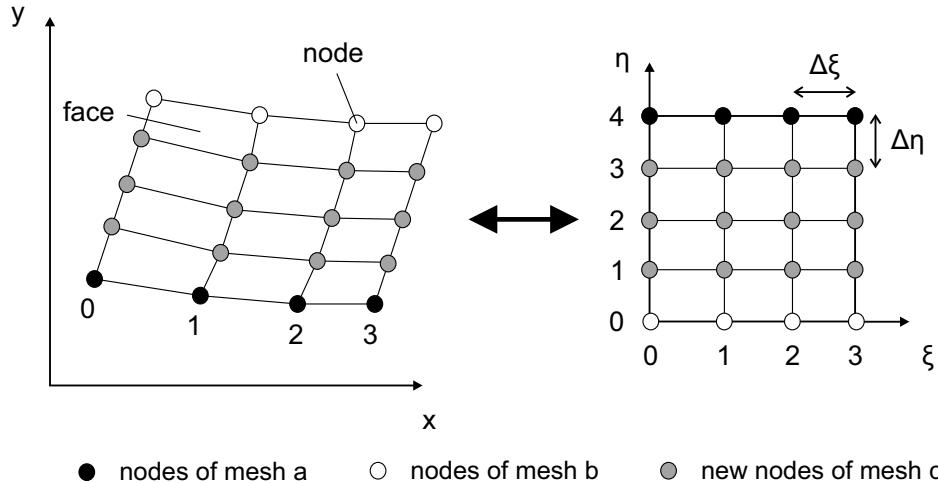


Figure 32: Concept of mesh generation using two constraint meshes of dimensionality  $d-1$  with  $d$  as the target dimension.

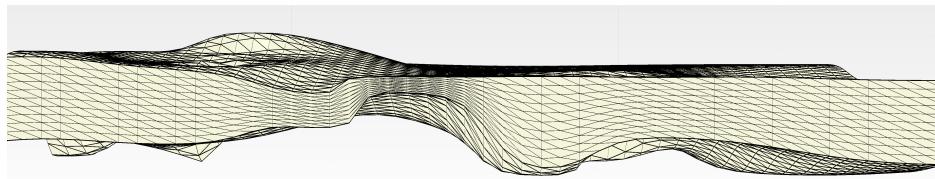


Figure 33: 3-D mesh generated with two bounding surfaces on the top and on the bottom.

between the two boundary meshes are created according to

$$\begin{aligned} R &= (i-1)(n_\zeta + 1) + 1 \\ C &= \text{mod}((i-1), (n_\iota + 1)) + 1, \end{aligned} \quad (11)$$

$$\begin{aligned} x &= \sum N_k(\zeta, \iota) x_k \\ y &= \sum N_k(\zeta, \iota) y_k \end{aligned} \quad (12)$$

where  $R$  is the row and  $C$  is the column of a node  $i$ ,  $n_\zeta$  is the number of

### 3.5.4 Mesh statistics

A mesh provides characteristic properties like dimensionality, count of nodes, faces and cells or bulk volume/area. Those statistics can be computed in the **Mesh statistics** tab. When the user presses "Compute" he or she will get all that information displayed immediately.

The volume of a mesh is defined as

$$V_m = \sum_{i=1}^n V_i, \quad (13)$$

where  $n$  is the number of elements and  $V_i$  is the volume of the  $i^{th}$  element. The volume of a tetrahedral element consisting of the nodes **a**, **b**, **c** and **d** the volume is

$$V_t = \frac{|(\mathbf{a} - \mathbf{d}) \cdot ((\mathbf{b} - \mathbf{d}) \times (\mathbf{c} - \mathbf{d}))|}{6}. \quad (14)$$

In order to calculate the volume of a hexahedral element, it will be subdivided into six tetrahedral elements whose cumulative volumes are the volume of the hexahedral element like

$$V_h = \sum_{i=1}^6 V_t^i. \quad (15)$$

### 3.5.5 Spatial interpolation and simulation

In the following section, we will explain how to interpolate values from one or more source data sets to a target mesh. To reduce the content in this section, we will cover the most important algorithms with examples. For more details on the theory behind the interpolation and simulation algorithms we will refer to the textbooks of Wackernagel (2003) and Webster & Margaret (2007).

All types of meshes can be used as both input and target of an interpolation. The selected **Source data sets** serve as constraints for the interpolation algorithms. The **Target data set** provides the nodes where the interpolated values will be calculated. The user can interpolate either the value, elevation, latitude/x or longitude/y of a mesh.

Spatial inter- and extrapolation can be generated with deterministic and stochastic techniques. All interpolations are based on the assumption that a known constraint point  $x_k$  with a value  $z(x_k)$  has a weight on a discrete point  $x_0$  in space with an unknown value  $z(x_0)$ .

#### Neighborhood

The number of constraints in each interpolation can be restricted by defining a neighborhood. Therefore, a maximum distance in x, y and z direction can be defined. Also, the total number of constraint can be reduced to a certain count.

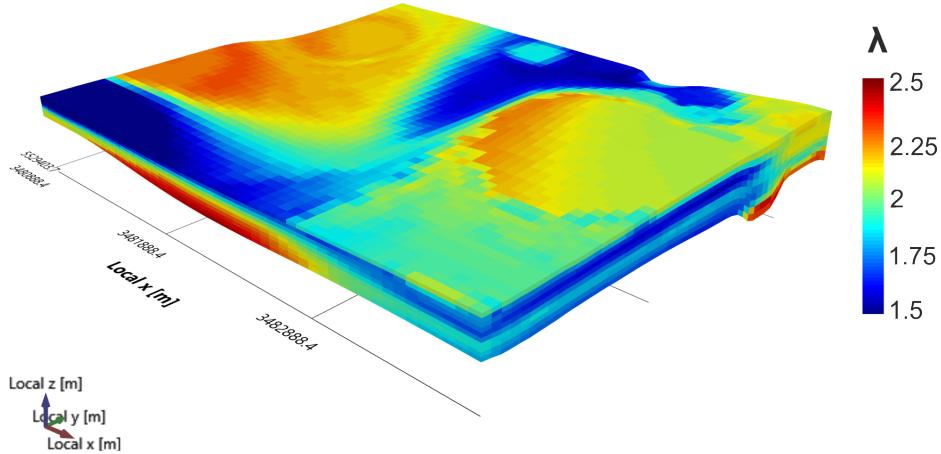


Figure 34: IDW interpolation result for thermal conductivity measurements in the Donnersberg formation located in central Germany.

### Inverse Distance Weighting

For deterministic interpolation we performed inverse distance weighting (IDW), p-value IDW and Shepard's IDW [35] interpolations. The IDW interpolation generally calculates an unknown value  $z(x_0)$  at point  $x_0$  by weighting the distance of that point to each known value point  $(x_k)$  in space. The underlying formula for IDW is:

$$z(x_0) = \frac{\sum_{k=1}^n 1/d_k^p \cdot z(x_k)}{\sum_{k=1}^n 1/d_k^p}, \quad (16)$$

where  $d$  is the Euclidean distance between the the point with the known value  $x_k$  and the point with the unknown value  $x_0$  and  $p$  is an exponent factor to bias the weights non-linearly. IDW is a reliable and widely applied method to interpolate static field parameters in one to three-dimensional space [28].

### Kriging

For stochastic interpolation we used simple kriging (SK), ordinary kriging (OK) and universal kriging (UK). Kriging is a common method to interpolate geological field parameters in space [23, 30]. Therefore, the value  $z(x_0)$  at an unknown point  $x_0$  is calculated by weighting the neighboring known values and building a linear combination of those like:

$$z(x_0) = \sum_{k=1}^n w_k \cdot z(x_k), \quad (17)$$

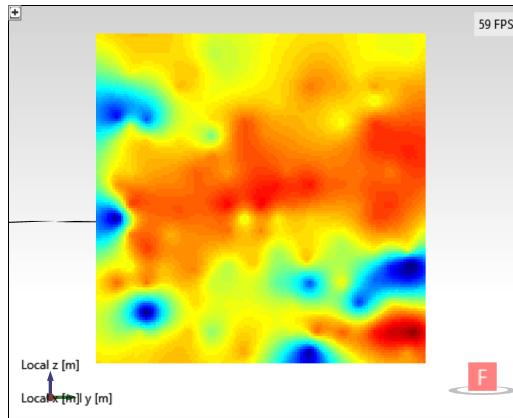


Figure 35: Simple kriging interpolation of the apparent permeability measured on 108 rock plugs on a rock slab of 0.5 x 0.5 m.

where  $w_k$  is the weight of the known point  $x_k$  with the value  $z(x_k)$ . While eq. 17 applies to SK and OK, UK

The kriging types primarily differ in the derivation of the weight vector. For all systems a set of linear equations has to be solved like it is outlined in the following paragraphs. The quality of kriging interpolation is dependent on the theoretical semivariogram and the goodness of fit to the experimental values.

### Simple Kriging weights

To obtain the simple kriging weights, a set of n equations have to be solved. In matrix notation, this set of equations can be written as:

$$\begin{pmatrix} c(x_1 - x_1) & \cdots & c(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ c(x_n - x_1) & \cdots & c(x_n - x_n) & 1 \end{pmatrix} \begin{pmatrix} w_1^{SK} \\ \vdots \\ w_n^{SK} \end{pmatrix} = \begin{pmatrix} c(x_1 - x_0) \\ \vdots \\ c(x_n - x_0) \end{pmatrix} \quad (18)$$

with  $c$  as covariance function,  $x_n$  as point with known value [39].

### Ordinary Kriging weights

To obtain the ordinary kriging weights, a set of n+1 equations have to be solved. In matrix notation, this set of equations can be written as:

$$\begin{aligned}
 & \begin{pmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \end{pmatrix} \\
 & = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix} \tag{19}
 \end{aligned}$$

with  $\gamma$  as theoretical semivariogram,  $x_n$  as point with known value and  $\mu$  as Lagrange parameter [39].

### Universal Kriging weights

To obtain the ordinary kriging weights, a set of  $n+1$  equations have to be solved. In matrix notation, this set of equations can be written as:

$$\begin{aligned}
 & \begin{pmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 & f_1(x_1) & \cdots & f_k(x_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 & f_1(x_n) & \cdots & f_k(x_n) \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ f_1(x_1) & \cdots & f_1(x_n) & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_k(x_1) & \cdots & f_k(x_n) & 0 & 0 & \cdots & 0 \end{pmatrix} \\
 & \cdot \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \\ \phi_0 \\ \vdots \\ \phi_k \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \\ f_1(x_0) \\ \vdots \\ f_k(x_0) \end{pmatrix} \tag{20}
 \end{aligned}$$

with  $\gamma$  as theoretical semivariogram,  $x_n$  as point with known value and  $\mu$  as Lagrange parameter  $f_i(x)$  as  $i^{th}$  polynomial [40].

### 3.5.6 Conditional Simulation

Another group of algorithms to produce spatial predictions are represented by conditional simulations. Most interpolation techniques do not represent the original parameter distribution adequately and induce a smoothing effect in the spatial distribution. To account for the spatial heterogeneity of a field parameter we implemented the sequential Gaussian simulation (SGS) algorithm. SGS is based on the results from the SK interpolation. In SK each interpolated point provides a SK variance  $\sigma_{SK}^2$  [40] like:

$$\sigma_{SK}^2 = C(0) - \sum_{k=1}^n w_k C(x_i, x_0), \quad (21)$$

where  $w_k$  is the kriging weight and  $C$  is the covariance function. SGS assumes that the kriging error is normally distributed with a mean of 0 and a variance of  $\sigma_{SK}^2$ . For each interpolated point  $x_i$  now a random value of the normal distribution  $\mathcal{N}(z(x_i), \sigma_{SK}^2)$  is drawn as  $z(x_i)$  using the Box-Muller transform that can be expressed as

$$z(x_0) = \sqrt{-2 \cdot \log(u_1) \cdot \cos(2\pi \cdot u_2)} \cdot \sigma + \mu \quad (22)$$

with  $u_1$  and  $u_2$  as random numbers  $\in [0, 1]$ ,  $\sigma$  as the standard deviation and  $\mu$  as arithmetic mean of the original distribution.

### 3.5.7 Measurement Error Integration

Integrating measurement errors into an interpolation can be achieved by estimating the measurement error precision  $\sigma_e$  with a variance of  $\sigma_e^2$  and incorporating it into the kriging system of linear equations like:

$$\begin{pmatrix} \gamma(x_1 - x_1) + \sigma_1^2 & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) + \sigma_n^2 & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix}, \quad (23)$$

what is the simple kriging system of linear equations [39] with error variance. In contrast to the conventional formula,  $\sigma_e^2$  was added in the diagonal of the matrix.

### 3.5.8 Cross-validation

Cross validation was conducted using point removal procedures called leave-p-out cross-validation (LpO CV). In LpO CV  $p$  randomly selected samples are removed from the input data set of size  $n$  with  $0 < p < n$  and the interpolation or simulation is performed without these samples. As measures of fit the root-mean-squared-error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (z(x_k) - z(x_0))^2} \quad (24)$$

and mean-absolute-error (MAE)

$$MAE = \frac{1}{n} \sum_{k=1}^n |z(x_k) - z(x_0)| \quad (25)$$

were used. While [6] questioned the status of the triangle inequality, that is required for a distance function metric, for the RMSE [6] showed that the RMSE satisfies it. If the model errors follow normal distribution RMSE is to favor over MAE [6].

## 3.6 Multivariate statistics

Multivariate analyses in GeoReVi comprise k-Means cluster analysis, principal component analysis, multidimensional scaling (Sammon mapping) and correlation analysis. The multivariate analysis can be performed in the **Multivariate** tab in the data analysis window. Here, the user can upload standard spreadsheets and perform basic data transformations on the individual columns. To add a data set to a multivariate analysis the user has to **right-click** on the selected data set and add it to the particular analysis.

### 3.6.1 Dimensionality reduction

As dimensionality reduction algorithm the principle component analysis (PCA) and Sammon mapping (SM) were implemented. Dimensionality reduction can be used e.g. to compress the observed data matrix, to detect multivariate outliers, to detect intrinsic correlation [39] or to proof the quality of a classification scheme applied on a data set.

#### Principal component analysis

The PCA is a statistical method which finds a linear projection of a data set  $X$  in relational format representing the data structure so that the variance is maximized in the lower-dimensional projection. Result from a PCA is a linear transformation, represented by a translation and rotation component. This transformation is also called *principal axis transformation*:

$$y_k = x_k - \bar{x} \cdot \mathbf{E} \quad (26)$$

where  $\mathbf{E}$  is a rotation matrix which is dependent on  $X$ . To determine  $\mathbf{E}$ , the variance ( $v_y$ ) in the target data set  $Y$  has to be maximized (eq. 27):

$$\begin{aligned} v_y &= \frac{1}{n-1} \sum_{k=1}^n y_k^T y_k \\ &= \frac{1}{n-1} \sum_{k=1}^n ((x_k - \bar{x}) \cdot \mathbf{E})^T \cdot (x_k - \bar{x}) \cdot \mathbf{E} \\ &= \frac{1}{n-1} \sum_{k=1}^n \mathbf{E}^T \cdot (x_k - \bar{x})^T \cdot (x_k - \bar{x}) \cdot \mathbf{E} \\ &= \mathbf{E}^T \left( \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^T \cdot (x_k - \bar{x}) \right) \cdot \mathbf{E} \\ &= \mathbf{E}^T \cdot C \cdot \mathbf{E} \end{aligned} \quad (27)$$

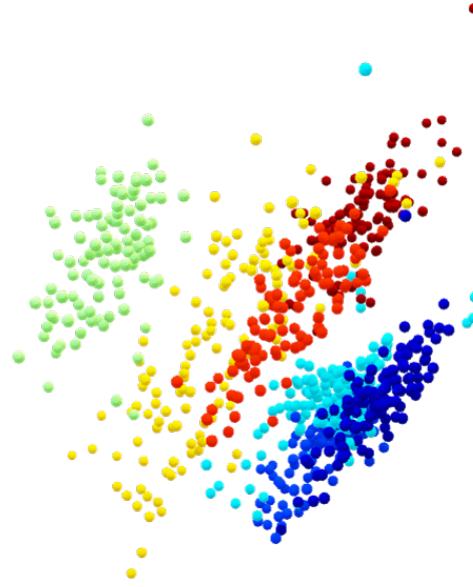


Figure 36: 3D representation of a Principal Component analysis.

where  $C$  is the covariance matrix (eq. ??) of  $X$  with the elements  $c_{ij}$ . Since the transformation matrix  $\mathbf{E}$  should be restricted to rotation, the following applies

$$\mathbf{E}^T \cdot \mathbf{E} = 1 \quad (28)$$

We will solve this optimization problem using the Lagrange optimization.

**Multidimensional scaling (Sammon Mapping)** SM aims to project a data set  $X$  into an  $n$ -dimensional representation  $Y$  minimizing the pairwise distances of  $X$  and  $Y$  so that  $d_{ij}^x \approx d_{ij}^y$ . Therefore, a representation  $Y$  is initiated randomly and Sammon's stress ( $E$ ) is calculated which is expressed as

$$E = \frac{1}{\sum_{i=1}^n \sum_{j=i+1}^n d_{ij}^x} \sum_{i=1}^n \sum_{j=i+1}^n \frac{(d_{ij}^y - d_{ij}^x)^2}{d_{ij}^x}, \quad (29)$$

where  $d$  is the Manhattan distance between the point pairs of  $X$  and  $Y$ . The SM algorithm minimizes the Sammon's stress gradient in a user-defined number of iterations.

### 3.6.2 Classification

GeoReVi covers classification domains for chronostratigraphy [33] , lithostratigraphy, petrography as well as facies, architectural element and depositional environment analyses [26] . The user-defined categories can be used to group the measurements for further analyses. Additionally, as an automated, multivariate classification method, k-Means clustering using Lloyd's algorithm [22] was implemented. This clustering method requires a predefined number of clusters, the data set should be subdivided into and categorized the data values by minimizing the sum of the squared differences from the data values  $\mathbf{x}$  to the cluster centers  $\mu$

$$c := \min_j \|\mathbf{x}^{(i)} - \mu_j\| \quad (30)$$

Another implemented automatic classification routine is the nearest-neighbor classification that assigns an unclassified observation to a class based on the nearest classified point from the input data set.

### 3.6.3 Correlation analysis

Correlation analysis can be performed on a data set with at least two outcome variables and are especially of relevance to correlate borehole geophysical logs with lab measurement results of core samples. As many field parameters provide mentionable physical relationships, the implementation of quantitative correlation analytical methods was necessary. To measure the linear correlation, the Pearson linear product-moment correlation coefficient ( $s_{pe}$ ) of two independent variables  $\mathbf{x}$  and  $\mathbf{y}$  was implemented.

$$s_{pe} = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\left(\sum_{k=1}^n x_k^2 - n \cdot \bar{x}^2\right) \left(\sum_{k=1}^n y_k^2 - n \cdot \bar{y}^2\right)} \quad (31)$$

Non-linear correlations can be quantified with Spearman's rank correlation ( $S_{sp}$ ) coefficient that is calculated with the ranks  $\mathbf{rg_x}$  and  $\mathbf{rg_y}$  of the independent variables  $\mathbf{x}$  and  $\mathbf{y}$ . The rank is defined as the index  $i$  of the value  $x_i$  when  $\mathbf{x}$  was sorted ascending. Accordingly, if all ranks are distinct integers  $s_{sp}$  can be calculated as

$$s_{sp} = 1 - \frac{6 \cdot \sum_{k=1}^n d_k^2}{n(n^2 - 1)} \quad (32)$$

with  $d_i$  as  $rg(x_i) - rg(y_i)$ .

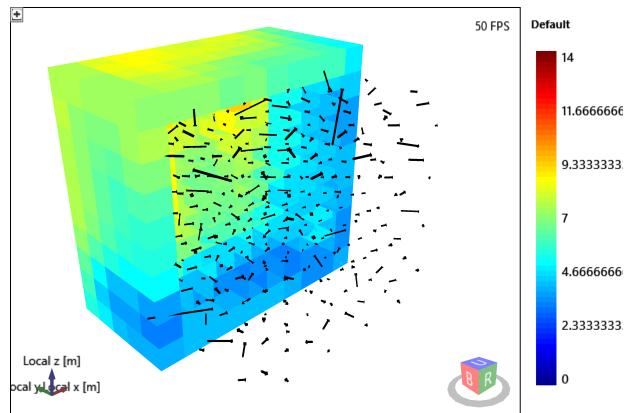


Figure 37: Vector field of the apparent permeability measured on a rock cube and interpolated with simple kriging.

## 3.7 Data visualization

Data visualization can be performed in different types of charts. GeoReVi includes line and scatter charts, bubble charts, ternary charts, histograms, box-whisker charts, semivariograms and 3-D charts. To add a mesh to a chart in the univariate analysis the user has to **right-click** → **Add mesh to chart**. When a mesh is added to a chart type it is displayed individually as a series. For first instructions, we refer to section 2 .

### 3.7.1 3-D chart

You can add meshes to the 3-D chart by **right-click** → **Add mesh to 3-D chart**. In the 3-D chart the user can adapt the chart width/height, background, colormap,

#### Vector field

Vector fields are displayed when the user selects the **Gradient** option. A vector field is based on the gradient  $\nabla$  of a field parameter  $f$  like

$$\nabla f = \left[ \frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \quad \frac{\partial f}{\partial z} \right]^T \quad (33)$$

## 4 Developing Plug-ins

This section is under preparation.

## 5 Personal data security

The database where GeoReVi stores user information is located on a local server or in the cloud which is dependent on the infrastructure provided by your administration. Passwords and user messages are encrypted and secured in a database management system. For internal communication, the implemented message service can be used. Messages are peer-to-peer encrypted and stored in the database.

## 6 Related works

VA is a widespread technology that is implemented in multiple software packages. [34] recently developed a VA framework for uncertainty analysis of reservoir models which can be considered the most similar software to GeoReVi. Domain experts can load and interactively compare reservoir models with regard to spatial field parameter uncertainty. However, [34] didn't overcome the durable data storage multi-user environment problems. SGeMS is another geostatistics application developed in C++, that covers a wide range of geostatistical modeling, simulation and visualization algorithms in one- to three-dimensional space [32]. However, SGeMS only provides a small range of statistical data analysis tools and does not provide database binding or geoscientific data formalization.

The .NET framework was successfully used for software development in numerous other geoscientific domains. Most recent approaches concentrated on hydrological system modeling and hydrological GIS [38, 37, 20]. In contrast to geothermal data management software packages like GeoData Manager [2], GeoReVi on the one hand side provides additional visualization functionality, but on the other hand side is not yet tested in commercial environments.

Past3 [18] constitutes a VA and KDD software related with paleontological problems. Nevertheless, Past3 holds a broad user community not only restricted to the paleontological domain thank to a wide range of universal statistical methods. In contrast to Past3, GeoReVi doesn't provide that wide range of statistical test and analysis methods yet. However, the database-binding and strongly normalized data model of GeoReVi provide fundamental standardization and data integrity which are not provided in any scientific software package with focus on reservoir characterization yet. Additionally, in contrast to Past3, GeoReVi provides spatial interpolation functionality in three-dimensional space.

[25] developed a Visual Basics based GUI named JeoStat to calculate semivariograms and to produce kriging interpolations. GeoReVi contains the same range of theoretical semivariograms and kriging functionality. Additionally, in contrast to JeoStat, GeoReVi provides controls to visualize results in 2D and 3D space.

## References

- [Ail12] Jia Ailin. “Advances and Challenges of Reservoir Characterization: A Review of the Current State-of-the-Art”. In: *Earth Sciences* (2012). DOI: [10.5772/26404](https://doi.org/10.5772/26404). URL: <https://doi.org/10.5772/26404>.
- [And13] Errol Anderson. “Developing a Universal Geothermal Data Management System”. In: *GRC Transactions* 37 (2013), pp. 1–8.
- [Are+15] Achim Aretz et al. “Outcrop analogue study of Permocarboniferous geothermal sandstone reservoir formations (northern Upper Rhine Graben, Germany): impact of mineral content, depositional environment and diagenesis on petrophysical properties”. In: *International Journal of Earth Sciences* 105.5 (2015), pp. 1431–1452. ISSN: 1437-3262. DOI: [10.1007/s00531-015-1263-2](https://doi.org/10.1007/s00531-015-1263-2). URL: <http://dx.doi.org/10.1007/s00531-015-1263-2> <http://link.springer.com/article/10.1007%2Fs00531-015-1263-2>.
- [Arm98] M. Armstrong. “Experimental Variograms”. In: *Basic Linear Geostatistics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 47–58. ISBN: 978-3-642-58727-6. DOI: [10.1007/978-3-642-58727-6\\_4](https://doi.org/10.1007/978-3-642-58727-6_4). URL: [https://doi.org/10.1007/978-3-642-58727-6\\_4](https://doi.org/10.1007/978-3-642-58727-6_4).
- [Bä+17] K. Bär et al. “P3 - International PetroPhysical Property Database”. In: *Proceedings, 42nd Workshop on Geothermal Reservoir Engineering* (Stanford, California). Ed. by Yolande Berbers and Willy Zwaenepoel. 2017, 1–6.
- [CD14] T. Chai and R. R. Draxler. “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature”. In: *Geosci. Model Dev.* 7.3 (2014), pp. 1247–1250. ISSN: 1991-9603. DOI: [10.5194/gmd-7-1247-2014](https://doi.org/10.5194/gmd-7-1247-2014). URL: <https://www.geosci-model-dev.net/7/1247/2014/>.
- [Cla+08] Jean-Baptiste Clavaud et al. “Permeability anisotropy and its relations with porous medium structure”. In: *Journal of Geophysical Research: Solid Earth* 113.B1 (2008), n/a–n/a. ISSN: 2156-2202. DOI: [10.1029/2007JB005004](https://doi.org/10.1029/2007JB005004). URL: <http://dx.doi.org/10.1029/2007JB005004>.
- [Cod70] E. F. Codd. “A relational model of data for large shared data banks”. In: *Commun. ACM* 13.6 (1970), pp. 377–387. ISSN: 0001-0782. DOI: [10.1145/362384.362685](https://doi.org/10.1145/362384.362685).

## REFERENCES

---

- [Cro+16] R. A. Crooijmans et al. “The influence of facies heterogeneity on the doublet performance in low-enthalpy geothermal sedimentary reservoirs”. In: *Geothermics* 64 (2016), pp. 209–219. ISSN: 0375-6505. DOI: <https://doi.org/10.1016/j.geothermics.2016.06.004>. URL: <http://www.sciencedirect.com/science/article/pii/S0375650516300517>.
- [Dü04] H. Dürrast. “Ultrasonic Laboratory Measurements on Spherical Samples — a Tool for the Investigation of Seismic Anisotropy in Rocks”. In: *Acoustical Imaging*. Ed. by W. Arnold and S. Hirsekorn. Dordrecht: Springer Netherlands, 2004, pp. 383–390. ISBN: 978-1-4020-2402-3. DOI: [10.1007/978-1-4020-2402-3\\_48](10.1007/978-1-4020-2402-3_48). URL: [https://doi.org/10.1007/978-1-4020-2402-3\\_48](https://doi.org/10.1007/978-1-4020-2402-3_48).
- [Fan02] John R. Fanchi. “Chapter 10 - Fundamentals of Reservoir Characterization”. In: *Shared Earth Modeling*. Ed. by John R. Fanchi. Woburn: Butterworth-Heinemann, 2002, pp. 170–181. ISBN: 978-0-7506-7522-2. DOI: <https://doi.org/10.1016/B978-075067522-2/50010-0>. URL: <http://www.sciencedirect.com/science/article/pii/B9780750675222500100>.
- [Fis+07] C. Fischer et al. “A 3D HIGH RESOLUTION MODEL OF BOUNDING SURFACES IN AEOLIAN-FLUVIAL DEPOSITS: AN OUTCROP ANALOGUE STUDY FROM THE PERMIAN ROTLIEGEND, NORTHERN GERMANY”. In: *Journal of Petroleum Geology* 30.3 (2007), pp. 257–273. ISSN: 1747-5457. DOI: <10.1111/j.1747-5457.2007.00257.x>. URL: <http://dx.doi.org/10.1111/j.1747-5457.2007.00257.x>.
- [Fit+15] P. J. R. Fitch et al. “An integrated and quantitative approach to petrophysical heterogeneity”. In: *Marine and Petroleum Geology* 63 (2015), pp. 82–96. ISSN: 0264-8172. DOI: <https://doi.org/10.1016/j.marpetgeo.2015.02.014>. URL: <http://www.sciencedirect.com/science/article/pii/S0264817215000434>.
- [GC10] Ji Ke Ge and Zuqin Chen. “Constructing Ontology-Based Petroleum Exploration Database for Knowledge Discovery”. In: *Applied Mechanics and Materials* 20-23 (2010), pp. 975–980. ISSN: 1662-7482. DOI: <10.4028/www.scientific.net/AMM.20-23.975>. URL: <https://www.scientific.net/AMM.20-23.975>.
- [GH18] H. Gudmundsdottir and R. Horne. “Reservoir Characterization and Prediction Modeling Using Statistical Techniques”. In: *43rd Workshop on Geothermal Reservoir Engineering* (Stanford, California). Stanford University, 2018.

---

## REFERENCES

- [GHH19] M. Gard, D. Hasterok, and J. Halpin. “Global whole-rock geochemical database compilation”. In: *Earth Syst. Sci. Discuss.* 2019 (2019), pp. 1–23. ISSN: 1866-3591. DOI: [10.5194/essd-2019-50](https://doi.org/10.5194/essd-2019-50). URL: <https://www.earth-syst-sci-data-discuss.net/essd-2019-50/>.
- [Gu+17] Y. Gu et al. “Using seismic data to estimate the spatial distribution of rock thermal conductivity at reservoir scale”. In: *Geothermics* 66 (2017), pp. 61–72.
- [HHR01] Oyvind Hammer, David Harper, and Paul Ryan. *PAST: Palaeontological Statistics Software Package for Education and Data Analysis*. Vol. 4. 2001, pp. 1–9.
- [HMG14] John A. Howell, Allard W. Martinius, and Timothy R. Good. “The application of outcrop analogues in geological modelling: a review, present status and future outlook”. In: *Geological Society, London, Special Publications* 387.1 (2014), pp. 1–25. DOI: [10.1144/sp387.12](https://doi.org/10.1144/sp387.12). URL: <http://sp.lyellcollection.org/content/387/1/1.abstract>.
- [HT19] Mutao Huang and Yong Tian. “An Integrated Graphic Modeling System for Three-Dimensional Hydrodynamic and Water Quality Simulation in Lakes”. In: *ISPRS International Journal of Geo-Information* 8.1 (2019), p. 18.
- [LE11] Pantelis Liolios and George Exadaktylos. “A relational rock mechanics database scheme with a hierarchical structure”. In: *Computers & Geosciences* 37.8 (2011), pp. 1192–1204. ISSN: 0098-3004. DOI: <https://doi.org/10.1016/j.cageo.2011.02.014>. URL: <http://www.sciencedirect.com/science/article/pii/S0098300411000896>.
- [Llo82] S. Lloyd. “Least squares quantization in PCM”. In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 129–137. ISSN: 0018-9448. DOI: [10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489). URL: <https://ieeexplore.ieee.org/document/1056489>.
- [Mal+19] T. Malvić et al. “Kriging with a Small Number of Data Points Supported by Jack-Knifing, a Case Study in the Sava Depression (Northern Croatia)”. In: *Geosciences* 9.1 (2019), p. 36. ISSN: 2076-3263. URL: <https://www.mdpi.com/2076-3263/9/1/36>.
- [Mat63] Georges Matheron. “Principles of geostatistics”. In: *Economic Geology* 58.8 (1963), 1246–1266.

## REFERENCES

---

- [MD17] B. A. Mert and A. Dag. “A Computer Program for Practical Semivariogram Modeling and Ordinary Kriging: A Case Study of Porosity Distribution in an Oil Field”. In: *Open Geosciences* 9.1 (2017), pp. 663–674. DOI: [10.1515/geo-2017-0050](https://doi.org/10.1515/geo-2017-0050). URL: <https://www.degruyter.com/view/j/geo.2017.9.issue-1/geo-2017-0050/geo-2017-0050.xml>.
- [Mia85] Andrew D. Miall. “Architectural-element analysis: A new method of facies analysis applied to fluvial deposits”. In: *Earth-Science Reviews* 22.4 (1985), pp. 261–308. ISSN: 0012-8252. DOI: [http://dx.doi.org/10.1016/0012-8252\(85\)90001-7](http://dx.doi.org/10.1016/0012-8252(85)90001-7). URL: <http://www.sciencedirect.com/science/article/pii/0012825285900017>.
- [Pop+99] Yuri A. Popov et al. “Characterization of rock thermal conductivity by high-resolution optical scanning”. In: *Geothermics* 28.2 (1999), pp. 253–276. ISSN: 0375-6505. DOI: [https://doi.org/10.1016/S0375-6505\(99\)00007-3](https://doi.org/10.1016/S0375-6505(99)00007-3). URL: <http://www.sciencedirect.com/science/article/pii/S0375650599000073>.
- [Rö6] W. Rühaak. “A Java application for quality weighted 3-d interpolation”. In: *Comput. Geosci.* 32.1 (2006), pp. 43–51. ISSN: 0098-3004. DOI: [10.1016/j.cageo.2005.04.005](https://doi.org/10.1016/j.cageo.2005.04.005). URL: <https://doi.org/10.1016/j.cageo.2005.04.005>.
- [Rü+15] Wolfram Rühaak et al. “Upscaling thermal conductivities of sedimentary formations for geothermal exploration”. In: *Geothermics* 58 (2015), pp. 49–61. ISSN: 0375-6505. DOI: <http://dx.doi.org/10.1016/j.geothermics.2015.08.004>. URL: <http://www.sciencedirect.com/science/article/pii/S0375650515001054>.
- [Rü5] W. Rühaak. “3-D interpolation of subsurface temperature data with measurement error using kriging”. In: *Environmental Earth Sciences* 73.4 (2015), pp. 1893–1900. ISSN: 1866-6299. DOI: [10.1007/s12665-014-3554-5](https://doi.org/10.1007/s12665-014-3554-5). URL: <https://doi.org/10.1007/s12665-014-3554-5>.
- [RB15] P. Ringrose and M. Bentley. *Reservoir Model Design*. 1st ed. Springer Netherlands, 2015, p. 249. ISBN: 978-94-007-5496-6. DOI: [10.1007/978-94-007-5497-3](https://doi.org/10.1007/978-94-007-5497-3). URL: <http://www.springer.com/de/book/9789400754966>.
- [RBW09] N. Remy, A. Boucher, and J. Wu. *Applied Geostatistics with SGEMS: A User’s Guide*. Cambridge: Cambridge University Press, 2009. ISBN: 9781107403246. DOI: [DOI:10.1017/CBO9781139150019](https://doi.org/10.1017/CBO9781139150019).

## REFERENCES

---

- URL: <https://www.cambridge.org/core/books/applied-geostatistics-with-sgems/35D45D41BF8FD9BB43E8DCD462E30975>.
- [Rn3] Chart or Table. 2013; updated. URL: <http://www.stratigraphy.org/ICSchart/ChronostratChart2017-02.pdf>.
- [Sah+18] Z. Sahaf et al. “A Visual Analytics Framework for Exploring Uncertainties in Reservoir Models”. In: *VISIGRAPP*. 2018.
- [She68] D. Shepard. “A Two-Dimensional Interpolation Function for Irregularly-Spaced Data.” In: *Proceedings of the 1968 ACM National Conference*. 1968, pp. 517–524. URL: <http://dx.doi.org/10.1145/800186.810616>.
- [TD12] Djebbar Tiab and Erle C. Donaldson. “Chapter 10 - Reservoir Characterization”. In: *Petrophysics (Third Edition)*. Ed. by Djebbar Tiab and Erle C. Donaldson. Boston: Gulf Professional Publishing, 2012, pp. 667–714. ISBN: 978-0-12-383848-3. DOI: <https://doi.org/10.1016/B978-0-12-383848-3.00010-4>. URL: <http://www.sciencedirect.com/science/article/pii/B9780123838483000104>.
- [Tia+18] Yong Tian et al. “A comprehensive graphical modeling platform designed for integrated hydrological simulation”. In: *Environmental Modelling & Software* 108 (2018), pp. 154–173. ISSN: 1364-8152. DOI: <https://doi.org/10.1016/j.envsoft.2018.08.011>. URL: <http://www.sciencedirect.com/science/article/pii/S1364815217312872>.
- [TZZ16] Yong Tian, Yi Zheng, and Chunmiao Zheng. “Development of a visualization tool for integrated surface water–groundwater modeling”. In: *Computers & Geosciences* 86 (2016), pp. 1–14. ISSN: 0098-3004. DOI: <https://doi.org/10.1016/j.cageo.2015.09.019>. URL: <http://www.sciencedirect.com/science/article/pii/S0098300415300595>.
- [Wac03] H. Wackernagel. *Multivariate Geostatistics*. Third Edition. Springer-Verlag Berlin Heidelberg GmbH, 2003, p. 388. DOI: <10.1007/978-3-662-05294-5>. URL: <https://doi.org/10.1007/978-3-662-05294-5>.
- [WM07] R. Webster and A. O. Margarit. *Geostatistics for Environmental Scientists*. 2nd ed. Wiley & Sons, Inc, 2007, p. 330. URL: <https://www.wiley.com/en-us/Geostatistics+for+Environmental+Scientists%2C+2nd+Edition-p-9780470028582>.

# Complete License

## GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc. 51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

### Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software—to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced

## REFERENCES

---

by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

### GNU GENERAL PUBLIC LICENSE TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be

## REFERENCES

---

licensed as a whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it. Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

- a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
- b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
- c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for non-commercial distribution and only if you received the program in object code

## REFERENCES

---

or executable form with such an offer, in accord with Subsection b above.) The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example,

## REFERENCES

---

if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and

## REFERENCES

---

reuse of software generally.

### NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

### END OF TERMS AND CONDITIONS

#### How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

one line to give the program's name and a brief idea of what it does.; Copyright (C) [year]; [name of author];

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

## REFERENCES

---

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Also add information on how to contact you by electronic and paper mail. If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

Gnomovision version 69, Copyright (C) year name of author Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type ‘show w’. This is free software, and you are welcome to redistribute it under certain conditions; type ‘show c’ for details.

The hypothetical commands ‘show w’ and ‘show c’ should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than ‘show w’ and ‘show c’; they could even be mouse-clicks or menu items—whatever suits your program.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a ”copyright disclaimer” for the program, if necessary. Here is a sample; alter the names:

Yoyodyne, Inc., hereby disclaims all copyright interest in the program ‘Gnomovision’ (which makes passes at compilers) written by James Hacker.

|signature of Ty Coon|, 1 April 1989 Ty Coon, President of Vice

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Library General Public License instead of this License.