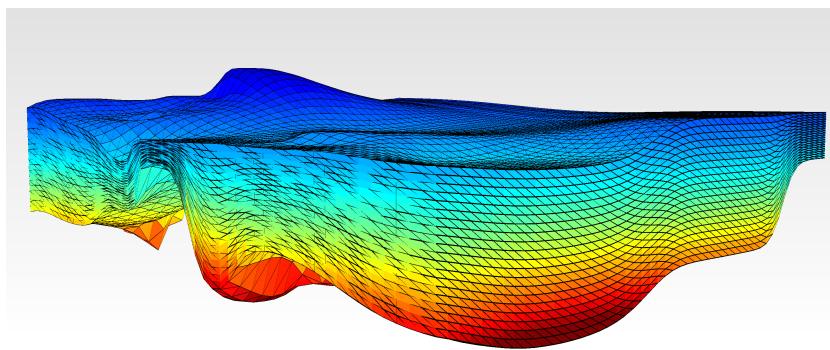


User Manual and Tutorial for GeoReVi

Geological Reservoir Virtualization

Date: November 13, 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

1	Introduction	8
1.1	Built with	8
1.2	Authors and Contributors	9
1.3	Hard- and software requirements	9
1.4	Installation	9
2	Quick start	10
2.1	Importing a local data set	10
2.2	Data transformation	11
2.3	Univariate statistics	12
2.3.1	Empirical distribution	13
2.3.2	Creating a line and bubble chart	13
2.3.3	Box-Whisker plot	14
2.3.4	Variography	14
2.4	Handling meshes	16
2.4.1	Generating custom meshes	20
2.4.2	Copying, removing and storing meshes	23
2.4.3	3D meshes	24
2.5	Performing Predictions	26
2.5.1	2-D predictions	26
2.5.2	Checking the interpolation for consistency	27
2.5.3	3-D predictions	29
2.5.4	Rescaling	31
2.5.5	Error integration	33
2.5.6	Complex geometries	34
2.6	Multivariate statistics	36
2.6.1	Correlation analysis	37
2.6.2	Regression analysis	38
2.6.3	Grouping	40
2.6.4	Principal component analysis	41
2.6.5	Self-organized Maps	43
2.6.6	Cluster analysis	45
3	In-depth guide and theoretical aspects	47
3.1	Introduction	47
3.1.1	Starting the application	48
3.1.2	Registration and Sign in	50
3.2	Obtaining a Bing Maps Key	51
3.3	Sample data	52

CONTENTS

3.4	Data management	53
3.4.1	Supported data types	53
3.4.2	Provided data set	57
3.4.3	Project Management	57
3.4.4	Objects of investigation	57
3.4.5	Lithologies	58
3.4.6	Rock samples	59
3.4.7	Measurements	59
3.4.8	Import to the database	60
3.5	Data analysis and visualization	62
3.5.1	Data transformation	63
3.5.2	Univariate data analysis	64
3.5.3	Variography	67
3.5.4	Indicator variography	68
3.5.5	Mesh generation	69
3.5.6	Mesh statistics	72
3.5.7	Spatial interpolation and simulation	73
3.5.8	Conditional Simulation	77
3.5.9	Measurement Error Integration	77
3.5.10	Cross-validation	78
3.6	Multivariate statistics	79
3.6.1	Dimensionality reduction	80
3.6.2	Classification	83
3.6.3	Correlation analysis	84
3.6.4	Regression analysis	85
3.7	Data visualization	87
3.7.1	3-D chart	87
4	Developing Plug-ins	88
5	Personal data security	89
6	Related works	90

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™](#) - Providing nice UI controls
- [Math.NET](#) - Math magic

1.2 Authors and Contributors

Adrian Linsel, Technische Universität Darmstadt, 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	Intel Core i5-7200U
	v3 @ 3.6 GHz	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 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**.

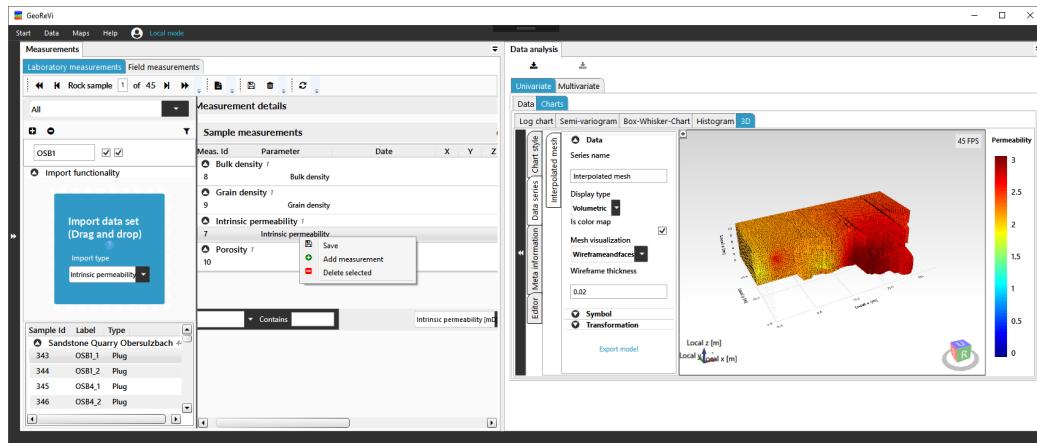


Figure 1: Measurements screen with a separate **Data analysis** window.

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

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 Data transformation

Meshes loaded as univariate data sets sometimes must be transformed to meet the requirements needed for particular statistical methods. GeoReVi provides a range of transformation methods like described in section 3.5.1. To perform a transformation, we click on the **Transformation** drop-down menu and select the type of transformation, that we want to perform. Exemplary,

we will perform a *Rescaling* transformation what will project our samples into the interval $[0,1]$. The results are displayed in the table below.

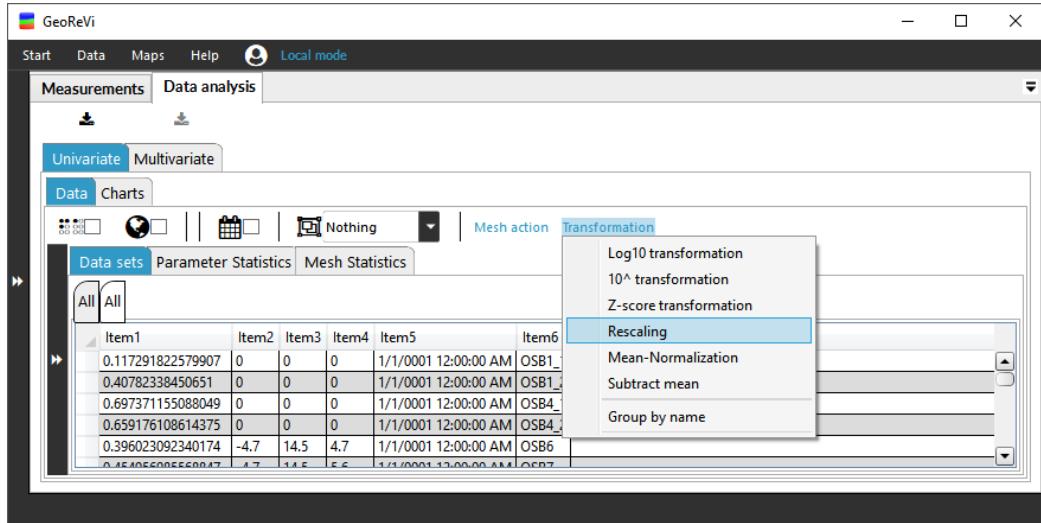


Figure 2: Performing a data transformation.

2.3 Univariate statistics

To get a first overview over the empirical distribution of the measurements, we import our *Thermal conductivity* data set again, go to **Univariate** and then 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. 3.

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*, *me-*

Parameter Statistics		
Compute		
<input type="checkbox"/> All		
<input checked="" type="radio"/> Descriptive statistics	<input type="radio"/> Heterogeneity measures	
Counts:	120	Median:
Range:	2.075	Standard deviation:
Maximum:	3.276	Variance:
Minimum:	1.201	Skewness:
Arithmetic mean:	2.16742083333333	Kurtosis:
Geometric mean:	2.140161356649	
Harmonic mean:	2.11045919830169	

Figure 3: Basic statistics result.

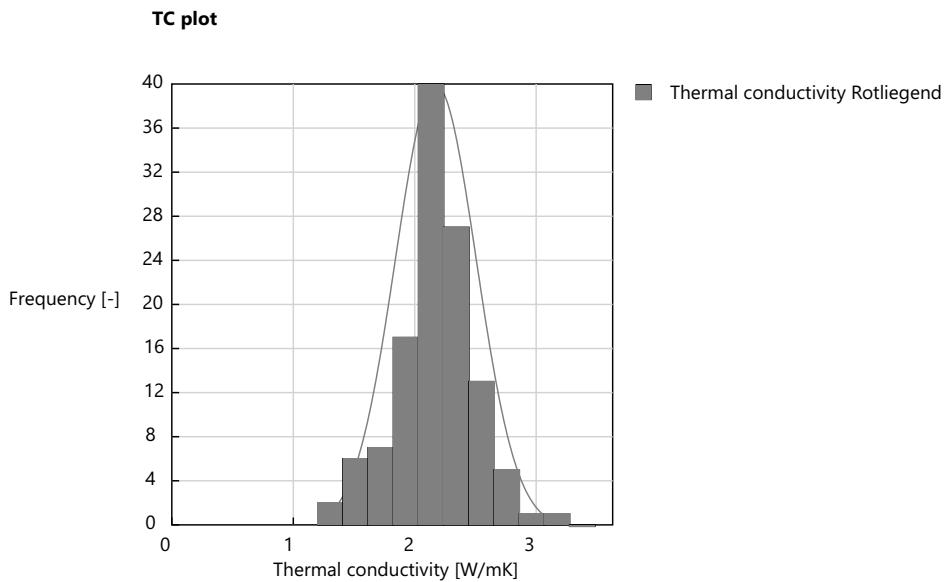


Figure 4: Empirical and theoretical distribution of the thermal conductivity measurement.

dian, standard deviation, variance, skewness, kurtosis, coefficient of variation and Dykstra-Parson coefficient.

2.3.1 Empirical distribution

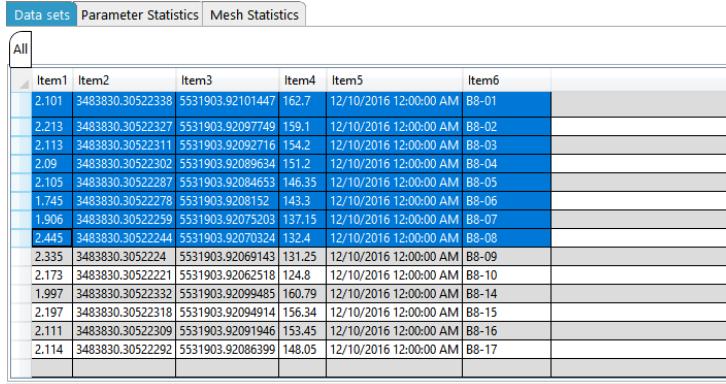
Empirical distributions can be visualized in histograms. Therefore, return to the **Data sets** and **right-click** on the mesh *Thermal conductivity Rotliegend* and press **Add data set to bar chart**. Navigate to **Charts → Histogram** and refresh the chart with the **Refresh** button in the top-left corner. Under **Data series → Theoretical distribution** make a tick in the field **Show** and refresh . The resulting distribution chart should look like fig. 4.



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. 4)

2.3.2 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



	Item1	Item2	Item3	Item4	Item5	Item6
2.101	3483830.30522338	5531903.92101447	162.7	12/10/2016 12:00:00 AM	B8-01	
2.213	3483830.30522327	5531903.92097749	159.1	12/10/2016 12:00:00 AM	B8-02	
2.113	3483830.30522311	5531903.92092716	154.2	12/10/2016 12:00:00 AM	B8-03	
2.09	3483830.30522302	5531903.92089634	151.2	12/10/2016 12:00:00 AM	B8-04	
2.105	3483830.30522287	5531903.92084653	146.35	12/10/2016 12:00:00 AM	B8-05	
1.745	3483830.30522278	5531903.9208152	143.3	12/10/2016 12:00:00 AM	B8-06	
1.906	3483830.30522259	5531903.92075203	137.15	12/10/2016 12:00:00 AM	B8-07	
2.445	3483830.30522244	5531903.92070324	132.4	12/10/2016 12:00:00 AM	B8-08	
2.335	3483830.3052224	5531903.92069143	131.25	12/10/2016 12:00:00 AM	B8-09	
2.173	3483830.30522221	5531903.92062518	124.8	12/10/2016 12:00:00 AM	B8-10	
1.997	3483830.30522332	5531903.92099485	160.79	12/10/2016 12:00:00 AM	B8-14	
2.197	3483830.30522318	5531903.92094914	156.34	12/10/2016 12:00:00 AM	B8-15	
2.111	3483830.30522309	5531903.92091946	153.45	12/10/2016 12:00:00 AM	B8-16	
2.114	3483830.30522292	5531903.92086399	148.05	12/10/2016 12:00:00 AM	B8-17	

Figure 5: Selecting rows in the **Data sets** view.

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 5. When pressing the **Delete** key, the selected rows will 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 fig. 8

2.3.3 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. 7.

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.

2.3.4 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

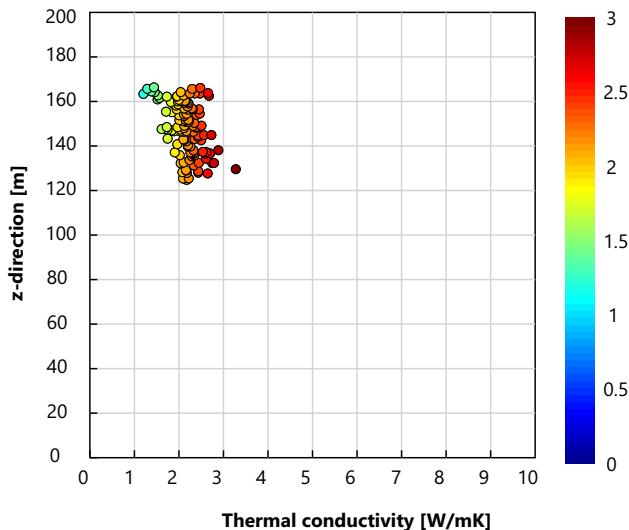


Figure 6: Bubble chart of the z-direction thermal conductivity measurements.

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.

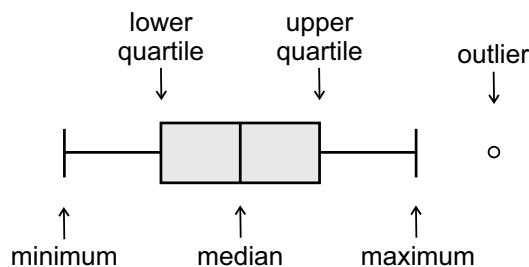


Figure 7: Schematic of a box-whisker plot.

2.4 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 hialal 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

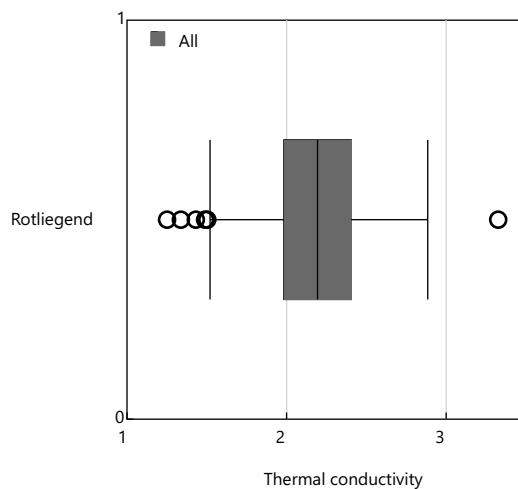


Figure 8: Box-Whisker chart of the thermal conductivity measurements with removed outliers.

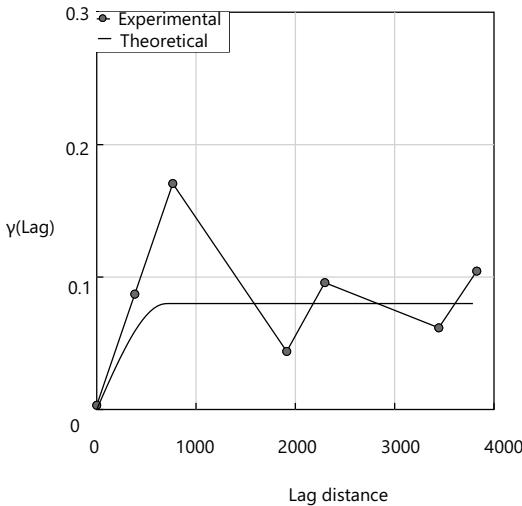


Figure 9: Semivariogram of the thermal conductivity measurement.

3D port are referenced by a metric system. When we refresh, the port should look like fig. 10.



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 θ . 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. 11.

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. 12. 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

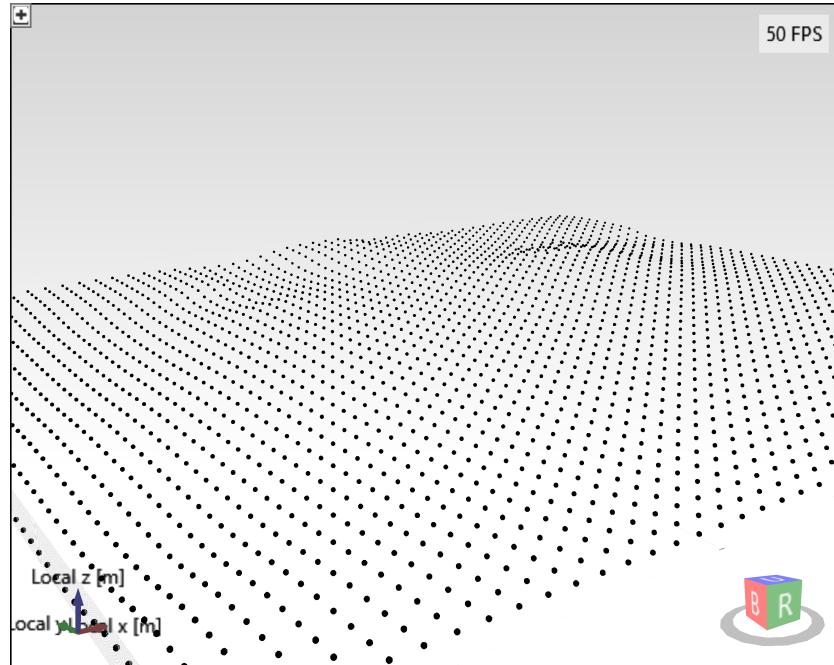


Figure 10: Point cloud visualization of a surface in GeoReVi

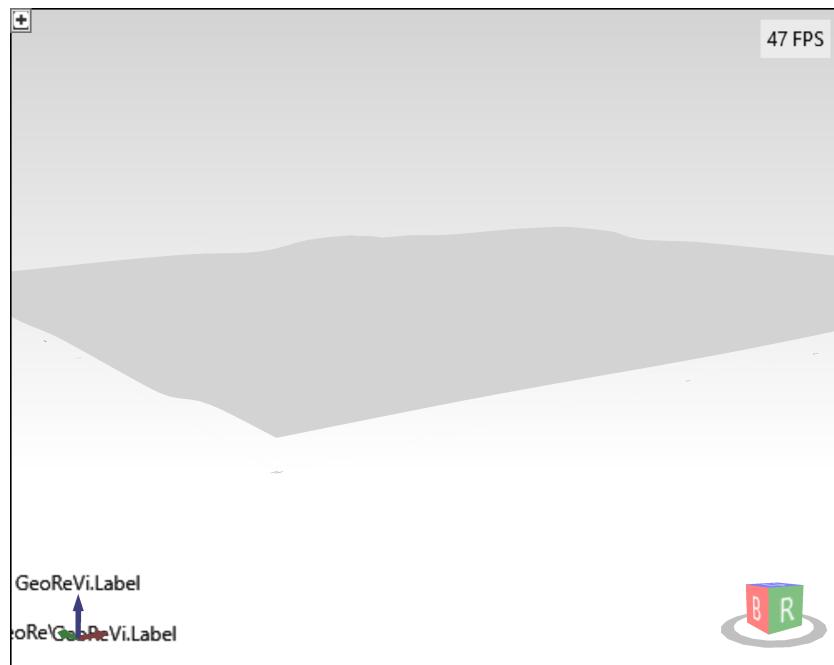


Figure 11: Faces visualization of a surface in GeoReVi

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. 13.

Surface interpolations react very sensitive to adaptions in the neighborhood and the p-value of an IDW interpolation. Fig. 14 shows three realizations of surfaces interpolated through a set of points. A small p-value leads to smooth interpolations, however, the distance from the interpolation points can be quite high. A high p-value leads to an accurate representation of the points, however, the surface tends to build *plateaus*, what is due to the high weight of near points. It is recommended to produce multiple realizations of a surface and to use first realizations as **Target data sets** for further realizations since those already provide a closer representation of the **Source data set** and hence, a neighborhood can be represented more accurately.

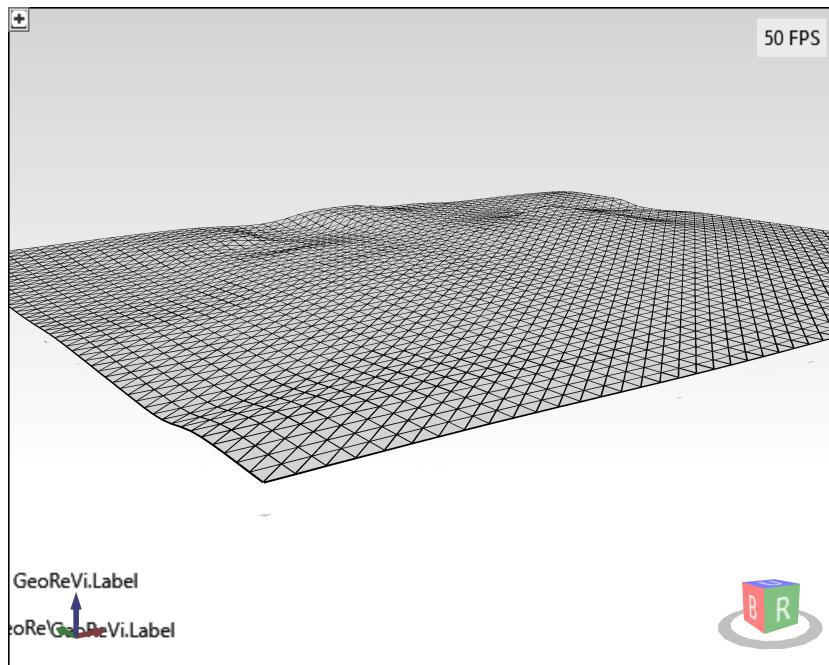


Figure 12: Faces visualization of a surface in GeoReVi

2.4.1 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 θ as **Bins z direction** to create a two-dimensional data set. We will leave the default values in the other boxes.



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 θ 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

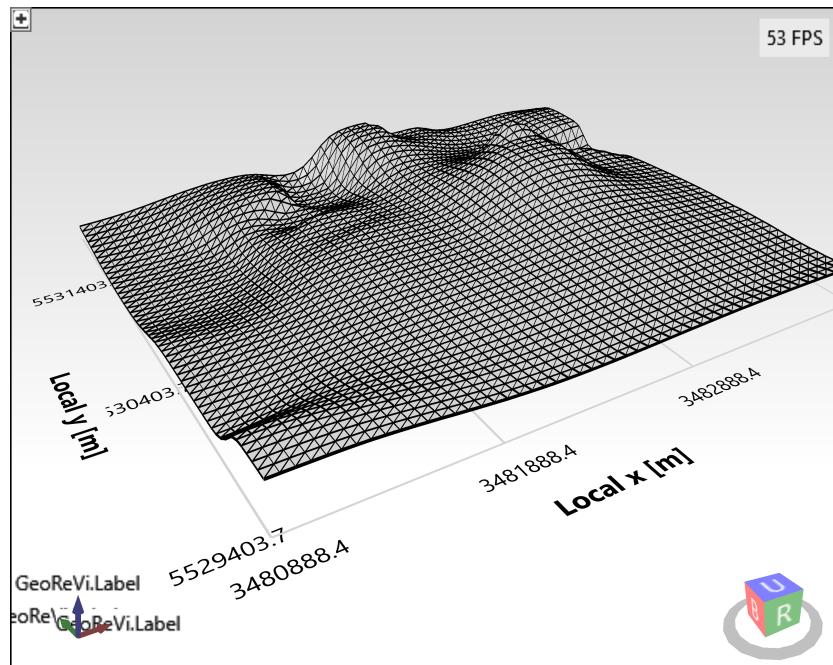


Figure 13: Exaggerated faces visualization of an IDW interpolated surface in GeoReVi

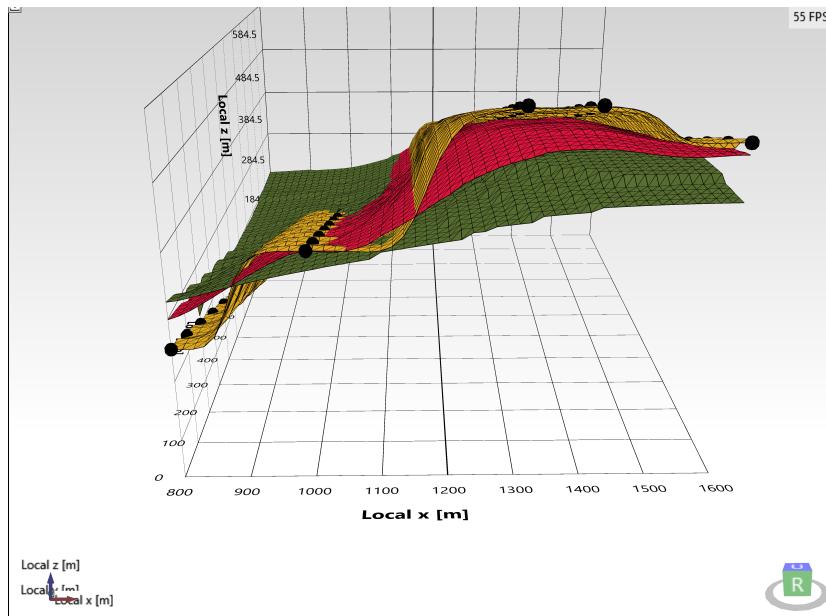


Figure 14: Exemplary IDW interpolation of an artificial data set. The green surface was produced with an IDW p-value of 1, the red one with 4 and the orange one with 8.

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. 15.

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 **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. 16.

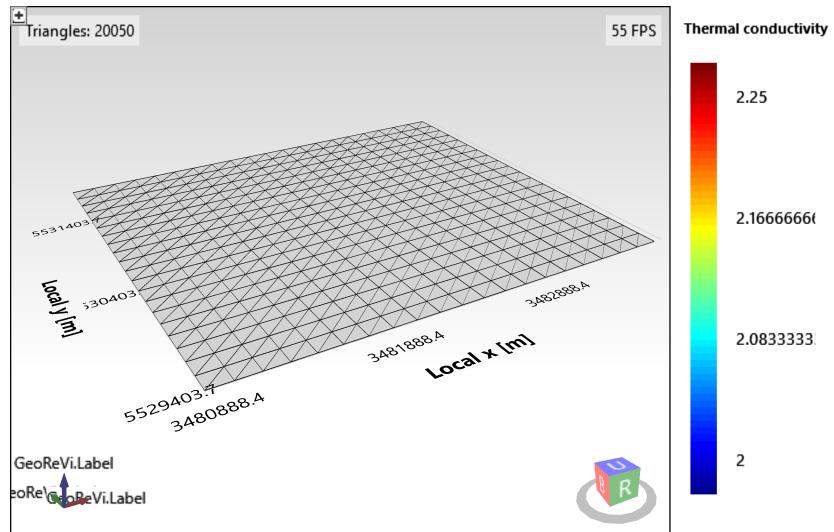


Figure 15: Custom surface.

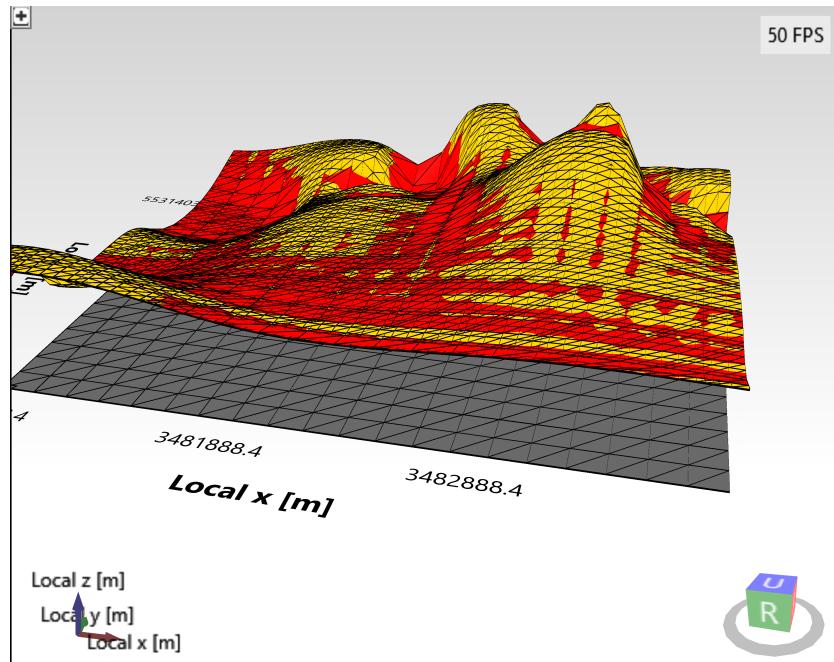


Figure 16: Comparison of the *Model base* (dark gray), *Base Rotliegend Interpolated* (yellow) and *Interpolated surface* (red) meshes.



```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <GeoReViFile xmlns="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
3   <@version>1.0</@version>
4   <@encoding>UTF-8</@encoding>
5   <@xsd:schemaLocation>http://www.w3.org/2001/XMLSchema-instance http://www.w3.org/2001/XMLSchema</@xsd:schemaLocation>
6 
7     <GeometricEntity>
8       <@id>geoRectangular</@id>
9       <@name>geoRectangular</@name>
10      <@type>rectangle</@type>
11      <@x>0.0</@x>
12      <@y>0.0</@y>
13      <@z>0.0</@z>
14      <@order>4</@order>
15      <@discretized>true</@discretized>
16      <@x1>0.0</@x1>
17      <@y1>0.0</@y1>
18      <@x2>1.0</@x2>
19      <@y2>1.0</@y2>
20      <@x3>1.0</@x3>
21      <@y3>0.0</@y3>
22      <@x4>0.0</@x4>
23      <@y4>1.0</@y4>
24      <@orderList>
25        <@id>geoRectangular_order1</@id>
26        <@id>geoRectangular_order2</@id>
27        <@id>geoRectangular_order3</@id>
28        <@id>geoRectangular_order4</@id>
29      </@orderList>
30      <@vertices>
31        <@id>geoRectangular_v1</@id>
32        <@id>geoRectangular_v2</@id>
33        <@id>geoRectangular_v3</@id>
34        <@id>geoRectangular_v4</@id>
35      </@vertices>
36      <@edges>
37        <@id>geoRectangular_e1</@id>
38        <@id>geoRectangular_e2</@id>
39        <@id>geoRectangular_e3</@id>
40        <@id>geoRectangular_e4</@id>
41      </@edges>
42      <@faces>
43        <@id>geoRectangular_f1</@id>
44      </@faces>
45    </GeometricEntity>
46    <GeometricEntity>
47      <@id>geoRectangular</@id>
48      <@name>geoRectangular</@name>
49      <@type>rectangle</@type>
50      <@x>0.0</@x>
51      <@y>0.0</@y>
52      <@z>0.0</@z>

```

Figure 17: File structure of the custom file format .GMSH to store GeoReVi meshes.

2.4.2 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. 17).

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 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.

2.4.3 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. 18. 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.



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.

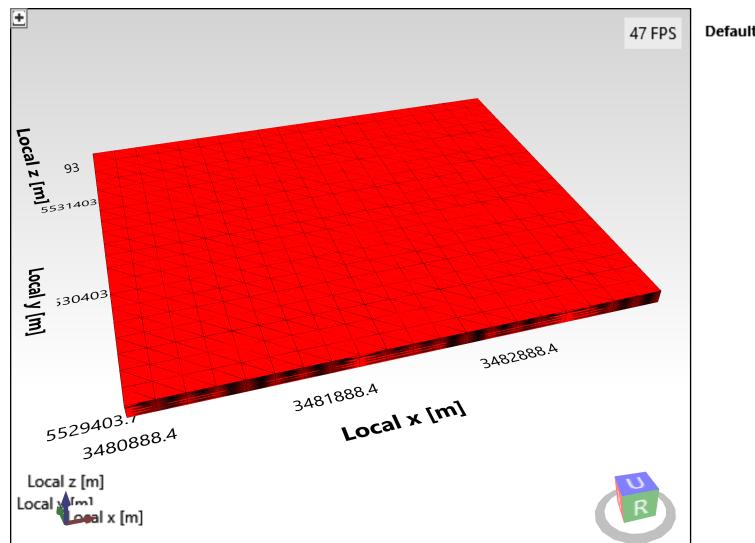


Figure 18: Three-dimensional regular, hexahedral mesh.

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. 19.

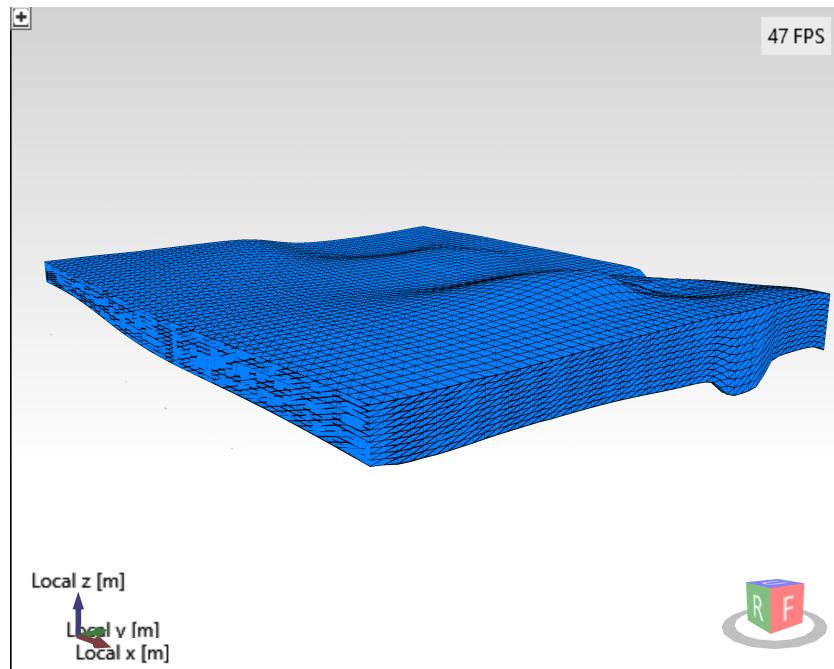


Figure 19: Stratigraphic grid of the Rotliegend group on the Sprendlinger Horst.

2.5 Performing Predictions

2.5.1 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 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

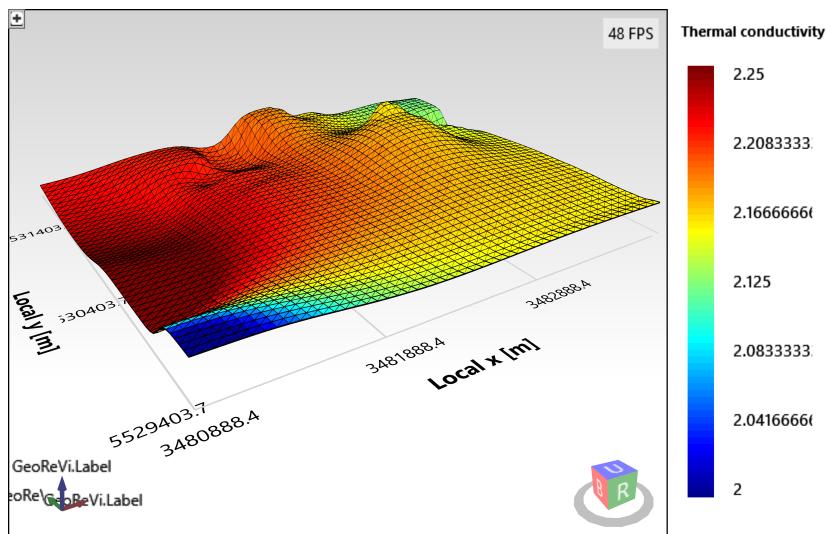


Figure 20: Exaggerated faces visualization of a surface in GeoReVi

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. 12 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. 20 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).

2.5.2 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. 21).

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.

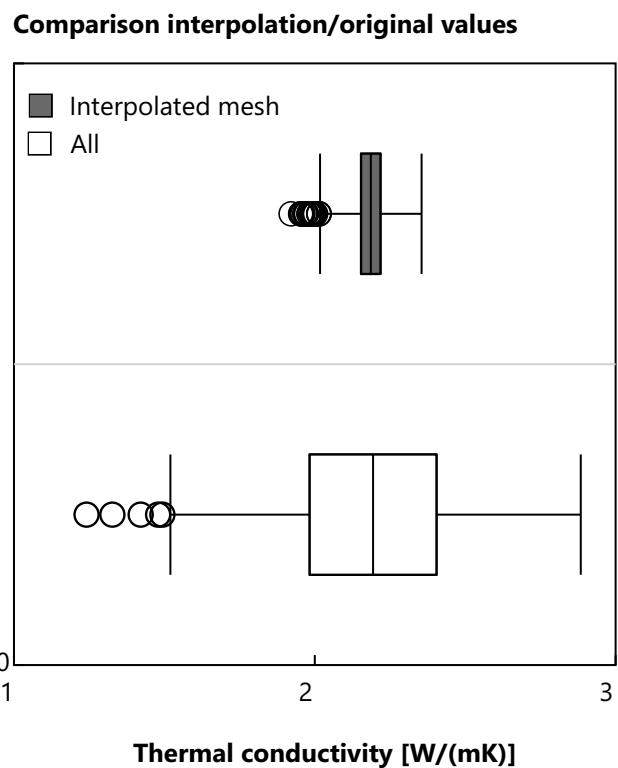


Figure 21: Comparison between the original distribution and the interpolated values.

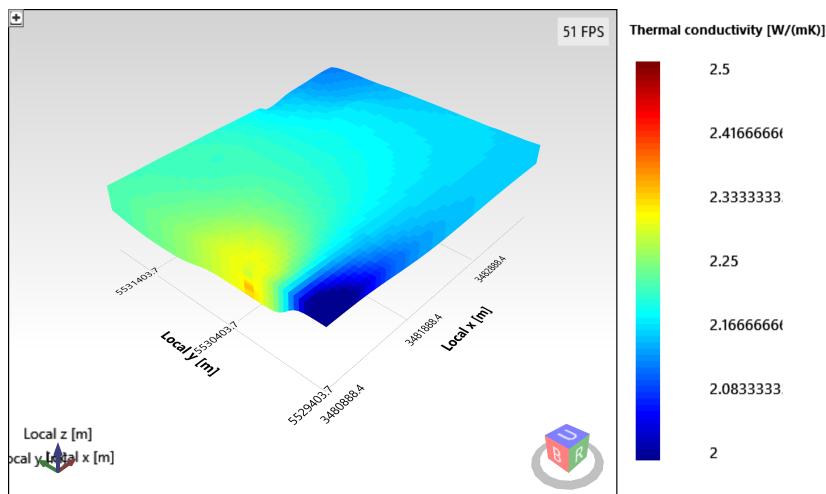


Figure 22: Three-dimensional property model created with IDW of the thermal conductivity in the study area.

2.5.3 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*. In the **Results** section, we can observe an RMSE of 0.25 which is reasonably low. When adding the result to the **3D chart** and exaggerating it by a factor of 5 it should look like fig. 26.

Defining a neighborhood Definition of a neighborhood can significantly enhance the spatial prediction. In a layered setting, an anisotropic neighborhood such as with an extended search radius in x and y direction will result in a more realistic representation of the observed geology. Therefore, we will define a neighborhood of 1500 in both x and y direction and a neighborhood of 30 in z direction. After computation we can observe an RMSE of 0.24 which is lower than the IDW result from the last section and the visualization reveals an anisotropic pattern now.

Stochastic interpolation

In probabilistic models, a field parameter distribution is considered to be

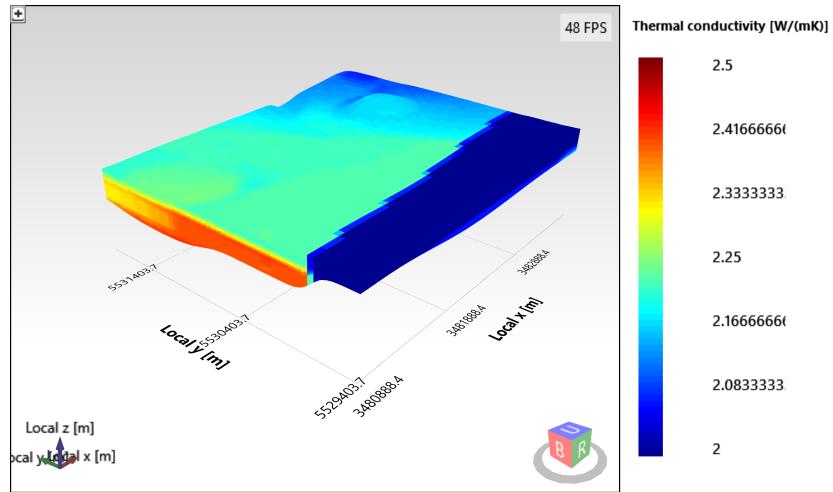


Figure 23: Three-dimensional property model created with SK of the thermal conductivity in the study area.

a realization of a random function. Therefore, different kinds of stochastic prediction algorithms from kriging class exist. Kriging aims to predict the value at an undetermined place in space by using a linear combination of kriging weights and corresponding known values. The weights are derived with the help of the theoretical semivariogram or the covariance function.

As we already defined a semivariogram for our *Thermal conductivity* data set in section 2.3.4 we will use the model, the range, the sill and the nugget from that chapter for our interpolation. Therefore, in the *Interpolation* menu we will define *Simple Kriging* as **Interpolation method**. Under **Variogram model** insert *0.08* for the **Sill**, *0* for the **Nugget** and *700* for the **Range**. As a neighborhood, we will define *2000* for both x and y direction and *30* for the z direction. The **Maximum count of neighbors** should be reduced to a reasonable number since each interpolated point requires to solve a number system of linear equations that corresponds at least to the number of included neighbors. When we now click on **Compute interpolation**, the simple kriging interpolation will be performed and the result will be added to the data sets. In the **Results** section, we can observe an RMSE of *0.26* which is slightly higher than the IDW result from the last section.

For performing an ordinary kriging (OK) interpolation, we simply select this option as **Interpolation method** and click **Compute interpolation** again.

Conditional 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

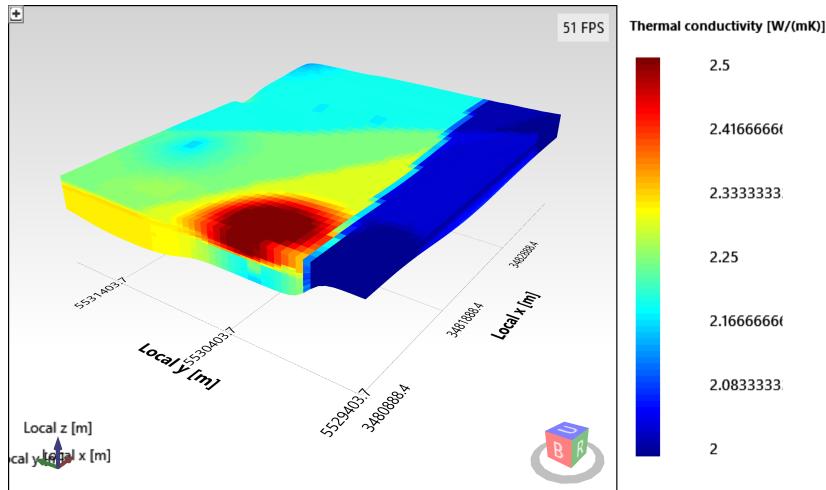


Figure 24: Three-dimensional property model created with SK of the thermal conductivity in the study area.

will again interpolate the **Value** with the **Euclidean distance** but this time with the **Sequential Gaussian Simulation** (SGS) method. This method simulates a random sample from a gaussian distribution 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. ???. Another method to simulate a random function is called *simulated annealing*. This method simulates the original histogram and sorts the randomly assigned points in a user-defined number of iterations to minimize the deviation from the original semivariogram. To produce a simulated annealing simulation we select that method as *Interpolation method* and select the same values as used before in the SGS. Additionally, we define a **Number of iterations**; of 10,000. Due to the big mesh size of our example, the result does not provide well patterns. To get a reasonable result, the number of iterations should exceed the number of node at least by 100 times.

2.5.4 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

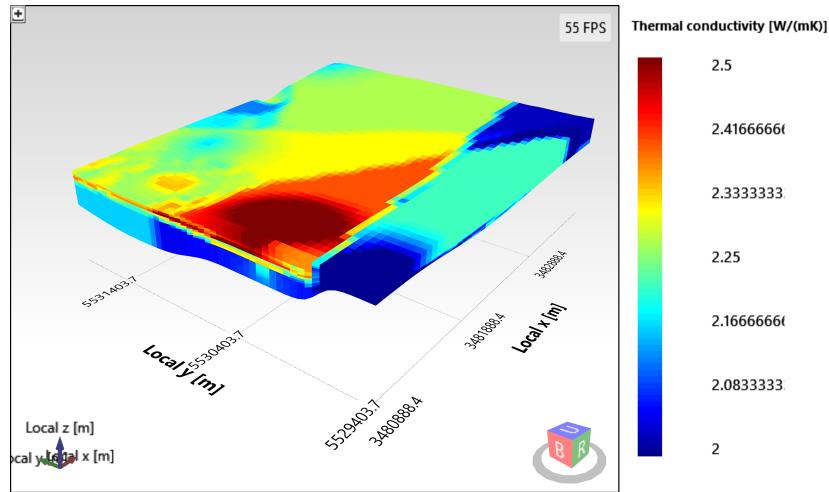


Figure 25: Three-dimensional property model created with OK of the thermal conductivity in the study area.

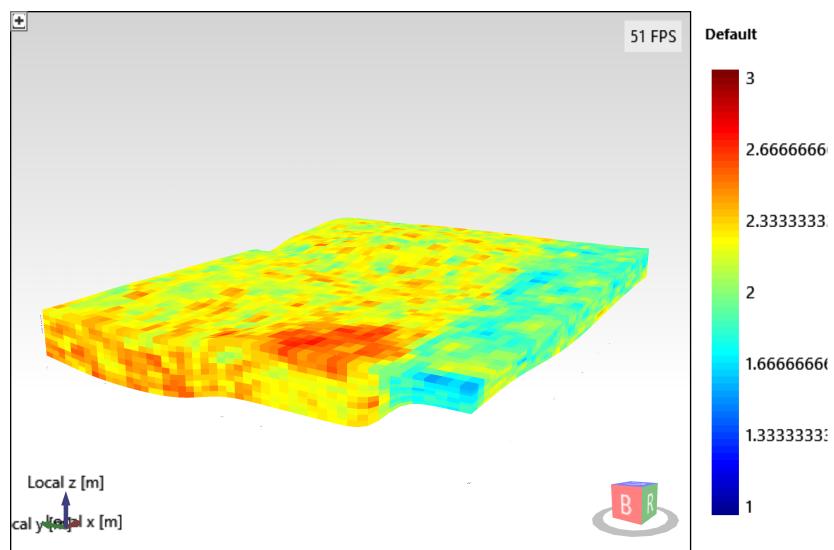


Figure 26: Three-dimensional property model generated by SGS of the thermal conductivity in the study area.

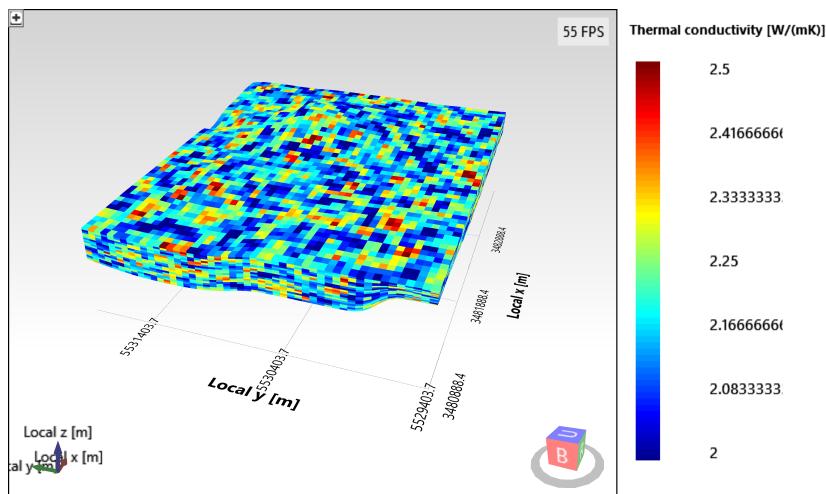


Figure 27: Simulated annealing realization using 10,000 iterations.

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.

2.5.5 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 40.

2.5.6 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* namely *FoldEastBase.csv*, *FoldEastMiddle.csv*, *FoldEastTop.csv*, *FoldWestBase.csv*, *FoldWestMiddle.csv* and *FoldWestTop.csv*. Those files represent surfaces belonging to a faulted fold comprising two lithological horizons. We will start to model the eastern side of the geological scene. Therefore, we will import *FoldEastBase.csv*, *FoldEastMiddle.csv* and *FoldEastTop.csv* as data sets.

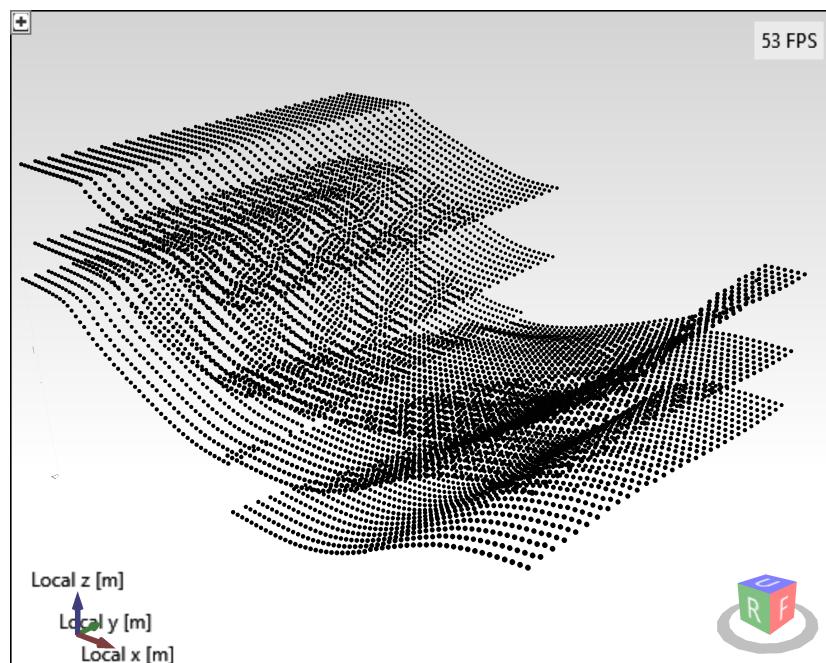


Figure 28: Complexly folded geological domain mesh as a scattered mesh.

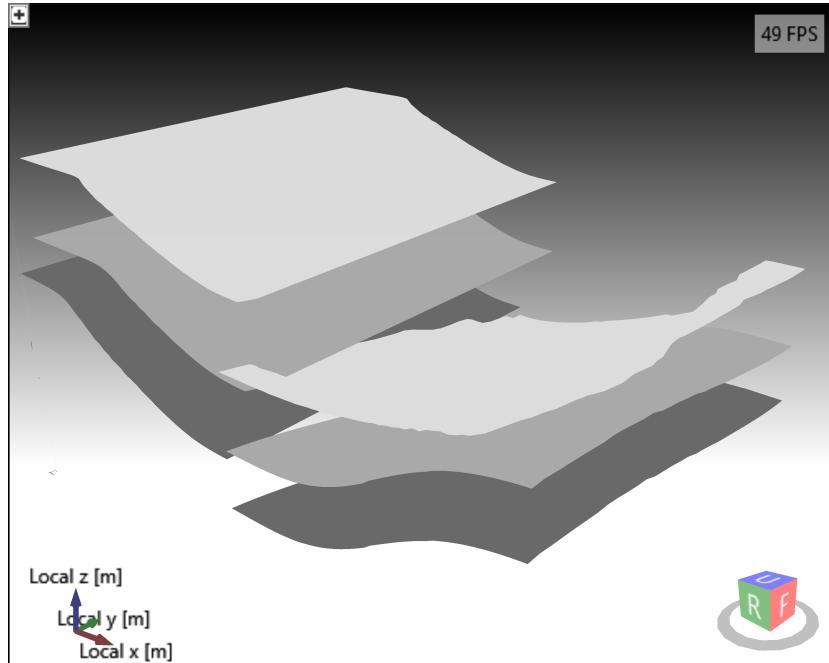


Figure 29: Complexly folded geological domain mesh.

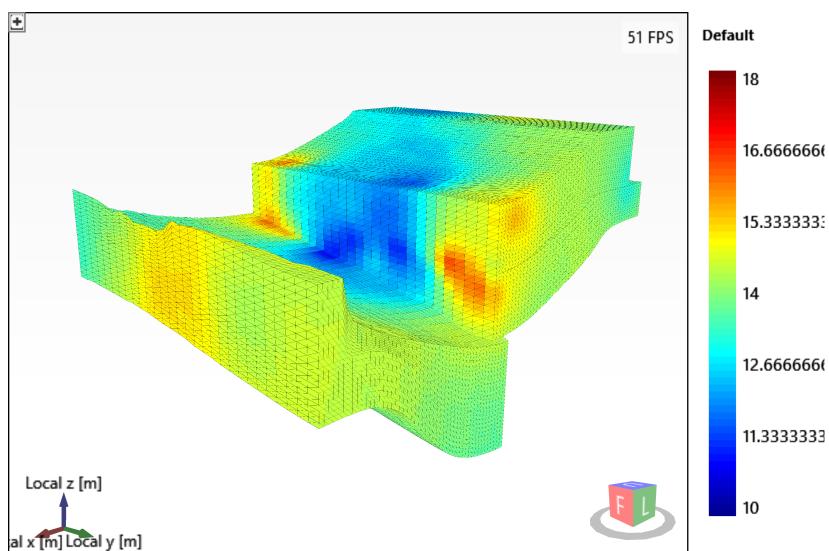


Figure 30: Parametrized using IDW and a random data set folded geological domain mesh.

2.6 Multivariate statistics

GeoReVi implements some algorithms to analyze multivariate data sets. Therefore, in the **Measurements → Data analysis** tabs we navigate to **Multivariate**. In the folder **Tutorials/FieldParameters** we provided a data sets called *Measurements_Obersulzbach.csv*. We will import this data set to the multivariate analysis tool by clicking on **Edit → Import table**. We select *.csv* in the file filter, select the table and click **Open**. Now, the table will be displayed as a data set in tabular format. The table contains measurements of a set of samples taken from an outcrop called *Sandstone Quarry Obersulzbach*. The outcrop exposes rocks from the Disibodenberg formation that belongs to the Rotliegend group in central Europe. The samples comprise homogeneous sandstones with a grain size in the middle to coarse sand fraction. In our laboratory in Darmstadt we measured the following parameters:

1. grain density
2. bulk density
3. intrinsic permeability
4. porosity
5. thermal conductivity
6. thermal diffusivity

Those parameters can be considered key parameters regarding the rock matrix controlling the quality of a hydrothermal system.

2.6.1 Correlation analysis

For getting a first overview about the correlations between the physical parameters, we perform a correlation analysis using the covariance, Pearson correlation coefficient and Spearman correlation coefficient. First two provide a measure for the linear relationship between two independent variables. Latter one also detects formal relationships. For conducting a correlation analysis, we **Right-click → Add to correlation**. When navigating to the **Correlation** tab in the **Statistics** menu, we will find our data set listed. Now, we define, how we want to treat *NULL* and *Ø* values. A common method is to exchange these values with the **Arithmetic mean** of the variables what we will define accordingly. After pressing **Compute** all previously mentioned parameters for each numeric data column in our data set will be calculated. The results can be seen in tabular format in the respective tab item. Exemplary results from the Pearson correlation coefficient should look like fig. 31. A clear trend should be observed between porosity and permeability, porosity and permeability, as well as grain and bulk density.

Results also can be visualized in a matrix chart that is displayed right next to the table.

a	A	B	C	D	E	F	G	H	I	J	K
A	1	0.93708	0.9621	-0.41707	-0.73419	-0.091502	-0.4677	0.68714	0.51969	0.058421	0.0026125
B	0.93708	1	0.99169	-0.59088	-0.6815	-0.13718	-0.47517	0.68336	0.06854	0.01076	0.0059322
C	0.9621	0.99169	1	-0.50647	-0.70916	-0.13245	-0.47847	0.68996	0.59101	0.0250503	0.011763
D	-0.41707	-0.59088	-0.50647	1	0.21373	0.15875	0.34169	-0.47116	-0.55659	0.12246	0.094623
E	-0.73419	-0.6815	-0.70916	0.21373	1	-0.068304	0.30987	-0.49822	-0.24151	-0.04041	-0.0081968
F	-0.091502	-0.13718	-0.13245	0.15875	-0.068304	1	0.77022	-0.14918	-0.31503	-0.50621	-0.68503
G	-0.4677	-0.47517	-0.47847	0.34169	0.30987	0.77022	1	-0.68445	-0.48662	-0.34599	-0.45647
H	0.68714	0.68336	0.68996	-0.47116	-0.49822	-0.14918	-0.68445	1	0.64136	-0.24143	-0.14283
I	0.51969	0.59101	0.59101	-0.55659	-0.24151	-0.31503	-0.49822	0.64136	1	-0.27545	-0.64397
J	0.058421	0.01076	0.0250503	0.12246	-0.04041	-0.50621	-0.34599	-0.24143	-0.27545	1	0.67344
K	0.0026125	0.0059322	0.011763	0.094623	-0.0081968	-0.68503	-0.45647	-0.14283	-0.064397	0.67344	1

b	Local x (m)	Longitude (WGS84)	Latitude (WGS84)	Local y (m)	Local z (m)	Grain density (g cm ⁻³)	Bulk density (g cm ⁻³)	Porosity (%)	Intrinsic permeability (m ⁻²)	Thermal conductivity (W m ⁻¹ K ⁻¹)	Thermal diffusivity (m ² s ⁻¹)
Local x (m)	1	0.937	0.962	-0.417	0.734	-0.091	-0.467	0.687	0.519	0.058	0.002
Longitude (WGS84)	0.937	1	0.992	-0.591	-0.681	-0.137	-0.475	0.683	0.61	0.011	0.006
Latitude (WGS84)	0.962	0.992	1	-0.506	-0.709	-0.132	-0.478	0.69	0.592	0.025	0.012
Local y (m)	-0.417	-0.591	-0.506	1	0.214	0.159	0.342	-0.472	-0.557	0.122	0.095
Local z (m)	-0.734	-0.681	-0.709	0.214	1	-0.068	0.31	-0.498	-0.242	-0.04	-0.008
Grain density (g cm ⁻³)	-0.091	-0.137	-0.132	0.159	-0.068	1	0.77	-0.149	-0.315	-0.506	-0.685
Bulk density (g cm ⁻³)	-0.467	-0.475	-0.478	0.342	0.31	0.77	1	-0.684	-0.487	-0.346	-0.456
Porosity (%)	0.687	0.683	0.69	-0.472	-0.498	-0.149	-0.684	1	0.641	-0.241	-0.143
Intrinsic permeability (m ⁻²)	0.52	0.61	0.592	-0.557	-0.242	-0.315	-0.447	0.541	1	-0.275	-0.064
Thermal conductivity (W m ⁻¹ K ⁻¹)	0.058	0.011	0.075	-0.058	-0.049	-0.035	-0.246	0.275	1	-0.57	0.675
Thermal diffusivity (m ² s ⁻¹)	0.002	0.006	0.012	0.095	-0.008	-0.685	-0.456	-0.143	-0.064	0.673	1

Figure 31: Pearson correlation coefficient of the data set and benchmarking of GeoReVi's results (b) to the well-established software Past3 (a).

2.6.2 Regression analysis

As often a clear trend between grain density and bulk density can be observed, we will conduct a regression analysis, whether this trend can be observed in the investigated outcrop as well. Therefore, we navigate to **Charts** and there to the **Line chart**. In the top part, we can find two drop-down menus where we can select the numerical properties to be visualized. First, we will select the *Grain density* and *Bulk density* parameters. When clicking on **Create chart** the data sets will be added to the line chart. Now, the chart can be adapted as needed. In order to perform the regression analysis, we will tick the **Show regression** checkbox in the **Regression** tab. After **Refreshing** the regression line will be shown. The user can select between *linear* and *curvilinear* and define the degree of the function that should be used for curve fitting.

It is obvious, that the observed trend does not fit a linear relationship completely. Thus, we will try to fit the regression curve according to a polynomial of higher order. Therefore, under **Data series → Regression** we will select the **Regression type** to be *curvilinear* and define a **Degree** of 3. After **Refreshing** the chart, we will observe a better goodness of fit with an R^2 value of .26 instead of .186 before (fig. 34).

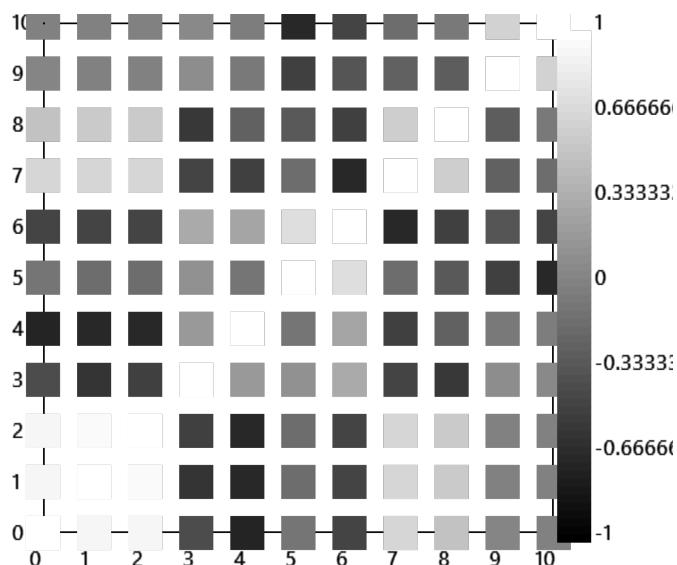


Figure 32: Visualization of the Pearson correlation coefficient.

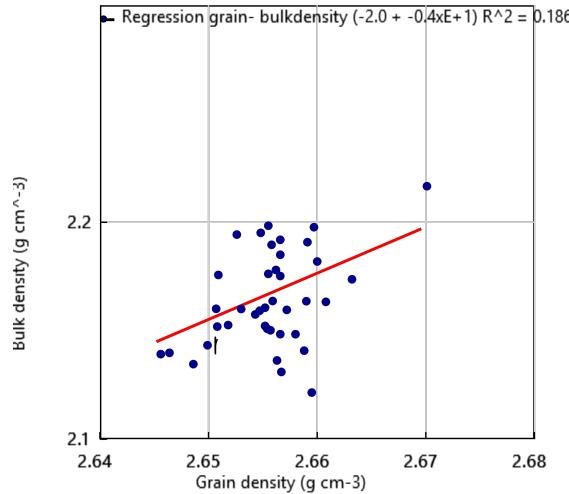


Figure 33: Linear regression analysis of grain and bulk density of the provided data set.

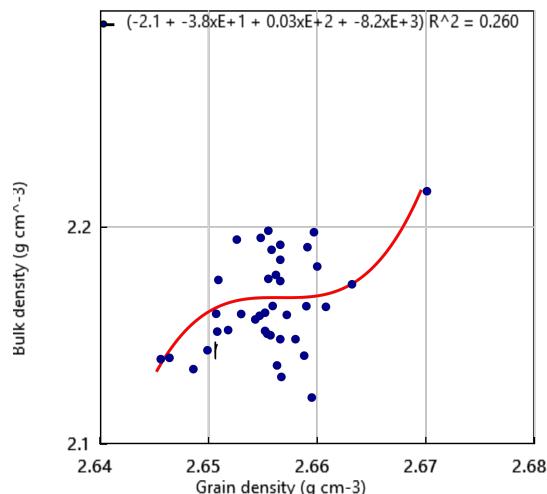


Figure 34: Non-linear regression analysis of grain and bulk density of the provided data set using a degree of 3.

2.6.3 Grouping

If groups are specified in a data set, it can be useful to split the data set up into these. Therefore, the training table contains a column called *Location relative to fault*. This column specifies whether the extracted rock sample is taken east or west of the central fault zone observed in our outcrop. We will select this column in the drop-down menu next to the **Group** button. When we eventually click the **Group** button, the data sets will be split up according to the unique names provided in the selected column. Results will be added to the data sets. The result in our case should contain a *West* and *East* data set.

2.6.4 Principal component analysis

Since often reservoir parameters provide physical high-dimensional interrelationships, we want to perform a dimensionality reduction algorithm. We **Right click** → **Add data set to PCA** for both data sets *West* and *East*. Now, we navigate to **Statistics** → **Principal Component Analysis** and we find our data set added to this analytical method. The PCA method will be defined as *Eigendecomposition*, the **Missing data treatment** as *Arithmetic mean* and the **Method** will be selected as *Standardized*.

When we press **Compute** we will receive the results of the PCA in a few seconds. At first, we will inspect the **Used data matrix**. Here, we find all columns that were used in the PCA. Please note that *Nan* and *0* values were replaced by the arithmetic mean of all values in the column.

Firstly, the eigenvalues of the principal components should be analyzed. In the tab menu displaying the calculator we can find the tab **Eigenvalues**. The eigenvalues derived from this analysis shows values of 5.24, 2.59 and 1.30 for the first, second and third principal component covering 47, 24 and 12% of the total variance (fig. 35).

When considering the eigenvectors of the PCA, we can see, that some properties are clustered. We should observe, that grain density and bulk density, thermal conductivity and thermal diffusivity as well as porosity and permeability are positively correlated to one another. Also we should observe, that the groups grain density/bulk density and thermal conductivity/thermal diffusivity are oriented conversely to one another. This confirms our findings from the correlation analysis, where grain density and thermal conductivity provide a Pearson correlation coefficient of -.506 and bulk density with

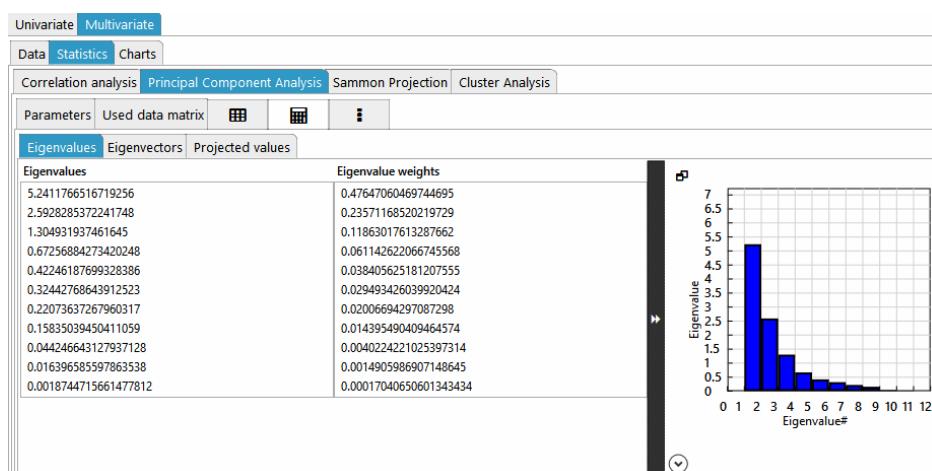


Figure 35: .

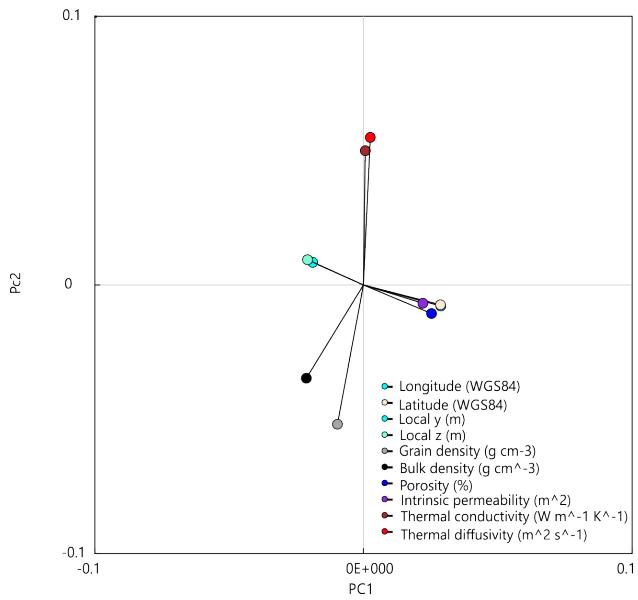


Figure 36: Biplot of the eigenvectors of PC1 and PC2 grouped by property.

thermal diffusivity of -.456.



If the data table does not show any values and the rows appear to be very thin, try to adapt the height of the topmost row. This will trigger a function that renders the data table accordingly.

Since the western and eastern side of the fault provide slightly different lithologies, the two clusters from *West* and *East* do not overlap much in the convex hull of the projected values considering PC1 and PC2 (fig. 37). This is an optimal result. However, when considering PC1 and PC3 that also has a significant weight on the total variance, the cluster begin to overlap slightly.

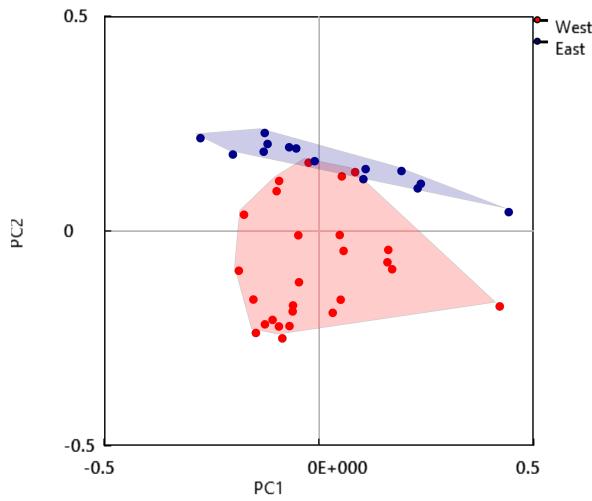


Figure 37: Values of the data sets *West* and *East* which were projected on the first and second principal component. Convex hulls are displayed for clearer differentiation.

2.6.5 Self-organized Maps

The aim of dimensionality reduction is to produce simplified representations of high-dimensional data sets. Another way to produce such projections are called self-organized maps (SOM). One method that belongs to the group of

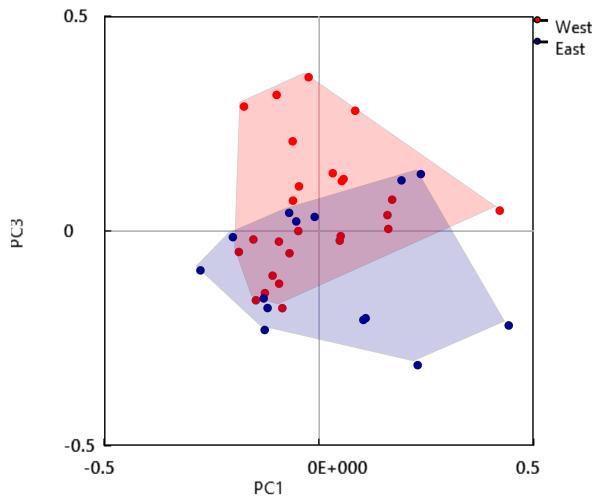
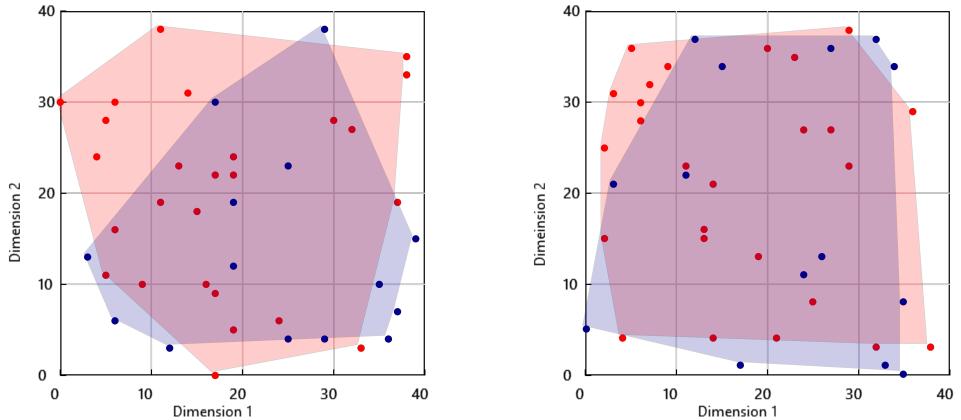
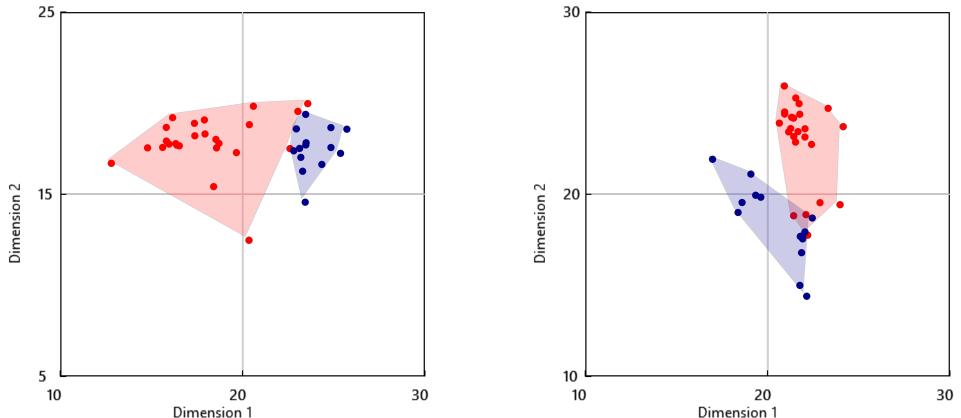


Figure 38: Values of the data sets *West* and *East* which were projected on the first and third principal component. Convex hulls are displayed for clearer differentiation.



(a) Sammon projection after 5 iterations. (b) Sammon projection after 50 iterations.



(a) Sammon projection after 500 iterations. (b) Sammon projection after 5000 iterations.

SOM is called Sammon projection (SM). The SM is a method that aims to truthfully represent clusters in high-dimensional data sets in order to provide meaningful visualizations in lowerdimensional space. To produce such representations, we will perform **Right click → Add data set to Sammon** for both data sets *West* and *East*. Now, we navigate to **Statistics → Sammon projection**. We will define maximum iterations of 5, 50, 500 and 5000 one after another and perform the projection. Results are visualized in tabular format and in scatter charts.



If you want to visualize a dimensionality-reduced data set in 3-D or 4-D you can copy the projected values into a separate .CSV or .XLSX file that you import into the **Univariate** data analysis form. Therefore, define the principal components 1-3 as X, Y and Z and optionally, define principal component 4 as *Value*.

2.6.6 Cluster analysis

As now, we could show, that the observed measurements cluster conversely in high-dimensional space. However, normally those relationships are not predefined by a classification. Thus, we want to perform a classification analysis now to categorize the observed rock samples into clusters. Therefore, we will perform **Right click** → **Add data set to cluster** for both data sets *West* and *East*. Now, we navigate to **Statistics** → **Cluster analysis** and define two data sets. In the **Parameter** menu, we will define *k-Means* as **Clustering method** and *2* as the number of

After clicking on **Compute** an unsupervised k-Means classification will be performed and the input samples will be assigned to the designated clusters. The results can be seen in table 1. With k-Means clustering we were able to classify 91% of the values correctly according to their positions relative to the fault zone.

Table 1: Rock samples used for cluster analysis and the true location with regard to the central fault zone in the outcrop. Erroneously classified rock samples are highlighted in red color.

Samples cluster 2	Relative position	Samples cluster 1	Relative position
OSB1_1	West	OSB25	West
OSB1_2	West	OSB26	West
OSB6	West	OSB28	West
OSB7	West	OSB30	East
OSB8	West	OSB31	East
OSB9	West	OSB32	East
OSB10	West	OSB33	East
OSB11	West	OSB34	East
OSB12	West	OSB35	East
OSB13	West	OSB36	East
OSB14	West	OSB37	East
OSB15	West	OSB38	East
OSB16	West	OSB39	East
OSB17	West	OSB40	East
OSB19	West	OSB41	East
OSB21	West	OSB42	East
OSB22	West	OSB43	East
OSB23	West	OSB44	East
OSB29	West	OSB27	West
OSB4_1	West		
OSB4_2	West		
OSB2_2	West		

3 In-depth guide and theoretical aspects

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 [39, 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 41 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

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

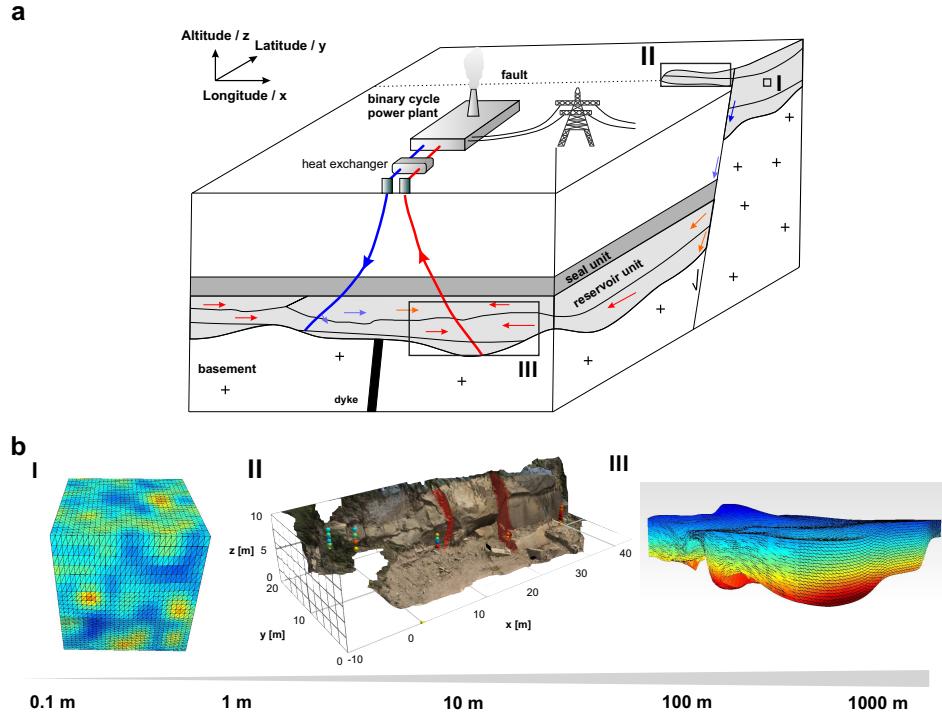


Figure 41: Concept of a multiscale outcrop analogue study.

are direction-dependent parameters and should be considered as tensors [30, 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 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

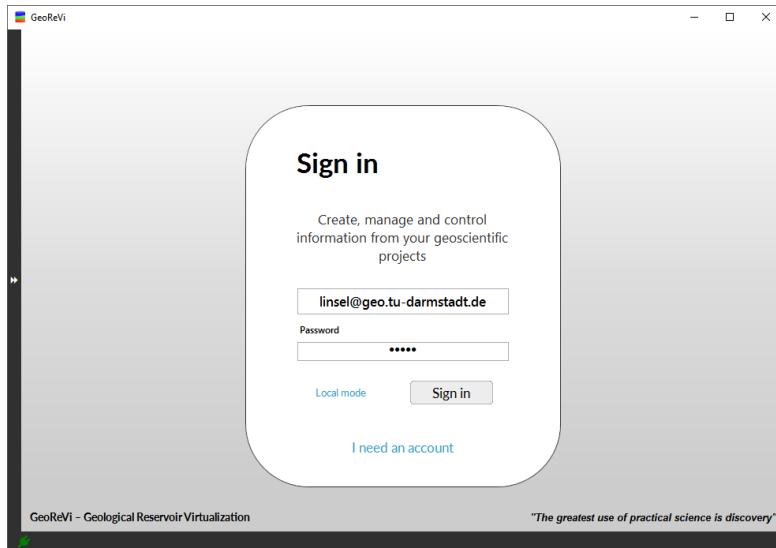


Figure 42: Login view.

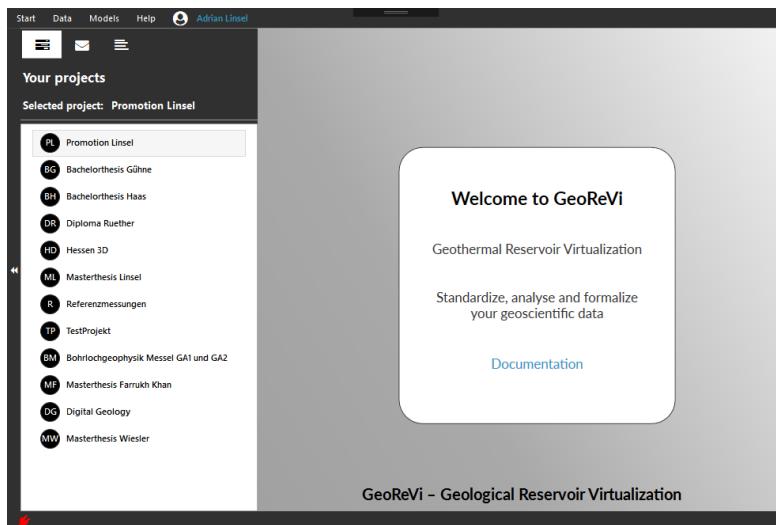


Figure 43: Home view with projects sidebar.

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

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 and that can be used to work through the tutorial.

A comprehensive data set is provided in the Disibodenberg project. 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, [35]) 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 [24] and the global geochemical database of [16] were used to extend GeoReVi's data model . However, most data models are developed within

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

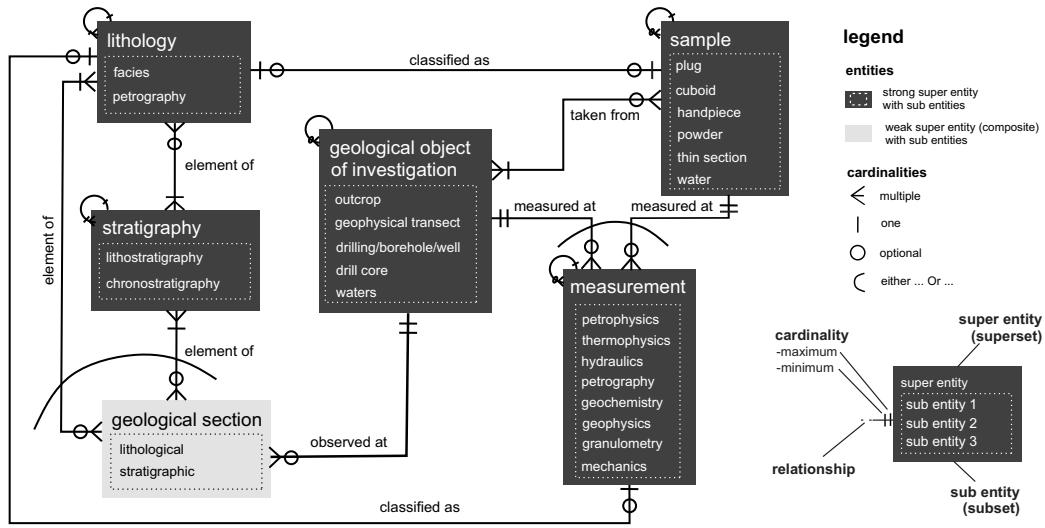


Figure 44: Core Entity-Relationship model of the GeoReVi database. To provide a better overview, attributes were excluded from the model.

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
- (d) hydro(geo)logical objects

3. Field/drill core samples

- (a) cylinders/plugs
- (b) cuboids
- (c) handpieces

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

- (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 γ -ray
- (b) total γ -ray
- (c) magnetic susceptibility
- (d) palaeo flow
- (e) bedding
- (f) lineaments
- (g) joints
- (h) temperature
- (i) sonic log
- (j) rock quality designation index

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

- (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. We will try to answer your requests as soon as possible.

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

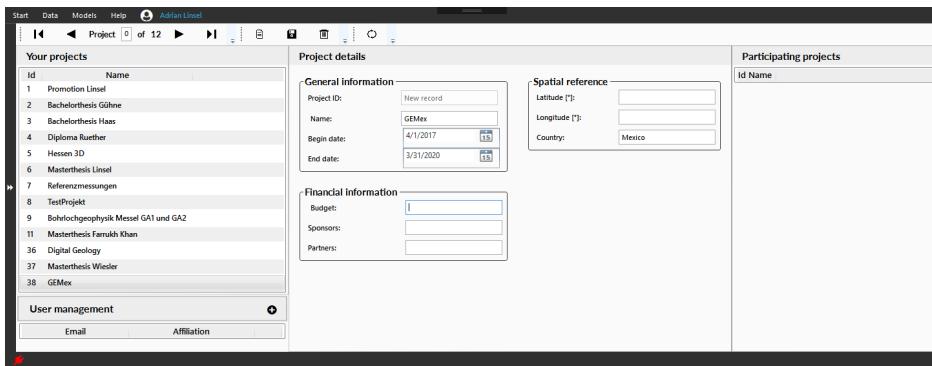


Figure 45: Projects view with a newly added project.

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 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

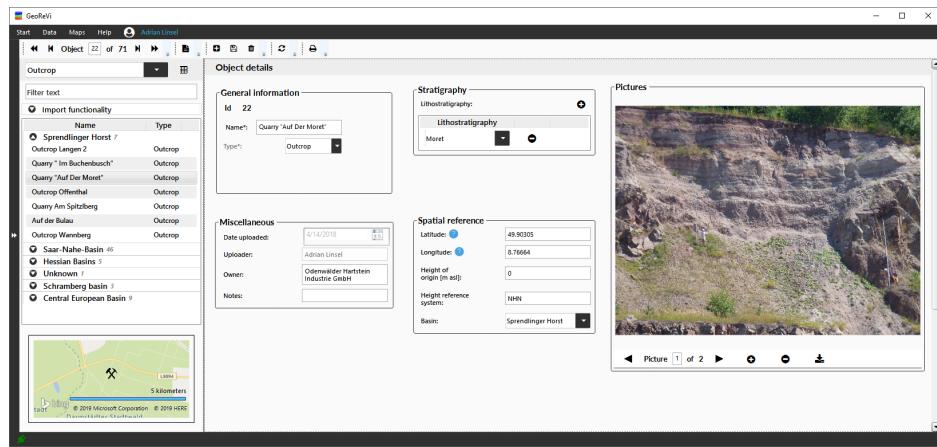


Figure 46: Objects of investigation form.

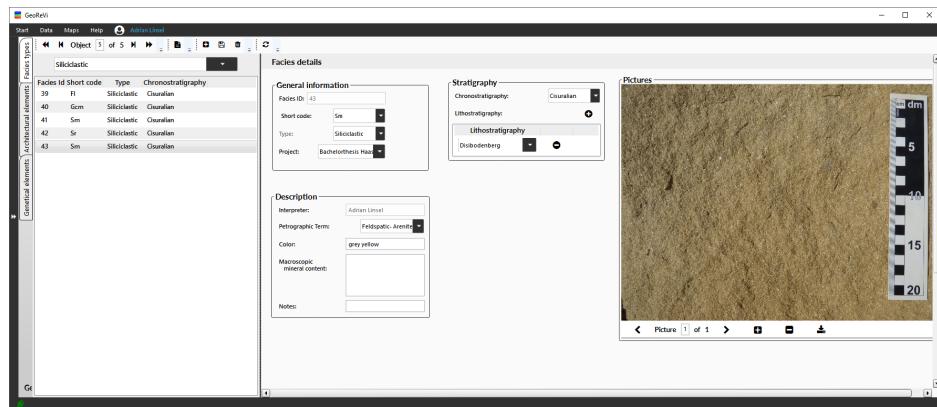


Figure 47: Lithologies form.

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 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

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.

3.5 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.

Sample name	ID	Lithotigraphy	Petrography	Lithofacies	Chronotigraphy	Object of investigation	Sample type	Architectural element	Depositional environment	Local x (m)	Longitude (°E)	Latitude
OSB9	387	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			-4.7	7467276	49.52363
OSB9	388	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			-4.7	7467276	49.52363
OSB7	389	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			-4.7	7467276	49.52363
OSB8	390	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			-4.7	7467276	49.52363
OSB4	391	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB42	392	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB41	393	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB10	394	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB11	395	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB12	396	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			0	7467113	49.52396
OSB13	397	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			0	7467113	49.52396
OSB14	398	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			0	7467113	49.52396
OSB15	399	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			34.5	7467113	49.52396
OSB16	400	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			34.5	7467113	49.52396
OSB17	401	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			34.5	7467113	49.52396
OSB18	402	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			34.5	7467113	49.52396
OSB19	403	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB20	404	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB21	405	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			30	7467113	49.52396
OSB22	406	Duisburg	Lithic-Arenite	Sandstein	Sandstone Quarry Oberurselbach	Plug	0			31.4	7467113	49.52396

Figure 48: Measurements form with an import area where .CSV and .XLSX files can be dropped for import.

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

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

$$y_i = \log(x_i) \quad (3)$$

exponential

$$y_i = 10^{x_i} \quad (4)$$

z-score

$$y_i = \frac{x_i - \bar{x}}{\sigma} \quad (5)$$

rescaling

$$y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (6)$$

mean-rescaling

$$y_i = \frac{x_i - \bar{x}}{x_{max} - x_{min}} \quad (7)$$

subtract-mean

$$y_i = x_i - \bar{x} \quad (8)$$

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.

When selecting the **All** checkbox, all measurements from all rock samples in the list will be loaded. Also, when the **Global** checkbox is ticked, the global instead of the local coordinates of the measurement will be provided. A way to filter measurements is to use the text box below the measurements listed. Here, the user can define specific text filters that should be applied on the loaded data set like. For example, if a user wants to load all sonic wave velocity values that are of type *s-wave* in the dropdown menu, the property *Wavetype [p-wave or s-wave]* must be selected and in the **Contains** textbox the text *s-wave* must be inserted like shown in fig. 49. Now all sonic wave velocities of type *s-wave* will be loaded into buffer memory.

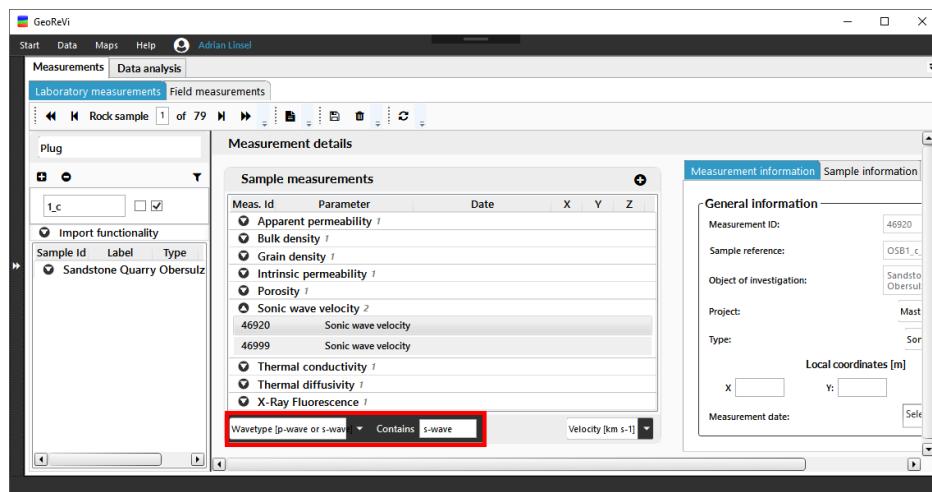


Figure 49: Filtering criteria to load a set of univariate measurements.

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

count of a sample

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

$$n \quad (9)$$

maximum

$$x_{max} \quad (10)$$

minimum

$$x_{min} \quad (11)$$

range

$$r = x_{max} - x_{min} \quad (12)$$

arithmetic mean

$$\bar{x} = \frac{\sum_{i=0}^n x_i}{n} \quad (13)$$

geometric mean

$$\bar{x}_g = \sqrt[n]{\prod_{i=0}^n x_i} \quad (14)$$

harmonic mean

$$\bar{x}_h = \frac{n}{\sum_{i=0}^n \frac{1}{x_i}} \quad (15)$$

sample variance

$$\sigma^2 = \frac{1}{n-1} \sum_{i=0}^n (x_i) \quad (16)$$

sample standard deviation

$$\sigma = \sqrt{\sigma^2} \quad (17)$$

skewness

$$s = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=0}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2}\right)^3} \quad (18)$$

kurtosis

$$k = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=0}^n (x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2}\right)^4} - 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)} \quad (19)$$

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 σ 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.

Oulier detection

Field parameters used in reservoir characterization are measured either in the laboratory or directly in the field. Each measurement method provides individual error ranges as well as erroneous measurement values due to operational errors or unexpected influence factors. Nevertheless, it is not straightforward to detect those erroneous measurements because often nature behaves strongly irregularly and in an unexpected manner.

For normally-distributed and log-normally-distributed random variables, the method of Inner-Quartile-Range (IQR) like described in [41] or [20] was implemented. Here, the difference between the first and third quartile is calculated. It is assumed, that each point exceeding the $n \cdot IQR$ distance, with $n=[0,2]$ from the arithmetic mean is classified as an outlier. The procedure is recursive which means, that the IQR has to be calculated again after outlier exclusion. This method is robust considering normally-distributed uni-variate data sets.

Outlier detection can be applied in box-whisker charts.

3.5.3 Variography

Originally defined by [27], 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_α and x_β 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_α and x_β can be expressed as points separated by a lag vector \mathbf{h} , dissimilarity can also be formulated as:

$$\gamma_{\alpha\beta}(\mathbf{h}) = \frac{1}{2} \cdot (z(x_\alpha + \mathbf{h}) - z(x_\alpha))^2 \in [0, \infty] \quad (20)$$

Semivariogram calculation results in a set of points representing the cumulative dissimilarity γ 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. 21).

$$\gamma(\mathbf{h}) = \frac{1}{2n_c} \sum_{\alpha=1}^{n_c} (z(x_\alpha + \mathbf{h}) - z(x_\alpha))^2 \in [0, \infty] \quad (21)$$

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 [43]. A theoretical variogram γ_{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, 33] like:

spherical

$$C_{sph}(\mathbf{h}) = \begin{cases} b \cdot \left(1 - \frac{3|\mathbf{h}|}{2a} + \frac{|\mathbf{h}|^3}{2a^3}\right) & \text{for } 0 \leq |\mathbf{h}| \leq a \\ 0 & \text{for } |\mathbf{h}| \geq a. \end{cases} \quad (22)$$

gaussian

$$C_{gau}(\mathbf{h}) = \begin{cases} b \cdot \exp\left(-\frac{|\mathbf{h}|^2}{a^2}\right) & \text{for } 0 \leq |\mathbf{h}| \leq a \\ 0 & \text{for } |\mathbf{h}| \geq a. \end{cases} \quad (23)$$

exponential

$$C_{exp}(\mathbf{h}) = b \cdot \exp\left(-\frac{|\mathbf{h}|}{a}\right) \quad \text{with } a, b > 0 \quad (24)$$

power

$$C_{pow}(\mathbf{h}) = b \cdot h^a \quad \text{with } a, b > 0 \quad (25)$$

cardinal sine

$$C_{lin}(\mathbf{h}) = \sin\left(\frac{|\mathbf{h}|}{a}\right) \cdot \left(\frac{|\mathbf{h}|}{a}\right)^{-1} \quad (26)$$

linear

$$C_{lin}(\mathbf{h}) = \begin{cases} b & \text{for } |\mathbf{h}| > 0 \\ 0 & \text{for } |\mathbf{h}| = 0 \end{cases} \quad (27)$$

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 [33, 17, 32]. Further on, the differences in range and sill in dissimilar directional semivariograms can quantify the geometric anisotropy of a field parameter [33]. 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.

3.5.4 Indicator variography

UNDER PREPARATION.

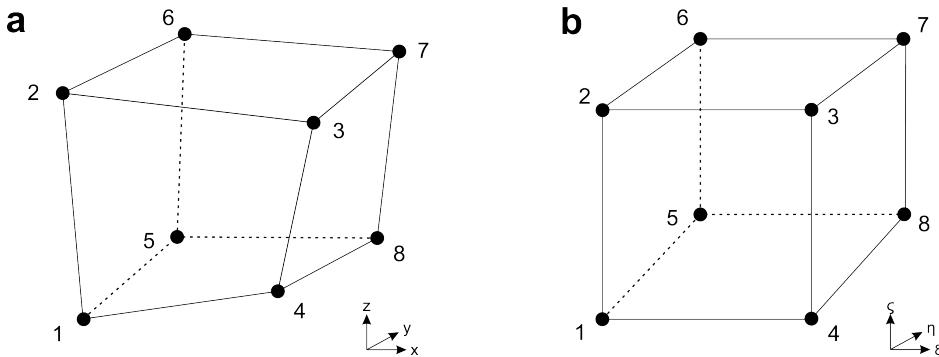


Figure 50: Conceptual representation of a hexahedral grid cell in GeoReVi as irregular shape (a) and with a local coordinate system (b).

3.5.5 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

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

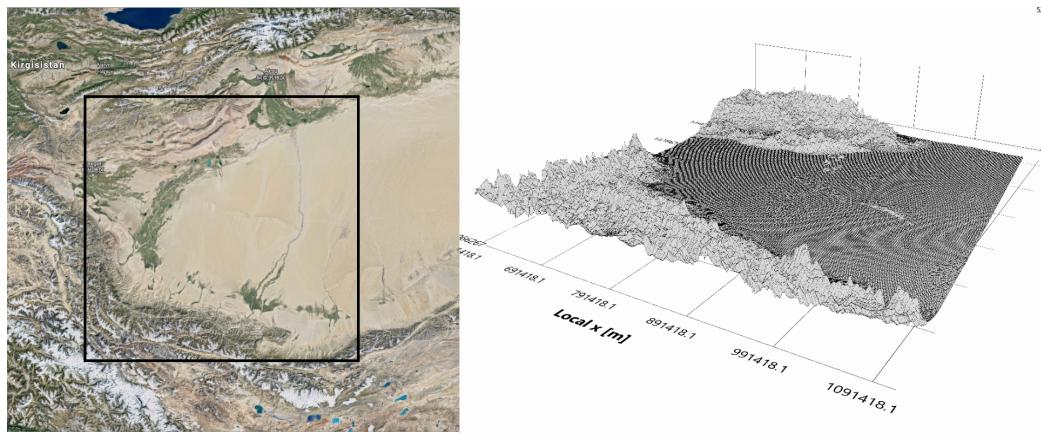


Figure 51: Digital elevation model of the area around the western part of the Xinjiang basin in China (15x exaggerated).

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. 52. Nodes from the two source data sets serve as boundary nodes for the new mesh. After projection, the points in

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

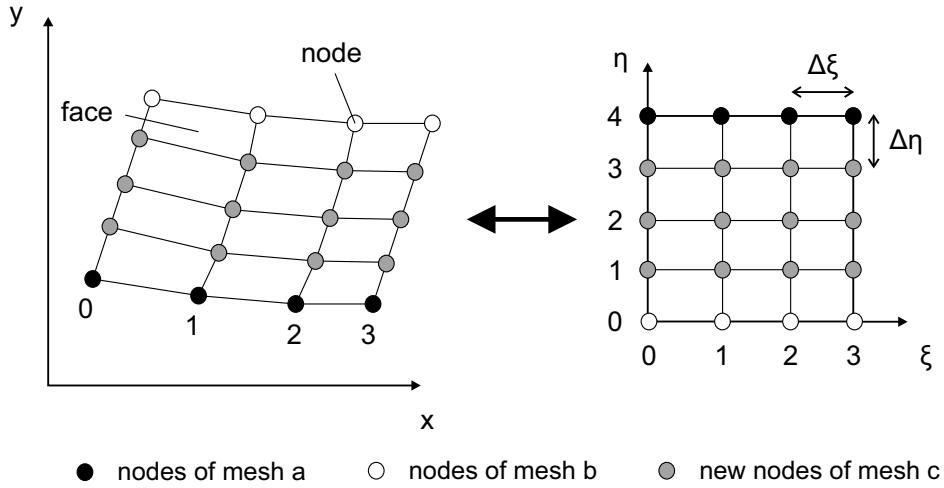


Figure 52: Concept of mesh generation using two constraint meshes of dimensionality $d-1$ with d as the target dimension.

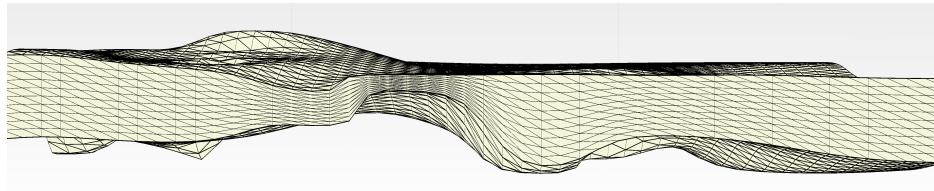


Figure 53: 3-D mesh generated with two bounding surfaces on the top and on the bottom.

between the two boundary meshes are created according to

$$\begin{aligned} R &= (i-1)(n_\zeta + 1) + 1 \\ C &= \text{mod}((i-1), (n_\iota + 1)) + 1, \end{aligned} \quad (28)$$

$$\begin{aligned} x &= \sum N_k(\zeta, \iota) x_k \\ y &= \sum N_k(\zeta, \iota) y_k \end{aligned} \quad (29)$$

where R is the row and C is the column of a node i , n_ζ is the number of

3.5.6 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

$$V_m = \sum_{i=1}^n V_i, \quad (30)$$

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

$$V_t = \frac{|(\mathbf{a} - \mathbf{d}) \cdot ((\mathbf{b} - \mathbf{d}) \times (\mathbf{c} - \mathbf{d}))|}{6}. \quad (31)$$

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

$$V_h = \sum_{i=1}^6 V_t^i. \quad (32)$$

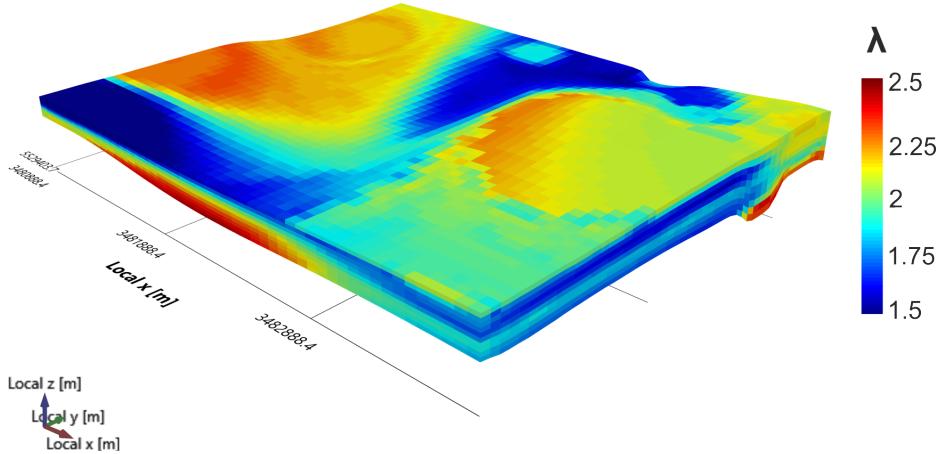


Figure 54: IDW interpolation result for thermal conductivity measurements in the Donnersberg formation located in central Germany.

3.5.7 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.

Inverse Distance Weighting

For deterministic interpolation we performed inverse distance weighting (IDW),

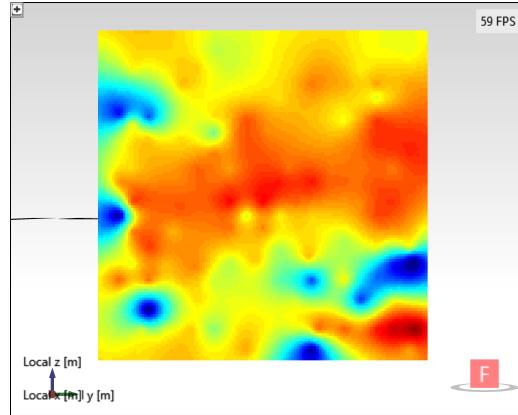


Figure 55: Simple kriging interpolation of the apparent permeability measured on 108 rock plugs on a rock slab of 0.5 x 0.5 m.

p-value IDW and Shepard's IDW [38] 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:

$$z(x_0) = \frac{\sum_{k=1}^n 1/d_k^p \cdot z(x_k)}{\sum_{k=1}^n 1/d_k^p}, \quad (33)$$

where d is the Euclidean distance between 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 [RN635].

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 [26, 31]. 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:

$$z(x_0) = \sum_{k=1}^n w_k \cdot z(x_k), \quad (34)$$

where w_k is the weight of the known point x_k with the value $z(x_k)$. While eq. 34 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:

$$\begin{pmatrix} c(x_1 - x_1) & \cdots & c(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ c(x_n - x_1) & \cdots & c(x_n - x_n) & 1 \end{pmatrix} \begin{pmatrix} w_1^{SK} \\ \vdots \\ w_n^{SK} \end{pmatrix} = \begin{pmatrix} c(x_1 - x_0) \\ \vdots \\ c(x_n - x_0) \end{pmatrix} \quad (35)$$

with c as covariance function, x_n as point with known value [43].

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:

$$\begin{pmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix} \quad (36)$$

with γ as theoretical semivariogram, x_n as point with known value and μ as Lagrange parameter [43].

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:

$$\begin{pmatrix}
 \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 & f_1(x_1) & \cdots & f_k(x_1) \\
 \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 & f_1(x_n) & \cdots & f_k(x_n) \\
 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\
 f_1(x_1) & \cdots & f_1(x_n) & 0 & 0 & \cdots & 0 \\
 \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 f_k(x_1) & \cdots & f_k(x_n) & 0 & 0 & \cdots & 0
 \end{pmatrix} \cdot \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \\ \phi_0 \\ \vdots \\ \phi_k \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \\ f_1(x_0) \\ \vdots \\ f_k(x_0) \end{pmatrix} \quad (37)$$

with γ as theoretical semivariogram, x_n as point with known value and μ as Lagrange parameter $f_i(x)$ as i^{th} polynomial [22].

3.5.8 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 σ_{SK}^2 [44] like:

$$\sigma_{SK}^2 = C(0) - \sum_{k=1}^n w_k C(x_i, x_0), \quad (38)$$

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 σ_{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

$$z(x_0) = \sqrt{-2 \cdot \log(u_1) \cdot \cos(2\pi \cdot u_2)} \cdot \sigma + \mu \quad (39)$$

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

3.5.9 Measurement Error Integration

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

$$\begin{pmatrix} \gamma(x_1 - x_1) + \sigma_1^2 & \cdots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) + \sigma_n^2 & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1^{OK} \\ \vdots \\ w_n^{OK} \\ \mu_{OK} \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix}, \quad (40)$$

what is the simple kriging system of linear equations [43] with error variance. In contrast to the conventional formula, σ_e^2 was added in the diagonal of the matrix.

3.5.10 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)

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (z(x_k) - z(x_0))^2} \quad (41)$$

and mean-absolute-error (MAE)

$$MAE = \frac{1}{n} \sum_{k=1}^n |z(x_k) - z(x_0)| \quad (42)$$

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 [43] 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*:

$$y_k = x_k - \bar{x} \cdot \mathbf{E} \quad (43)$$

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. 44):

$$\begin{aligned} v_y &= \frac{1}{n-1} \sum_{k=1}^n y_k^T y_k \\ &= \frac{1}{n-1} \sum_{k=1}^n ((x_k - \bar{x}) \cdot \mathbf{E})^T \cdot (x_k - \bar{x}) \cdot \mathbf{E} \\ &= \frac{1}{n-1} \sum_{k=1}^n \mathbf{E}^T \cdot (x_k - \bar{x})^T \cdot (x_k - \bar{x}) \cdot \mathbf{E} \\ &= \mathbf{E}^T \left(\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^T \cdot (x_k - \bar{x}) \right) \cdot \mathbf{E} \\ &= \mathbf{E}^T \cdot C \cdot \mathbf{E} \end{aligned} \quad (44)$$

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

$$\mathbf{E}^T \cdot \mathbf{E} = 1 \quad (45)$$

We will solve this optimization problem using the Lagrange optimization.

Self-organized maps (Sammon Mapping)

Sammon mapping aims to project a data set X into an n-dimensional representation Y minimizing the pairwise distances of X and Y so that $d_{ij}^x \approx d_{ij}^y$.

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

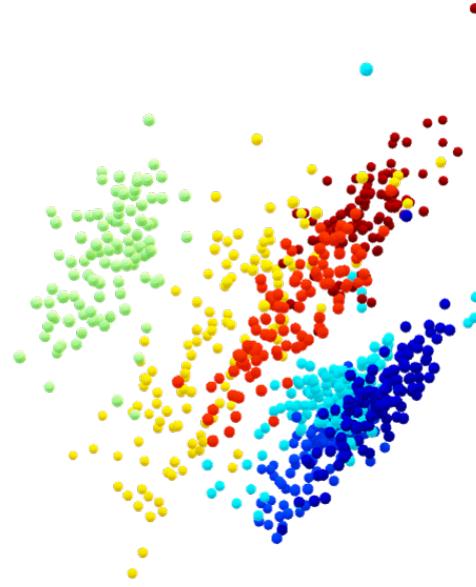


Figure 56: 3D representation of a Principal Component analysis.

Therefore, a representation Y is initiated randomly and Sammon's stress (E) is calculated which is expressed as

$$E = \frac{1}{\sum_{i=1}^n \sum_{j=i+1}^n d_{ij}^x} \sum_{i=1}^n \sum_{j=i+1}^n \frac{(d_{ij}^y - d_{ij}^x)^2}{d_{ij}^x}, \quad (46)$$

where d is the Manhattan distance between the point pairs of X and Y [37]. The SM algorithm minimizes the Sammon's stress gradient in a user-defined number of iterations. In order to minimize the projection error, [37] applied a steepest-descent iteration. Here, the mapping error is stated as

$$E(m) = \frac{1}{c} \sum_{i < j}^n (d_{ij}^* - d_{ij}(m))^2 / d_{ij}^*, \quad (47)$$

with

$$c = \sum_{i < j}^n (d_{ij}) \quad (48)$$

3 IN-DEPTH GUIDE AND THEORETICAL ASPECTS

and

$$d_{ij}(m) = \sqrt{\sum_{k=1}^d (y_{ik}(m) - y_{jk}(m))^2}. \quad (49)$$

The projected data set at iteration step $m + 1$ can be derived as

$$y_{pg}(m + 1) = y_{pg}(m) - \alpha \cdot \Delta_{pg}(m) \quad (50)$$

where α is the so called **Magic factor** that can be chosen to be between 0.3 and 0.4 and

$$\Delta_{pg}(m) = \frac{\partial E(m)}{\partial y_{pq}} / \left\| \frac{\partial^2 E(m)}{\partial y_{pg}(m)^2} \right\|. \quad (51)$$

The derivates can be derived as

$$\frac{\partial E(m)}{\partial y_{pq}} = \frac{-2}{c} \sum_{j=1}^n \left(\frac{d_{pj}^* - d_{pj}}{d_{pj} d_{pj}^*} \right) (y_{pq} - y_{iq}) \quad (52)$$

and

$$\frac{\partial^2 E(m)}{\partial y_{pg}(m)^2} = \frac{-2}{c} \sum_{j=1}^n \frac{1}{d_{pj}^* d_{pj}} \cdot \left[(d_{pj}^* - d_{pj}) - \frac{(y_{pq} - y_{iq})^2}{d_{pj}} \left(1 + \frac{d_{pj}^* - d_{pj}}{d_{pj}} \right) \right]. \quad (53)$$

Since, those formulations provide significant performance draw-backs, we used an implementation of [23] where the corrections of the target vector $y_{pg}(m + 1)$ can be derived as

$$\Delta y_{pg} = \lambda \cdot \frac{d_{ij} - |r_i - r_j|}{|r_i - r_j|} \cdot (r_i - r_j) \quad (54)$$

where

$$\Delta r_j = -\Delta r_i \quad (55)$$

3.6.2 Classification

GeoReVi covers classification domains for chronostratigraphy [35] , lithostratigraphy, petrography as well as facies, architectural element and depositional environment analyses [29] . 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 [25] 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 μ

$$c := \min_j ||\mathbf{x}^{(i)} - \mu_j|| \quad (56)$$

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.

$$s_{pe} = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\left(\sum_{k=1}^n x_k^2 - n \cdot \bar{x}^2 \right) \left(\sum_{k=1}^n y_k^2 - n \cdot \bar{y}^2 \right)} \quad (57)$$

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

$$s_{sp} = 1 - \frac{6 \cdot \sum_{k=1}^n d_k^2}{n(n^2 - 1)} \quad (58)$$

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

3.6.4 Regression analysis

Regression aims to find a fitting function between samples of two or more random variables. In GeoReVi, two types of regression are implemented, namely the linear and curvilinear regression. A linear regression tries to fit a linear function of the form

$$y = b_0 + b_1 x, \quad (59)$$

where y and x are the random variables, b_0 is the y-axis interception and b_1 is the gradient. In order to find a function where the squared sum of the Δy deviations is minimized like

$$\sum_{k=1}^n (\Delta y_k)^2 = \sum_{k=1}^n (y_k - (b_0 + b_1 x_k))^2 \quad (60)$$

we need to calculate the gradient as

$$b_1 = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^n (x_k - \bar{x})^2}. \quad (61)$$

Accordingly, b_0 can be calculated as

$$b_0 = \bar{y} - b_1 \bar{x}. \quad (62)$$

For curvilinear regression, a function of a degree ≥ 1 will be approximated for a discrete set of values. A second-degree polynomial function for instance would provide the form

$$y = b_0 + b_1 x + b_2 x^2 \quad (63)$$

Thus, we would need to find $n + 1$ regression coefficients, where n is the degree of the polynomial function. In general, the regression model yields

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_n x_i^n \text{ with } i = 1, 2, \dots, n. \quad (64)$$

The estimation of the regression coefficients is aimed through solving a system of linear equations as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1^1 & \cdots & x_1^m \\ 1 & x_2^1 & \cdots & x_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & \cdots & x_n^m \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad (65)$$

where x and y are the samples.

Quality assessment

Results from a regression analysis can be assessed through jack-knifing, cross-validation or bootstrapping. Regressions in GeoReVi are directly evaluated using the coefficient of determination (R^2) which is calculated as

$$R^2 = 1 - \frac{s_{res}}{s_{tot}} \in [0, 1], \quad (66)$$

where

$$s_{res} = \sum_{k=1}^n (y_k - f_k)^2 \quad (67)$$

is the explained sum of squares whereas

$$s_{tot} = \sum_{k=1}^n (y_i - \bar{y})^2 \quad (68)$$

is the total sum of squares.

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. In the following sections, we will show some special features that can be used to enhance the individual visualizations.

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 ∇ of a field parameter f like

$$\nabla f = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \quad \frac{\partial f}{\partial z} \right]^T \quad (69)$$

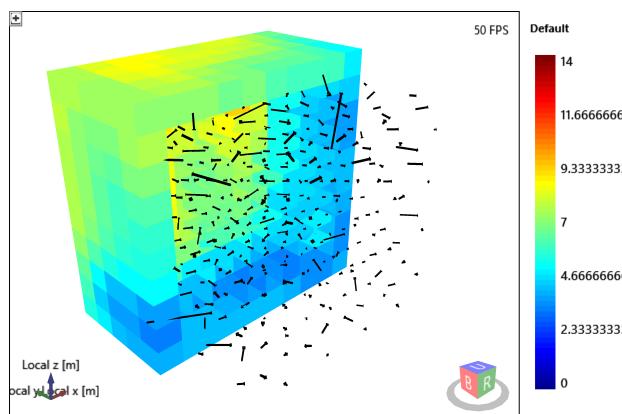


Figure 57: Vector field of the apparent permeability measured on a rock cube and interpolated with simple kriging.

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. [36] 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, [36] 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 [34]. 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 [42, 40, 21]. 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.

[28] 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.
- [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] J.-B. Clavaud et al. “Permeability anisotropy and its relations with porous medium structure”. In: *Journal of Geophysical Research: Solid Earth* 113 (2008), pp. 1–10. 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).

- [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>.
- [Dö04] 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. DOI: 10.1007/978-1-4020-2402-3_48. URL: 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. Data 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. DOI: [10.1016/j.geothermics.2016.11.007](https://doi.org/10.1016/j.geothermics.2016.11.007). URL: <http://dx.doi.org/10.1016/j.geothermics.2016.11.007>.
- [HHR01] Oyvind Hammer, David Harper, and Paul Ryan. *PAST: Paleontological 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>.
- [HSS16] Christian Heumann, Michael Schomaker, and Shalabh. “Measures of Central Tendency and Dispersion”. In: *Introduction to Statistics and Data Analysis : With Exercises, Solutions and Applications in R*. Cham: Springer International Publishing, 2016, pp. 37–66. ISBN: 978-3-319-46162-5. DOI: [10.1007/978-3-319-46162-5_3](https://doi.org/10.1007/978-3-319-46162-5_3). URL: https://doi.org/10.1007/978-3-319-46162-5_3.
- [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.
- [HW94] Gordon Hudson and Hans Wackernagel. “Mapping temperature using kriging with external drift: Theory and an example from scotland”. In: *International Journal of Climatology* 14.1 (1994), pp. 77–91. ISSN: 0899-8418. DOI: [10.1002/joc.3370140107](https://doi.org/10.1002/joc.3370140107). URL: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3370140107>.
- [Koh01] Teuvo Kohonen. “Variants of SOM”. In: *Self-Organizing Maps*. Vol. 30. Berlin, Heidelberg: Springer, 2001.

- [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>. 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.
- [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>. 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] Y. 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>.
- [Rü15] 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>. URL: <https://doi.org/10.1007/s12665-014-3554-5>.

REFERENCES

- [R⁺15] W. 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>.
- [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). 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.
- [Sam69] J. W. Sammon. “A Nonlinear Mapping for Data Structure Analysis”. In: *IEEE Transactions on Computers* C-18.5 (1969), pp. 401–409. DOI: [10.1109/T-C.1969.222678](https://doi.org/10.1109/T-C.1969.222678).
- [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] Y. 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>.

- [Tuk77] J.W. Tukey. *Exploratory Data Analysis*. Pearson, 1977, p. 712.
- [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](https://doi.org/10.1007/978-3-662-05294-5). URL: <https://doi.org/10.1007/978-3-662-05294-5>.
- [WM07] R. Webster and A. O. Margaret. *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

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

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

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.

jone 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.