Sub-Cellular Heterogeneity Toolkit Documentation

Developed by the Leeds Computational Physiology Lab @ University of Leeds

Contacts:

Dr. Michael A. Colman (m.a.colman@leeds.ac.uk)

Maxx Holmes (ed12m3h@leeds.ac.uk)

Getting Started

The Sub-Cellular Heterogeneity Toolkit was developed to study the sub-cellular non-uniform structure within cardiovascular cells. The premise of this toolkit is that subcellular heterogeneity on the scale of microns can be modelled as a spatial random field – that is, a region of space in which each point takes some value in a distribution of real numbers. The parameters of this spatial random field can be determined from confocal and super-resolution microscopy images using a variogram fitting protocol with an appropriate kernel – but in order to do this, images much be processed suitably for this analysis.

This toolkit contains tools to appropriately process single images and stacks of images for a variogram analysis. This analysis extracts the average expression, the expression of variation, and the correlation length-scale of variation across the image which can be used to produce an analogous spatial random field which represents the spatial parameters of the analysed image. This presents a method to quantifiably measure spatial heterogeneity in structure, and to model this heterogeneity computationally.

The last part of this toolkit is a Spatial Random Field generator, which can produce a spatial random field given analysed or custom parameters. One may produce any number of unique spatial random fields with the same parameter set, allowing us to study both the impact of spatial heterogeneity, but also the structural distribution itself.

This toolkit is compatible with our Multi-Scale Cardiac Simulation Framework (MSCSF) which can also be found on https://github.com/michaelcolman

This documentation will go over each tab in the toolkit in turn and explain how they should be used. If you have any questions, or feedback – please submit them to either of the emails provided and we will get back to you shortly.

I hope you find use in our tool and can find ways to implement this method in your own studies.

Contents

Setup

Requirements to run this toolkit:

- Python 3.6 (or greater)
- The following Python3 libraries:
    - PyQt5
    - Numpy
    - Matplotlib
    - Scipy
    - Imutils
    - Pillow
    - Gstools
    - Pylab
    - Pandas
    - CSV
    - Datetime
    - Sklearn

Unix users will be able to make use of the Install_Requirements.sh script, using the command "./Install_Requirements.sh"

Alternatively, these packages can be installed using the pip package manager for python.

ID & Load Tab
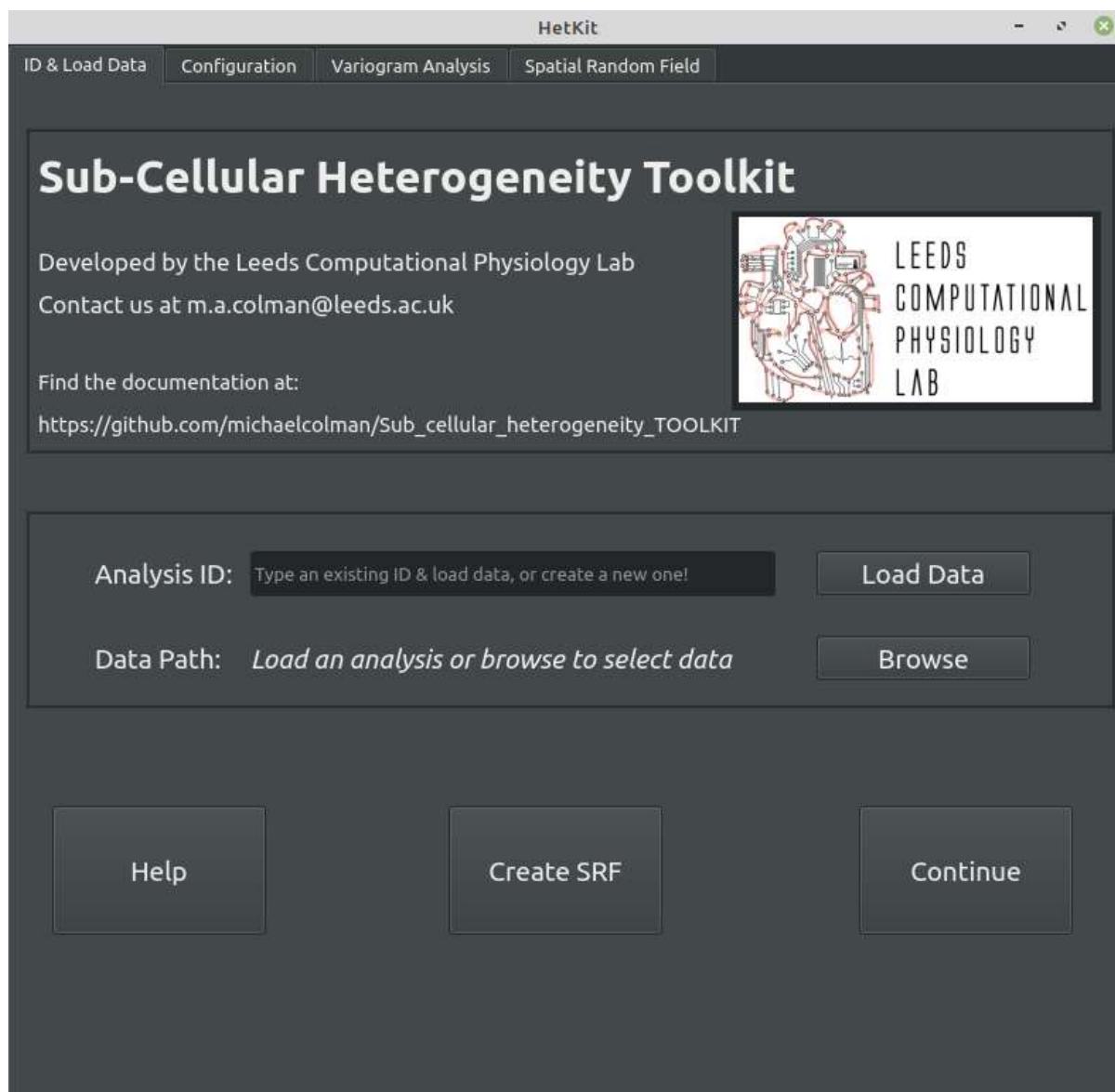
Key Features

Analysis ID: An identifier for an analysis

Load Data: This button will search the database for an existing analysis, and if one is found, it will load the previous configurational settings

Browse: This will open a file explorer window for the user to select an image to load. Accepted filetypes are: JPEG, TIFF, JPG and PNG

Help: This button will open the browser on the repository where you can find documentation and examples

Create SRF: This will move the user to the Spatial Random Field Generation screen.

Continue: This button will save all your settings to the database and move you to the configuration screen

Configuration Tab
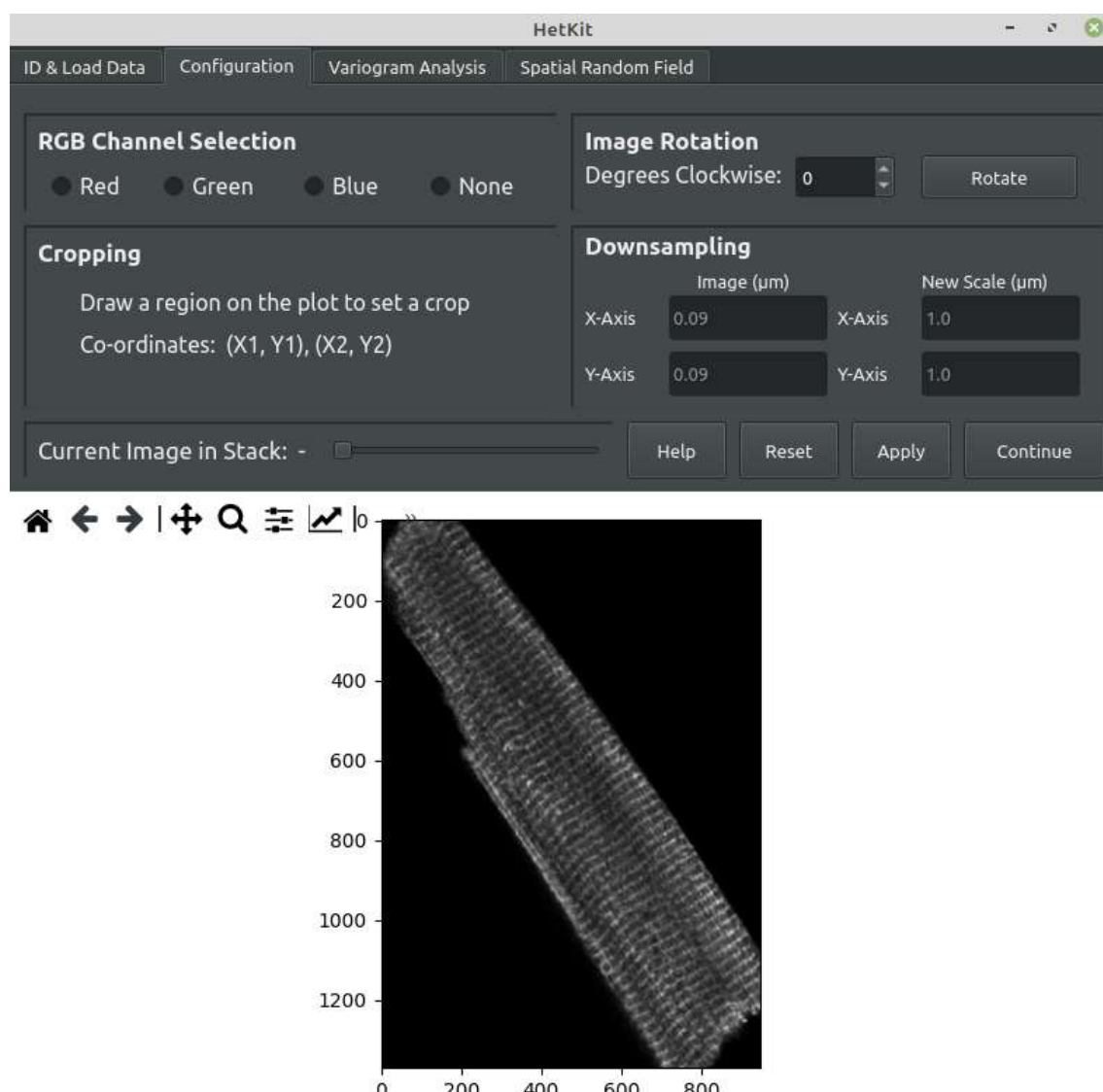
Key Features

RGB Channel Selection: Select the channel you wish to run the analysis on
Image Rotation: Enter an integer degree between -360 and 360 to rotate the image clockwise
Cropping: Use the right mouse button to drag and create a rectangle on the image to designate an area to crop the image to
Downsampling: Enter the X and Y scales for the original image (pixel size) and the new scale you wish to downsample the image to
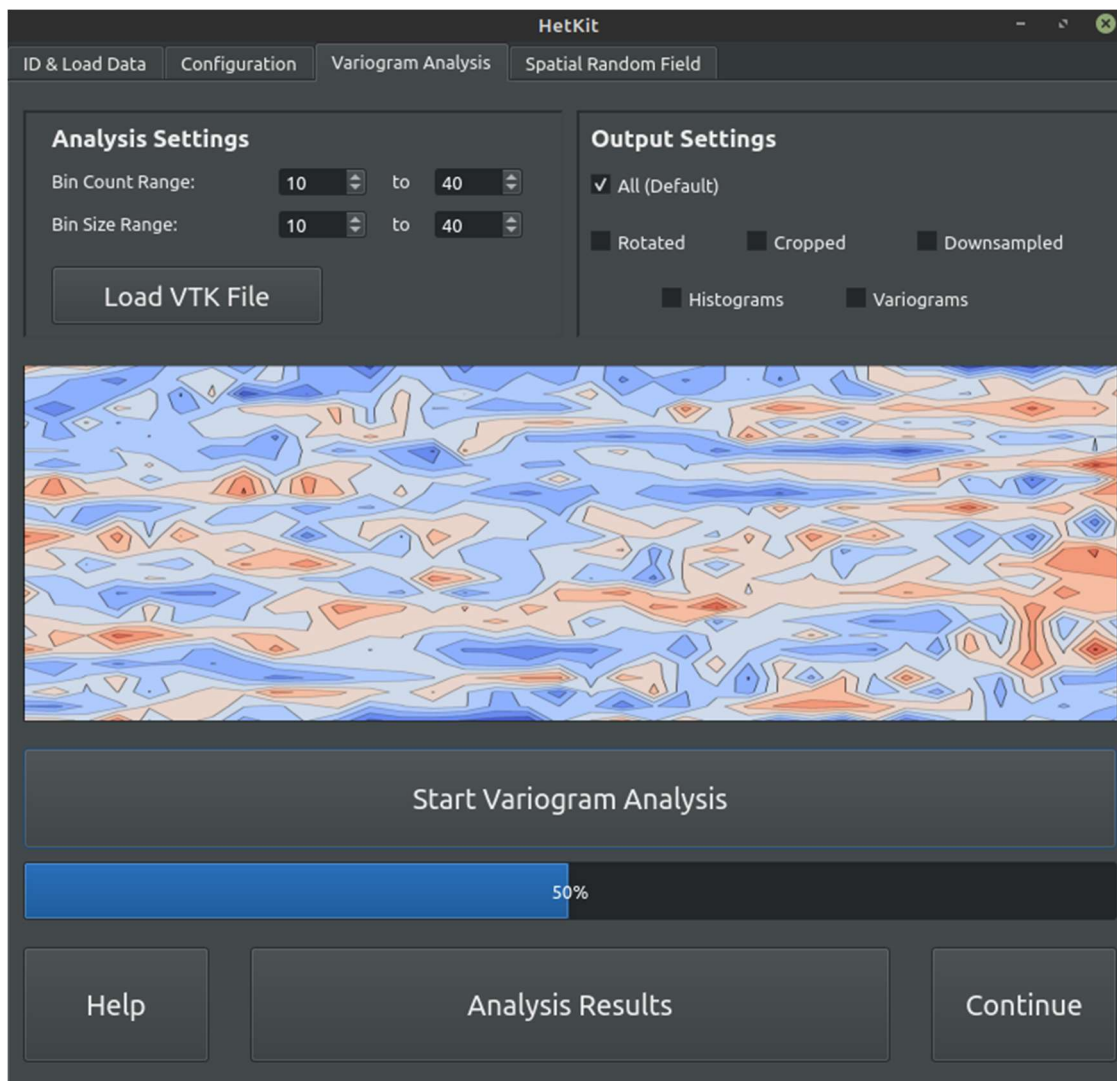**Current Image Slider**: Select which image to display on the canvas below
Help: Open the GitHub repository to find documentation and examples
Reset: Reload the original image onto the canvas
Apply: Apply all settings to the image and display it on the canvas
Continue: Save all settings to the database and move the user to the variogram tab

Variogram Analysis Tab

Key Features

Bin Count Range: The range of total bins to run the analysis on
Bin Size Range: The range of bin sizes to run the analysis on
Load VTK File: Load a .vtk file to analyse instead of a configured image / stack
Output Settings: Check to select which outputs you would like from your analysis
Start Variogram Analysis: Begin the variogram fitting protocol on the chosen dataset
Help: Opens the GitHub Repository where documentation can be found
Analysis Results: Open the directory in which the analysis results for this ID are stored
Continue: Save settings and continue on to the Spatial Random Field tab

Spatial Random Field Tab

Key Features

SRF Name: Enter an output name for the SRF

SRF Number: Enter the number of SRFs you wish to produce

Range (3SD): This defines the range of data, 3SD = 99% of the data is contained within plus or minus the range specified from the mean expression of the SRF.

Mean Expression: The cell-total average expression of the SRF

Lengthscales: The longitudinal and transverse correlation lengthscales of the SRF

Open Analysis Results: Opens the analysis results

Generate Preview: Produce a smaller, 2D random field with those lengthscales for the display canvas

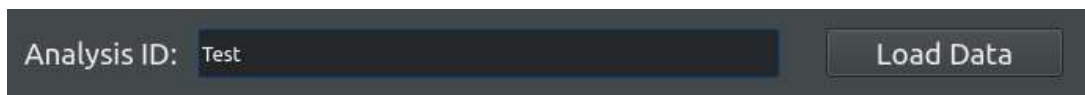**Generate Spatial Random Field:** Produce SRFs to the specifications set above

Open SRF Location: Open the SRF directory, where the user can find their SRFs

Help: Open the GitHub repository in browser
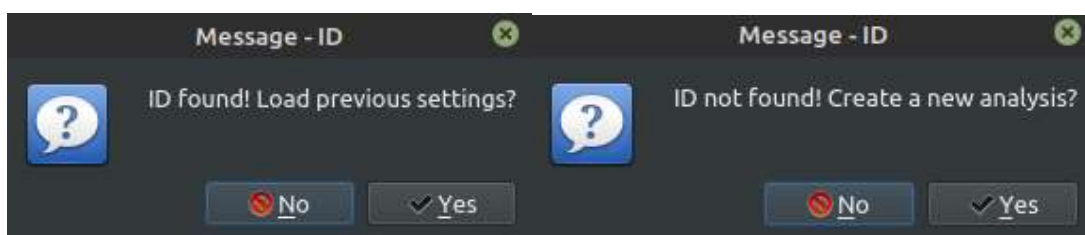
How to… Process an Image

STARTING A NEW ANALYSIS

Firstly, a new analysis should be created in the database so you can reload these settings and give an identifier to your outputs. To do this, enter an ID into this field.
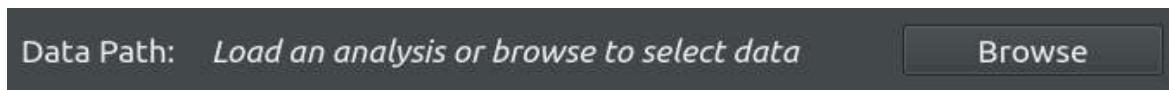

Once you have entered your ID, click the "Load Data" button to save your ID into the database, or load a previous analysis using the same ID!


One of these two windows will pop up, simply confirm yes or no.
Loading an ID will load in previous data paths and configurational settings, however for this guide, we will create a new ID.


Click on the "Browse" button to select any data you wish to load. This data needs to be in the form of an image. Most image types are accepted, including .tiff files! A file explorer will open, allowing you to choose one or more files to load into the program.
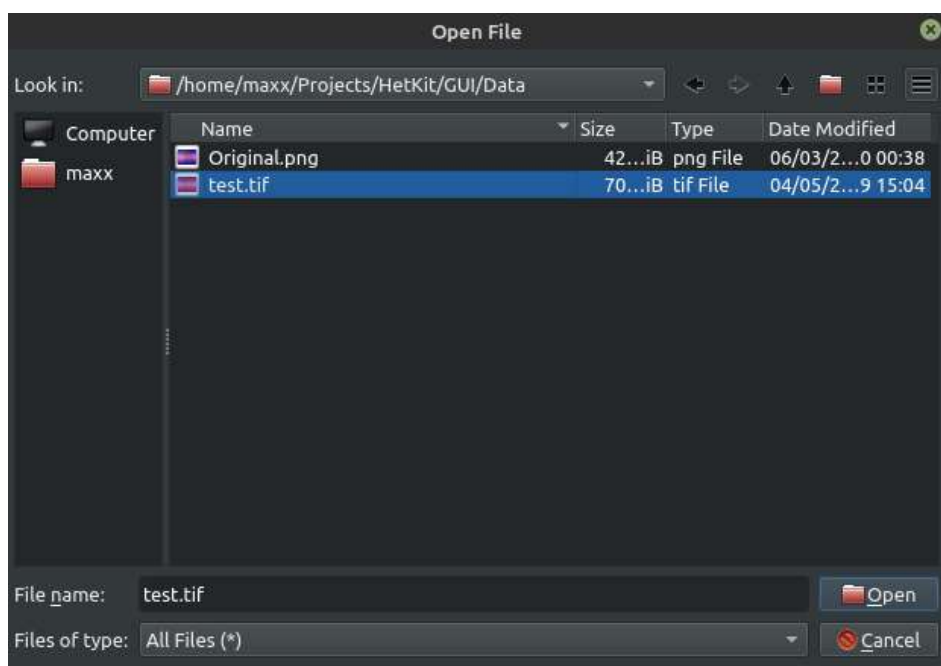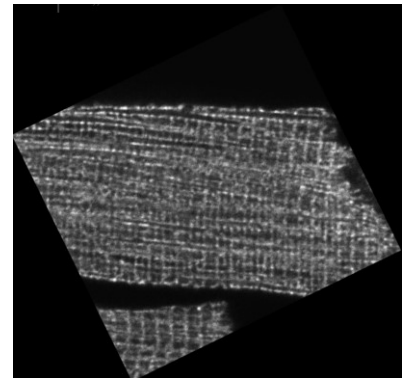
Once you are happy with your selection press the "Continue" button as shown below!

The program will now move forward to the "Configuration" tab! On this tab, you have four options for processing your image.
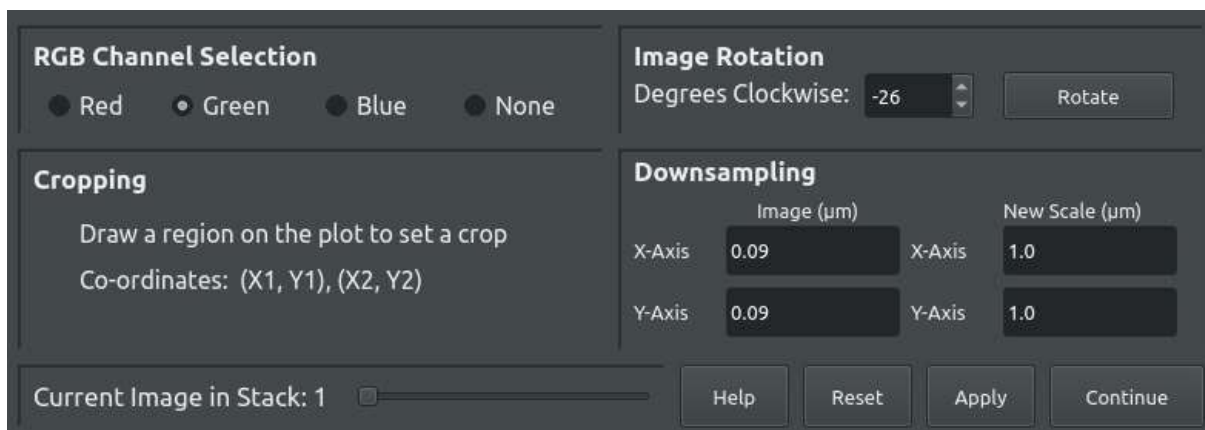

First, select an RGB channel to use. Clicking any of these buttons will select the respective channel and display the new image in the canvas below.

Next, determine a rotation for your data. You want to align the transverse and longitudinal axes of the cell with the x and y dimensions, such as shown on the right! Clicking the "Rotate" button, will rotate the dataset and display it on the canvas below.

Once the image is rotated, you can then draw a rectangle on the canvas to determine an area to be cropped, by clicking, holding, and dragging the left mouse button. It is ideal to select an area not containing any background, while still capturing as much data as possible, such as on the right.

Finally, input the original scale of the image in µm, and the desired scale of the image in µm. A value of around 1 µm is ideal.
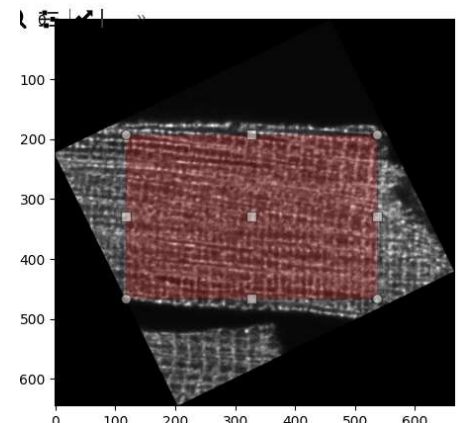
The "Current Image" slider allows you to move between images in your stack and test that these settings work on each image in your stack. It is important that each images undergoes the same processing procedure, as they are integrated over during the variogram fitting protocol.

Clicking the "Apply" button will apply all your configuration settings to the dataset and display the image on the canvas. If you are happy with the processed dataset, you may click "continue" to save your options to the database and continue. If you would like to make changes, you are able to use the "Reset" button, to display the original image again.

Once you have pressed "Continue" you will be moved to the "Variogram Analysis" tab.

There are several settings you should pay attention to before you start the variogram fitting process. The first is found under "Analysis Settings".

There are two settings to adjust here, "Bin Count" and "Bin Size" which correspond to the bins used within the fitting protocol. Values between each of these ranges are iterated over during the fitting process to maximise the chance of a successful fit. It is recommended to use a large range, however, a larger range will increase the computational time significantly. Larger datasets will require higher ranges, whereas smaller datasets will require smaller ranges.
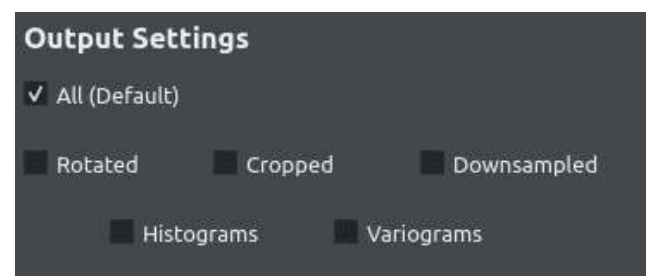
Analysis Settings

| Bin Count Range: | 5 | to | 50 |
| Bin Size Range: | 5 | to | 50 |

Next, is are the "Output Settings" which are checkboxes, in which the user is able to determine which outputs are given. In most cases, you should use the "All" default option, however if you want to reduce computational time, or do not need the variograms or histograms of the images plotted for you, you are able to skip these outputs to save time.
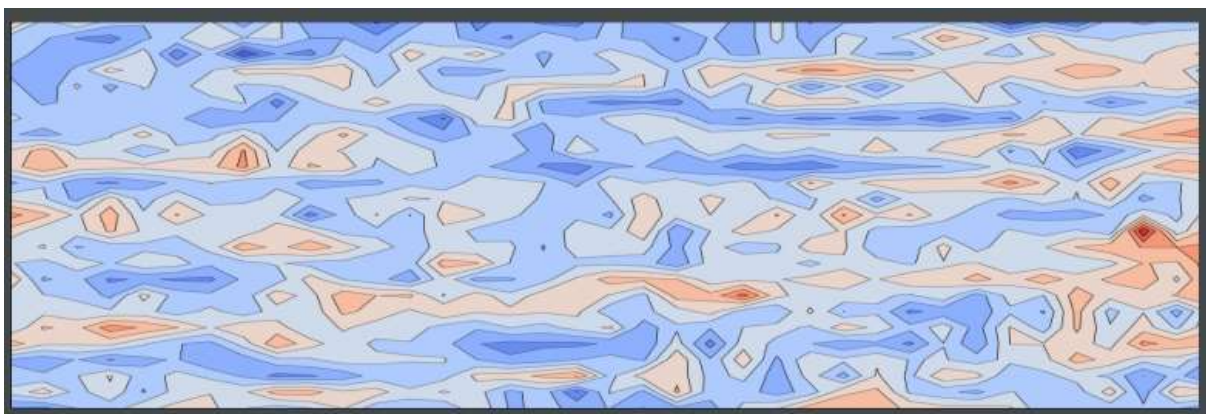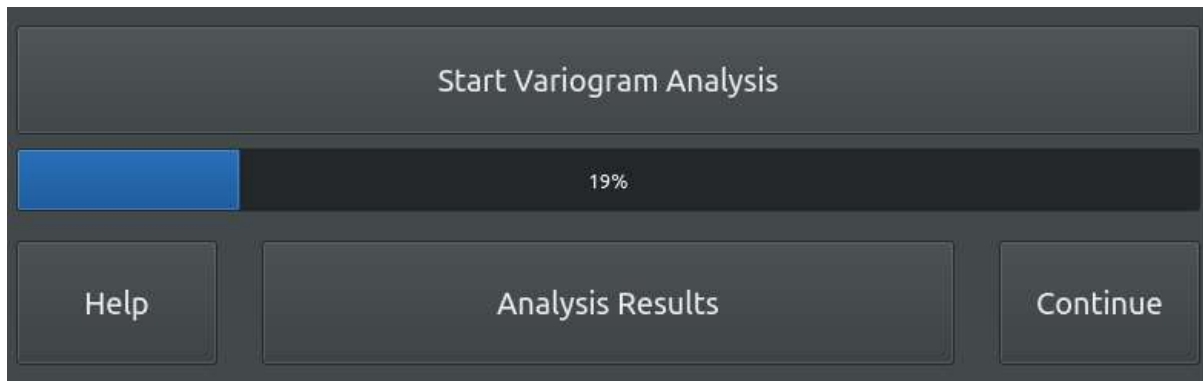
Output Settings

✓ All (Default)

☐ Rotated   ☐ Cropped   ☐ Downsampled

☐ Histograms   ☐ Variograms

The image above in a contour map of the data you have just processed in the "Configuration" tab to give you an idea of how heterogeneous that dataset is.

Once the settings have been filled out, the user may press the "Start Variogram Analysis" button to begin the fitting protocol. While the program is doing this, you are unable to do anything else, however the progress bar will fill up as the process is underway.

Once the process is over, you may either use "Analysis Results" to open up the directory containing the results in an explorer window, or alternatively, you may press "Continue" to move onto the "Spatial Random Field" tab, however this is covered in another section.

How to… Process a VTK file

There are several reasons why you may wish to process a .vtk file, and so there is an option created to load a .vtk file directly into the "Variogram Analysis" tab.
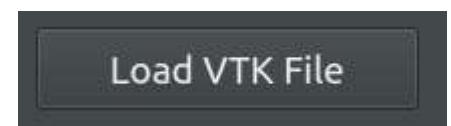
Before you do this, it is required to set up an analysis ID in the database by entering an ID into the field below and clicking "Load Data"
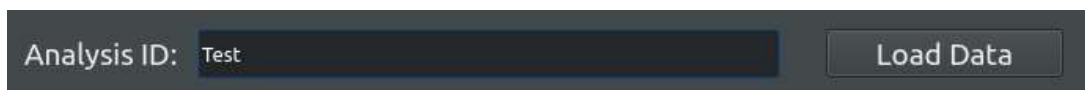
Once this has been done, you may click directly onto the "Variogram Analysis" tab at the top of the window

Once on the "Variogram Analysis" tab, proceed to click the "Load VTK File" button.

Once the file is loaded, you are able to configure the variogram settings as shown in the previous section, and click the "Start Variogram Analysis" button. One difference to note is that in the outputs, the final scales results of this analysis will **NOT BE SCALED TO THE DIMENSIONS OF THE VTK FILE** and instead are scaled to a value of 1 µm. It is recommended to calculate the scaled values of these results to ensure a correct solution. Results can be accessed by clicking the "Analysis Results" button, or used to produce spatial random fields by clicking "Continue"

How to… Create a Spatial Random Field

One of the main purposes of this toolkit is to produce a spatial random field (SRF). To access this feature, you may either go through the analysis process, or more simply, click the "Spatial Random Field" tab on the top of the window. No pre-requisite is needed to perform this action.

This toolkit uses a Gaussian distribution to produce a Gaussian random field.

Firstly, choose a name for your SRF outputs, and select the number of SRFs you wish to produce. The SRF will have a numerical denominator of _X which separates the individual SRFs with the same name.

Select a proportional variance – the difference of the maximum and minimum value from the mean, as a proportion of that mean (i.e. 0.50 = +- 50%).
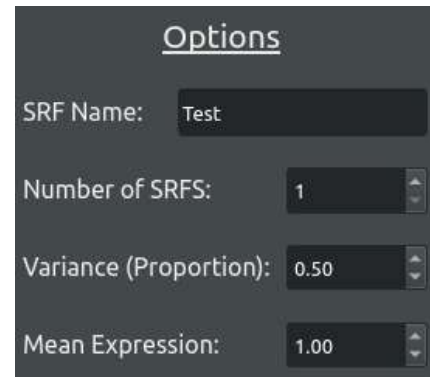
Select a mean expression for your output. This is the total field-average expression of the SRF.

Next, select a transversal and longitudinal correlation lengthscale. These can vary from anywhere between 1.00 and 100.00, and do not need to be the same. Alternatively, if you have performed a variogram analysis and have an ID loaded in, you can press the "Open Analysis Results" button to open the results, and find the final values of X and Y correlation lengthscales obtained by the analysis to enter into these fields.

Lastly, choose the dimensions of your SRF.

These tools are compatible with the Multi-Scale Cardiac Simulation Framework, developed by Michael Colman (available at github.com/michaelcolman) and as such, there are two checkboxes with the "full" and "part" cell sizes used in this framework. Checking either box will automatically fill the X, Y and Z dimension fields below.

Alternatively, you may enter any integer value you wish into those fields.
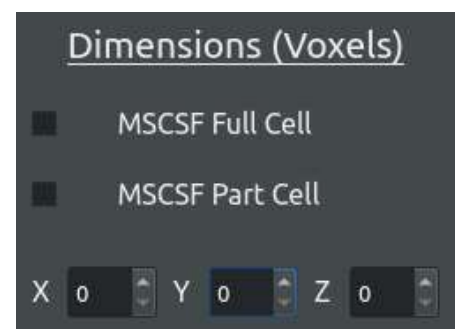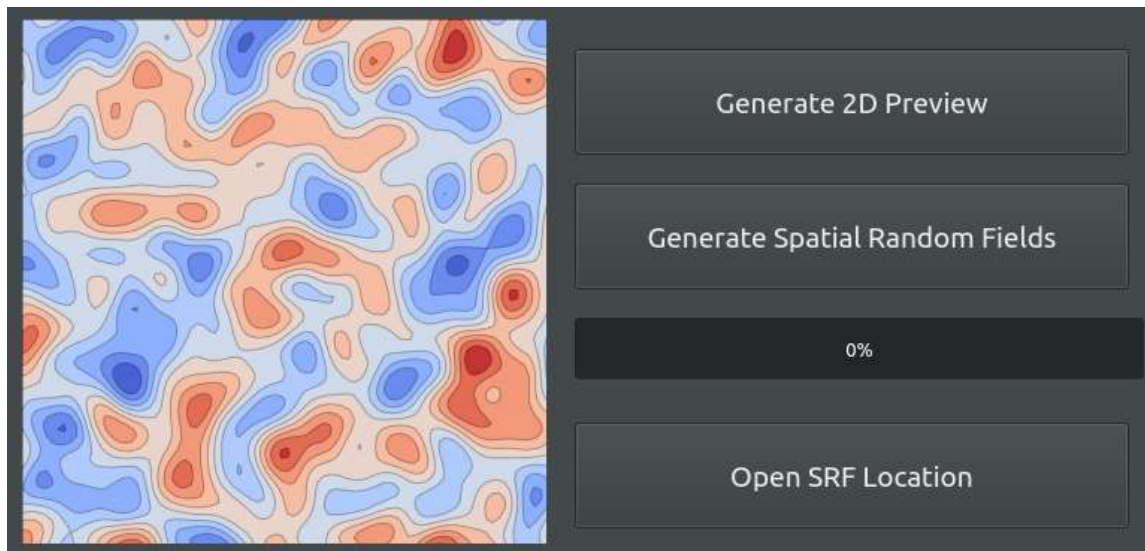
Once you have selected your settings there are two options:
The "Generate 2D Preview" button, quickly creates a 2D SRF and displays it on the left canvas, whereas the "Generate Spatial Random Fields" button, will begin the process of generating the random fields according to your specifications.

While either of these processes are ongoing, you will be unable to do anything else with the toolkit, however the progress bar will notify you of the current progress, and when it reaches 100%, the program will be active again.

The "Open SRF Location" button, will open the directory containing any produced SRFs in a file explorer.

How to… Understand the Outputs

VTK Files

.vtk files are spatial geometry files which start with a header containing information on the geometry dimensions, spacing, origin, and data type. These can be directly loaded into Paraview, geometry visualisation software that is free to use, to investigate these geometries.

Dat Files

.dat files are used to output data in matrices.

Histogram Files

Histogram files (a type of .dat) contain histogram information on the distribution of pixel intensity for an image slice, or a spatial random field.

Analysis/Variograms/Data_BS_X_BC_Y.dat

These files contain the variogram fitting procedure data for each combination of bin space (BS) and bin count (BC).
Column 0:   Bin Center
Column 1:   Bin Value

Analysis/Variograms/Data_X/Y.dat

These files contain the structured variogram fitting data for the X and Y axes. There is only one of each kind of these files as they are independent of bin space and bin count.
Column 0:   X/Y_S, an equivalent "Bin Center" derived from the structured mesh grid
Column 1:   Gamma_X/Y, an equivalent Bin Value.

Analysis/Variograms/Full_Results.dat

This file contains the full variogram fitting results from every iteration of the protocol.
Column 0:   ID
Column 1:   Bin Space
Column 2:   Bin No
Column 3:   Isotropic Variation
Column 4:   Isotropic Lengthscale (μm)
Column 5:   Tranverse Variation
Column 6:   Tranverse Lengthscale (μm)
Column 7:   Longitudinal Variation
Column 8:   Longitudinal Lengthscale (μm)

Analysis/Variograms/Final_Results.dat

This file contains the filtered, averaged and scaled variogram results.
These may not be the best way to interpret the results, and as such, the full results are included.

Column 0:   Result Type (Mean / Scaled)
Column 1:   Isotropic Variation
Column 2:   Isotropic Lengthscale (μm)
Column 3:   Tranverse Variation
Column 4:   Tranverse Lengthscale (μm)
Column 5:   Longitudinal Variation
Column 6:   Longitudinal Lengthscale (μm)

How to… Reset the Database

This program utilised a JSON database to save and load identifiers and associated configurational settings, and loads this on start up.

If you wish to clear the database, simply delete the file "HetKit_Database.json" and the program will generate a new one upon startup.

Tips

The variogram fitting procedure is tricky on smaller datasets.
Try slightly adjusting your configurational settings to make sure there are no background regions in your processed dataset. Alternatively, try changing the bin count and bin size settings.