# Data-Model Comparison

[File Structure](#_Toc473728802)

[R Setup](#_Toc473728803)

[Input Data](#_Toc473728804)

[Observations](#_Toc473728805)

[Model Data](#_Toc473728806)

[Mask](#_Toc473728807)

[Calculation Settings](#_Toc473728808)

[Scripts](#_Toc473728809)

[FE Mesh](#_Toc473728810)

[Sphere](#_Toc473728811)

["Egg" Anisotropy](#_Toc473728812)

[Smoothness - Range](#_Toc473728813)

[Output](#_Toc473728814)

[Running the Calculation](#_Toc473728815)

## File Structure

* Code
* Contains R code and also Matlab code for generating mesh, along with the output mesh files.
* Scripts
* Fit\_eg.R – example wrapper script that specifies all data locations and calculation options and then run the calculations and save the output.
* install\_packages.R installs the required packages
* a setwd path for convenience
* Observation\_data
* observations at discrete points
* Model\_data
* gridded model output, reformed as long vector
* mask.txt - giving location in vectorised grid of data points to be included in the calculation
* Output
* FE surface gridded onto 200x200 grid for plotting

Observation\_data, Model\_data and Output have separate sub-directories for each "analysis".

## R Setup

* Set the root path of the file tree
* setwd ("/home/bridge/fb7897/31\_Methods/04\_DMC")
* paths within the scripts are relative to this root
* Need to install all the necessary packages the first time the system is used.
* source("Scripts/install\_packages.R")
* this should prompt you for a CRAN mirror site and maybe to install the packages in a personal directory.

## Input Data

The data supplied in the example produce the results shown in Figure 7.3a of my thesis.

### Observations

Observations are listed in a text file with one row for each point

* Header line 'x y z std'
* Each row contains lon (degrees), lat(degrees), data value, uncertainty.
* Observation files are stored in Observation\_data directory

The data supplied here in the sub-directory P3+\_SST\_anom are from the PRISM3 extended dataset and are anomalies from the modern temperatures at the same locations. The various files lambda*\_<nn>.*txt have different values of uncertainty to test the impact on the results. Section 7.1 explains more about the PRISM sea surface temperature dataset, the different proxies used and how the uncertainties were made up from the confidence scores supplied by PRISM.

### Model Data

Model data is supplied as a vector of gridded output starting with the first column of data, so looping around lons first, then lats. This can be created using IDL file um2DMC.pro. This has options to take absolute temperatures or anomalies, also an option to sample the UM at discrete points as though it were observations.

The model data section has a line to define the resolution of the model grid – some examples are given.

The data supplied here are sea surface temperature anomalies between the Pliocene and modern climate model results. The 8 variants are the same model setup except for increasing values of atmospheric CO2 concentration. There are plots of each file in the directory.

#### Mask

Mask could be a land-sea mask or more generally, whatever points on the model input are to be used in the likelihood comparison. This is specified as a list of the vector indices of the required points and is created by um2krig.pro alongside the model data for land-sea masks or other masks could be fed to it at this point, e.g. ocean basin.

The mask is output at the end for checking, but is quite distorted by interpolation to and from FE grid at different resolutions, so is only useful for checking the mask is in the area it was expected to be. Fit\_eg.R does this at the point where Res is to be regridded, which works but has to be repeated uselessly within each loop of surface fitting. There is another version that does it at the more logical point at the start where the lon/lats of the FE grid are calculated, but the indices appear to have changed by the time they are used. Possible solution may be to do the lon/lat part and store a mask variable in the data frame, but not determine the indices until later on when they are needed.

Examples of masks applied can be seen in Figure 7.5.

## Calculation Settings

### Scripts

Scripts of the format of Fit\_eg.R are stored in the Scripts directory. The specification of input data and settings is all in the top section of the script.

### FE Mesh

The method uses an ellipsoid FE mesh generated separately by MATLAB. Although a sphere is a special case of an ellipsoid, it is generated separately.

#### Sphere

A unit spherical mesh is generated by the MATLAB file: Generate\_Mesh.m.

The sphere should not need changing as there is no benefit in using different sized spheres. To change the range of influence of data points, use the smoothness parameter.

* Sphere1\_K.csv
* Sphere1\_M.csv
* p.csv
* tri.csv

#### "Egg" Anisotropy

|  |  |
| --- | --- |
|  | The eccentricity of the ellipsoid can be varied as a sensitivity test. The point of this mesh is essentially to allow different "smoothness" values latitudinally and longitudinally. This is difficult to achieve directly in this mathematical construct.  The egg mesh is generated by MATLAB file: Generate\_Mesh\_egg.m. This has a parameter "stretch" which is the square of the equatorial radius. The default is 0.05, i.e. the equatorial radius is much smaller than the polar one (√0.05 = 0.2236).  Note the plot created by MATLAB is generated in lines 63-64 of distmeshsurface.m in the distmesh sub-directory.  MATLAB is called from R with the appropriate radius derived from “stretch” and produces the following files:   * Egg\_K.csv * Egg\_M.csv * p\_egg.csv * tri-egg.csv |

### Smoothness - Range

The smoothness parameter is the last argument to Prec\_from\_SPDE\_wrapper: default l=0.3.

## Output

The name of the output directory is formed automatically by Fit\_eg.R from the various imput parameters.

For each set of obervations read in, there is a 200x200 grid interpolated from the statistical model of those observations, for plotting purposes. Plots of this output are shown in Figure 7.4.

Results\_matrix.txt is the summary of the likelihood paramters for each combination of models and observations read in.

## Running the Calculation

Once the necessary packages have been set up, the code should all run from Fit\_eg.R. This should call Matlab to generate an egg mesh if specified, do all the calculations and save the output. There are 10 sets of observations and 8 models, so this takes about an hour on my clunky PC.

It is currently set to generate a sphere mesh, which doesn’t give very good answers as the call to Matlab wasn’t behaving, though it definitely did used to, so it should be something to do with my current computer setup. As the sphere never changes, it is not generated from the script, so the Matlab bit was skipped.