Table of Contents

STARBME Usage

After install STARBME, below main menu will show after activating the icon of STARBME.

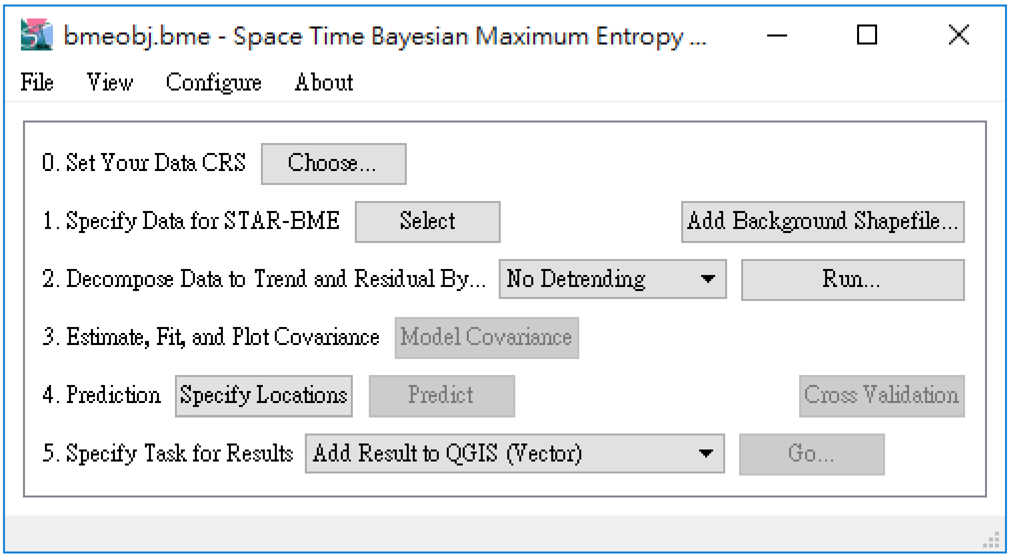


Figure 1 Main menu of STARBME

1. Set CRS(Coordinate Reference System)

Before specifying the data, the CRS (Coordinate Reference System) must be set to arrange your spatial relationship correctly. Here are some CRS settings for commonly use:

I. In Taiwan, the TWD97-TM2 reference system is the most commonly used CRS, and this CRS has a world-wide EPSG code: 3826. But this CRS is not included in QGIS1.X, so if you want to set your data with this CRS on QGIS 1.X, you may custom a new CRS with the following texts:

   +proj=tmerc +lat_0=0 +lon_0=121 +k=0.9999 +x_0=250000 +y_0=0 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs

II. If the spatial information of your data is just longitude and latitude, a popular CRS you may consider: the WGS84 reference system. It also has its own EPSG code, which is 4326.

2. Specify data

Your first step is to specify to STAR-BME the data that you want to work with. The input data can be categorized into two kinds:

- *Hard data*: These are sampled or observed values on different sites and at different time instances if you are processing space-time data.
- *Soft data*: These are uncertain observations for which you can quantify their uncertainty. These are also sampled or observed values on different points and at

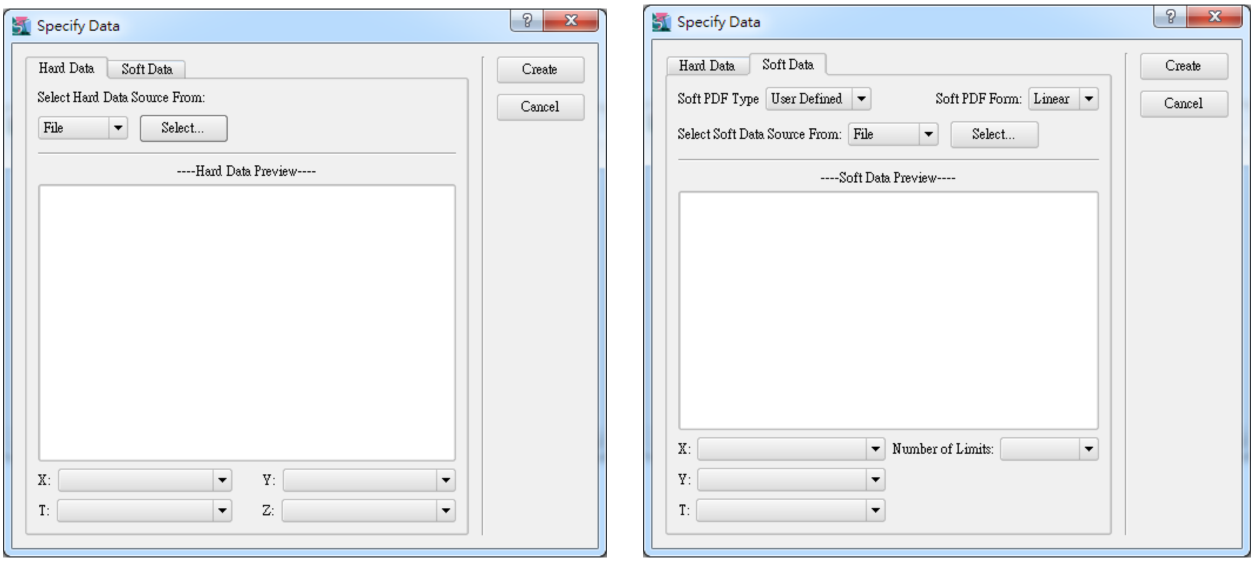
different time instances if you are processing space-time data.

There is more information about definitions of hard data and soft data in our references (Yu, et al., 2009; Yu and Wang, 2010).

Acceptable file types for your input data include the .csv and .shp file types. To run STAR-BME analytical tasks, you must specify at least one kind of data (that is, only hard data, only soft data, or both hard and soft data).

Start by specifying a coordinate reference system (CRS) by clicking the 'Choose' button next to the 'Set Your Data CRS' text area (Figure 1) of the main STAR-BME menu. This is a necessary step to specify the same CRS for your data and your map canvas.

Specify each kind of available data: First click on the 'Select' button next to the 'Specify Data for STAR-BME' text area (Figure 1) of the main STAR-BME menu. Then, work under the appropriate tab for hard or soft data in the new pop-up window (Figure 2.1).



(a) (b)

Figure 2.1

Inside the both data type selection tabs, you need to select data source. In the drop down menu, we have three choices, and after the drop down menu choosing, the 'Select' button would popup different actions. If the 'File' option was chosen, we suppose that the data source comes from text file (.txt) or comma-separated values (.csv). After clicking the 'Select' button, the 'Load File Dialog' window would ask to fill the file name first, delimiters and the start row number. If the 'Layer' option was chosen, the data source would be supposed from layers which QGIS preloaded. Therefore, the data source must be selected from the drop down list of layers. If the 'Shape File' option was chosen, there would be a file-loading window to load shape files (.shp) by clicking 'Select' button.

When you specify a hard data file, each observation record should be written in a different line or row of the input file. Each record contains the x-coordinate ('X'), y-coordinate ('Y'), time instance ('T', if time axis is necessary), and the value of the corresponding observation ('Z'). In any single file, the values of each one of those arguments must be consistently registered in the same column. Then, specify the column where each one of the 'X', 'Y', 'T', and 'Z' arguments is placed in the input file, as shown in Figure 2.1.a. Any two rows of input data must have different sets of x-coordinates, y-coordinates and time instances (if time axis is necessary).

In a similar manner, when you specify a soft data file, each observation record should be written in a different line or row of the input file. Each record must contain the x-coordinate ('X'), y-coordinate ('Y'), and time instance ('T', if time axis is necessary) of the corresponding observation. In any single file, the values of each one of those arguments must be consistently registered in the same column. Then, specify the column where each one of the 'X', 'Y', and 'T' arguments is placed in the input file, as shown in Figure 42.1.b. Any two rows of input data must have different sets of x-coordinates, y-coordinates and time instances (if time axis is necessary).
Note: For spatial-only studies, you must set the 'T' argument to 'None'.

Unlike hard data where you specify a true observed value, for soft data you must specify uncertainty about your observations in the form of probability density functions (PDFs). The present version of STAR-BME supports specification of two theoretical distributions, 'Gaussian' and 'Uniform'.

- If your soft data are normally distributed, select the 'Gaussian' soft data type, and specify the columns for the 'Mean' and the 'Variance' values in your input file.
- If your soft data follow uniform distribution (for example, if they range within an interval), then select the 'Uniform' soft data type, and specify the columns for the 'Lower bound' and the 'Upper bound' values in your input file.

Once you finish specifying your data, click the 'Create' button in the 'Specify Data' window. This action creates a temporary .bme object that STAR-BME uses in the data analysis (Figure 2.2).

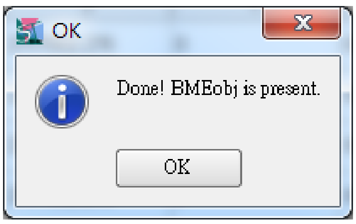


Figure 2.2

After specifying your input data and create the .bme object, your data are displayed on the QGIS main window as vector layers. You can use any QGIS tool to explore your data. If you have a background layer that you want to use when mapping your data, you can join it (if it is a shapefile (.shp)) by clicking the 'Add Background Shapefile' button in the main STAR-BME window. The background will be show in the QGIS main window with just outline shape. This outline vector may help you figure out the result of estimation.

3. Compute Trend and Residual From Data

In general, the values of space-time attributes have a more elaborate dependence in space and time than being simply dependent on a given distance in space or time. In geostatistical terms, this is expressed by saying that a spatiotemporal process is generally nonhomogeneous (spatial) and non-stationary (temporal)[1]. For the actual task of space-time prediction in a following step, it is necessary to work with a homogeneous and stationary process. To get there, we decompose the original process into a trend (also known as 'mean trend' or 'surface trend'; it is roughly a general average of the process and represents long-term variations) and residuals (homogeneous and stationary components of the process; they represent variation across space and time at the study scale of interest). This is known as the detrending step in the analysis. Following the prediction step at a later point, the trend is restored in the predicted values, thus performing effectively spatiotemporal prediction in any process regardless its homogeneity and stationarity status.

The current version of STAR-BME provides two methods to detrend a process, namely kernel smoothing and STMean estimation. In the main STAR-BME window, make a selection from the drop-down menu in step 2 of data decomposition.

- 'Kernel Smoothing' will show a related window pops up and prompts you to specify the spatial and temporal range parameters b_s and b_t, respectively (Figure 2.3). These are necessary for the kernel smoother. Select either a 'Gaussian' or a 'Quadratic' kernel from the window drop-down menu.
- 'STMean' will give you a convenient result without choosing any parameters. A spatiotemporal trend is estimated from your data by a preloaded model, so that user just need to specify the original data carefully, then the spatiotemporal detrend process will be completed automatically.

For any of the previous two options, click the 'Run' button in the main STAR-BME window to perform the trend estimation. A message pops us to update you when this estimation is complete. Alternatively, you might select the 'No Detrend' option in the step 2 menu of the main STAR-BME window. In this case, you assume that the spatiotemporal process represented by your input data is already a homogeneous and stationary process. This might be a valid assumption if you have prior information that your data exhibit no significant trends in their values across space and time.

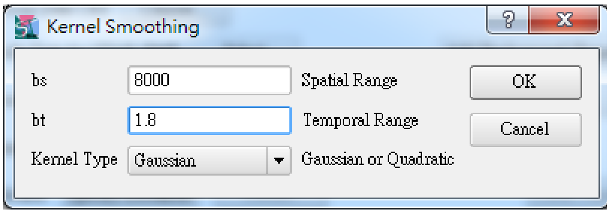


Figure 2.3

[1] In Statistics, the terms 'stationary' and 'homogeneous' are used in different contexts. Traditionally, Statistics has had a focus on spatial-only analysis. In Statistics, the term 'stationarity' is used for what is here referred to as 'spatial homogeneity', and there is no corresponding term for the temporal dimension.

4. Covariance Analysis

To predict your attribute at locations and time instances where you have no observations, you need to explore how the attribute values are correlated in space and time. You work with the detrended (or zero trend) residuals, and you perform this task in two steps: First, you examine sequentially all pairs of your data that are separated by a series of spatial and temporal distance lags. The covariance of the pair values at these lags produces a series of covariance values with distance, and this is an empirical measure of the similarity of attribute values. Then, you fit a suitable mathematical function to the empirical covariance to obtain a theoretical model that describes this behavior for data separated by any distance in space-time. This function is the covariance model, and you need this to proceed to the prediction task. In the following two steps you can see how to obtain a covariance model with STAR-BME.

4.1 Empirical Covariance Estimation

In the present step, STAR-BME estimates an empirical covariance from the relationship of pair-wise space-time points in different spatial and temporal lags. Intuitively, one expects that similarity among values decreases with distance and time form a point of origin. Also, one is primarily interested in the behavior close to the point of origin; estimating the covariance accurately at smaller lags is more important than at larger ones.

To start covariance analysis, push the 'Model Covariance' button on the main window of STAR-BME. A new window pops up titled 'Covariance Analysis'. Work first on the upper part of this window to estimate the empirical covariance from your detrended data (Figure 2.4). This estimation requires a set of parameters for space and one set for time. The required parameters are:

- *Distance limit*: Specifies the farthest distance in space or time that STAR-BME should look for data pairs. This is typically about half the longest distance within the area you work in. By default, STAR-BME considers the limit to be two-thirds of the largest space and time distances observed among all pair-wise points in space and time, respectively.
- *Number of lags*: Designates in how many distance brackets you wish to break down the distance limit. Since one cannot practically obtain the covariance from pair-wise points at an exact distance value, the covariance is estimated on a lag basis from all pairs that fall into each lag distance. The number of lags indicates how many times STAR-BME estimates the empirical covariance at an equal number of distances from the origin. It is recommended to design your analysis so that you have at least several lags. When you have only some tens of data points, use a smaller amount of lags so that each distance lag has enough pair-wise points to estimate their covariance; in larger data sets, this issue does not exist and you can use larger lag numbers. STAR-BME provides default 8 lags based on the number of pairs formed by your data.
- *Lag tolerance*: Indicates the distance on each side of a lag limit, within which pairs are assumed to belong to this lag. This can be typically equal to half of the distance between two consecutive lags. STAR-BME provides initial values on the basis of the initial lag numbers.

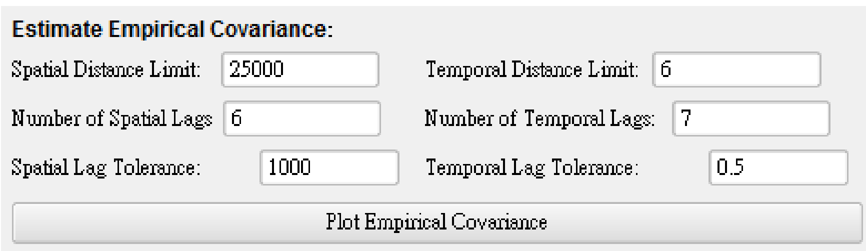
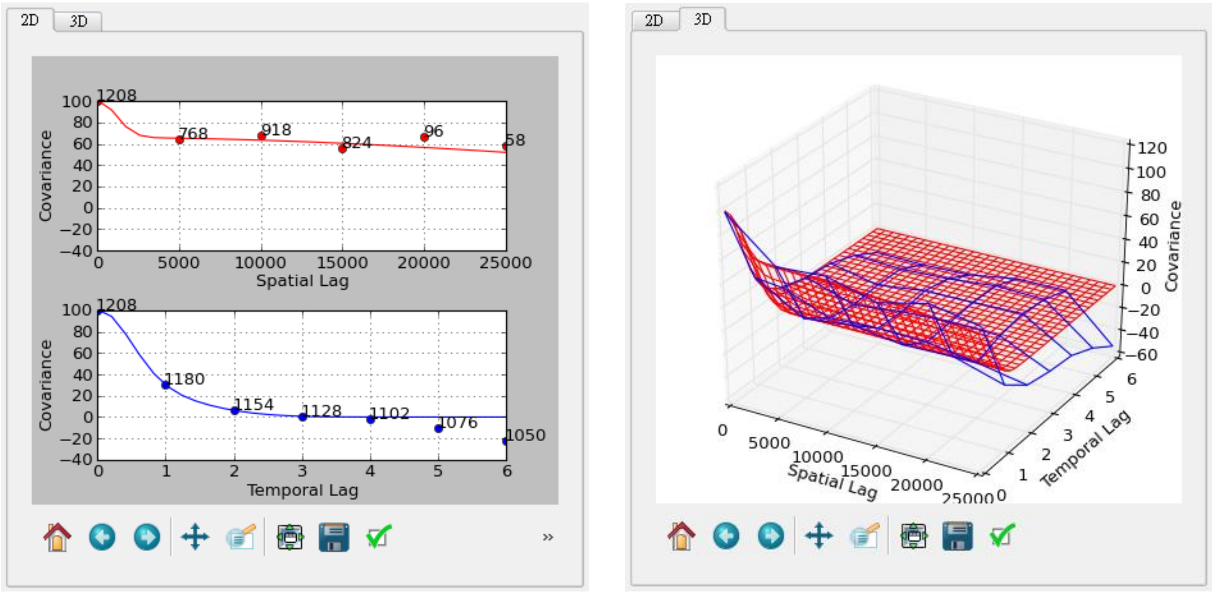


Figure 2.4

After you specify all of the above parameters, click on the 'Plot Empirical Covariance' button to produce plots of the empirical correlation analysis of your data

(Figure 2.5). In spatiotemporal studies, the 2-D views display the marginal covariances at distances t=0 (for the spatial covariance) and s=0 (for the temporal covariance). Spatiotemporal studies also offer 3-D views of the space-time covariance, where the horizontal axes are the spatial and temporal distances, and the vertical axis is the covariance value. Requesting to plot the empirical covariance also saves temporarily the empirical estimates in the .bme object.



(a) (b)

Figure 2.5

In the plots of Figure 2.5.a, the red and blue lines represent the marginal spatial (at t=0) and temporal (at s=0) covariance functions, respectively. Figure 2.5.b illustrates a 3-D surface where the grid nodes are space-time locations where the empirical spatiotemporal covariance is estimated.

4.2 Covariance Model Fitting

In the present step, STAR-BME fits a theoretical model to the empirical spatiotemporal covariance function that was estimated in the previous step. For this task, use the lower part of the 'Covariance Analysis' window under the title 'Fit Covariance Model'.

You can fit a model visually, or choose to perform an automated fit by means of one of the fitting methods provided by STAR-BME. A visual fitting might have the advantage that in some cases you can obtain a fit that feels generally better, compared to a strict parameter fitting process that primarily aims to satisfy a mathematical criterion. In any case, the fit quality is assessed by the Akaike Information Criterion (AIC) which is a number based on the model parameters and expresses how well the theoretical covariance fits the empirical estimates; smaller AIC values indicate a more accurate fit.

You can nest up to 3 different components in a covariance model to achieve a more accurate fit. Each component is a permissible covariance model form. In STAR-BME, you can choose among the following covariance forms in the form drop-down menus: 'Gaussian', 'Exponential', 'Spherical', 'Holecos' (for cosine hole effect), and 'Nugget' (for nugget effect). The nugget effect is a contribution to the model variance, and represents smaller-scale variations or measurement inaccuracies. The other model forms are each characterized by the following two parameters:

- *Sill* (or *scale*): The variance contribution of the component to the total variance of the process.
- *Range*: Specifies the spatial or temporal extent of the behavior described by the given form. Start by selecting the number of nested models in your model from the drop-down 'Nest Number' menu.

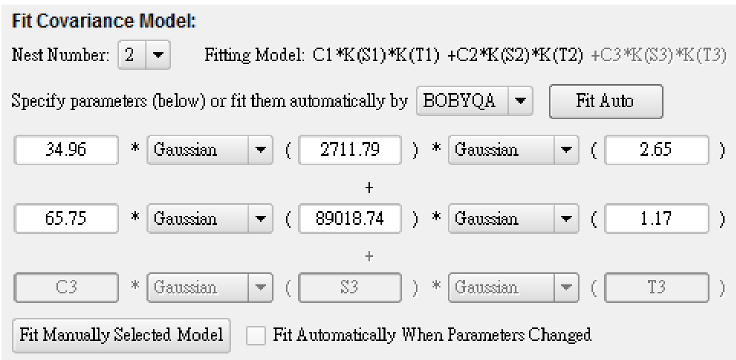


Figure 2.6

If you perform a visual model fit, then specify explicitly the parameters values for each model in the appropriate boxes, according to the fitting model representation in the window. Specifically, in an N-component nested model, the sum of the component variances C_i, i=1,…,N must be equal to the maximum covariance variance value C (assumed to be the process variance). In addition, each i-component has its own spatial and temporal ranges $K_{s,i}$ and $K_{t,i}$, respectively (Figure 2.6). The Concept of these ranges can be referred in papers (Yu, et al., 2009; Yu and Wang, 2010).

After you specify the model parameter values, push the 'Fit Covariance Model' button in the bottom of the screen. This action produces another pop-up window that shows you the model AIC value for the parameter values you specified (Figure 2.7), and also causes your theoretical model to display alongside the empirical covariance in the plots on the right hand side of the 'Covariance Analysis' window. You can perform this process repeatedly, try different model and form combinations, and continue until you achieve a fit that you deem satisfactory.

Figure 2.7

If you prefer an automated model fitting, then you can select between two different methods in the related drop-down menu, or use them both:

- Bound Optimization by Quadratic Approximation ('BOBYQA' option; based on weighted least squares)
- Particle Swarm Optimization ('PSO' option)

Each option pops up different windows after clicking the 'Auto Fit' button. In the BOBYQA method, you need to minimize the objective function. The corresponding PSO method window also shows the objective function, and includes another tab where you can set parameter values for the specific optimization method. For either method, click the 'Compute' button to perform the fitting. Once the fitting computations are complete, a 'fitting' window will show the AIC value of covariance model from the selected method, and your fitted theoretical model parameters will display alongside the empirical covariance in the plots on the right hand side of the 'Covariance Analysis' window. If you try to use both of the methods, it is recommended that use the PSO method first, to find an acceptable combination of parameters; then the BOBYQA method would help to optimize the combination by clicking 'Initial Guess From Last Fitting' button to copy the parameters into the pop-up window.

In general, exercise your good judgment and use your intuition to define the theoretical covariance model for your prediction. The empirical covariance that is computed according to the instructions in the previous subsection is but an estimate itself; thus, there is little meaning in attempting to overfit the empirical estimate. Also, small variations among different covariance models might have little or no effect in prediction. In overview, focus on selecting a covariance model that reproduces overall well the main characteristics of the empirical estimate.

Upon completing the covariance analysis, push the 'Ok' button at the bottom of the 'Covariance Analysis' window. STAR-BME confirms that it has a covariance model for the current analysis and takes you back to the main STAR-BME window.

5. Prediction

In this final step, you will use the result from the previous steps to predict your spatial-temporal value at those locations and time instances yet sampled. First you need to specify the space-time locations where you want to predict. In the step 4 of the STAR-BME window, click the 'Specify Locations' button. In the 'Prediction Locations' window, click the button 'Select Location Data Source' and specify the file source. Inside the drop-down menu, there are several options for you to specify the spatial or temporal positions to predict:

- The 'File' option allows you set the positions in text files (.txt, .csv, etc.).
- The 'Layer' option let you choose the layers preloading in QGIS, either polygons or points.
- The 'Shape File' option let you select those shape files (.shp) that haven't been loaded in QGIS.
- But when the 'Shape File (With Time Input)' option is selected, the shape file you select would be simply loaded, and you just need to set the maximum and minimum of Time T, and how many T's you want to predict.
- If the 'Grid Input' option is chosen, the spatial range and time intervals can be fully set by user, or click the 'Set By Data Boundary' button to get the spatial and temporal limits from data specified.

In most of options mentioned above, you must specify the column where each one of the 'X', 'Y', and 'T' arguments is placed in the input as needed to indicate which input columns contain the x-coordinate ('X'), y-coordinate ('Y'), and time instance ('T', if time axis is necessary), respectively. Eventually, push the 'Create' button to register the specified prediction locations with STAR-BME.

By completing the above process and returning to the main STAR-BME window, you can see that the 'Predict' button in step 4 is activated. Push the 'Predict' button to specify prediction settings and start the prediction task. Upon pushing the 'Predict' button, a 'Prediction' pop-up window appears (Figure 2.8). The 'Order' drop-down menu asks you to specify whether you want to assume a 'Zero Mean' process that corresponds to use detrended residuals, or a 'Constant Mean' process that applies when you skip the detrending process and assume that your data have a constant mean.

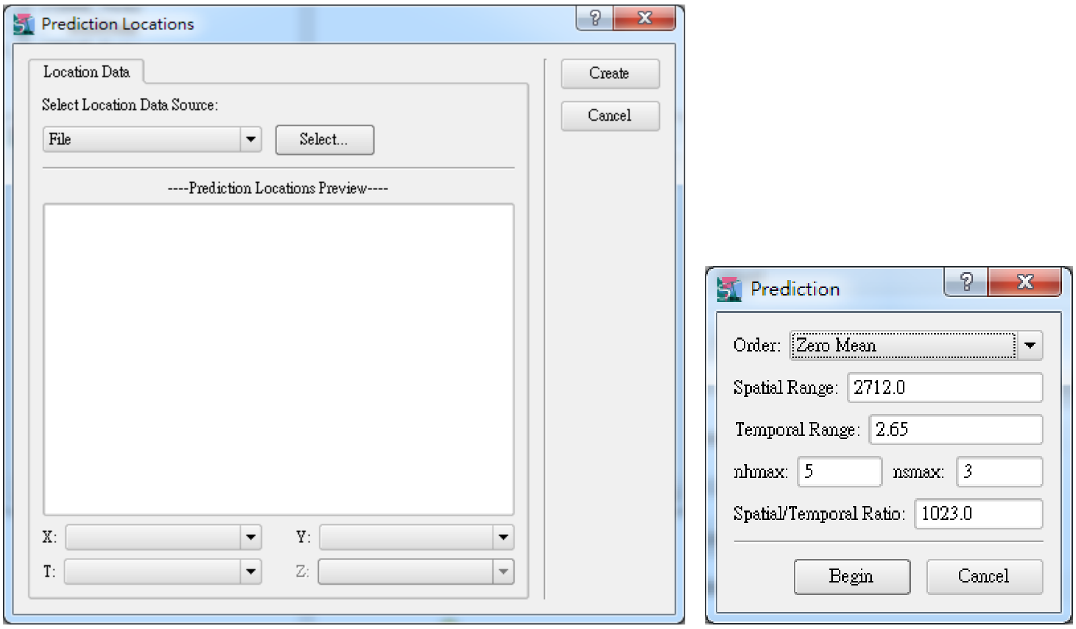


Figure 2.8

neighboring data that are considered. By means of the specified covariance model, STAR-BME looks for neighboring data that can affect the predicted value at the current prediction location. The following parameters guide the prediction process to select the desired amount of nearest neighbors to the current prediction location, and to regulate consumption of computational resources in this task.

- In the spatial and temporal range fields, specify the maximum spatial and temporal radii, respectively, around a prediction location within which to search for neighboring data. These values need to be comparable to the range values in your covariance model, so that prediction locations can correlate with observed data.
- In the ‘nhmax’ and ‘nsmax’ fields specify the maximum number of hard and soft data neighbors to consider, respectively. These parameters enable you to account for the spatial and temporal density of your observed data. In general, you can use up to a few soft data neighbors without considerable impact on the computational burden. Requesting a much higher number of soft data neighbors in the ‘nsmax’ field might also lead to numerical issues.

The spatial/temporal ratio is an important parameter in space-time studies that defines how the spatial and temporal dimensions are associated in the spatiotemporal continuum. This parameter provides the key about the space-time geometry that governs your analysis. STAR-BME provides an initial value that is based on your analysis general characteristics.

Push the ‘Begin’ button at the bottom of the ‘Prediction’ window to start the prediction process.

5.1 Cross Validation

The ‘Cross Validation’ button allows user to evaluate the covariance model parameters by using the root mean square error (RMSE). After specifying the prediction locations (but haven’t clicked the ‘Predict’ button), the ‘Cross Validation’ button can be clicked for the function.

After clicking the button, the ‘Cross Validation Dialog’ Window will pop up, and the Sample Type part will be the same with what type of data used. If the used data includes hard data and soft data, other options for only hard data or only soft data can be checked.

The number of Sample Size represents those sampled data can be estimated by other samples with the variance model. The sample size would not necessarily be the whole amount of samples, but if the size less than the amount of samples, the samples would be randomly picked for validation.

The Sample Boundary Clipped can be checked if user just wants to validate some specific spatial or temporal interval. If checked, the maximum and minimum of X, Y and T must be decided.

At the bottom of the Dialog, there are four buttons with various functions. The ‘Run’ button will start the cross validation process and pop up a window with result, which are the values included root mean square error, mean, standard deviation, median, minimum value, maximum value. The ‘View Spatial Result’ and ‘View Temporal Result’ buttons will show the validated sample distributed spatially or temporally with their separated error.

6. Output result

In the output process, there are several options can be selected.

- "Add Result to QGIS (Vector)" - Add result in vector data to QGIS layer.
- "Add Result to QGIS (Raster)" - Add result in raster data to QGIS layer.
- "Add Result to QGIS (Raster with Mask)" - Add result in raster data with Mask to QGIS layer.
- "Export Result to File" - Export result as delimited text files (.txt) or comma separated values text files (.csv).
- "Export Result to PNG" – Batching export result as images in PNG type (.png). The exportation would call the Layout Composer (An function of QGIS) and

user must arrange the layout of those images. For more detailed information of the Composer, you can check the page: http://docs.qgis.org/2.0/en/docs/user_manual/print_composer/print_composer.html