

Covariance Matrices

The PPCOV Suite

A GMDSI tutorial

by Rui Hugman and John Doherty



PUBLISHED BY

The National Centre for Groundwater Research and Training
C/O Flinders University
GPO Box 2100
Adelaide SA 5001
+61 8 8201 2193

DISCLAIMER

The National Centre for Groundwater Research and Training, Flinders University advises that the information in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice.

PREFACE

The Groundwater Modelling Decision Support Initiative (GMDSI) is an industry-funded and industry-aligned project focused on improving the role that groundwater modelling plays in supporting environmental management and decision-making.

Over the life of the project, GMDSI will produce a suite of tutorials. These are intended to assist modellers in setting up and using model-partner software in ways that support the decision-support imperatives of data assimilation and uncertainty quantification. Not only will they focus on software usage details. They will also suggest ways in which the ideas behind the software which they demonstrate can be put into practice in everyday, real-world modelling.

GMDSI tutorials are designed to be modular and independent of each other. Each tutorial addresses its own specific modelling topic. Hence there is no need to work through them in a pre-ordained sequence. That being said, they also complement each other. Many employ variations of the same synthetic case and are based on the same simulator (MODFLOW 6). Utility software from the PEST suite is used extensively to assist in model parameterization, objective function definition and general PEST/PEST++ setup. Some tutorials focus on the use of PEST and PEST++, while others focus on ancillary issues such as introducing transient recharge to a groundwater model and visualization of a model's grid, parameterization, and calculated states.

The authors of GMDSI tutorials do not claim that the workflows and methodologies that are described in these tutorials comprise the best approach to decision-support modelling. Their desire is to introduce modellers, and those who are interested in modelling, to concepts and tools that can improve the role that simulation plays in decision-support. Meanwhile, the workflows attempt to demonstrate the innovative and practical use of widely available, public domain and commonly used software in ways that do not require extensive modelling experience nor an extensive modelling skillset. However, users who are adept at programming can readily extend the workflows for more creative deployment in their own modelling contexts.

We thank and acknowledge our collaborators, and GMDSI project funders, for making these tutorials possible.

Rui Hugman

John Doherty

CONTENTS

1. Introduction	5
Software Executables	5
2. Covariance Matrices	5
3. PPCOV	6
3.1 Pilot Point Files (A Brief Recap)	6
3.2 Geostatistical Structure File	8
3.3 Run PPCOV	12
4. PPCOV_SVA	13
4.1 MKPPSTAT	14
4.2 PPCOV_SVA	15
5. Final Remarks	17

1. INTRODUCTION

This document is part of series of tutorials which demonstrate workflows for parameter estimation and uncertainty analysis with PEST/PEST++. These are not the only (or necessarily the best) workflows; their purpose is to take the reader through the basics of how to accomplish common tasks. The present document is a tutorial on how to construct covariance matrix files using PPCOV* utilities from the PEST Groundwater Utility suite. These covariance matrix files can be used for regularisation during calibration (see [GMDSI tutorial on calibration](#)), and to define parameter probability distributions prior to linear or nonlinear uncertainty analysis (see [GMDSI tutorial on linear uncertainty analysis](#)).

[The PEST roadmaps](#) are a useful complement to this series of tutorials. Going through **Roadmap 7: Pilot Point Workflow** prior to commencing this tutorial may be helpful. The roadmap provides some background concepts. The current tutorial provides a step-by-step demonstration of how to use PEST-suite software to implement these concepts.

Completed versions of all files generated during this tutorial can be found in the folder named *completed*. These are useful if you need to troubleshoot your own files, or if you wish to jump into the tutorial at a specific point. That being said, it is recommended that you at least read through the parts of the tutorial that you do not complete yourself.

Software Executables

A number of programs are used throughout this tutorial. To make completion of this tutorial easier, executable versions of these programs have been placed in relevant work folders; these are *.exe files. This is generally not recommended practice, for data files and executable files should be kept separate! Preferably, executable files (i.e. programs) should be located in a folder that is cited in your computer's PATH environment variable. Doing this allows them to be run from a command prompt that is open to any other folder without including the full path to these executables in the command to run them.

2. COVARIANCE MATRICES

A variance-covariance matrix, often referred to as a covariance matrix, is a square matrix that provides covariances between pairs of elements of a random vector. Diagonal elements of this matrix provide element variances of each element. Off-diagonal elements of the matrix provide covariances of each element with every other element. Any covariance matrix is symmetric and positive semi-definite.

Thus a covariance matrix of model parameters describes the variance of each parameter and the covariance of that parameter with that of every other parameter. For cases in which parameters represent points in a spatial (or temporal) parameter field (e.g. pilot points), it is reasonable to expect that parameters which are spatially (or temporally) closer together are more likely to be similar than those which are far apart. A covariance matrix provides a means with which to characterize this relationship.

When employing spatially (or temporally) inter-dependent parameters such as pilot-points, covariance matrices are applied for two purposes in the PEST ecosystem.

1. When undertaking inversion, they are used in regularisation.
2. When undertaking uncertainty analysis they are used for generation of parameter ensembles.

The PPCOV* suite of utility programs supplied with the PEST suite enables construction of covariance matrices based on one or a number of variograms. These matrices are housed in ASCII files (referred to as “matrix files” or MAT files) which can be read by PEST or other utilities of the PEST suite. As well as holding the array of numbers that comprises the matrix, a MAT file also lists the names of matrix rows and columns. The contents of a matrix can sometimes be related to the parameters that feature in a calibration, uncertainty analysis or data processing task through these names; on other occasions parameter-to-matrix-element linkages are provided through other mechanisms.

PEST documentation also refers to a “vector file”. This file holds a matrix that contains only a single column. Such a matrix is, in fact, a vector.

Matrix files are written by many of the utility programs of the PEST suite. As is apparent from their simple format, they can also be easily read and written by self-prepared programs. Matrix files which hold small matrices can be prepared using a text editor. The current tutorial demonstrates the use of the PPCOV* suite of programs to construct covariance matrix files for sets of pilot-point parameters.

3. PPCOV

PPCOV constructs covariance matrices based on spatially uniform variograms. PPCOV reads a “pilot point file” and a “geostatistical structure file. On the assumption that each pilot point represents a model parameter, it writes a covariance file for these parameters based on geostatistical structures pertaining to various sets (i.e. zones) of these points. Points within the pilot point file can belong to one such set, or many different sets, each potentially being characterised by a different structure. Pilot points within each set are assumed to be statistically independent of pilot points in other sets; thus covariances between pilot point parameters belonging to different groups are assumed to be zero.

The use of PPCOV requires:

1. a “pilot point file”, and
2. a geostatistical “structure file”.

The PPCOV utility constructs covariance matrices for two-dimensional pilot points. The PPCOV3D utility does the same for three-dimensional pilot points. Usage of the two utilities is very similar. The latter requires that pilot point files and geostatistical structures listed in the structure file be of the three-dimensional type; hence pilot points must be endowed with z coordinates as well as x and y coordinates, and geostatistical structures must specify spatial correlation in the vertical direction. The current tutorial demonstrates the use of PPCOV for two-dimensional pilot points. The same workflow can easily be adapted to use of PPCOV3D. See Part A and Part B of the PEST Groundwater Utilities Manual for more details on the use of PPCOV3D.

3.1 Pilot Point Files (A Brief Recap)

Pilot point files have been described in previous [GMDSI tutorials on the use of PLPROC; they are also](#) described in Part A of the PEST Groundwater Utilities Manual. In the tutorial folder you will find a file named *ppoints.dat*. This is a PLPROC list file which provides specifications for pilot points. As will be explained further on, this file structure differs slightly from that of a “pilot point file” used by utilities of the PPCOV* suite, as well as by other utilities from the PEST suite.

The pilot point file *ppoints.dat* is the same file that was used to parameterize hydraulic conductivity (Kx) and recharge rates (rch) of a groundwater model in [other tutorials from this series](#). The model has three layers. The same 523 pilot points are used to parameterize each layer. There are two

distinct geological “zones” in layer 1 of the groundwater model that was featured in this tutorial. There is no zonation in layers 2 and 3 of the model.

1 Open *ppoints.dat* in your text editor of choice and inspect it.

Data residing in pilot point files are arranged in columns. The first column contains point identifiers (i.e., pilot point names), the second and third columns contain point *x* and *y* coordinates. (In a three-dimensional pilot point file, the fourth column would contain *z* coordinates; however this is not the case here). Following these is a column containing zone numbers. The zone number column is followed by columns containing the parameter values assigned to pilot points. Utilities from the PPCOV* suite ignore values assigned to the last of these columns. However these utilities do use zone numbers from the fourth column to separate pilot points into sets.

In the PEST configuration used in other tutorials, *ppoints.dat* is read by PLPROC which then interpolates values from pilot points to the model grid. Column headers in *ppoints.dat* explain what is housed in each column. As previously mentioned, the fourth column contains zone numbers. As you can see, pilot points can pertain to either zone 1 or zone 4; these correspond to the two aforementioned geological zones in layer 1. The remaining columns house values for hydraulic conductivity in layers 1 to 3 (*kx1*, *kx2* and *kx3*) and recharge rate in layer 1 (*rch*).

Unfortunately, utilities from the PPCOV* suite require that some alterations be made to *ppoints.dat* before they can generate covariance matrices based on the contents of this file. In particular:

- A pilot point file cannot contain column headers.
- As previously mentioned, covariance between pilot point parameters from different sets (i.e. zones) are assumed to be zero. However, in our case, no zones are considered in layer 2 and 3; therefore covariances between pilot point parameters pertaining to *kx2* and *kx3* cannot be assumed to be zero.

We need to prepare two pilot point files that express zonation as it prevails within our model. One pilot point file that contains pilot points grouped into zones (for pilot point parameters in model layer 1); and a second pilot point file in which all pilot points belong to the same zone (for pilot point parameters in model layers 2 and 3).

Let us start by preparing two pilot point files for use with the PPCOV* suite.

- 2 Make a copy of *ppoints.dat*. Name the copy *ppoints1.dat*.
- 3 In *ppoints1.dat*, delete the first line which contains the column headers. Pilot points in this file are still assigned to either zone 1 or zone 4. Therefore, we will use this “pilot point file” for pilot point parameters in layer 1. Save the file.
- 4 Make a copy of *ppoints1.dat*. Name the copy *ppoints23.dat*.
- 5 In *ppoints23.dat*, replace the values in the fourth column (the zone number column) so that they are all the same. It does not matter what value you assign, as long as it is an integer, and that all rows feature this same integer. For the purposes of this tutorial, replace every value in the fourth column with “23”. (Twenty-three was chosen arbitrarily to represent layer 2 and 3). This is now the “pilot point file” pertaining to pilot point parameters in layers 2 and 3. Save the file.

We now have two pilot point files. Although these contain pilot points with the same spatial coordinates, they correspond to different zonation. For the purposes of this tutorial, the values contained in columns to the right of the zone (4th) column are irrelevant; they are ignored by utilities of the PPCOV* suite.

3.2 Geostatistical Structure File

Part A and Part B of the PEST Groundwater Utilities documentation describe the structure file format. They also provide some background on relevant geostatistical concepts. A structure file stores variables that specify the properties of one or more variograms assigned to one of more “geostatistical structures”. Each such geostatistical structure can contain up to 5 nested variograms and a nugget. Each such structure is normally (but not necessarily) assigned to a particular zone within the model domain.

Use of a geostatistical structure to characterise the spatial variation of a hydraulic property assumes that values taken by that property are spatially correlated, and that the degree of correlation between values at two different points is dependent only on their separation (i.e. how far apart they are). Furthermore, it is assumed that this inter-point distance-dependence can be described by one or more nested variograms; an optional uncorrelated component (a “nugget”) of the hydraulic property field can also be represented.

Let us construct a geostatistical structure file. This file must contain structures for each of the “zones” within the model domain. These are:

- zone 1 and zone 4 in layer 1,
- zone 2 in layer 2 and
- zone 3 in layer3.

Let us get started.

6 Open a new blank text file in your preferred text editor. Name this file *struct.dat*.

A structure file is made up of two types of segments: “structure segments” and “variogram segments”. Each of these segments must have unique names. A structure segment must end with the words END STRUCTURE; a variogram segment must end with the words END VARIOGRAM.

A “structure segment” specifies the geostatistical components which characterize spatial variability of a given family of pilot-points. It can reference one or several “variogram segments”.

A “variogram segment” provides specifications for a variogram referenced in a structure segment.

7 Recall that we have three model layers. Our geological expert knowledge informs us that each layer has a distinct geostatistical structure. Furthermore, layer 1 is divided into two zones. Therefore we will need to create four “structure segments” (as each set of parameters is characterized by different geostatistical properties) and four “variogram segments” (as each structure references only 1 variogram).

8 Type the following into *struct.dat*.

```
STRUCTURE zone1
TRANSFORM log
NUMVARIOGRAM 1
VARIOGRAM var1 1
END STRUCTURE
```

9 The first line initiates a structure segment and assigned it the name *zone1*.

10 The second line states that the geostatistical structure pertains to the log (to base 10) of the parameter and not its native value.

If the variograms and nugget which comprise a geostatistical structure pertain to the native value of a parameter, then TRANSFORM should be set to “none”. However if they pertain to the log (to base

10) of a hydraulic property, then TRANSFORM must be provided as “log”. In our case (as you may recall from the [GMDSI calibration tutorial](#)), all pilot point parameters are log-transformed in our PEST setup.

11 The third line specifies how many variograms contribute to the structure. In our case we will use only one. In the groundwater context it is rare that spatial variability is characterised in enough detail to require more than one variogram for its specification.

12 The fourth line cites the name of a variogram (*var1*) and the contribution made to the overall structure by that variogram. The sum of the contributions made by all variograms, plus the nugget, is equal to the sill of the nested variogram comprising the structure. The sill is equivalent to parameter variance; variance is the square of standard deviation. However, it is often useful to set the sill to 1. When undertaking uncertainty analysis, this allows us to use the same covariance matrix file for various parameter types that share the same spatial distribution (i.e. the same pilot points and zonation) but have different variances; so-called “variance multipliers” are used to accommodate variabilities of different parameter types which employ the same covariance matrix. For example, in the present case this would apply for *kx1* and *rch*. (The reasons for this become apparent in the [GMDSI tutorial on linear uncertainty analysis](#)). Note however that, at the time of writing, PEST does not allow the use of variance multipliers on covariance matrices that are used for regularisation purposes.

You may be wondering what the *var1* variogram is. It is the name of a variogram segment which we still need to specify. Let us do that now.

13 Add the text below to *struct.dat*.

```
VARIOGRAM var1
VARTYPE 2
BEARING 90
A 3000
ANISOTROPY 2
END VARIOGRAM
```

14 The first new line initiates a variogram segment and assigns it the name *var1*. This is the variogram referenced by the structure segment named *zone1*.

15 VARTYPE specifies the type of variogram; this must be supplied as either “1”, “2”, “3” or “4” indicating a spherical, exponential, Gaussian or power variogram respectively. We have chosen exponential.

See Part A of the PEST Groundwater Utilities Manual for the definition of each of these variogram types.

16 BEARING specifies the angle (in degrees) between north and the axis of anisotropy of the random field characterised by the variogram. As we have specified “90”, the axis of anisotropy will be east west. This is the direction to which the value of “a” applies; the value of “a” in the orthogonal direction is determined by dividing by the value of ANISOTROPY.

17 The “A” keyword pertains to the “a” variable appearing in the first three variogram type equations. It is proportional to the “range” of the variogram (see Part A of the manual for the Groundwater Utilities for details). A little care must be taken in selection of an appropriate value for the parameter “a”.

In general, depending on the geometry of a study area and the number of pilot points used in the parameterisation process, a variogram range that is roughly equal to 2 to 3 times the average distance between pilot points is suitable for most occasions. For the spherical variogram (type “1”) the range

is approximately equal to “a”. For the exponential (type “2”) and Gaussian (type “3”) variograms the range is approximately equal to 3a.

As we are using an exponential variogram, “a” corresponds to roughly a third of the variogram “range”. Larger values will result in a “smoother”, more homogenous parameter field. Lower values will result in more “granular” parameter fields. In our case an a of 3000 corresponds to a range that is roughly 1.2 times the average distance between pilot points.

18 Finally, the ANISOTROPY keyword specifies the anisotropy ratio. By assigning an anisotropy of 2 the variogram “range” along the axis of anisotropy will be twice as large as the range at a 90-degree angle. In the present case the latter range is obtained from the former range through division by 2.

19 Your *struct.dat* file should look something like the following:

```
STRUCTURE zone1
TRANSFORM log
NUMVARIOGRAM 1
VARIOGRAM var1 1
END STRUCTURE

VARIOGRAM var1
VARTYPE 2
BEARING 90
A 3000
ANISOTROPY 2
END VARIOGRAM
```

20 Go ahead and add in “structure segments” for zone 4 in layer 1. Name the segment *zone4*. Create a corresponding variogram segment named *var4*.

21 Set BEARING, ANISOTROPY and A in *var4* to 0, 1.2 and 1000 respectively.

22 Lastly, add structure segments for the zones in layers 2 and 3. Name them *zone2* and *zone3* respectively. Create corresponding “variogram segments” named *var2* and *var3*.

23 Set BEARING, ANISOTROPY and A in *var2* to 45, 1.5 and 2000 respectively.

24 Set BEARING, ANISOTROPY and A in *var3* to 111, 3 and 5000 respectively.

25 Now we have defined all four geostatistical structures. The values of anisotropy, bearing and range for these structures are different. Your *struct.dat* file should look something like this:

```

#Zone 1
STRUCTURE zone1
TRANSFORM log
NUMVARIOGRAM 1
VARIOGRAM var1 1
END STRUCTURE

VARIOGRAM var1
VARTYPE 2
BEARING 90
A 3000
ANISOTROPY 2
END VARIOGRAM

#Zone 4
STRUCTURE zone4
TRANSFORM log
NUMVARIOGRAM 1
VARIOGRAM var1 1
END STRUCTURE

VARIOGRAM var1
VARTYPE 2
BEARING 0
A 1000
ANISOTROPY 1.2
END VARIOGRAM

# Zone2
STRUCTURE zone2
TRANSFORM log
NUMVARIOGRAM 1
VARIOGRAM var2 1
END STRUCTURE

VARIOGRAM var2
VARTYPE 2
BEARING 45
A 2000
ANISOTROPY 1.5
END VARIOGRAM

# Zone 3
STRUCTURE zone3
TRANSFORM log
NUMVARIOGRAM 1
VARIOGRAM var3 1
END STRUCTURE

VARIOGRAM var3
VARTYPE 2
BEARING 111
A 5000
ANISOTROPY 3
END VARIOGRAM

```

We are now ready to use PPCOV to construct the covariance matrix files.

3.3 Run PPCOV

Now that we have prepared pilot point files and a geostatistical structure file, we are ready to run PPCOV to construct covariance matrices of the pilot point parameters. Recall that we are using the same set of 2D pilot points for the three model layers. Each layer has a different geostatistical structure and zonation. Therefore, we wish to create a parameter covariance matrix for each layer.

We shall start with layer 1, which has two zones (zone 1 and zone 4).

26 Open a command line window in your working folder, type the following and then press <enter>:

```
ppcov
```

27 You will be prompted for the name of a pilot point file. Reply with the name of the pilot point file we prepared in step 2 (*ppoints1.dat*) and press <enter>:

```
Enter name of pilot point file: ppoints1.dat
```

28 You will be prompted for a minimum allowable separation. Type 0 and press <enter>: That is the number zero, not the letter "o".

```
Enter minimum allowable separation for points in same zone: 0
```

29 You will be prompted for the name of the structure file. Type the name of the file you previously prepared (*struct.dat*) and press <enter>:

```
Enter name of structure file: struct.dat
```

Next you are going to be prompted for the structure names in *struct.dat* that should be associated with each zone number in the fourth column of *ppoints1.dat*.

30 First you will be prompted for the structure to use for pilot point zone 1. Recall that we named the structure segment for zone 1 as *zone1*. Respond accordingly and press <enter>:

```
Enter structure to use for pilot point zone 1: zone1
```

31 And then the same for zone 4:

```
Enter structure to use for pilot point zone 4: zone4
```

32 Next, you will be prompted for the name of the matrix file that PPCOV must write. This file will house the covariance matrix for parameters belonging to layer 1. So let's name it *cov1.mat*. Reply accordingly and press <enter>:

```
Enter name for output matrix file: cov1.mat
```

33 Finally, you will be prompted for a prefix for parameter names. This can be a useful device for associating covariance matrix elements with parameter names featured in a PEST dataset. However, it is not a requirement as these associations can be made in other ways as well. Simply reply by pressing <enter>:

34 If all went well, your screen should display the following:

```
Filling covariance matrix....  
- file cov1.mat written ok.
```

```
Warning: in any future processing of this covariance matrix, sensitivities  
for parameters with a log-variogram must be taken with respect to the log  
of the parameters.
```

35 Repeat steps 26 to 33 for layer 2 and for layer 3. When prompted for the structure name (step 30) respond with *zone2* and *zone3*, respectively. Name the output matrix files (step 31) *cov2.mat* and *cov3.mat*, respectively.

By now you should have three covariance matrix files (*cov1.mat*, *cov2.mat* and *cov3.mat*). These files can now be used to assign Tikhonov regularisation to a PEST control file, or to characterize prior parameter uncertainty. You can inspect them in a text editor if you wish. Although useful, they are not particularly exciting. The first row specifies the dimensions of the matrix. If you scroll down to the bottom of the file you will see the list of parameter names (in this case, the pilot point names). If you compare *cov1.mat* to *cov2.mat* or *cov3.mat*, you will notice that some of the elements of *cov1.mat* are zero. As previously mentioned, this is due to the statistical independence of pilot point parameters which belong to different zones.

If you are interested in seeing a pictorial representation of these matrices the MAT2SRF utility supplied with PEST can be used to convert these matrices to SURFER grid files, which can then be contoured and coloured in SURFER or other GIS software such as QGIS.

36 Return to the command line window, type the following command and press <enter>:

```
mat2srf cov1.mat cov1.grd
```

If you inspect your working folder you will see a new file named *cov1.grd*. This is a SURFER grid file. If you wish, you can import it into your GIS visualisation software of choice (i.e. SURFER, QGIS, etc.) and generate a pictorial representation such as is shown in Figure 1.

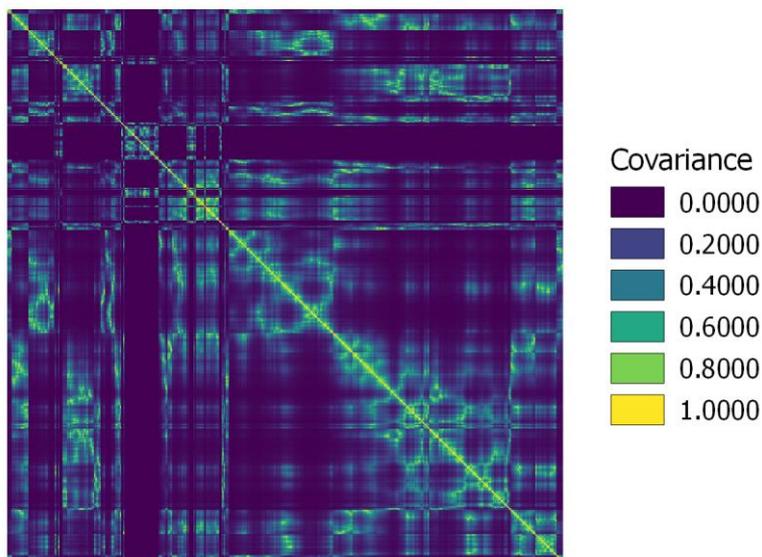


Figure 1 - Pictorial representation of the covariance matrix stored in *cov1.mat*. Brighter colours denote larger covariance between parameter pairs.

Files generated up to this point can be found in the folder *completed\t1-ppcov*.

4. PPCOV_SVA

A more sophisticated approach to construction of a covariance matrix is offered by the PPCOV_SVA and PPCOV3D_SVA utility programs. (SVA stands for “spatially varying anisotropy”). This approach is useful where the spatial density of pilot points varies throughout a model domain (as it usually does). PPCOV_SVA and PPCOV3D_SVA build a covariance matrix whose correlation range reflects local pilot point separations; the spatial correlation length rises where pilot points are locally far apart

and decreases where they are closer together. This encourages the emergence of heterogeneity at the same scale as that at which observations comprising a calibration dataset can support it (provided pilot point spatial density reflects observation spatial density). It can also allow the alignment of this heterogeneity to vary with location. This can be useful when calibrating a model whose domain follows a sinuous feature such as an alluvial valley (see for example the [GMDSI Biggenden Worked Example](#)).

In a way, PPCOV_SVA and PPCOV3D_SVA are easier to set up than their homogeneous counterparts. Unlike the later, the former do not require that the user provide a geostatistical structure file. Instead, the MKPPSTAT utility (from the Groundwater Utility suite) creates an input file for PPCOV_SVA. MKPPSTAT3D does the same for PPCOV3D_SVA. Both of these programs read pilot point files. They evaluate the local spatial density of pilot points at the location of each pilot point. They then write a file which provides variogram specifications for each pilot point; these specifications reflect the proximity (or otherwise) of neighbouring pilot points.

Construction of covariance matrix files using PPCOV_SVA or PPCOV3D_SVA requires only that a user provide pilot point files such as those prepared in Chapter 3.1 of this document

4.1 MKPPSTAT

37 Return to the command line window (or open a new one if you closed it), type the following and press <enter>:

```
mkppstat
```

38 You will be prompted for the name of a pilot point file. Let us start with pilot points in layer 1. Reply as below and press <enter>:

```
Enter name of pilot point file: ppoints1.dat
```

39 You should see a message saying that specifications of 523 pilot points were read from this file. You are then prompted to provide the number of these to use when computing the local average inter-pilot-point separation. Let us use the average distance to the closest 10 pilot points. Reply as below and press <enter> (user-provided responses are marked in **bold**):

```
Enter name of pilot point file: ppoints1.dat  
- data for 523 pilot points read from pilot point file ppoints1.dat
```

```
Enter no. of pilot points to compute local ave. pp. sepn.: 10
```

40 Next you will be prompted for a factor by which to multiply the average distance between pilot points. The outcome of this multiplication is the value of the variogram “a” for each pilot point.

41 As was discussed previously, a variogram range that is roughly equal to 2 to 3 times the local separation between pilot points is suitable for most occasions. For the spherical variogram (type “1”) the range is approximately equal to “a”. For the exponential (type “2”) and Gaussian (type “3”) variograms the range is approximately equal to 3a. We will be using an exponential variogram. A factor of 1.5 is often reasonable. Reply as below and press <enter>:

```
Enter factor of ave. separation for local variogram "a" value: 1.5
```

42 Finally you will be prompted for the name of a “statistical specification file” which MKPPSTAT must write. Reply as below and press <enter>:

```
Enter name for pilot point statistical spec. file: layer1.dat
```

43 You should see a message saying that *layer1.dat* was written OK.

44 Repeat steps 37 to 42 for *ppoints23.dat*. Name the respective statistical specification file *layer2.dat*.

45 Make a copy of *layer2.dat*. Name the copy *layer3.dat*.

46 Inspect *layer1.dat*, *layer2.dat* and *layer3.dat* in a text editor, or spreadsheet software.

As you can see, the pilot points statistical specification file written by MKPPSTAT assigns the same nugget, sill, anisotropy and bearing values to all pilot points, these values being 0.0, 1.0, 1.0 and 0.0 respectively. However the variogram “a” value (proportional to its range) is calculated by MKPPSTAT as being inversely proportional to local pilot point spatial density. It is thus smaller for pilot points whose nearest neighbours are close, and greater for pilot points whose nearest neighbours are further away.

So we now have statistical specification files for each layer. These files account for variations in the spatial density of pilot point emplacement. However, anisotropy is still homogenous throughout the model domain. If we wish to have spatially varying anisotropy, we must add it ourselves. This task is easy to accomplish in any spreadsheet software, or even most text editors (or even better, programmatically or in a GIS) by replacing values in the *hanis* (anisotropy) and *bearing* (bearing) columns of the statistical specification files written by MKPPSTAT.

47 In *layer1.dat*, replace values in the *hanis* columns that pertain to pilot points in zone 1 with 1.5. This maintains a north-south anisotropy axis with an anisotropy value of 1.5.

48 Still in *layer1.dat*, replace values in the *hanis* and *bearing* columns that pertain to pilot points in zone 4 with: 2 and 90, respectively. Thus in zone 4 hydraulic properties are more continuous in the east-west direction than in the north-south direction; meanwhile the anisotropy ratio is greater than in zone 1.

49 In *layer2.dat*, replace values in the *hanis* and *bearing* columns with 1.8 and 45.

50 Leave *layer3.dat* as it is.

As is evident, it is a relatively simple matter to assign spatially varying anisotropy as a function of pilot point location in ways that are most appropriate for a particular modelling context.

4.2 PPCOV_SVA

Once we have the necessary pilot point and statistical specification files in place, we are ready to use PPCOV_SVA to construct covariance matrices which reflect spatially varying parameter correlations and variabilities.

51 Return to the command line window (or open a new one if you closed it), type the following and press <enter>:

```
ppcov_sva
```

52 Let us start with the covariance matrix for pilot point parameters of layer 1. At the prompt, provide the relevant statistical specification file. Reply as below and press <enter>:

```
Enter name of pilot points statistical specs file: layer1.dat
```

53 By default, MKPPSTAT adds column headers to the statistical specification file which it writes. At the next prompt, respond affirmatively when asked if the first line should be skipped. Once you have done this, you should see a message stating that data for 523 pilot points have been read.

```
Skip a line at the top of this file? [y/n]: y
```

54 As in step 28, respond to the next prompt for the minimum allowable distance between pilot points with zero. Reply as below and press <enter>:

```
Enter minimum allowable separation for points in same zone: 0
```

55 As previously mentioned, we will use an exponential variogram. Responds with the letter “x” as shown below and press <enter>:

```
Is overall variogram spherical, exponential or Gaussian? [s/x/g]: x
```

56 At the next prompt provide the name of the covariance matrix file which PPCOV_SVA must write. Name it *covsva1.mat* as shown below.

```
Enter name for output matrix file: covsva1.mat
```

57 Finally, as previously, do not ask PPCOV_SVA to add a prefix to parameter names. Simply press <enter>. (If your curiosity commands you to see what happens if you ignore this suggestion, the rest of this tutorial will not be impacted.)

```
Enter pilot point prefix for parameter name (<Enter> if none):<enter>
```

58 If all went well, you should see the following message.

```
Filling covariance matrix....  
Using SVD to assure positive definiteness of matrix....  
- file covsval.mat written ok.
```

59 Inspect your working folder. You should see a new covariance matrix file *covsva1.mat*.

60 Repeat steps 51 to 57 for pilot point file *ppoints23.dat* with the statistical specification file for each layer (*layer2.dat* and *layer3.dat*). Name the resulting covariance matrix files *covsva2.mat* and *covsva3.mat*.

That is all there is to it. Again, if you wish to display the covariance matrices pictorially use MAT2SRF. Figure 2 displays the covariance matrix for pilot points in layer 1 generated using PPCOV (left) and PPCOV_SVA (right). The difference between the two is largely due to different variogram range (“a”) which was used. The smaller range which we used for the spatially varying anisotropy results in lower covariance between pilot point parameters. (This is not a characteristic of spatially varying anisotropy! Merely a result of the values of “a” which were used throughout the tutorial).

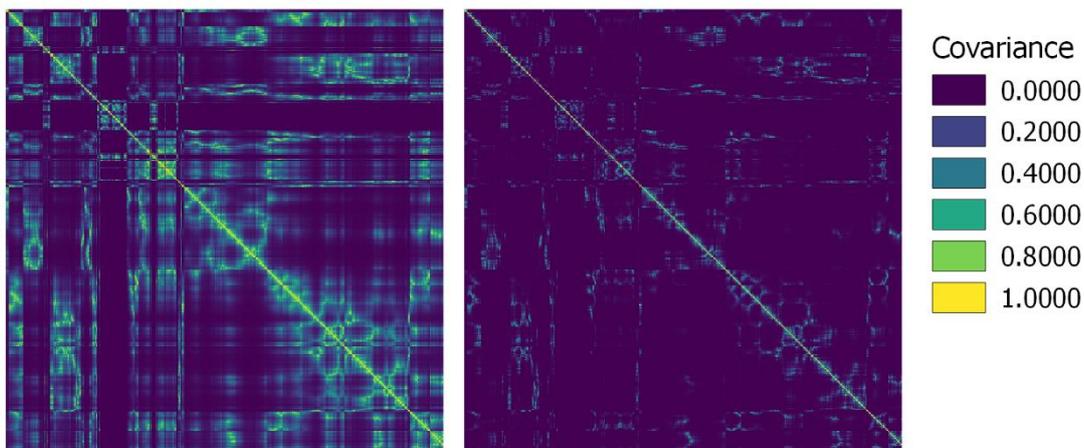


Figure 2 - Pictorial representation of covariance matrices stored in (left) *cov1.mat* and (right) *covsva1.mat*.

Files generated up to this point can be found in the folder *completed\t1-ppcov_sva*.

5. FINAL REMARKS

As has been mentioned, covariance matrix files play two roles in the PEST ecosystem. These are:

1. ensuring parameter reasonableness during history-matching and
2. characterizing the statistical distribution of parameters during uncertainty analysis.

In the former case, these files are referenced in PEST control files alongside the names of regularisation observation groups. In the second case they are cited within “covariance matrix” blocks of parameter uncertainty files. For examples of their application see the GMDSI tutorials on [calibration](#) and [linear uncertainty analysis](#).

Covariance matrix files used in other GMDSI tutorials were constructed in this manner using PPCOV_SVA (although the spatial distribution of anisotropy was defined with some more complex patterns just for fun). The files generated as part of the current tutorial can be used interchangeably with those supplied with other tutorials should you wish to explore the effect that different variograms or settings may have. (Hint: the covariance matrix files supplied with the [calibration](#) and [linear uncertainty analysis](#) tutorials are the same as those used to generate “reality”). Most of the steps taken here can be automated using batch files. Alternatively it is relatively simple to accomplish all of these tasks with simple scripts in programming languages such as Python.



gmdsi.org

CRICOS NO. 00114A



NATIONAL CENTRE FOR
GROUNDWATER
RESEARCH AND TRAINING

RioTinto