Non-Gaussian Bayesian Geostatistical Modelling

M.B. Palacios and M.F.J. Steel

The code to obtain the results in the paper *Non-Gaussian Bayesian Geostatistical Modelling* has been programmed in Fortran90. The main folder `NonGaussian` contains the source files, the executable code, the input files and the output files, organised in the following subfolders:

- Programs: source files in Fortran 90
- Debug: executable code
- Datafiles: data sets and parameter files (INPUT)
- Results: files with the results created from the code (OUTPUT)
- Readme: documents with information about the code

1 Programs

The folder Programs contains the main code called `GLGmodel.f90`, the common subroutines for Gaussian and Non-Gaussian models, the subroutines to make calculations separately for the GLG model and the Gaussian model in folders `NGaussian` and `Gaussian` respectively. Note that the names for the Gaussian subroutines begin with the letter g.

Four different parts can be distinguished in the main code `GLGmodel.f90`.

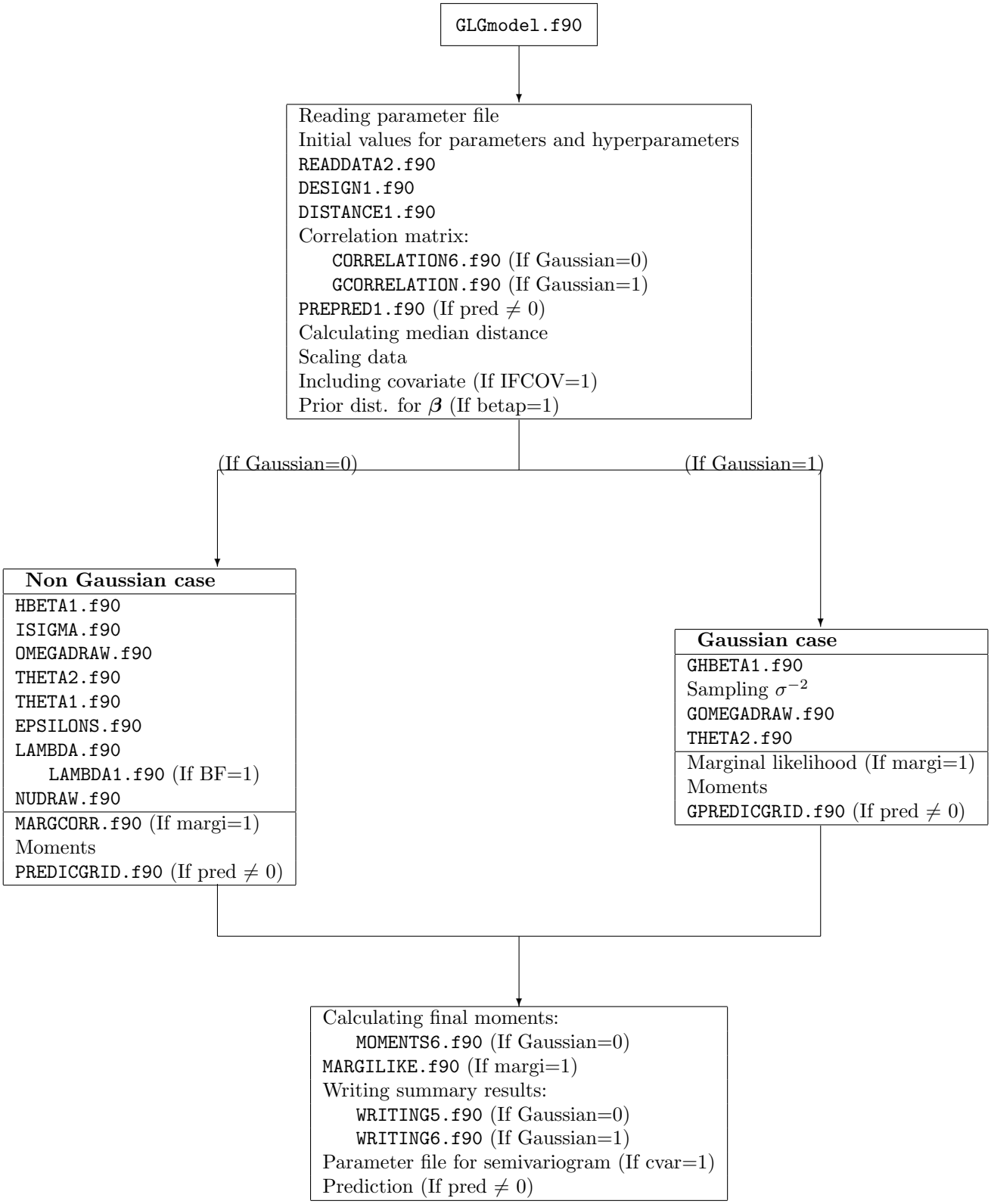
1.- Definitions and first calculations
Definition of the variables and some initial values
Reading the parameter file and the data file
Calculation of the Design matrix, the Distance matrix and the Correlation matrix

2.- GLG model
Sampling from the posterior distributions for the GLG model
Intermediate calculations for posterior moments, prediction, marginal likelihood and Bayes factors for $\lambda_i = 1$

3.- Gaussian model
Sampling from the posterior distributions for the Gaussian model
Intermediate calculations for posterior moments, prediction and marginal likelihood

4.- Final results
Calculation of Formula 13 (Bayes factors for $\lambda_i = 1$)
Calculating posterior moments for the parameters
Calculation of the Marginal likelihood
Writing results
Writing the file to calculate the semivariogram
Prediction

1.1 Flow Chart



1.2 List of Subroutines

The subroutines have been organized in the following way taking into account the part of the code they reference. Whatever case, Gaussian or Non Gaussian, will be considered, the code allows to calculate the marginal likelihood, the moments of the predictive posterior distribution over a grid of points, and the full predictive posterior distribution for a set of points. Also, it can write a parameter file to calculate semivariograms, and evaluate Equation (11) for use in calculating the Bayes factor for $\lambda_i = 1$.

Source files
- GLGmodel.f90

Header files
- DESIGN1.f90
- DISTANCE1.f90
- MATER.f90
- READDATA2.f90

Non Gaussian case
- CORRELATION6.f90
- EPSILONS.f90
- HBETA1.f90
- ISIGMA.f90
- LAMBDA.f90
- MOMENT6.f90
- NU.f90
- OMEGADRAW.f90
- THETA1.f90
- THETA2.f90
- WRITING5.f90

Gaussian case
- GCORRELATION.f90
- GHBETA1.f90
- GOMEGADRAW.f90
- GTHETA1.f90
- GTHETA2.f90
- WRITING6.f90

Prediction
- GPREDICGRID.f90
- PREDICGRID.f90
- PREPRED.f90

Margilike
- MARGCORR.f90
- MARGLIKE.f90

Bayes factor
- LAMBDA1.f90

2 Variables

The most important variables defined in the code have been classified in groups and described in the following tables.

The dimension variables define the maximum dimension of storage for the objects used in the code. These values can be changed directly in the main program.

Description of the dimension variables:

Parameter	Description
N=100	Maximum number of observations
MV=10	Maximum number of variables
KH=10	Maximum number of trend parameters
MC1=10	Maximum number of clusters
GRID=450	Maximum number of points to predict over a grid
npdf=40	Maximum number of points to evaluate the predictive posterior distribution

Most of these variables specifying the data are read from the parameter file. The number of trend parameters, K, does not include the coefficient for the extra covariate. If such a covariate is added, K changes to K+1 automatically.

Description of the variables related to the data:

Variable	Description
DATAFILE	Name of the data set
P	Number of observations
K	Number of trend parameters (INPUT: without covariate OUTPUT: with covariate)
K1	Trend degree for the coordinate x
K2	Trend degree for the coordinate y
CV	Variable column in the data set
COV	Covariate column in the data set
MC	Number of clusters
MI, MA	Minimum and maximum cluster size
NCLUS	Cluster sizes
Z	Vector containing the data
DEN	Constant to scale the data

Some matrices such as the design matrix and the distance matrix are fixed during the execution of the code. However, the correlation matrix and its inverse are updated continuously.

Description of matrices:

Matrix	Description
X	Design matrix
DIST	Distance matrix
CORR	Correlation matrix
ICORR	Inverse correlation matrix
XP	Design matrix of the points to predict
PDIST1	Distance matrix between the points to predict and the sampled points

The variables containing a vector of parameters are defined with capital letters. Their moments are also calculated in the code and shown as results.

Description of parameters and moments:

Parameter	Moments		Description (all moments are posterior)
BET	Mbet	Varbet1	$\boldsymbol{\beta}$, E[$\boldsymbol{\beta}$],cov[$\boldsymbol{\beta}$]
gamma1	Msig	Vsig	σ^{-2}, E[σ^{-2}], var[σ^{-2}]
	Ms	Vs	E[σ], var[σ]
omega2	Momega	Vomega	ω^2, E[ω^2], var[ω^2]
tau2	Mtau	Vtau	τ^2, E[τ], var[τ]
phi, thet1	Mrang	Vrang	$1/\theta_1$, θ_1, E[θ_1], var[θ_1]
rho	Mrho	Vrho	ρ, E[ρ], var[ρ]
thet2	Mthet2	Vthet2	θ_2, E[θ_2], var[θ_2]
lnu	Mlnu	Vlnu	ν, E[ν], var[ν]
EPSI	Mepsi	Vepsi	$\boldsymbol{\epsilon}$, E[$\boldsymbol{\epsilon}$], var[$\boldsymbol{\epsilon}$]
LAM	Mlambda	Vlambda	$\boldsymbol{\lambda}$, E[$\boldsymbol{\lambda}$], var[$\boldsymbol{\lambda}$]
snu	Msnu	Vsnu	$\sigma^2\exp(\nu)$, E[$\sigma^2\exp(\nu)$], var[$\sigma^2\exp(\nu)$]
tsnu	Mtsnu	Vtsnu	$\omega/\exp(\nu/2)$, E[$\omega/\exp(\nu/2)$], var[$\omega/\exp(\nu/2)$]

The values of the control variables indicate the different options of the code. All of them must be specified in the parameter file.

Description of control variables:

Variable	Value	Description
IFCOV	0	No including covariate
	1	Including covariate
Gaussian	0	Non-Gaussian model
	1	Gaussian model
betap	1	Second prior on $\boldsymbol{\beta}$ (Equation (18))
omgig	0	Exponential dist. for the prior on ω^2
	1	GIG dist. for the prior on ω^2
nugig	0	Exponential dist. for the prior on ν
	1	GIG dist. for the prior on ν
PRIORTHETA	0	Prior independence between θ_1 and θ_2
	1	Prior independence between ρ and θ_2
margi	1	to calculate the marginal likelihood
BF	1	to calculate Bayes factors
BF1	1	to save the drawn $\boldsymbol{\lambda}$
cvar	1	to calculate the semivariogram
vari	0	Classical semivariogram of Matheron
	1	Cressie and Hawkins robust semivariogram
pred	1	Prediction in points
	2	Prediction over a grid

Other variables have been defined in the main code and for each subroutine to do intermediate calculations and present other results.

3 Datafiles

Data set	Parameter file
topoclus	topo.par
hotclusz	hot.par
coldclusz	cold.par

Description of the data sets:

- topoclus: 52 topographic elevations, Davis (1973)
- hotclusz: minimum and maximum temperature during a hot week in May 2001 in 63 locations of the Spanish Basque Country. The altitude of the stations is used as covariate.
- coldclusz: minimum and maximum temperature during a cold week in December 2001 in 65 locations of the Spanish Basque Country. The altitude of the stations is used as covariate.

Originally, coordinates of x and y for hot and cold data were given in UTM unities. To build these data sets, the coordinates x and y were scaled subtracting the minimum value and dividing by 10000. The same scales were applied to the coordinates of the points to predict, zp1 and zp2. However, the original coordinates can be used since the code automatically scales x, y and the optional additional covariate to build the design matrix.

The variable of interest in topo data is scaled dividing by DEN=100. Hot and cold data is scaled dividing by DEN=10.

The parameter file contains the information needed to run the code such as the data set, initial values, values for the hyperparameters, and control variables. The name of this file must be indicated at the beginning of the code. There are three options; one for each parameter file contained in the subdirectory `datafiles`:

```
OPEN(UNIT=4,FILE="datafiles/topo.par.txt",STATUS="OLD", ACTION="READ", IOSTAT=ios)
OPEN(UNIT=4,FILE="datafiles/hot.par.txt",STATUS="OLD", ACTION="READ", IOSTAT=ios)
OPEN(UNIT=4,FILE="datafiles/cold.par.txt",STATUS="OLD", ACTION="READ", IOSTAT=ios)
```

Note: To open a particular parameter file, a symbol must be added at the beginning of the unnecessary sentences to “comment out” these lines.

Description of the parameter file:

Variables	Description
DATAFILE	Data file
P CV MC	Number of observations, Variable column, Number of clusters
MI MA	Minimum cluster size, Maximum cluster size
(NCLUS(I),I=1,MC)	Cluster sizes
K1 K2 K	degree trend x, degree trend y, Number of trend parameters
DEN	Scale data
IFCOV COV	Covariate(0,1), Covariate column
ISEED	Seed
simu	Number of simulations
burnin	Burnin
Gaussian	1:Gaussian case, 0: GLG case
gamma1	Initial value of σ^{-2}
omega2	Initial value of ω^2
thet2	Initial value of θ_2
thet1	Initial value of θ_1
lnu	Initial value of ν
Hyperparameters	
c1 betap	Prior dist. for $\boldsymbol{\beta}$, betap=1 Equation (18)
c2 c3	Prior dist. for σ^{-2}
omgig	1: GIG, 0: Exp
clam45 c4 c5 c45	$\mathrm{GIG}(\lambda, \delta, \gamma); \mathrm{Exp}(\lambda)$
nugig	1: GIG, 0: Exp
clam67 c6 c7 c67	$\mathrm{GIG}(\lambda, \delta, \gamma); \mathrm{Exp}(\lambda)$
c8	Prior dist. for θ_1
c9	Prior dist. for θ_2
PRIORTHETA	1: Prior independence between ρ and θ_2
Marginal likelihood	
margi delta	1: to calculate the marginal likelihood, delta
Bayes factors	
BF IB NC	1: to calculate Bayes factors, $\lambda_i = 1$, Number of the cluster
BF1 NF	1: to save the drawn lambdas, Number of obs.s to save their lambdas
IB1 NC1	Observation, Number of the cluster
⋮ ⋮	
Semivariogram	
cvar np	1:to calculate the semivariogram, Number of lags
minlag maxdis	Minimum number of pairs per lag, Percentage of the maximum distance
vari	0:Matheron, 1: Cressie&Hawkins
pred npr ngrid	Prediction (1=points, 2=grid), Number of points to predict, Grid size
pl plmin plmax	Points in pdf, Minimum value, Maximum value
zp1, zp2, zp3	X, Y, Covariate of points to predict
⋮ ⋮ ⋮	

4 Results

The subdirectory Results contains the files with the results generated by the code. Some files are always obtained when the code is run, while others depend on the options indicated in the parameter file.

Description of the result files:

Files	Description
resul.txt	Summary of the results
resulmat.txt	Saved drawn values of the parameters
betas.txt	Saved drawn $\boldsymbol{\beta}$
varx.txt	If betap=1 Writing the moments mz, vz and varx of Equation (18)
lambdas.txt	If BF1=1 Saved drawn $\boldsymbol{\lambda}$ to calculate the numerator of Equation (10)
margilike.txt	If margi=1 Saved marginal likelihood
outfile.txt	If vari=1 Parameter file to calculate the semivariogram
ppd.txt	If pred=1 Posterior predictive distribution at prediction points
ppdmeansd.txt	If pred=2 Posterior predictive means and standard deviations at grid points