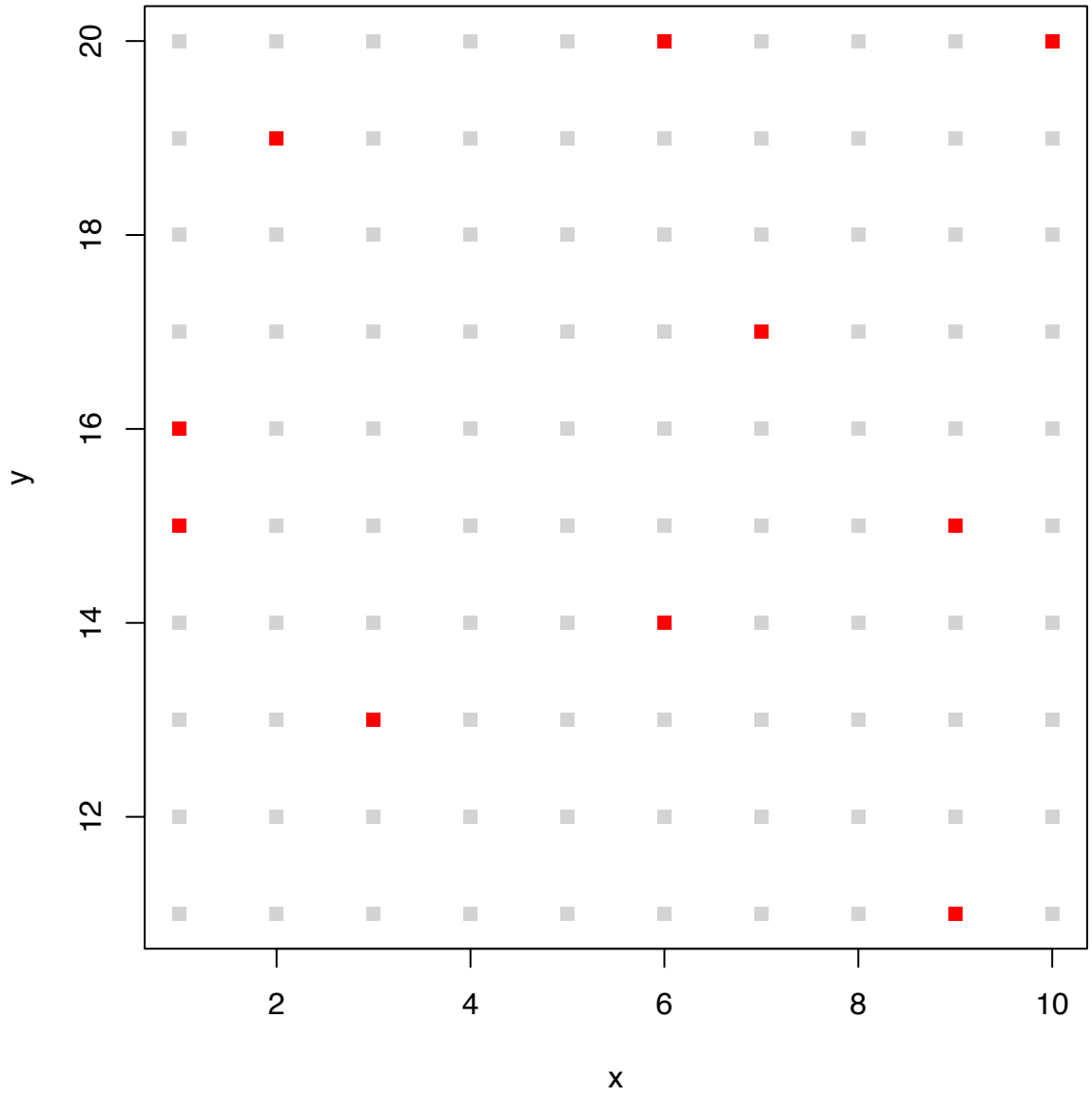
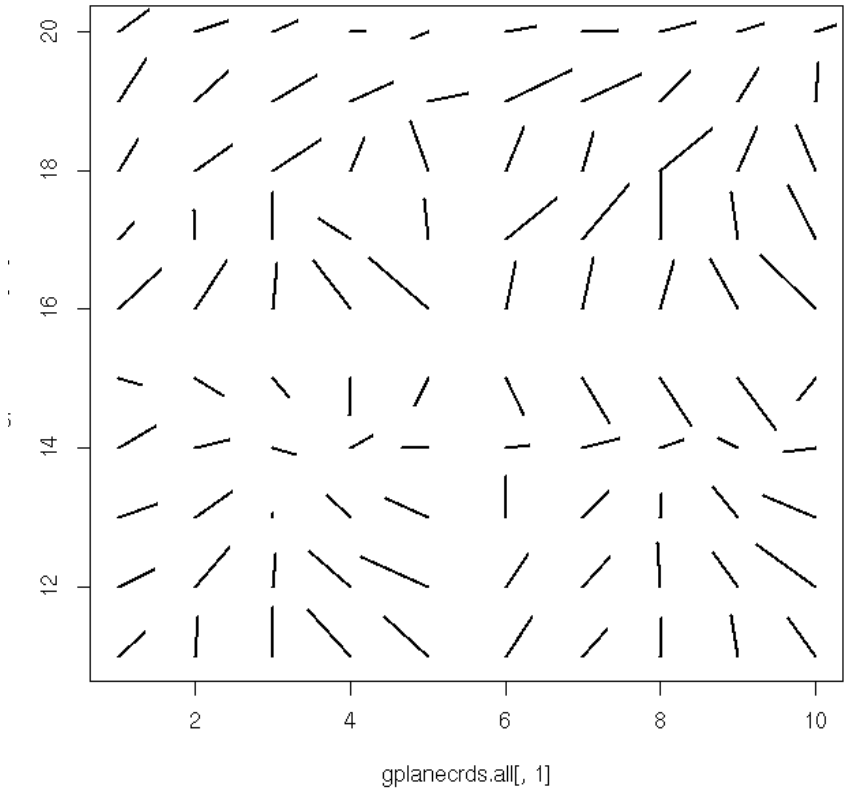
Deformation simulation Summary

Wendy Meiring

PASI 2014


Geographic Locations – Red Randomly Withheld
y
20
18
16
14
12
2
4
6
8
10
x

Transformation G-plane to D-plane visualized at monitoring locations

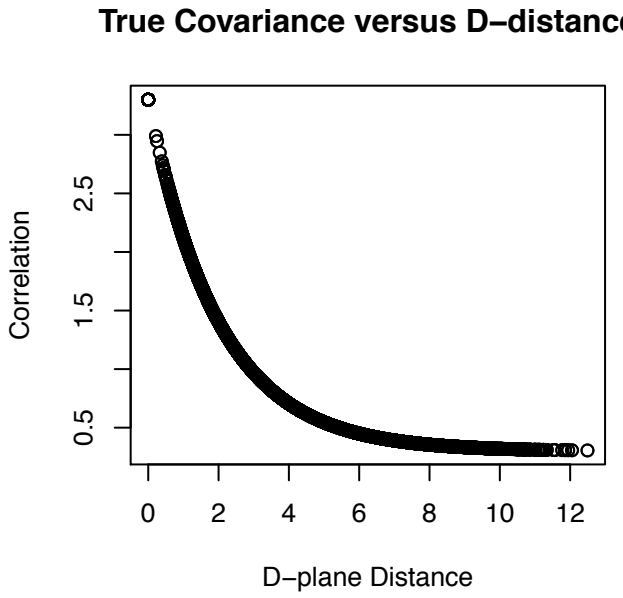


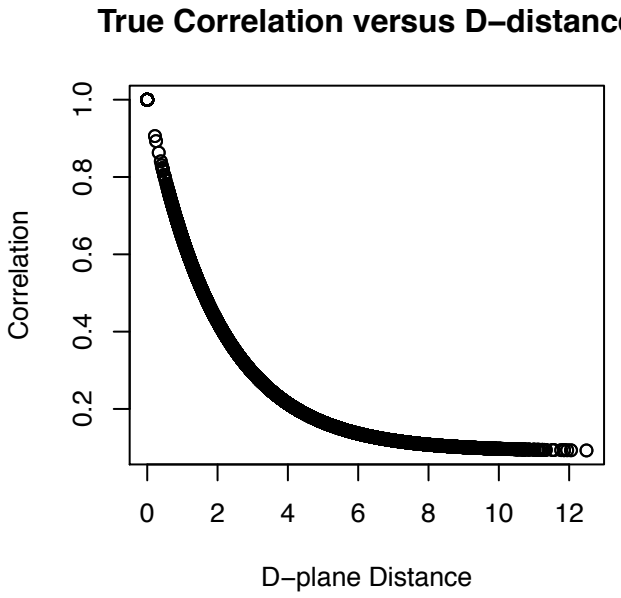
Motivation for need for Single Summary for Differences from True Correlation Matrix

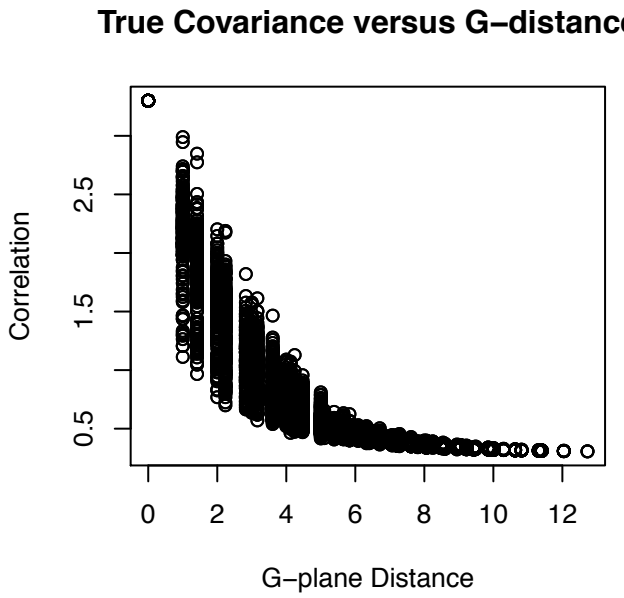
- True correlation matrix amongst withheld sites

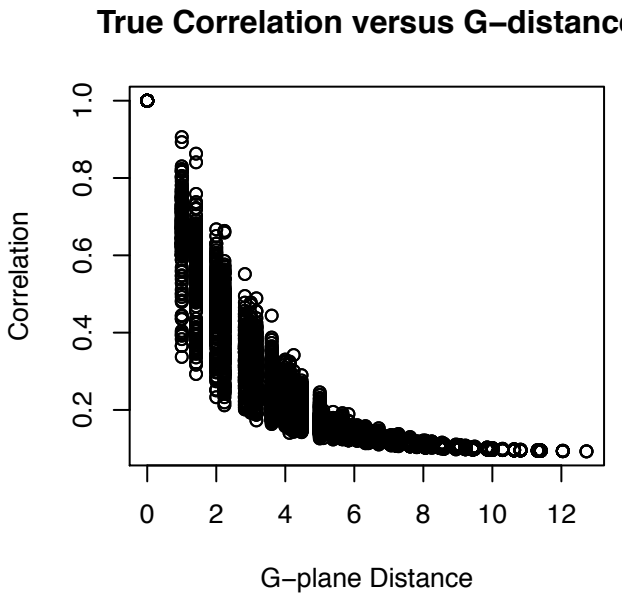
- > [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
- [1,] 1.00 0.13 0.24 0.11 0.32 0.10 0.13 0.10 0.10 0.10
- [2,] 0.13 1.00 0.25 0.35 0.12 0.23 0.12 0.13 0.11 0.10
- [3,] 0.24 0.25 1.00 0.16 0.27 0.15 0.22 0.12 0.14 0.12
- [4,] 0.11 0.35 0.16 1.00 0.11 0.48 0.12 0.18 0.12 0.10
- [5,] 0.32 0.12 0.27 0.11 1.00 0.11 0.21 0.10 0.12 0.14
- [6,] 0.10 0.23 0.15 0.48 0.11 1.00 0.13 0.30 0.14 0.10
- [7,] 0.13 0.12 0.22 0.12 0.21 0.13 1.00 0.15 0.34 0.25
- [8,] 0.10 0.13 0.12 0.18 0.10 0.30 0.15 1.00 0.21 0.11
- [9,] 0.10 0.11 0.14 0.12 0.12 0.14 0.34 0.21 1.00 0.22
- [10,] 0.10 0.10 0.12 0.10 0.14 0.10 0.25 0.11 0.22 1.00

True covariance and correlation versus D-plane and G-plane distance

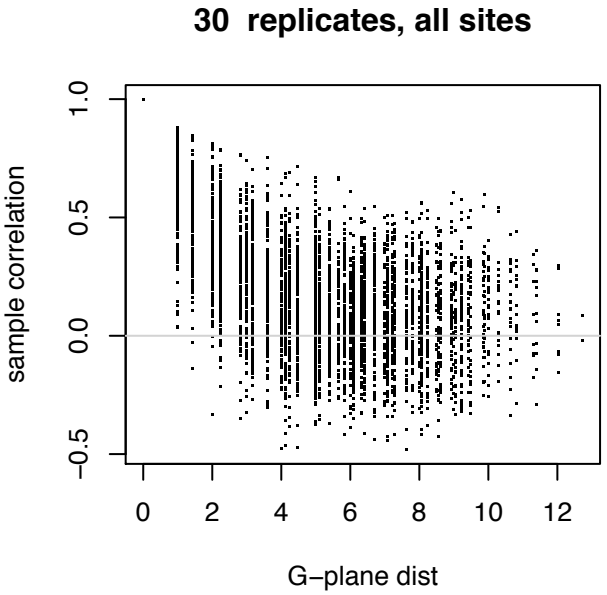


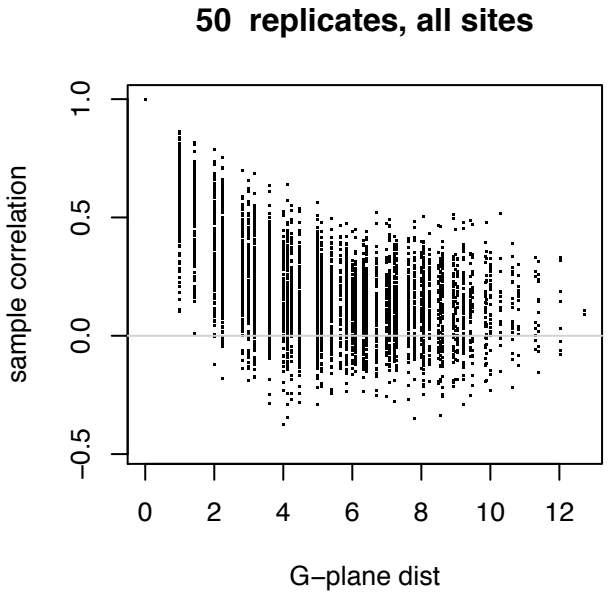


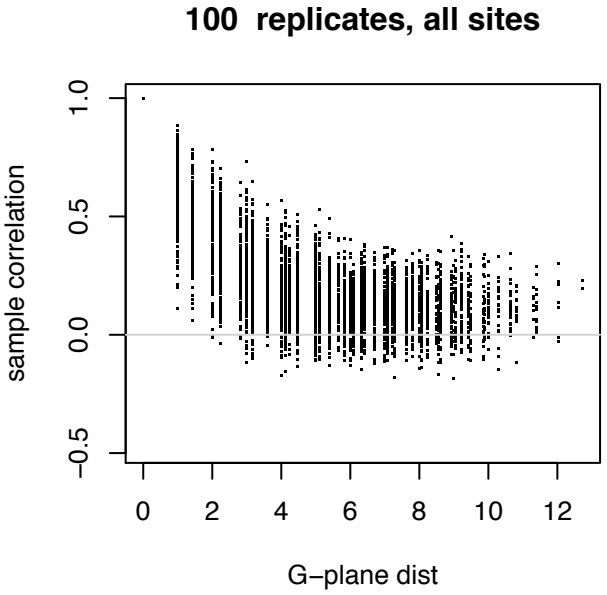


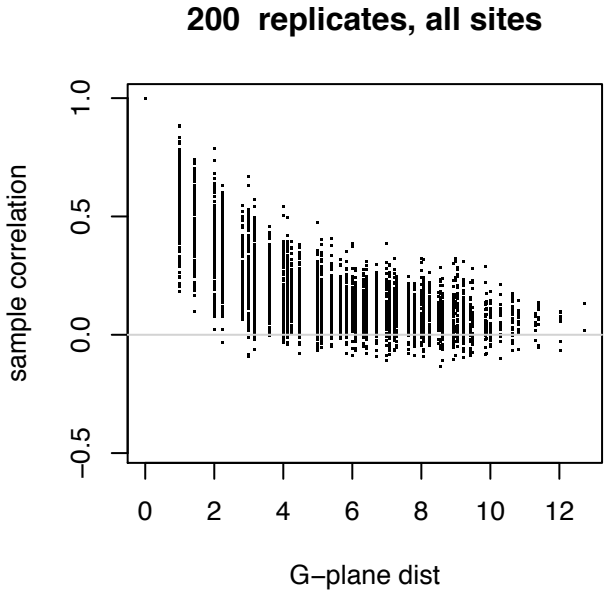


Pairwise Empirical Sample Correlation versus Geographic Distance









Practical guidance

- Two (if 2D) sites fixed to control rotation and scaling. These should be relatively far apart from each other, otherwise alternating algorithm might not converge or take a long time to converge.
- Fortran code – old, but does have hard-coded derivatives. Ideas for publically available reliable fortran optimizer?

Have modified EnviroStat code to use the penalized WLS criterion used in my talk – I think EnviroStat uses penalized OLS currently but still need to confirm.

Model Comparison Questions

- Have started by calculating average Frobenius norms between true correlations and estimated.

  (An idea from Metrics Brainstorming Group)
- Also need to compare covariances. In this simulation mean was zero and variance was constant.
- Only 10% of sites omitted – but already difficult to visualize comparison.
- Next steps –

bootstrap based estimation of uncertainty and get coverage of bootstrap-based confidence intervals.

- Possibly simulate multiple sets of realizations from same model, and estimate coverage.

Another Simulation: Deformation Simulation 1

Two hundred independent replications of a mean zero non-stationary Gaussian spatial process were simulated at 100 points in 2D.

The true covariance of the process is isotropic in a deformed space - but non-isotropic in geographic space. The variance was constant.

Data for 10 of the 100 sites are withheld.

Three csv files are uploaded for analysis, together with a pdf figure showing locations of all 100 sites:

- "GeographicLocProvided.csv" contains the locations of 90 sites for which data are provided.
- "DeformationDataProvided.csv" contains the 200 simulated observations for 90 sites.
- "GeographicLocWithheldSites.csv" contains the locations of the 10 withheld sites.
- GeographicLocationsRedWithheld.pdf shows the locations of all 100 sites - indicating which are withheld.

Goal: estimate the true spatial correlation structure, both between the locations where observations have been provided, and also between the withheld locations.

200 replications from the same process have been provided. However, it may be of interest to run certain methods with subsets of these replications to study how estimation improves with increasing numbers of replications.

The data may be read in using

```
obs.loc  <- read.csv( "GeographicLocProvided.csv", header=T)
                                      # gives a 90 by 2 data frame of locations where observations available

pred.loc <- read.csv( "GeographicLocWithheldSites.csv", header=T)
                                    #   a 10 by 2 data frame of locations where data are withheld

dat.loc    <- read.csv(    "DeformationDataProvided.csv", header=T)
                                       #    a 90 by 200 data frame with observations for 90 sites.
                                       #    The columns correspond to the 200 realizations
```