On the simple and partial Mantel tests with spatial data

Gilles Guillot [1]
Joint work with François Rousset [2]

[1]Department of Informatics and Mathematical Modelling
Technical University of Denmark

[2]Institut des Sciences de l'Évolution
CNRS, Montpellier, France

May 2012

The (simple) Mantel test

Mantel N., **The detection of disease clustering and a generalized regression approach**, *Cancer Research*, 27, 209-220, 1967.

The (simple) Mantel test

Mantel N., **The detection of disease clustering and a generalized regression approach**, *Cancer Research*, 27, 209-220, 1967.

- Goal: *"identifying subtle time-space clustering of disease, as may be occurring in leukemia"*
- Data: $(x_i, y_i)_{i=1,\dots,n}$ observations of a space-time point process
- Idea:
  - transform data so as to get two univariate variables
  - compute correlation of transformed data
  - assess significance of correlation by some permutation method

The simple Mantel test: detailed algorithm

The simple Mantel test: detailed algorithm

- Compute $D^x = (|x_i - x_j|)_{i,j}$ and $D^y = (|y_i - y_j|)_{i,j}$
- Compute the empirical correlation r between D^x and D^y
- For `iter =1,N`
  - draw a random permutation τ of $1, ..., n$
  - compute $D^x_\tau = (|x_{\tau(i)} - x_{\tau(j)}|)_{i,j}$
  - compute the empirical correlation r_τ between D^x_τ and D^y
- If $|r|$ larger than some quantile estimated from the r_τ values: report that there is *"subtle time-space clustering of disease"*

The partial Mantel test

The partial Mantel test

Smouse, P.E., J.C. Long, R.R. Sokal, **Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence**, Systematic Zoology, 35(4), 627-632, 1986.

The partial Mantel test

Smouse, P.E., J.C. Long, R.R. Sokal, **Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence**, Systematic Zoology, 35(4), 627-632, 1986.

- x_i and y_i observations of p and q variables for n statistical units.
- still attempts to assess the dependence between x and y
- need to *"filter out"* or *"control for"* the effect of a third variable z (e.g. z_i spatial coordinates of obs. i)

The partial Mantel test: detailed algorithm

The partial Mantel test: detailed algorithm

- Compute $D^x = (|x_i - x_j|)_{i,j}$, $D^y = (|y_i - y_j|)_{i,j}$ and $D^z = (|z_i - z_j|)_{i,j}$
- Compute residuals $\tilde{D}^x$ of linear regressions $D^x \sim D^z$
- Compute residuals $\tilde{D}^y$ of linear regressions $D^y \sim D^z$
- Compute the empirical correlation r between $\tilde{D}^x$ and $\tilde{D}^y$
- For `iter =1,N`
  - draw a random permutation τ of $1, ..., n$
  - compute $\tilde{D}^x_\tau$ as above for permuted x_i values
  - compute the empirical correlation r_τ between $\tilde{D}^x_\tau$ and $\tilde{D}^y$
- Assess significance of r by comparing to quantiles of r_τ.

Mantel put into orbit

Mantel put into orbit

Mantel (Cancer Res., 1967) and Sokal (Sys. Zool., 1979) claimed that

- the approach was general
- could be used to assess dependence between matrices of "distance"

Mantel put into orbit

Mantel (Cancer Res., 1967) and Sokal (Sys. Zool., 1979) claimed that

- the approach was general
- could be used to assess dependence between matrices of "distance"

Features of the method

- deals with multivariate data
- synthetize data into a single numerical value
- does not seem to rely on any distributional assumption

Posterity of Mantel's work

Posterity of Mantel's work

- Simple Mantel test [Mantel, 1967]: $\geq$ 5000 ISI citations
- Partial Mantel test [Smouse et al., 1986]: $\geq$ 1000 ISI citations
- Implemented in most ecology computer programs
- Countless number of articles using the Mantel tests citing other supporting references
- Routinely used in landscape genetics: x genotypes, y environmental variables, z geographical coordinates
- Practice strongly rooted:

Posterity of Mantel's work

- Simple Mantel test [Mantel, 1967]: $\geq$ 5000 ISI citations
- Partial Mantel test [Smouse et al., 1986]: $\geq$ 1000 ISI citations
- Implemented in most ecology computer programs
- Countless number of articles using the Mantel tests citing other supporting references
- Routinely used in landscape genetics: x genotypes, y environmental variables, z geographical coordinates
- Practice strongly rooted:

Pr. XXX, Assoc. Editor J. of XXX:

Posterity of Mantel's work

- Simple Mantel test [Mantel, 1967]: $\geq$ 5000 ISI citations
- Partial Mantel test [Smouse et al., 1986]: $\geq$ 1000 ISI citations
- Implemented in most ecology computer programs
- Countless number of articles using the Mantel tests citing other supporting references
- Routinely used in landscape genetics: x genotypes, y environmental variables, z geographical coordinates
- Practice strongly rooted:

Pr. XXX, Assoc. Editor J. of XXX:

"Referee 3 pointed out some issues with the Mantel tests but they are so widely used in lansdcape genetics that this comment can be disregarded."

Is the Mantel test a statistical test?

Is the Mantel test a statistical test?

What is a statistical test in Biology?

- A method that returns a numerical value between 0 and 1
- The lower the best

Is the Mantel test a statistical test?

What is a statistical test in Biology?

- A method that returns a numerical value between 0 and 1
- The lower the best

More formal definition involves...

- A null hypothesis
- A method to derive a p-value
- Some additional distributional assumptions

Are the Mantel tests appropriate?

Are the Mantel tests appropriate?

A common implementation:

Are the Mantel tests appropriate?

A common implementation:

- x_i mutivariate genotype or phenotype.
  Due to population history and limited mixing in space x is spatially-autocorrelated
- y_i multivariate descriptor of landscape (elevation, temperature, vegetation cover).
  Due to bio/geo-physical laws y is spatially-autocorrelated
- Interest in testing H_0: x and y are independent

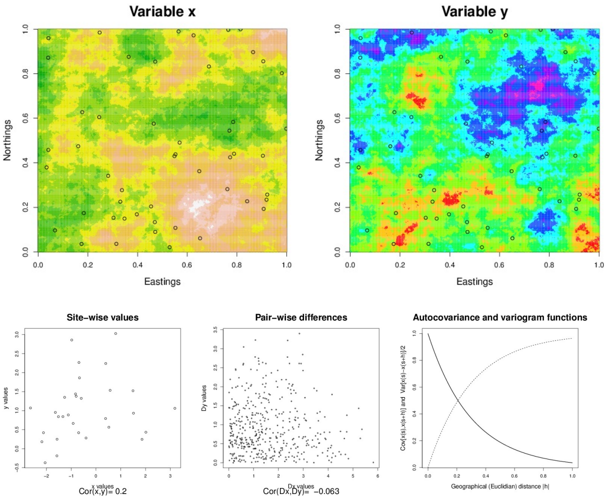
A simulation study

A simulation study

Simulation to mimic the situation of one phenotypic variable and one environmental variable.

- $s_1, ..., s_n$ n=50 sites in $[0,1]^2$
- $x(s_1), ..., x(s_n)$ values of a GRF with expo. covariance
- $y(s_1), ..., y(s_n)$ values of a GRF with expo. covariance
- x and y independent
- common scale param. κ

Example of simulated data



Simulation study (cont')

- simulation above repeated for 200 realizations of x and y
- p-values for simple Mantel test
- p-value for partial Mantel test with matrix D^s entered to "control the effect of space".
- common scale param. κ vaying from 0 to 0.7
- plot of ordered p-values against quantiles of a uniform distribution
- Under H_0, the p-values should be uniformly distributed [Schweder and Spjøtvoll, 1982]

Qq-plots of p-values obtained on simulated data

Qq-plots of p-values obtained on simulated data

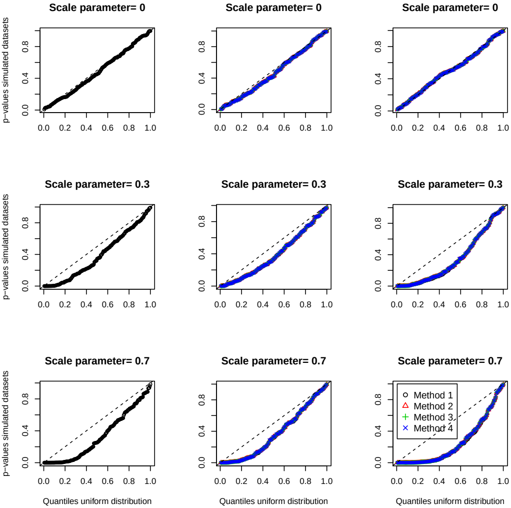


Figure: Left: simple Mantel test. Middle: partial Mantel test, no drift. Right: partial Mantel test, RFs with linear trend.

What's wrong with the Mantel tests?

What's wrong with the Mantel tests?

Mantel tests are based on permuation of one of the data vector entries

- Permutation of x values breaks the potential dependence between x and y
- Also breaks the spatial structure of x!!

The Mantel test fallacy:

$$cor(D^x_\tau, D^y) \overset{\mathcal{L}}{\neq} cor(D^x, D^y)$$

Alternative approaches

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].
- Extension to categorical data [Cerioli, 2002]

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].
- Extension to categorical data [Cerioli, 2002]
- Restricted permutations:

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].
- Extension to categorical data [Cerioli, 2002]
- Restricted permutations:
  - for clumpped geostatistical data: within-population permutation

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].
- Extension to categorical data [Cerioli, 2002]
- Restricted permutations:
  - for clumpped geostatistical data: within-population permutation
  - lattice data: shift permutation

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].
- Extension to categorical data [Cerioli, 2002]
- Restricted permutations:
  - for clumpped geostatistical data: within-population permutation
  - lattice data: shift permutation
- Testing in a GLMM framework

Conclusion

Conclusion

- Mantel tests are flawed in presence of structure in the data
- Conclusion extends to other form of structure (phylogentic trees)

Conclusion

- Mantel tests are flawed in presence of structure in the data
- Conclusion extends to other form of structure (phylogentic trees)
- A clear warning is timely

Conclusion

- Mantel tests are flawed in presence of structure in the data
- Conclusion extends to other form of structure (phylogentic trees)
- A clear warning is timely
- Needs further work on the side of computer program development

Conclusion

- Mantel tests are flawed in presence of structure in the data
- Conclusion extends to other form of structure (phylogentic trees)
- A clear warning is timely
- Needs further work on the side of computer program development

Research report:

Guillot, G. and Rousset, F., On the simple and partial Mantel tests in presence of spatial auto-correlation, arXiv:1112.0651v1,(2012).

Conclusion

- Mantel tests are flawed in presence of structure in the data
- Conclusion extends to other form of structure (phylogentic trees)
- A clear warning is timely
- Needs further work on the side of computer program development

Research report:

Guillot, G. and Rousset, F., On the simple and partial Mantel tests in presence of spatial auto-correlation, arXiv:1112.0651v1,(2012).

Thank you!

References

[Cerioli, 2002] Cerioli, A. (2002).
Testing mutual independence between two discrete-valued spatial processes: A correction to Pearson chi-squared.
Biometrics, 58:888–897.

[Clifford et al., 1989] Clifford, P., Richardson, S., and Hemon, D. (1989).
Assessing the significance of the correlation between two spatial processes.
Biometrics, 45(1):123–134.

[Dutilleul et al., 1993] Dutilleul, P., Clifford, P., Richardson, S., and Hémon, D. (1993).
Modifying the t test for assessing the correlation between two spatial processes.
Biometrics, 49:305–314.

[Guillot and Rousset, 2012] Guillot, G. and Rousset, F. (2012).
On the use of the simple and partial Mantel tests in presence of spatial auto-correlation.
arXiv:1112.0651v1.

[Mantel, 1967] Mantel, N. (1967).
The detection of disease clustering and a generalized regression approach.
Cancer Research, 27:209–220.

[Richardson and Clifford, 1991] Richardson, S. and Clifford, P. (1991).
Testing association between spatial processes.
Lecture Notes-Monograph Series, 20:295–308.

[Schlather et al., 2004] Schlather, M., Ribeiro, P., and Diggle, P. (2004).
Detecting dependence between marks and locations of marked point processes.
Journal of the Royal Statistical Society, series B, 66:79–93.

[Schweder and Spjøtvoll, 1982] Schweder, T. and Spjøtvoll, E. (1982).
Plots of p-values to evaluate many tests simultaneously.
Biometrika,, 69(3):493–502.

[Smouse et al., 1986] Smouse, P., Long, J., and Sokal, R. (1986).
Multiple regression and correlation extensions of the Mantel test of matrix correspondence.
Systematic Zoology, 35(4):627–632.