# Modeling of the spatial covariance structure of the brain using variograms with a non-Euclidean metric

R. R. Nandy[1], Y. Gel[2], D. Cordes[1]

[1]Department of Radiology, University of Washington, Seattle, Wa, United States, [2]Statistics, George Washington University, Washington, DC, United States

## Introduction

The identification of the structure of the spatial dependence among voxels in the human brain is an important but difficult problem in MRI/fMRI. A popular statistical technique to analyze spatial data is based on the use of variograms [1], which analyzes the effect of spatial separation on the observed quantities. If the spatial process is stationary, then the spatial covariance function only depends on the spatial distance and parametric model fitting is possible. However, the assumption of stationarity is certainly not valid for the spatial image of the whole brain due to the inhomogeneity arising from the presence of three different physical structures in the brain (gray matter, white matter and CSF) which gives rise to discontinuity at the boundaries of the surfaces of these three constituents of the brain. However, stationarity may be a valid assumption in a segmentation based approach, where the brain is segmented separately into gray matter, white matter and CSF. This makes each individual segment more homogeneous so that it is possible to assume stationarity. This however creates additional problems as all the individual segments are not simply connected any more. This means there may be holes inside the segments, which are most visible in gray matter. Thus there are pairs of points in gray matter which satisfy the following characteristic. For each such pair, the points may be close to one another in terms of the conventional Euclidean distance, but the length of the shortest path connecting them through the gray matter is much longer. If the gray matter is considered separately, it may not be appropriate to treat the points in the pair to be close to one another in terms of the stationarity of the data. We propose the use of a non-Euclidean distance function, which measures the shortest distance between two points from paths that lie entirely within the segmented region, so that the assumption of stationarity remains valid.

## Theory and methods

Let $X(\boldsymbol{u})$ be the observed value of a voxel with coordinates $\boldsymbol{u}$. Then $X(\boldsymbol{u})$ may be considered as a spatial random process with a covariance structure $\Omega$. Due to the segmentation, it is reasonable to assume that for each individual segment, the process is stationary. We will ignore the spatial trend, if any, for the sake of simplicity. From the stationarity assumption, the covariance only depends on the distance between the two points and may be expressed as $\mathrm{cov}(X(\boldsymbol{u}),X(\boldsymbol{v}))= \Omega(d(\boldsymbol{u}-\boldsymbol{v}))$, where $d$ is the distance function. The covariance function can be obtained from the semi-variogram $\mathrm{V}(\boldsymbol{u},\boldsymbol{v})=0.5\mathrm{E}[X(\boldsymbol{u})-X(\boldsymbol{v})]^2$. The semi-variogram can be estimated directly from the data and it is possible to perform a parametric model fitting to the semi-variogram as a function of distance (from stationarity). The same fit easily extends to the covariance function using the relation $\mathrm{V}(\mathbf{u},\mathbf{v})=\mathrm{cov}(0)-\mathrm{cov}(\|\mathbf{u}-\mathbf{v}\|)$. A popular choice for the parametric fit to the semi-variogram is the exponential function $\mathrm{V}(\boldsymbol{u},\boldsymbol{v})=\rho+\sigma^2(1-e^{-d/r})$, where $\sigma^2$ (sill), $r$ (range) and $\rho$ (nugget) are the parameters to be estimated to fit the curve. The quantity $d$ is the distance between voxels with locations u and v. Sill $\sigma^2$, range $r$ and nugget $\rho$, as well as the trend $\mu$ may be estimated by the non-linear least squares method applied to the estimated semi-variogram. It should be noted that by definition the semi-variogram function takes the value 0 when $d$ is 0. But it may not be the case when $d$ is very small but not identically 0. This is the reason for the appearance of the nugget in the model. We should also mention that the choice of exponential covariance function is made only for the sake of simplicity. There are numerous other models which may give us a better fit.

We have already mentioned that the conventional Euclidean metric is not an appropriate choice in the present scenario. We now describe how we measure the distance function with a non-Euclidean metric. Ideally, the distance function for a pair of points within a segment should be the minimum length of all paths that connect the points and lie entirely within the segment. However, it is not feasible from a computational standpoint to calculate such a metric for all possible pairs of points in the segment of interest. Instead we make use of the fact that we have a finite number of voxels which may be considered as a subset of a finite rectangular graph with nodes at voxel locations. Note that this is only a subset of the rectangular graph due to the missing nodes at locations where there the voxels are missing. We can now make it a connected graph by joining adjacent vertical and horizontal nodes. Each single edge is assigned a length of one voxel unit. We can then compute the shortest distance between any two points in this graph along the edges [2]. The algorithm can be further refined by including adjacent diagonal edges as well, but this comes at a higher computational price and is not performed here.

## Results

For computational simplicity, instead of the whole brain, we only consider a 256×256 segmented gray matter slice in the sagittal plane from an image acquired with a standard SPGR protocol. In Figure 1, we plot the semi-variogram for the data. For very large values of distance, the semi-variogram appears to be unstable. There are two reasons behind that. Large distance in general corresponds to voxels which are far apart from one another and predictive ability is expected to be low in such a case. More importantly, since we are only considering a single slice, there may be points with large non-Euclidean distances among them which would be significantly reduced if the paths are considered in 3D. Variogram analysis using a 3-D data is expected to improve the results significantly. Nevertheless, we observe a nice exponential behavior if we restrict ourselves to distances less than 80 voxel units, which is more than sufficient for any sort of prediction involving the data. In Figure 2, we plot the semi-variogram restricted to distances less than 80 voxel units and the best nonlinear least square fit of the exponential model. The fit was excellent with an $R^2$ value of 0.9431984 which corresponds to a $p$-value of 0. The estimated values of nugget, sill and range are, respectively, 105.0899, 118.8801 and 23.70773 . In Figure 3, we plotted the semi-variogram using Euclidean distance which appears to be scattered without a definite structure.

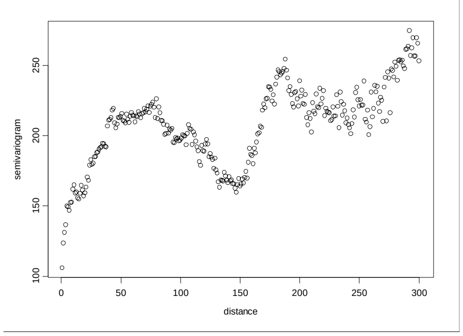


**Figure 1.** Semi-variogram using non-Euclidean metric

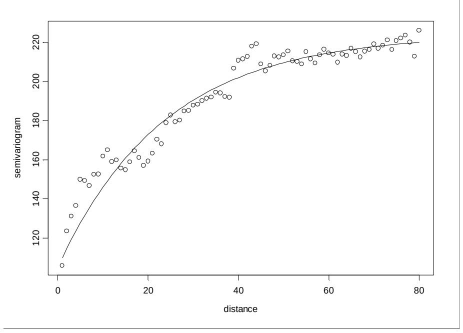


**Figure 2.** Fitted exponential for distances upto 80

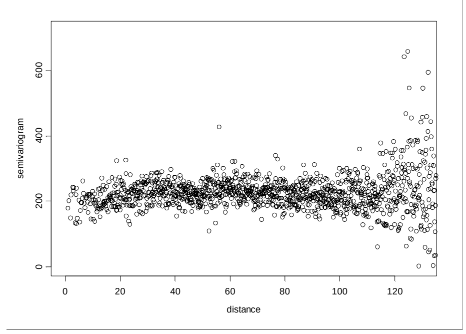


**Figure 3.** Semi-variogram using Euclidean metric

## References

1. H. Wackernagel (1998), "Multivariate Geostatistics", Springer-Verlag
2. S. Pettie, V. Ramachandran, "An optimal minimum spanning tree algorithm.", Proc. 27th International Colloquium on Automata, Languages and Programming (ICALP'2000)