Notes on Distance Dependence of Pair Correlation Statistics

We use two correlation statistics:
pair-products of trial z-scores: $\{z_i\}$
pair-products of the "zero-mean trial variances": $\{ z_i^2 - 1 \}$

In all cases we look at sums of the products as

ppnet = $\sum_{t,\,i<j} \sum z_{i,t}\, z_{j,t}$

and

ppcov = $\sum_{t,\,i<j} \sum \left(z_{i,t}^2 - 1\right)\left(z_{j,t}^2 - 1\right)$

where (i,j) label RNGs for one second of network output and t labels time in seconds.

ppnet and ppcov are both zero-mean statistics with positive skewness and bounded variances of N_T and $4\,N_T$, respectively, where N_T is the total number of products in the sums. By the central limit theorem, they approach gaussian distributions for $N_T \gg 1$. This condition always holds for the cases we consider. In order to assess deviations, Δ, of the correlation sums, Z-scores can be formed as

$$Z_{Net} = \Delta_{Net} \Big/ \sqrt{N_T}$$

$$Z_{Cov} = \Delta_{Cov} \Big/ \sqrt{4\,N_T}\ .$$

These results are valid for standard normal trials, $\{z_i\}$. The GCP data are derived from binomial B[200,1/2] generators, which alters the variance of ppcov slightly. The variance of ppcov for GCP data is $3.96\,N_T$, and

$$Z_{Cov}^{gcp} = \Delta_{Cov} \Big/ \sqrt{3.96\,N_T}\ .$$

Working directly with pair-products allows us to investigate if distance is implicated in the measured experimental correlations. In effect, the correlation strengths

pp = $z_i\, z_j$

can be written as

$$pp = s(\vec{r}_i, \vec{r}_j)$$

where the vectors $\vec{r}$ denote the network node positions. An obvious formulation is

$$pp = s(d_{i\,j})$$

where $d_{i\,j}$ is the geometric distance between RNGs i and j. In this case the analysis problem is to determine the functional form of s(d). Intuition suggests a monotonic decrease of s versus d. This may be modeled analytically as linear or gaussian fall-offs, for example. Other examples can be cited, but more complicated models are probably not warranted due to the noisiness of the data. The null hypothesis for distance, s(d) = constant, is of primary importance, since it implies that RNG networks do not require geographical deployment in order to measure an effect. The principal aim of the distance analysis is to establish a confidence interval against a null distance effect.

There are many choices for modeling a distance effect. If a distance effect obtains, the data may be able to distiguish between some of these. Two possibilities are worth mention:

$$s = s(\overrightarrow{r_{i\,j}}),$$

where $\overrightarrow{r_{i\,j}} = \vec{r}_i - \vec{r}_j$. In this case the distance enters only through the RNG-RNG separation and is independent of where the RNGs are located on the earth. This could describe a situation of entanglement of some sort, for which correlations depend only on the proximity of RNGs. The global consciousness could then be imagined as a time-dependent scalar field, with constant spatial amplitude. Entanglement would imply that data correlations are not associated with superposed deviations of individual RNGs. The individual RNGs loose their identity, as far as the correlations are concerned. In a sense, individual RNG deviations become complementary to network correlations.

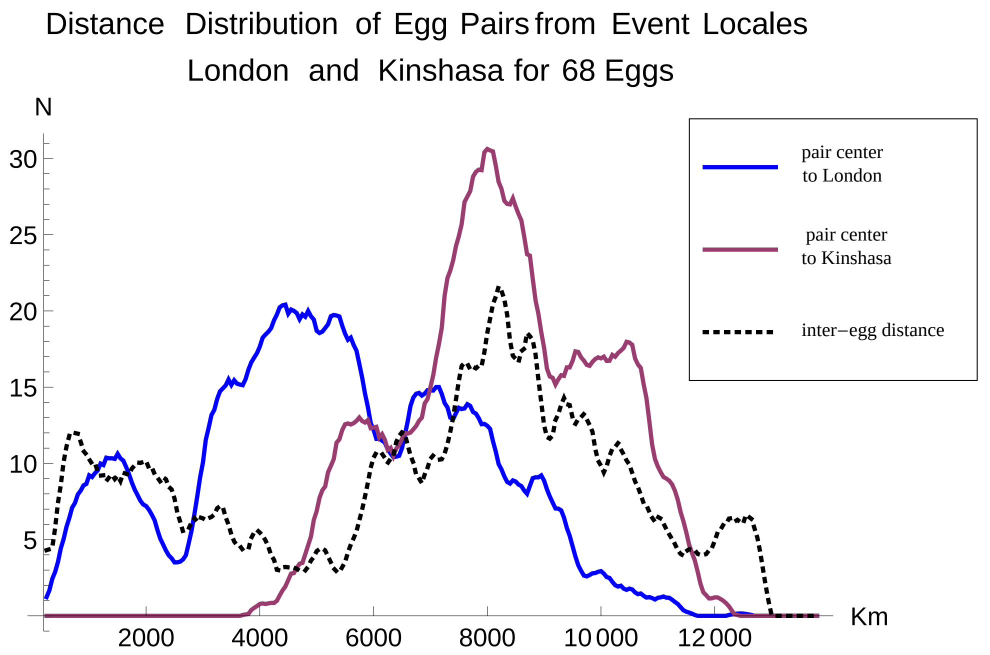
A second possibility is

$$s = s(\vec{r}_i, \vec{r}_j; f(\overrightarrow{r_{source}})).$$

The source distribution, f_s, is a local "global consciousness" which effects the RNG behaviour in some way. The network deviations become correlated because f_s has a large spatial extent and thus simultaneously effects many RNGs in the same way. The source could be associated with an event's locale, the geographical distribution of populations effected by an event, or other factors depending on how one wants to model global consciousness. In order for f_s to produce a distance effect, it needs extend over distances large enough to encompass many network RNGs, but still decay significantly on the scale of the earth's diameter.

Other approaches may involve a combination of these possibilities. One could also image "distance" measures which include psychological, historical or other factors.

In thinking about physical distance, it is important to remember that the distribution of pair distances varies over time and will depend on the details of the model as well. For example, the figure below shows the distribution of distances for 68 RNGs for three different senarios: geometric RNG-RNG pair separations and the distance of RNG pair mid-points from locales in London and Kinshasa. The pair separation distribution samples distances from 0 up to the earth diameter, while the locale distributions are biased toward distances greater or lesser than the earth radius.



The uncertainty in how to model distance will lead to a loss of power in statistical tests of a distance effect. The best we can do as a start is to assume a simple model $s(d_{i\,j})$ where the distances d_{ij} are taken to be the geometric separation of RNG nodes. Tests based on this model will be sensitive to both source and entanglement models, but will loose power to the extent that the model is a poor match to the data.

Results for the test follow.

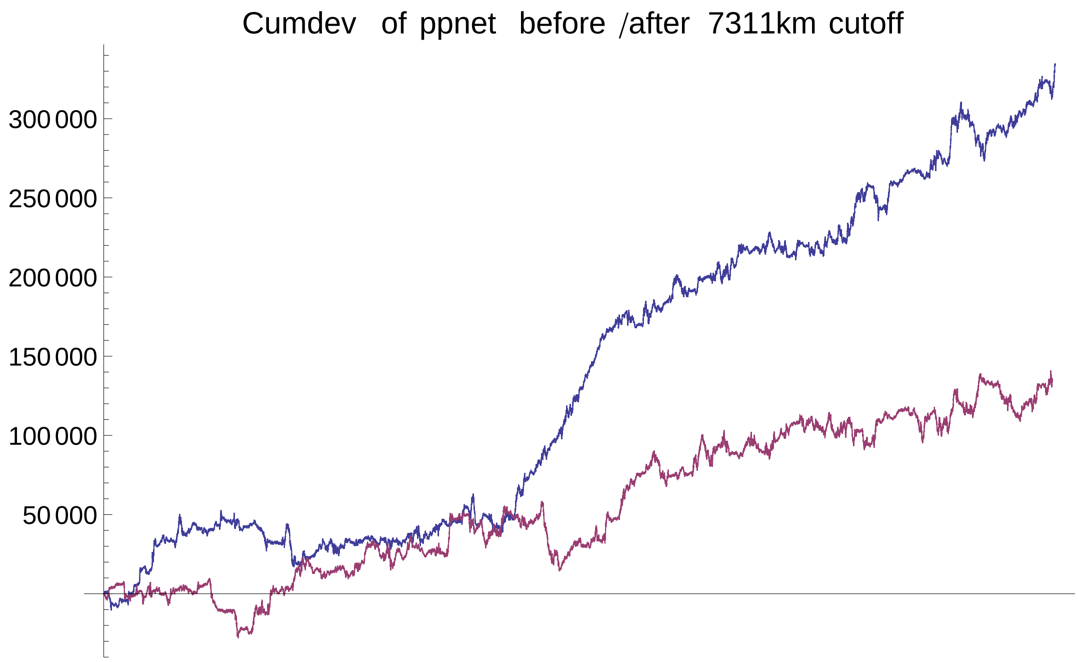
To begin we select a data set of 245 events which excludes events that are formally rejected, collected during periods network instability, of durations longer that 1 day or less that 15 minutes, or occur for a network with less than 17 online nodes. The data set contains 1.14×10^{10} correlation pair products. The average correlation strengths are 4.10 and 3.93×10^{-5}, for ppnet and ppcov, respectively. These values yield Z-scores of

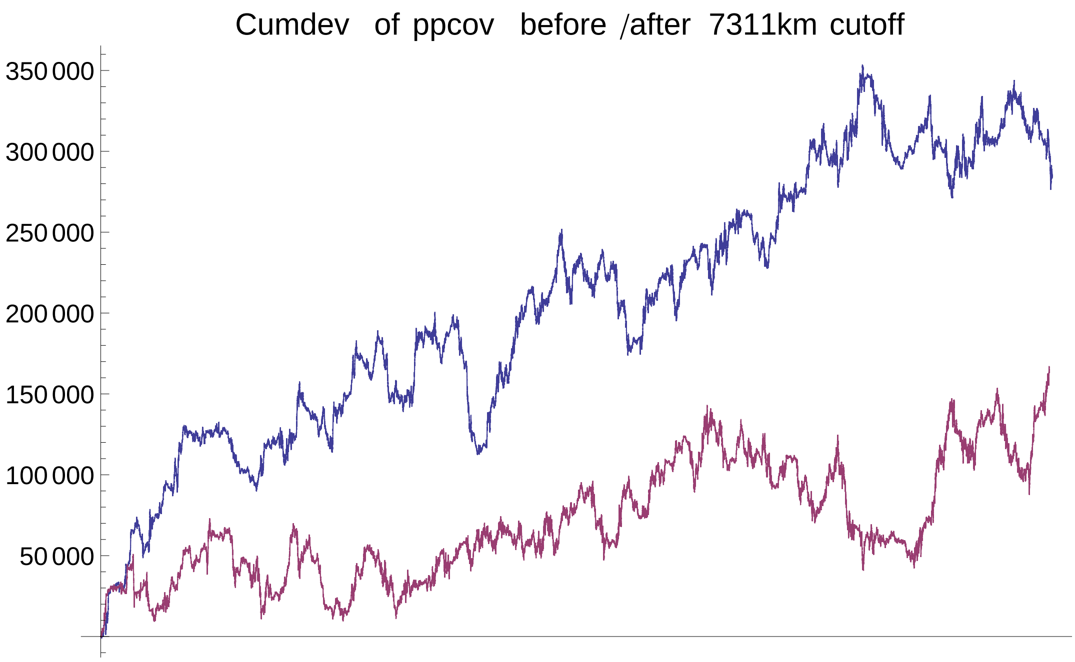
$Z_{Net} = 4.38$

$Z_{Cov}^{gcp} = 2.10$,

in agreement with the Z-scores calculated directly from the netvar and covar statistics for these data.

To study the distance effect, each pair-product is assigned a distance equivalent to the geometric separation of the RNG node locations. A preliminary test can be made by dividing the data into equal sets of low and high distance and comparing the total correlation strengths. A distance effect would appear as larger strength for the low distance set. Cumulative deviations for the two statistics are shown below. The cumdevs sum the data in the chronological order of events. Note the larger variance in the ppcov data.

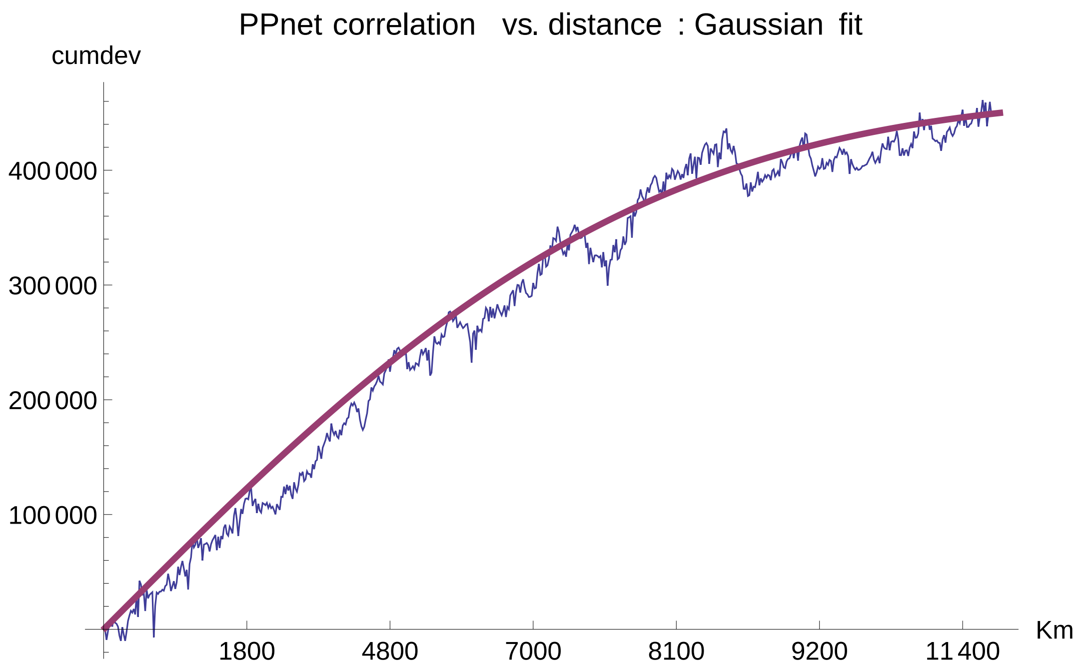


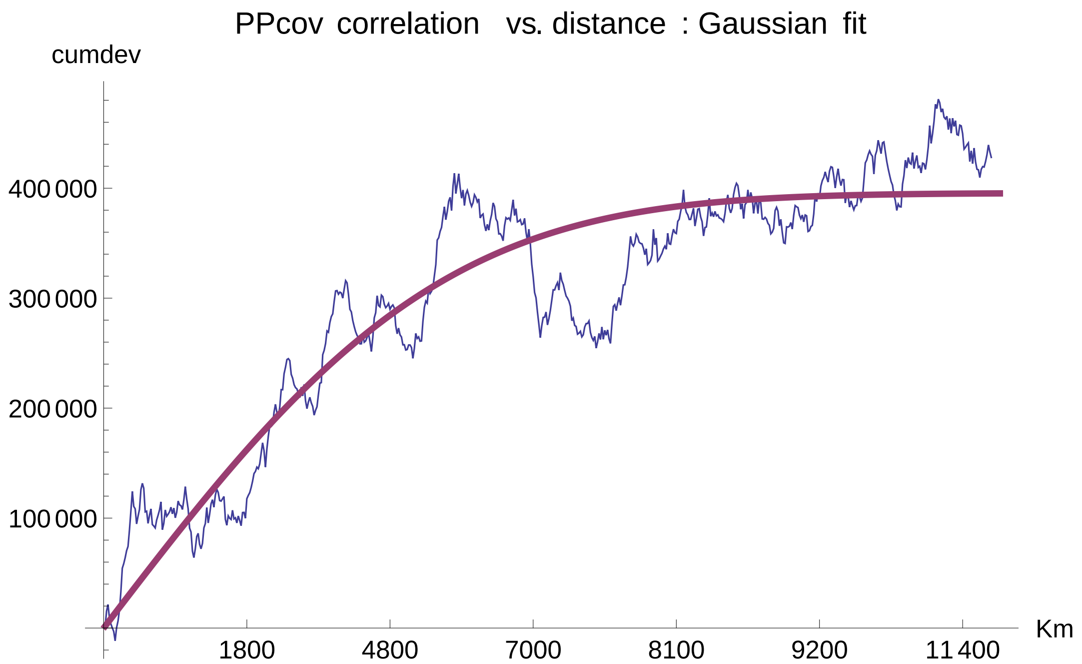


A mean difference test for the distance split yields Z-scores of 1.87 and 0.55 for ppnet and ppcov, respectively, and a combined score of Z = 1.71. This marginally significant evidence for a diffuse distance effect in the event data.

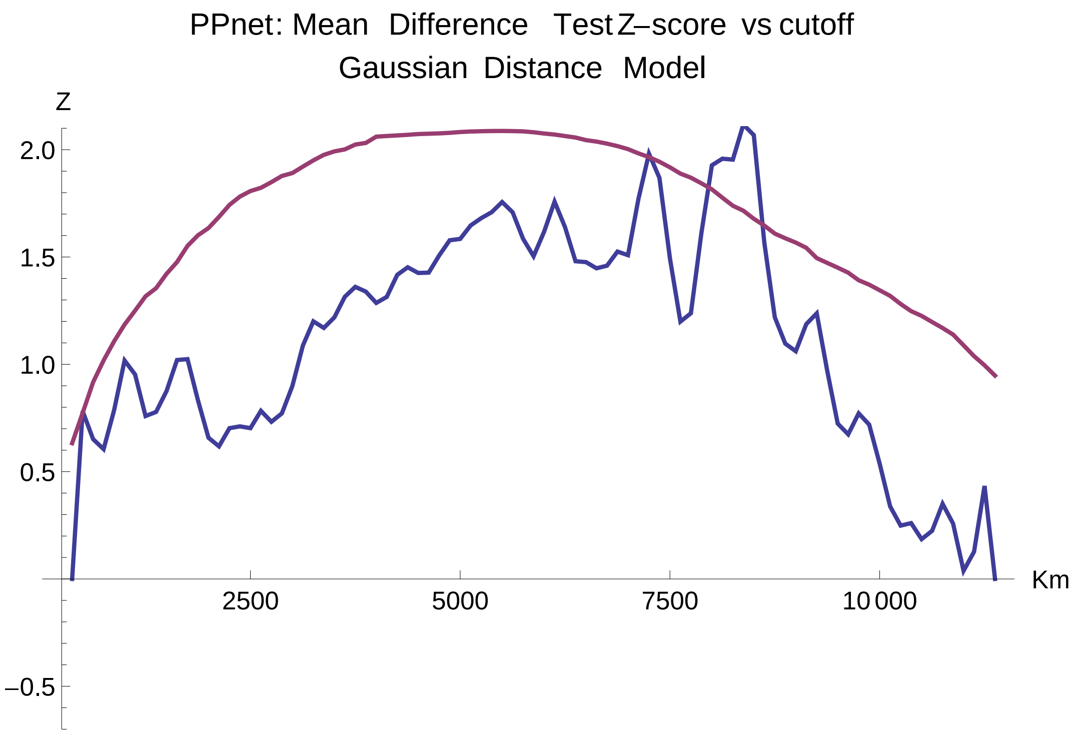
An alternate approach is to perform a linear regression of the average correaltion strength at a given distance. Regression fits should yield a negative slope, with the significance of the slope parameter indicating the P-value for a distance effect. Regressions of the average correlation strength on distance, weighted by the number of pair-products at each distance yield negative slopes with P-values of 0.06 and 0.08 (ppnet, ppcov) for Z-scores of 1.55 and 1.39. The Stouffer Z of these results is Z = 2.08 . Interestingly, the decay distances, given by the distances at which the regressions decline to zero correlation strength are different for the two statistics. The regression intercepts are 16,800km and 12,000 for (ppnet, ppcov). Alternatively, the distances at which the regressions decay to 1/2 the maximum correlation strength are 8400km and 6000km.

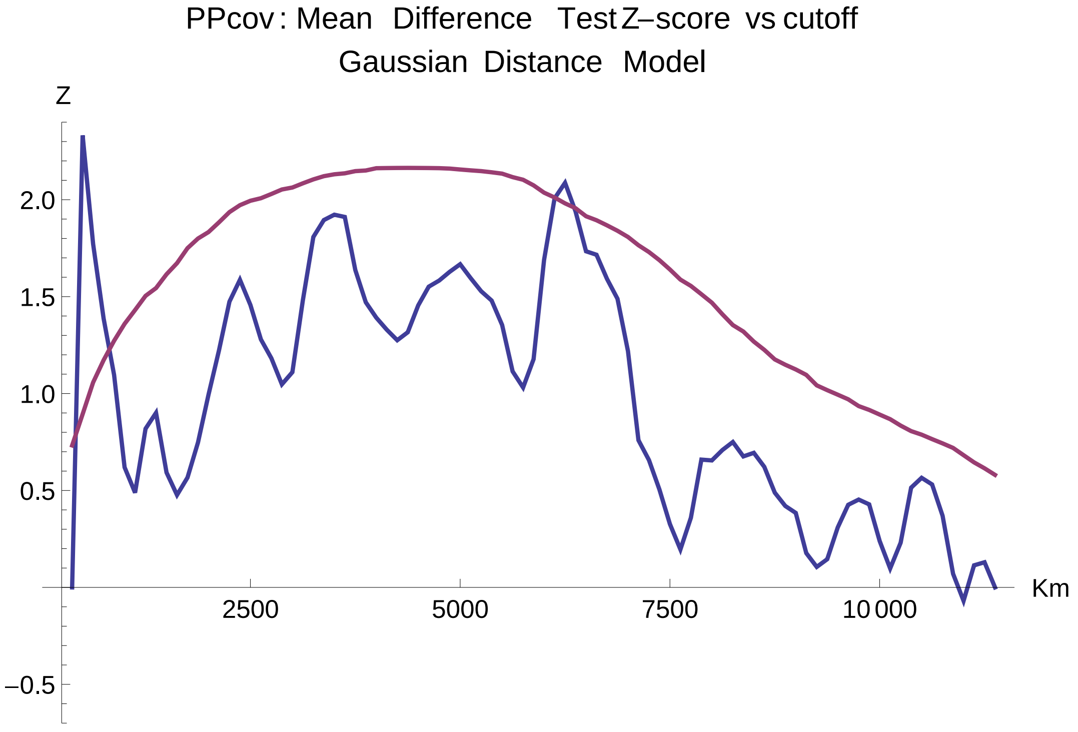
As a check, a non-linear regression using a gaussian fall-off finds decay distances at 1/2 maxima to be 8410km and 5240km, respectively, in reasonable agreement with the linear fits. The gaussian fits can be compared to the full cumdevs to give a sense of the change in correlation strengths as distance increases. In the following plots the correlation products are sorted by distance and plotted as a cumulative deviation. The bold curve is the cumdev of the gaussian fit. Note again that the ppcov has a larger variance, resulting in a noisier plot. Under the null hypothesis, the cumdev plots approximate straight lines, with slopes proportional to the average correlation strength. The decreasing slopes of the gaussian cumdevs signals the decline in correlation strength with distance.





Next we can examine the mean difference test Z-scores across all cutoffs and compare to the gaussian models. The models reproduce the shape of the test Z-scores and somewhat overestimate the magnitudes. The maxima of the model curves indicate cut-off regions where the difference test has the most power, if we assume the models reasonably fit the data. In this case a cut-off around 6000km should be optimal for the ppnet statistic and 4000km for ppcov. Using the experimental Z's from these regions, we estimate a Z of ≈ 1.6 for both statistics. The Stouffer Z for the two is Z ≈ 2.25.





The last part of the distance analysis looks at resampling of the mean difference tests...