The whitening of ambient seismic noise

Cornelis Weemstra[1,2], Willem Westra[3], Roel Snieder[4] and Lapo Boschi[5]

[1]*Spectraseis AG, Zürich, Switzerland*
[2]*ETH, Zürich, Switzerland*
[3]*University of Iceland, Reykjavik, Iceland*
[4]*Colorado School of Mines, Golden, CO, USA*
[5]*UPMC, Université Paris 6, Paris, France*

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Spectraseis

SEVENTH FRAMEWORK PROGRAMME

MARIE CURIE ACTIONS

Motivation & Problem statement

- Recent focus on the estimation of subsurface attenuation using the ambient seismic field (Lawrence and Prieto, 2011; Prieto et al., 2009; Weemstra et al., 2013)
- Methodology based on normalized spatial autocorrelation (SPAC) (Aki, 1957)
- In practice it involves fitting a damped Bessel function ($J_0 e^{-\alpha r_{xy}}$) to the time-and azimuthally-averaged cross-spectra
- It is now well established, however, that the source distribution has a major effect on the decay of the normalized cross-spectrum.(Tsai, 2011; Weaver, 2011)
- The averaging process mitigates the effect of azimuthal variations in the distribution of sources; the effect of the radial distribution of sources remains a problem
- Nevertheless, obtained results seem to be geologically meaningful

Notwithstanding the points listed above, an additional issue exists: the attenuation studies required the damped Bessel function to be scaled by a factor < 1 (Lawrence, 2012; Weemstra et al., 2013). **We aim to explain this observation**. We suspect that the need for this scaling factor can well be explained by a discrepancy between the general theoretical description of the averaging and normalization process and its actual practical implementation. That is to say, there is a subtle difference between them:

"Whitened complex coherency" (Spectral whitening):

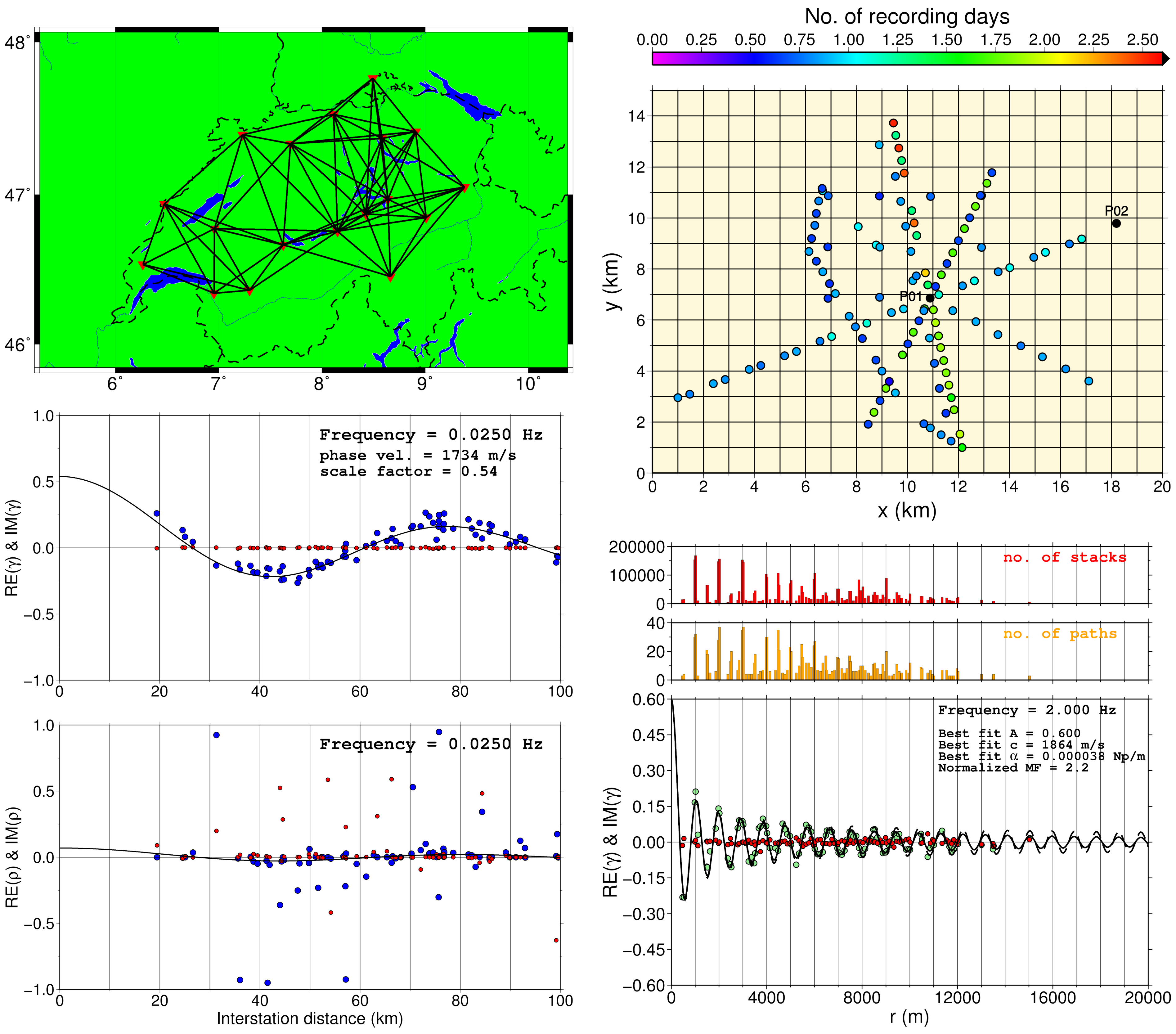
$$\gamma(r,\omega) \equiv Av\left[\frac{\hat{C}_{xy}(\omega)}{\sqrt{\hat{C}_{xx}(\omega)}\sqrt{\hat{C}_{yy}(\omega)}}\right]$$

"Averaged complex coherency" (SPAC normalization):

$$\rho(r,\omega) \equiv \frac{Av\left[\hat{C}_{xy}(\omega)\right]}{\sqrt{Av\left[\hat{C}_{xx}(\omega)\right]}\sqrt{Av\left[\hat{C}_{yy}(\omega)\right]}}$$

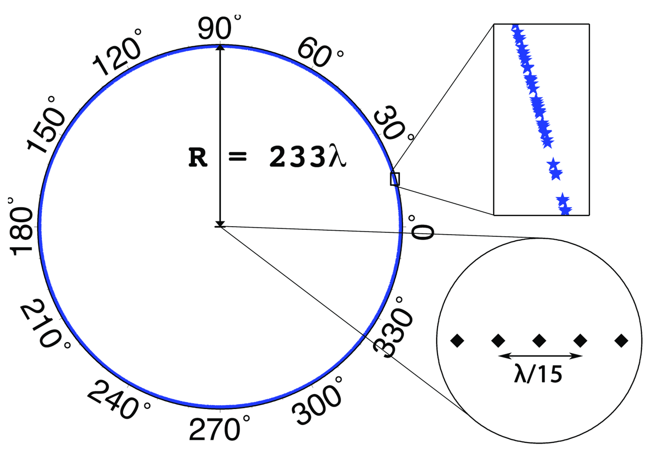
$$\gamma \neq \rho$$

The Difference in Practice



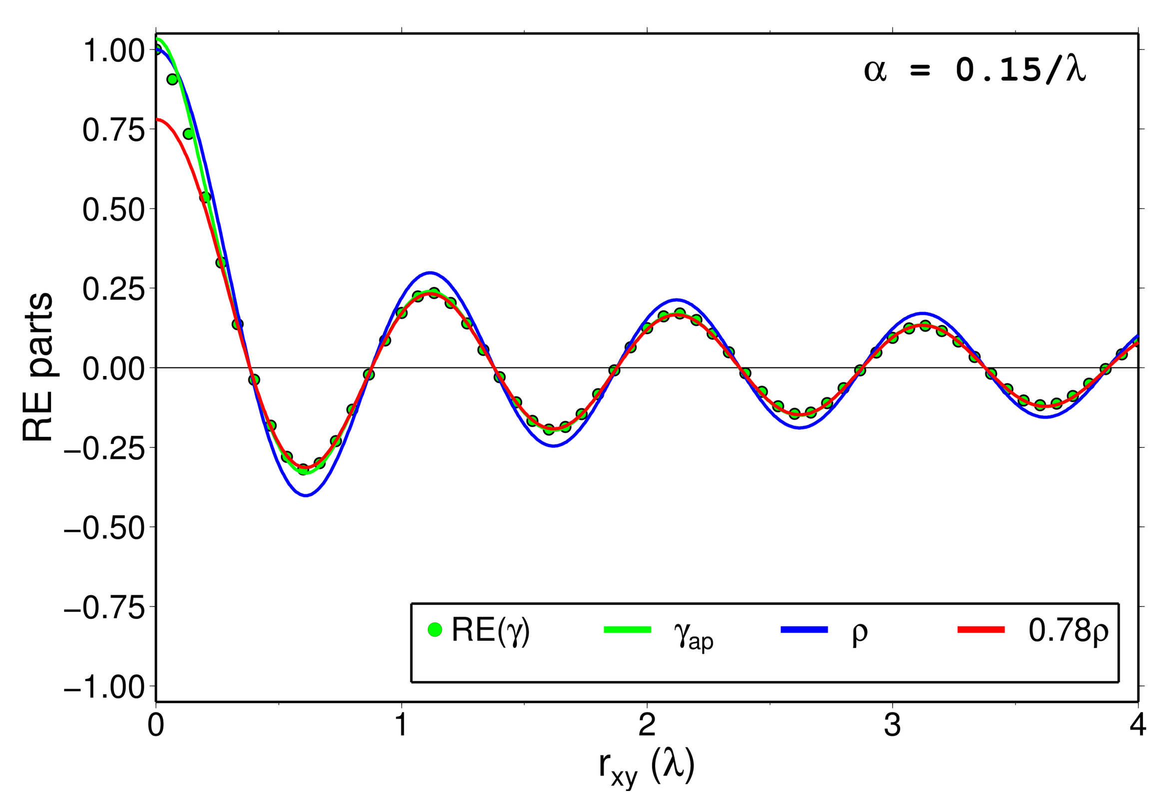
Cross-spectra as function of station separation for two different frequencies of two different datasets. The two datasets measure the ambient seismic field at different scales. The cross-spectra on the left are obtained from 1 year of cross-correlations (2006) for an array of stations in Switzerland. Real (blue) and imaginary (red) parts of the whitened complex coherency γ are plotted as function of interreceiver distance. Cross-spectra are calculated based on γ (top) and ρ (bottom). A zeroth-order Bessel function of the first kind (J_0) is fitted to the data, which results in an estimate of phase velocity in case of γ. On the right side, modified from Weemstra et al. (2013), we see cross-spectra computed from an array of Broad-Band Ocean-Bottom Seismometer recordings deployed on the bottom of the North Sea. Cross-spectra are based on ~ 1 day of coincident recordings. Real (green) and imaginary (red) parts of the whitened complex coherency γ are given as function of interreceiver distance. In this case, a damped Bessel function is fitted to the real parts (solid line) to obtain a measure of phase velocity and attenuation. In both cases we require the Bessel function to be multiplied with a scaling factor in order to obtain a good fit.

Numerical Validation



The setup of the numerical experiment. Sources, in blue, are randomly placed on a circle with radius 233λ (see also the top right box that zooms in on a short stretch of this ring). The bottom right zooms in on the very center of the experiment and shows a blow-up of the line of receivers

- Two million randomly placed sources on a circle (N=2 000 000)
- Ten thousand realizations (M=10 000)
- $\alpha = 0.15\lambda$



Analytical obtained expressions for ρ (blue) and γ (green) for an attenuating medium ($\alpha = 0.15\lambda$). The numerical behavior of γ is depicted by the green dots and a linearly downscaled version of the analytical behavior of ρ is given by the red line.

A Monochromatic Displacement Model

In order to evaluate the difference between the behavior of ρ and γ in a diffuse field, we use the model for such a field described by Tsai (2011). We allow for attenuation and therefore base our discussion on the damped wave equation. The Green's function associated with the two-dimensional (2D) damped wave equation is approximated, for a single frequency, by,

$$G^{(0)}(x;s,\omega) \approx \frac{i}{4}H_0\left(\frac{r_{sx}\omega}{c}\right)e^{-\alpha r_{sx}},$$

where ω is the angular frequency, c the phase velocity and α the attenuation coefficient. The approximation holds for weak attenuation, i.e. $\omega/c \gg \alpha$. H_0 is a 0-order Hankel function of the first kind and $r_{sx} \equiv |s-x|$ is the distance from the source at s to the receiver at x. The total displacement at x due to sources at s_j, where $j = 1, ..., N$, is given by

$$u(x,\omega) = \sum_{j=1}^{N} A_j e^{i\phi_j} G^{(0)}(x;s_j,\omega) = \frac{i}{4}\sum_{j=1}^{N} A_j e^{i\phi_j} H_0\left(\frac{r_{jx}\omega}{c}\right)e^{-\alpha r_{jx}},$$

where the amplitude of the source at s_j is denoted A_j, its phase ϕ_j and the distance between that source and the receiver r_{jx}. We assume the phases ϕ_j to be random variables homogeneously distributed between 0 and 2π. This displacement model enables us to derive expressions for the cross correlation $\hat{C}_{xy}$, and autocorrelations $\hat{C}_{xx}$ and $\hat{C}_{yy}$, and, eventually, approximate the behavior of ρ and γ. To simplify the evaluation of ρ and γ, we first isolate the source phases, i.e. we write the displacement at x as,

$$u(x,\omega) = \sum_{j=1}^{N} f_{jx} e^{i\phi_j} \qquad \text{with the phase independent part described by } f_{jx}\text{, i.e.,} \qquad f_{jx} \equiv \frac{i}{4}A_j H_0\left(\frac{r_{jx}\omega}{c}\right)e^{-\alpha r_{jx}}$$

We isolate these source phase, because we model a 'diffuse wavefield' by averaging cross-correlations over different realizations where the source phases are assumed to change randomly from one realization to the other. The cross-correlation is defined below. One could think of such a realization, as a single time-window. During this time-window, the phase ϕ_j of a source at s_j is assumed to remain stable. The power of the sources is assumed constant between realizations, i.e. constant in time.

The cross-spectrum of the Fourier-decomposed recordings $u(x,\omega)$ and $u(y,\omega)$ for a single frequency ω is given by,

$$\hat{C}_{xy} = u(x)u^*(y) = \left(\sum_{j=1}^{N} f_{jx}e^{i\phi_j}\right) \times \left(\sum_{k=1}^{N} f^*_{ky}e^{-i\phi_k}\right) = \sum_{j=1}^{N}\sum_{k=1}^{N} f_{jx}f^*_{ky}e^{i\phi_j - \phi_k},$$

where the complex conjugate of a variable z is denoted z^*.

This sum can be split in two summations: one sum over N cross-correlations of signal associated with the N sources and another sum over $N(N-1)$ cross-correlations of signal associated with different sources. The latter sum is over the so-called *cross-terms*. disentangling these two summations gives,

$$\hat{C}_{xy} = \sum_{j=1}^{N} f_{jx}f^*_{jy} + \sum_{j=1}^{N}\sum_{k\neq j} f_{jx}f^*_{ky}e^{i(\phi_j-\phi_k)}$$

The first summation does not depend on the phases of the sources and hence does not differ from one realization (time-window) to the other. Note that this does not mean that it is independent of the source distribution. We will refer to this term as the *coherent term* and it will be denoted $\hat{C}_{xyC}$. The second summation, however, does change from one realization (time-window) to the other. Each of the cross-terms in this summation has a different random phase and, also, a different amplitude. We will refer to this term as the *incoherent term* and it will be denoted $\hat{C}_{xyI}$.

Computation of ρ and γ

The ensemble average, denoted by $Av[...]$, is generally calculated over a large number of windows. In our formulation, the ensemble average will tend to its expected value for a large number of windows,. Since we assume the phases to be independent identically distributed random variables, the expected value of the cross-correlation, denoted $E\left[\hat{C}_{xy}\right]$, can be computed by integrating from 0 to 2π over $\phi_1, \phi_2, ..., \phi_N$. The expected value of $\hat{C}_{xy}$ is computed,

$$E\left[\hat{C}_{xy}\right] = \sum_{j=1}^{N} f_{jx}f^*_{jy} + \frac{1}{(2\pi)^N}\int_0^{2\pi}\sum_{j=1}^{N}\sum_{k\neq j} f_{jx}f^*_{ky}e^{i(\phi_j-\phi_k)}\, d\phi_1 d\phi_2 ... d\phi_N.$$

Because the integrands, i.e. $\phi_1, \phi_2, ..., \phi_N$ traverse a circle in the complex plane from 0 to 2π, integration yields zero for all elements of $\hat{C}_{xyI}$. Consequently, only the coherent term $\hat{C}_{xyC}$ survives. Similarly, the expected values of the autocorrelations $\hat{C}_{xx}$ and $\hat{C}_{yy}$ coincide with $\hat{C}_{xxC}$ and $\hat{C}_{yyC}$, respectively. We therefore conclude, in agreement with Tsai (2011), that,

$$\rho = \frac{\hat{C}_{xyC}}{\sqrt{\hat{C}_{xxC}\,\hat{C}_{yyC}}}.$$

Tsai (2011) shows how the source distribution and subsurface attenuation determine the decay of the real part (and imaginary part) of this identity with distance between x and y.

We now turn to γ. As the phases are prescribed to be independent identically distributed variables, the expected value of $\frac{\hat{C}_{xy}(\omega)}{\sqrt{\hat{C}_{xx}(\omega)}\sqrt{\hat{C}_{yy}(\omega)}}$, i.e. γ, is computed,

$$\gamma = \frac{1}{(2\pi)^N}\int_0^{2\pi}\frac{\hat{C}_{xyC} + \hat{C}_{xyI}(\phi_1,\phi_2,...,\phi_N)}{\sqrt{\hat{C}_{xxC} + \hat{C}_{xxI}(\phi_1,\phi_2,...,\phi_N)}\sqrt{\hat{C}_{yyC} + \hat{C}_{yyI}(\phi_1,\phi_2,...,\phi_N)}}\, d\phi_1 d\phi_2 ... d\phi_N.$$

In order to be able to evaluate these integrals, we will rewrite the integrand as a Taylor series in the incoherent terms. Based on our numerical results and the behavior of γ computed from real recordings, we suspect the solution for ρ not an unreasonable proxy for γ and therefore pretend the incoherent terms in the integrand to be small. To that end, we introduce the auxiliary parameter ϵ,

$$\frac{\hat{C}_{xyC} + \hat{C}_{xyI}}{\sqrt{\hat{C}_{xxC} + \hat{C}_{xxI}}\sqrt{\hat{C}_{yyC} + \hat{C}_{yyI}}} \rightarrow \frac{\hat{C}_{xyC} + \epsilon\,\hat{C}_{xyI}}{\sqrt{\hat{C}_{xxC} + \epsilon\,\hat{C}_{xxI}}\sqrt{\hat{C}_{yyC} + \epsilon\,\hat{C}_{yyI}}}.$$

Notice that $\epsilon = 0$ implies that the solution for γ coincides with that for ρ.

A Taylor expansion in the small parameter ϵ yields a power series that quantifies the deviation from the solution for ρ: a so-called perturbation series. The right-hand side of the mapping above we denote γ_ϵ. The Taylor series about $\epsilon = 0$ is given by,

$$\gamma_\epsilon = \rho\left[\gamma_{\epsilon^0} + \gamma_{\epsilon^1}\epsilon + \gamma_{\epsilon^2}\epsilon^2 + \mathcal{O}\left(\epsilon^3\right)\right]$$

The coefficients of this power series are denoted $\gamma_{\epsilon^0}, \gamma_{\epsilon^1}, \gamma_{\epsilon^2}, ...$. Coefficients are explicitly computed up to degree two. The formulation as a power series in ϵ, i.e. a perturbation with respect to the solution for ρ, is simply a vehicle to be able to evaluate the integrals over $\phi_1, \phi_2, ..., \phi_N$. We don't actually expect the incoherent terms to be small for individual realizations. That is to say, if we set ϵ to 1, the perturbation associated with the amplitude of the incoherent terms might not be so small. Importantly however, we do anticipate the expected value of the perturbation to be small. With this rationale we therefore set $\epsilon = 1$. This implies that we can compute γ by simply calculating the expected value of the coefficients $\gamma_{\epsilon^0}, \gamma_{\epsilon^1}, \gamma_{\epsilon^2}, ...$. The independent identically distributed source phases ϕ_j allow us to follow the same procedure as for the earlier calculation of the expected value of the cross-correlation: the expected value of γ_ϵ (with $\epsilon = 1$) is obtained by integrating the coefficients over the phases ϕ_j from 0 to 2π, i.e.,

$$\gamma = E\left[\gamma_{\epsilon=1}\right] = \frac{\rho}{(2\pi)^N}\int_0^{2\pi}\gamma_{\epsilon^0} + \gamma_{\epsilon^1} + \gamma_{\epsilon^2} + \mathcal{O}\left(\gamma_{\epsilon^3}\right)\, d\phi_1 d\phi_2 ... d\phi_N.$$

We neglect the expected values of terms that or of of higher order than the second degree and denote this approximation γ_{ap}.

We find that the expected values of γ_{ϵ^0} and γ_{ϵ^1} coincide with 1 and 0, respectively. The expected value of γ_{ϵ^2} can be approximated by,

$$E\left[\gamma_{\epsilon^2}\right] \cong -\frac{1}{4} + \frac{1}{4}\frac{\hat{C}_{xyC}\,\hat{C}_{yxC}}{\hat{C}_{xxC}\,\hat{C}_{yyC}}.$$

We therefore obtain

$$\gamma_{ap} = \frac{3}{4}\rho + \frac{|\hat{C}_{xyC}|^2}{\hat{C}_{xxC}\,\hat{C}_{yyC}}\frac{1}{4}\rho.$$

It is useful to note that, in case of an isotropic distribution of sources, $\hat{C}_{xyC}$ is purely real (Tsai, 2011), which implies that $|\hat{C}_{xyC}|^2 = \hat{C}_{xyC}\hat{C}_{yxC}$. Using the result for ρ we therefore recognize that $|\hat{C}_{xyC}|^2/\hat{C}_{xxC}\hat{C}_{yyC} = \rho^2$. Hence, we find for an isotropic distribution of sources,

$$\gamma_{ap} = \frac{3}{4}\rho + \frac{1}{4}\rho^3.$$