6 ADDITIONAL SUPPORTING INFORMATION

6.1 NK dependence on the frequency spectra

Consider a time-domain cross-correlation with zero mean associated with two filtered noise sequences in the frequency bandwidth (FBW) $[f_{min},\ f_{max}]$. We define $\phi(t)$ as the time-domain energy density of the cross-correlation. Over the elementary time length L_0 (eq. 3), the time-domain energy variance v^2 is given by

$$v^2 = \int_0^{L_0} t \cdot \phi(t) dt = \frac{1}{2} L_0^2 \cdot \langle \phi \rangle, \tag{11}$$

where $\langle \phi \rangle$ is the mean energy density over L_0. The energy density $\phi(t)$ is equivalent to $m_k^i(t,\tau)$ in eq. 7, for a given pair (i,k). Eq. 11 shows that the mean time-domain energy density $\langle \phi \rangle$ is proportional to the ratio $v^2/{L_0}^2$, which is equivalent to the standardized time-domain variance ${\sigma_k^i}^2$. That is, ${\sigma_k^i}^2$ in eq. 9 is proportional to $\langle m_k^i(t,\tau) \rangle$. Then, the ratio of standardized input variances σ_B^2/σ_A^2, respectively associated to the cross-correlations B and A, is given by

$$\frac{\sigma_B^2}{\sigma_A^2} = \frac{\langle \phi_B \rangle}{\langle \phi_A \rangle} = \frac{v_B^2/{L_0^B}^2}{v_A^2/{L_0^A}^2}. \tag{12}$$

We show below that the ratio σ_B^2/σ_A^2 can be calculated in the frequency domain under the assumption that both sequences are white. The frequency-domain energy density of a white sequence is $E(f) = 1$. So, in the FBW $[f_{min},\ f_{max}]$, the frequency-domain energy variance V^2 is given by

$$V^2 = \int_{f_{min}}^{f_{max}} f \cdot E(f) dt = \frac{1}{2}(f_{max}^2 - f_{min}^2). \tag{13}$$

For the two cross-correlations A and B, we obtain using eqs 3 and 13:

$$\frac{V_B^2}{V_A^2} = \frac{(f_{max}^B)^2 - (f_{min}^B)^2}{(f_{max}^A)^2 - (f_{min}^A)^2} = \frac{{L_0^A}^2}{{L_0^B}^2} \cdot \frac{(n_B^2 - 1)}{(n_A^2 - 1)}, \tag{14}$$

where $n_A = f_{max}^A/f_{min}^A$ and $n_B = f_{max}^B/f_{min}^B$.

For each cross-correlation, the time- and frequency-domain variances v^2 and V^2 given in eqs 11 and 13, respectively, are tied through the Gabor (or Heisenberg) uncertainty relation $v^2V^2 \geq 1/(4\pi)$, so we obtain

$$\frac{v_A^2}{v_B^2} = \frac{V_B^2}{V_A^2}. \tag{15}$$

Using eqs 12, 14, and 15, we obtain

$$\frac{\sigma_B^2}{\sigma_A^2} = \frac{(n_A^2 - 1)}{(n_B^2 - 1)}. \tag{16}$$

Now, under the same statistical confidence level ϵ^2 (eq. 9), we can relate the associated products N_AK_A and N_BK_B, for the respective sequences A and B, with their relative FBWs (eq. 16) to obtain

$$\frac{N_BK_B}{N_AK_A} = \frac{(n_A^2 - 1)}{(n_B^2 - 1)}. \tag{17}$$

Eq. 17 is an exact expression. A proxy for fast track calculation is

$$\frac{N_BK_B}{N_AK_A} = \frac{\left[1 + 1/n_B^2 - 1/n_A^2 + \mathcal{O}(n^{-4})\right]}{(n_B/n_A)^2} \approx (n_A/n_B)^2, \tag{18}$$

which is valid when $1/n_A^2$ and $1/n_B^2$ are both smaller than 1 or when $1/n_A^2 \approx 1/n_B^2$.

Table 1. Predicted and observed K-ratios (definition in text) associated with the tolerance thresholds $\epsilon = 0.01$ for synthetic data (Fig. 1e) and $\epsilon = 0.02$ for field data (Figs 2d, e). Double horizontal lines separate synthetic (above) from field data (below). For field data, *P* and *R* stand for *with* and *without* preprocessing (1-bit + whitening), respectively. For *R* data, predicted values were estimated using the Additional Supporting Information.

FBW (Hz)	Predicted (Eq. 10)	Observed	Predicted / Observed
[0.2, 0.4] *vs.* [0.2, 1.0]	0.35	33/72 = 0.46	0.76
[0.2, 0.4] *vs.* [0.2, 0.8]	0.46	39/72 = 0.55	0.84
[0.2, 0.6] *vs.* [0.2, 1.0]	0.58	33/52 = 0.63	0.92
[0.2, 0.4] *vs.* [0.2, 0.6]	0.61	52/71 = 0.72	0.85
[0.2, 0.6] *vs.* [0.2, 0.8]	0.73	39/52 = 0.75	0.97
[0.2, 0.8] *vs.* [0.2, 1.0]	0.79	33/39 = 0.85	0.93
[0.05, 0.4] *vs.* [0.10, 0.2] *R*	2.01	49/22 = 2.23	0.90
[0.05, 0.4] *vs.* [0.05, 0.1] *R*	2.34	40/22 = 1.82	1.29
[0.05, 0.1] *vs.* [0.10, 0.2] *R*	1.17	49/40 = 1.23	0.95
[0.05, 0.4] *vs.* [0.10, 0.2] *P*	4.58	34/12 = 2.83	1.62
[0.05, 0.4] *vs.* [0.05, 0.1] *P*	4.58	34/12 = 2.83	1.62
[0.05, 0.1] *vs.* [0.10, 0.2] *P*	1.00	34/34 = 1.00	1.00

6.2 How to adjust NK values for non-white noise

Consider two filtered non-white noise sequences and the problem of adjusting correlation window length and number of stacks when changing from one FBW to another in correlation studies. To give an approximate solution to this problem, below we first define the *equivalent* white spectrum for a given filtered non-white noise, and then, using the two equivalent spectra, we apply eq. 10.

Let $E(f)$ be the energy density spectrum of a non-white noise filtered in the FBW $[f_{min}, f_{max}]$. We define its equivalent white noise energy density spectrum $E_{eq}(f)$ as the box-shaped distribution reproducing the first two moments of $E(f)$, with the additional constraint of amplitude equal to 1. Then the parameters equivalent FBW (b_{eq}) and central frequency (f_C) of $E_{eq}(f)$ are given by:

$$b_{eq} \cdot 1 = \gamma_0 = \int_{f_{min}}^{f_{max}} E(f) df, \tag{19}$$

$$f_C = (1/\gamma_0) \int_{f_{min}}^{f_{max}} f\, E(f) df. \tag{20}$$

$E_{eq}(f)$ extends over the equivalent FBW $[f_C - b_{eq}/2, f_C + b_{eq}/2]$, which might be different from $[f_{min}, f_{max}]$. Thus each equivalent ratio n to be used in eq. 10 (or its approximation eq. 18) is given by

$$n = \frac{f_C + b_{eq}/2}{f_C - b_{eq}/2}. \tag{21}$$

6.3 Predicted and observed K-ratios

Along the line N=K (Figs 1 and 2) we can use the K-ratio K_B/K_A, associated to the respective values of K where the decay curves cross the same threshold ϵ, to measure the relative decay rate. Table 1 shows that reasonable ratios among the predicted (eq. 10) and observed (Figs 1e, 2d, and 2e) K-ratios are obtained in most cases both for synthetic and field data.

The largest differences between predicted and observed K-ratios are obtained for the field data (with and without preprocessing) when comparing very different relative FBWs (that is, when

$(n_B^2 - 1)/(n_A^2 - 1)$ is large). In this case, one possibility is that the threshold $\epsilon = 0.02$ is attained at values of N=K smaller than our prediction for the relatively narrow FBW, or when the opposite happens for the relatively wide FBW. We did not further investigate the reasons why predicted and observed K-ratios depart from 1, both for synthetic and field data. Nonetheless, a detailed study should analyze at least two causes: 1) failures in approximating a non-white noise spectrum by its equivalent white-noise spectrum and 2) whether spectral leakage associated with bandpass filtering can bias the results.