Supplementary material of

"A procedure to clean, decompose and aggregate univariate time series"

Part I: Outliers

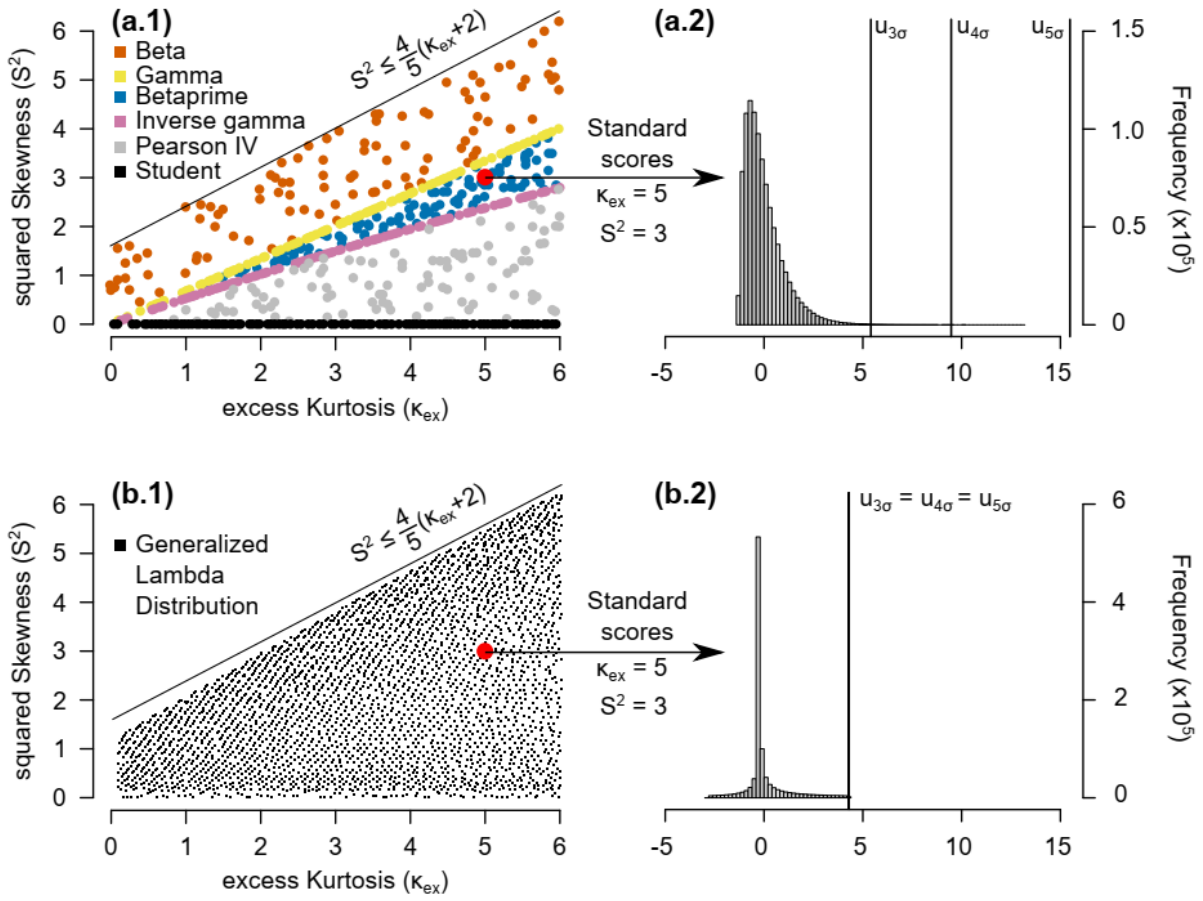

Fig. S1. Location of the 600 random distributions picked among the Pearson family in the (κ_{ex}, S^2) space (panel **a.1**). Example of the three upper thresholds based on the boxplot rule ($u_{3\sigma}$, $u_{4\sigma}$ and $u_{5\sigma}$) to detect outliers for a Betaprime distribution with a skewness of 5 and a kurtosis of $\sqrt{3}$ (panel **a.2**). The Generalized Lambda Distribution (GLD) system is shown for comparison in the (κ_{ex}, S^2) space (panel **b.1**). Similar figure than **a.2** except for a distribution coming from the GLD system with a skewness of 5 and a kurtosis of $\sqrt{3}$ (panel **b.2**).

- Figure 1 shows that the Generalized Lambda Distribution (GLD) system produces non-realistic outlier thresholds because each distribution from this system is bounded. The Pearson family is therefore preferred over the GLD system to model outlier behaviors.

Part II: the *past* procedure

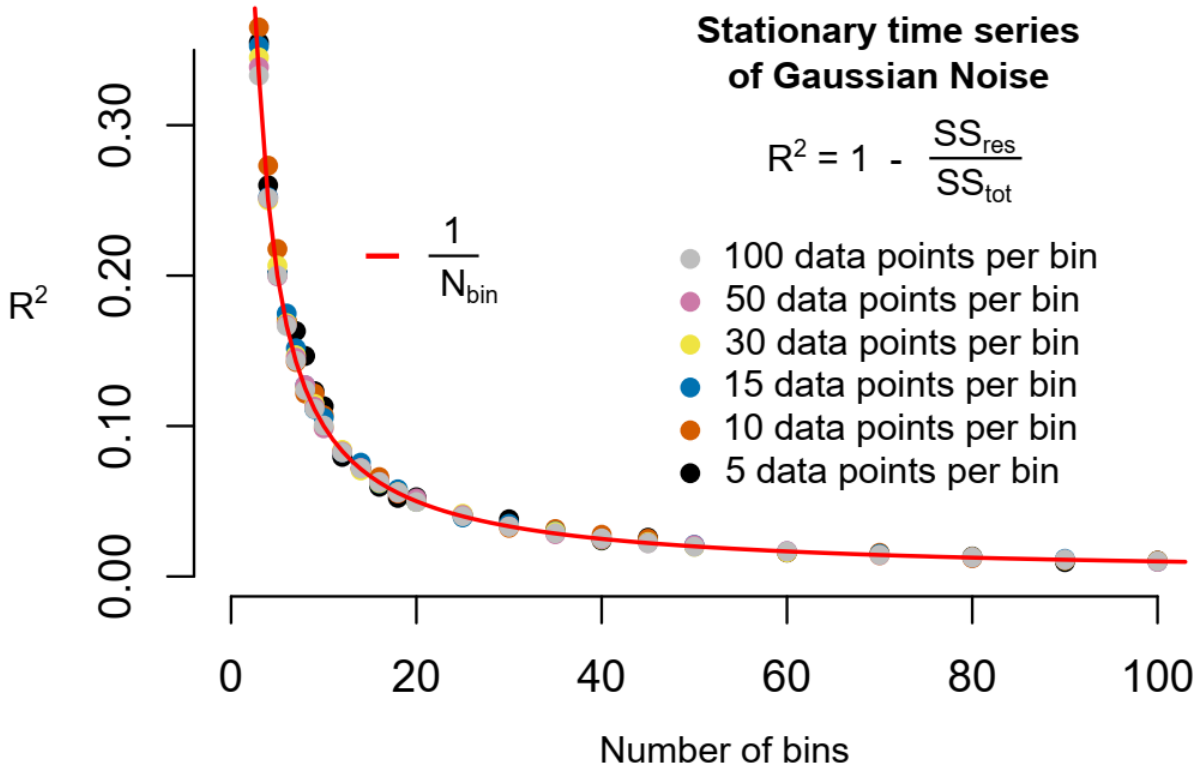

Fig. S2. The coefficient of determination (R^2) has been calculated for multiple stationary time series of Gaussian noise y, with $SS_{res} = \sum(y_i - S_i)^2$, $SS_{tot} = \sum(y_i)^2$ and S the cyclic component calculated with the *past* procedure.

- Considering Fig. S2, an inverse relationship appears between the coefficient of determination calculated on a pure Gaussian noise and the number of bin used (related to the sample size). This relationship is independent from the number of points per bin (illustrated by different colors). Theoretically, a stationary timeseries has a null cyclicity ($S = 0$, $R^2 = 0$). While this is observed for a large number of bins ($N_{bin} \gg 100$), a bias of ${N_{bin}}^{-1}$ exists at a smaller amount and needs to be corrected. This justifies the definition of the stacked cycles index as $SCI = R^2 - N_{bin}^{-1}$.