Appendices to: A Piecewise Gaussian Process Model for Evaluating Training and Performance of Time-Series Pipelines

May 7, 2025

1 Appendix

From the paper, we have the following model for the signal amplitude at time k:

$$s(k) \sim f_s(k|\mathcal{M}), k = 0, 1, 2, ..K - 1 \tag{1}$$

and we can describe the GP model as

$$f_s(k) = \mathcal{N}(\mu(k); \Sigma) \tag{2}$$

where Σ is the covariance matrix of additive noise.

1.1 Entropy

The entropy of signal amplitude at each time point is calculated across an ensemble of 100,000 signals. The entropy is calculated as:

$$H = -\sum_{i=1}^{N} p_i \log(p_i) \tag{3}$$

where p_i is the probability of the i^{th} amplitude value, and N is the number of unique amplitude values. The entropy of the signal amplitudes is shown in Figure 1. The high entropy suggests that the instantaneous signal amplitude at each time index is very variable, despite the model mean process, $\mathcal{M}$ having a well defined temporal origin. Amplitude levels were quantised into 100 bins after removing the mean signal amplitude and binning across the range of the resulting amplitude.

The theoretical limits represented with the dashed lines are, in each of the two noise cases, calculated by assuming uniform amplitude distribution between min and max levels, after removing the time-varying mean estimated across the ensemble of 100,000 sample signals. We suggest that for some use cases, changes should be made to the generator to ensure that the fluctuations in entropy-of-amplitudes are minimal, perhaps by changes to the distributions of temporal offsets, starting times or waveform width distributions.

1.2 Entropy

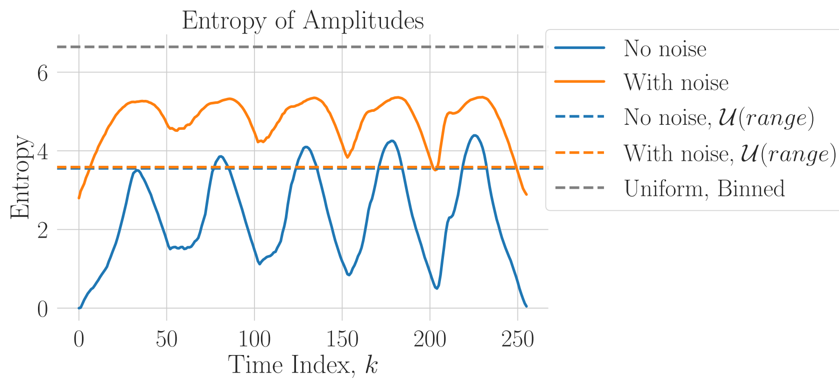

Figure 1: Entropy of amplitudes of signal waveforms calculated for each time point across an ensemble of 10,000 signals. The high entropy suggests that the signal amplitude is very variable, despite the model mean process, $\mathcal{M}$ having a well defined temporal origin. See main Appendix text, and main paper, for further details. The limits corresponding to uniform continuous distributions given the observed (empirical) amplitude limits are shown, as is the Shannon entropy limit for a uniform distribution of amplitudes obtained by a uniform distribution of amplitudes over bins.

1.3 Covariance Structure

The covariance matrix of the signal amplitudes is estimated from 100,000 realisations of the signal model. The eigenvalues of the covariance matrix are shown in Figure 2. The strong temporal structure in the signal model is evident from the eigenvalues, which decay relatively slowly. This suggests that the signal model has a strong temporal structure, which is not captured by the mean process, $\mathcal{M}$.

1.4 Histograms of Signal Amplitudes

The histograms of signal amplitudes at three time points over the ensemble are shown in Figure 3. The heavy tails and distinct one-sidedness of the distribution are evident, despite being outwardly Gaussian. This suggests that the signal model has a strong temporal structure, which is captured by the mean process, $\mathcal{M}$, which is itself generated by a Markov-N_w process.

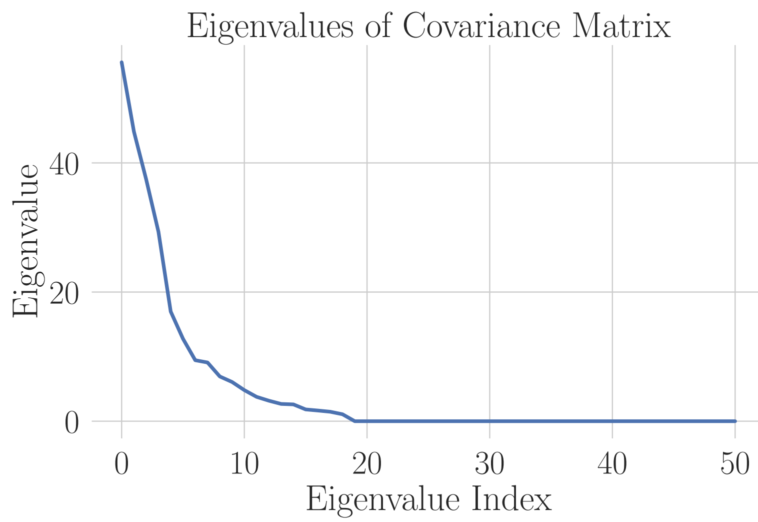

Figure 2: The eigenvalues of the covariance matrix of the signal amplitudes, estimated from 100,000 realisations; this suggests strong temporal structure in the signal model.

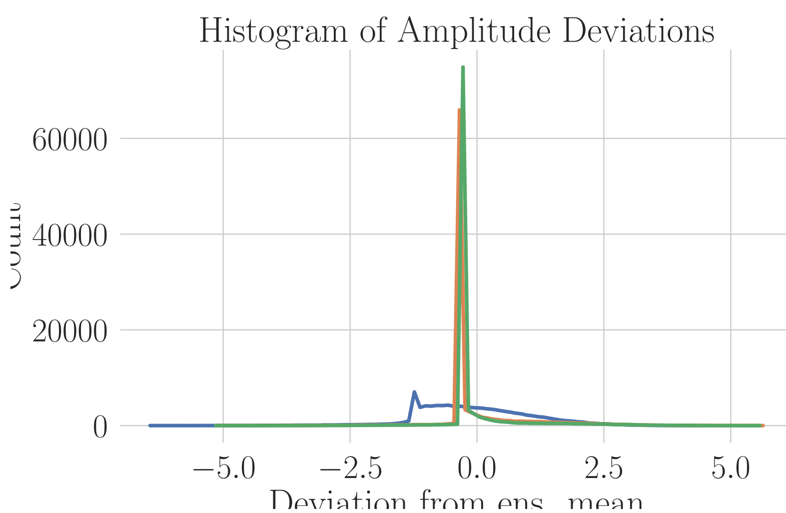

Figure 3: The histograms of signal amplitudes at three time points over the ensemble. Note the heavy tails and distinct one-sidedness of the distribution, despite being outwardly Gaussian.