Notes on MLE significance[[1]](#footnote-1)

November 2024

The goal of this exercise is to create cumulative histograms of changes in the Maximum Likelihood Estimator, δMLE, when comparing a more complex noise model with the simpler, null-hypothesis noise model. Here, I created a time-series with both 0.5 mm of white noise and 1 mm/yr0.5 random walk (RW) (also, in a separate exercise, 1 mm/yr0.25 flicker noise (FL)). This becomes the underlying noise model and is termed the ‘null hypothesis’. The time series consists of daily samples spanning 20 years. I used the program est\_noise to estimate the amplitudes of white noise and random-walk noise along with the rate (it should be zero). The value of MLE is saved (along with the other parameters including rate, and the amplitudes of the noise estimates). Next, I used est\_noise to estimate the parameters power-law noise (PL), noting that RW noise is a special case of PL where the index, n, of the power-law, 1/fn, is not fixed to 2. Again, MLE for this iteration is saved and compared with that from RW, thus δMLE. For PL, there is one addition degree of freedom, the power law index, relative to the null model. Next, est\_noise is used to compute the parameters for the generalized, Gauss-Markov (GGM) noise and its MLE is saved and compared with that from the null model. GGM has 2 more degrees of freedom relative to the null model and one more degree of freedom relative to the PL model. The final set of tests are for band passed-filtered noise (BP) with a range of 1 to 4 poles in combination with the RW model. The table provides a summary on the number of unknowns for each of the 5 noise models being tested.

|  |  |
| --- | --- |
| **Noise Model** | **Number of noise parameters** |
| Random Walk (RW) + white noise (WN) **NULL** | 2 |
| Power law (PL) + WN | 3 |
| Gauss-Markov + WN | 4 |
| Band passed with 1 pole (BP1), RW + WN | 3 |
| Band passed with 1 to 4 poles (BPx) + RW +WN | 4 |

The above set of calculations were run 1000 times using a different set of simulations of RW + WN noise. (or FL + WN). From these simulations, δMLEs were tabulated and sorted to provide plots of the cumulative histograms for each difference of MLEs. The expectation is that with more model parameters than the null model, then the MLE should be larger than that from the null model. Then from the cumulative histograms, the 68%, 95%, and 99% confidence levels can be identified for which the null model can be rejected in favor of the more complex model. These comparisons are carried out for PL-null, GGM-null, GGM-PL, (BP1+null)-null, and (BPx+null)-null. Figures 1 and 2 show these histograms where null is either RW in Figure 1 or FL in Figure 2.

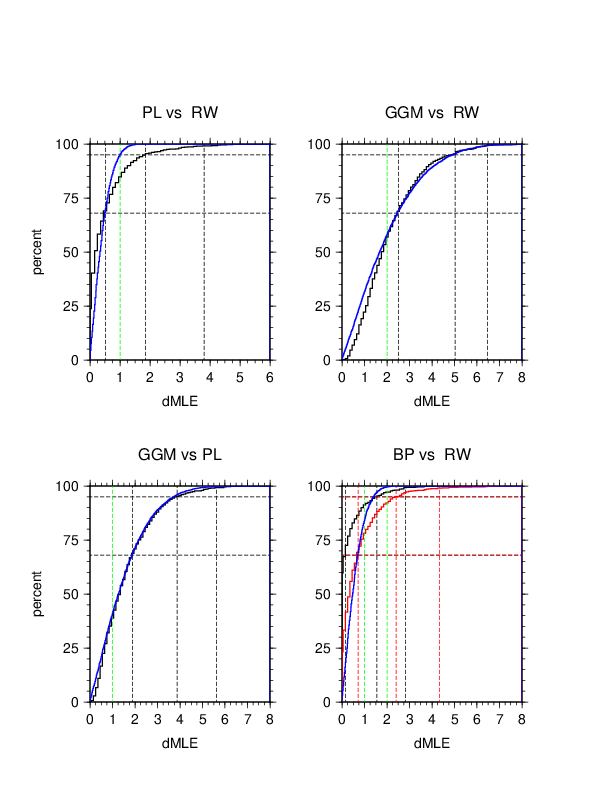
The upper left of Figure 1 shows the cumulative histogram of the PL model compared with the null, RW model. In each plot, I’ve identified the 68%. and 95%, levels as horizontal, black, dashed lines. The 68% level intersect the cumulative histogram when δMLE=0.51 indicating that 68% of the simulations had a δMLE that fell between 0 and 0.51. And, for 99% of the simulations, the change of MLE >= 3.80.

Figure . Change in MLE for PL, GGM, and BP noise models relative to the null, RW model of noise

Consequently, for any simulation that δMLE>3.80, that represents 0.01 probability that the null hypothesis (RW) is better than the PL model, or one can reject the null model in favor of the PL model. The black, vertical, dashed lines show where the δMLE where cumulative histogram matches the 68%, 95%, and 99% levels. This identification is carried through with the GGM vs RW, and the GGM vs PL comparisons. For the comparison of the BP models with RW, I’ve combined two different sets, the black cumulative histogram is the comparison with BP having only a single pole (BP1) and the red cumulative histogram is for the case where the poles can range from 1 to 4 (BPx). The identification of the 68. 95, and 99% levels of δMLE are color coded with the vertical black lines being those for BP1 and the red being for BPx.

For many of the comparisons of the BP models with RW, the computed change in MLE was negative indicating that est\_noise failed to correctly iterate to the appropriate value of the amplitude of BP noise; est\_noise should have converged to zero for the amplitude of BP noise which would have resulted in δMLE=0. Instead, for those negative values of δMLE, I change them to be zero when counting for the cumulative histogram. Consequently, the two cumulative histograms associated with BP noise have a large step at zero.

For reference, I plot in blue the cumulative histogram of a one-sided, normal distribution. That distribution has been normalized such that its 68% level matches that from cumulative histogram of δMLE. For the cases of GGM-RW and GGM-PL, the cumulative histogram of δMLE closely matches the distribution from a one-sided normal distribution. For the other two cases, PL-RW and BP-RW, the distribution of δMLE have heavier tails than that predicted by a one-sided, normal distribution.

Also, noted with a dashed, vertical green line is the point where δMLE corresponds to AIC=0. AIC has often been used as a metric to decide which noise model is most appropriate. For instance, with the comparison of PL versus RW, δMLE=1 corresponds to AIC=0 and, from the cumulative histogram, that corresponds to approximately 0.85 probablity that RW can be rejected in favor of PL. For the same comparison with GGM versus RW, δMLE=2 corresponds to AIC=0. This corresponds to an unimpressive 0.55 probability that RW can be rejected in favor of GGM. The point here is that AIC doesn’t consistently quantify the probability of selecting the “best” noise model.

Below, I tabulate the 68, 95, and 99% values of δMLE used for hypothesis testing.

|  |  |  |  |
| --- | --- | --- | --- |
|  | δMLE needed to reject the null hypothesis where underlying noise is **RW+WN** | | |
| Noise test | **68%** | **95%** | **99%** |
| PL-RW | 0.51 | 1.86 | 3.80 |
| GGM-RW | 2.51 | 5.00 | 6.47 |
| GGM-PL | 1.89 | 3.87 | 5.63 |
| BP1-RW | 0.16 | 1.54 | 2.81 |
| BPx-RW | 0.72 | 2.41 | 4.34 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | δMLE needed to reject the null hypothesis where underlying noise is **FL+WN** | | |
| Noise test | **68%** | **95%** | **99%** |
| PL-RW | 0.54 | 2.30 | 4.30 |
| GGM-RW | 2.51 | 4.78 | 6.64 |
| GGM-PL | 1.81 | 3.81 | 5.35 |
| BP1-RW | 0.16 | 1.50 | 3.25 |
| BPx-RW | 0.54 | 2.35 | 4.09 |

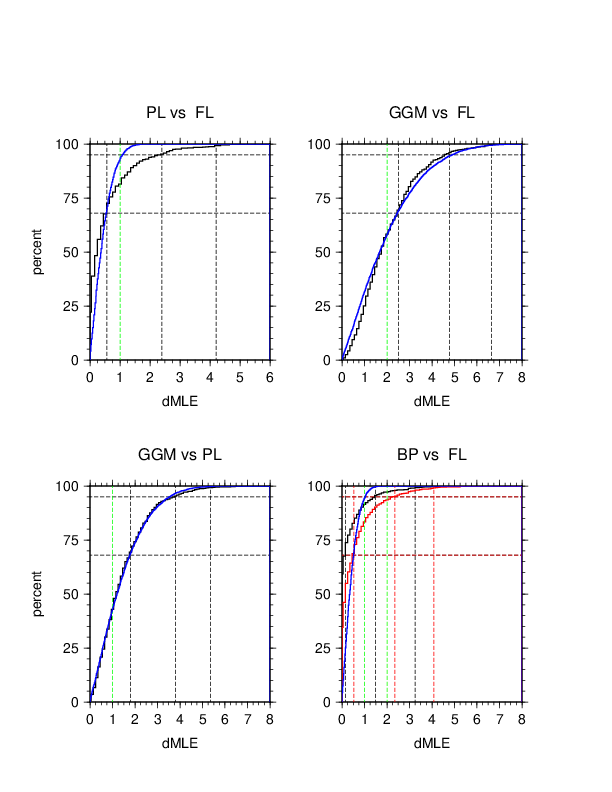


Figure . Change in MLE for PL, GGM, and BP noise models relative to the null, FL model of noise

Note that the cumulative distribution curves resemble each other in the two figures; there are some small differences in the δMLE levels for probabilities as noted in the two tables.

1. The script used for this exercise can be found /home/langbein/CREEP/Noise/SIMULATAE01on thepub. There are three scripts; simulateGGM08.sh does the actual simulations, loop\_simGGM08.sh specifies the core and the number of iterations, and BigLoop.sh fires off each “loop” onto specific core. [↑](#footnote-ref-1)