Workable Model of Non-Gaussian Noise

Roberto Conte, Maurizio Longo, Stefano Marano
Vincenzo Matta, Innocenzo M. Pinto, Maria Principe

LSC-VIRGO Meeting, Hannover, 22-25 October 2007

Albert-Einstein-Institut
Hannover

Goal, Motivation, Rationale

Motivation : after power lines, violin modes and glitch removal, IFO noise is a “breathing Gaussian”.

Rationale : spherically invariant (aka exogenous-Gaussian) is possibly the *simplest* available model capturing *short-time-gaussian (non – stationary)* noise features but still manageable as concerns

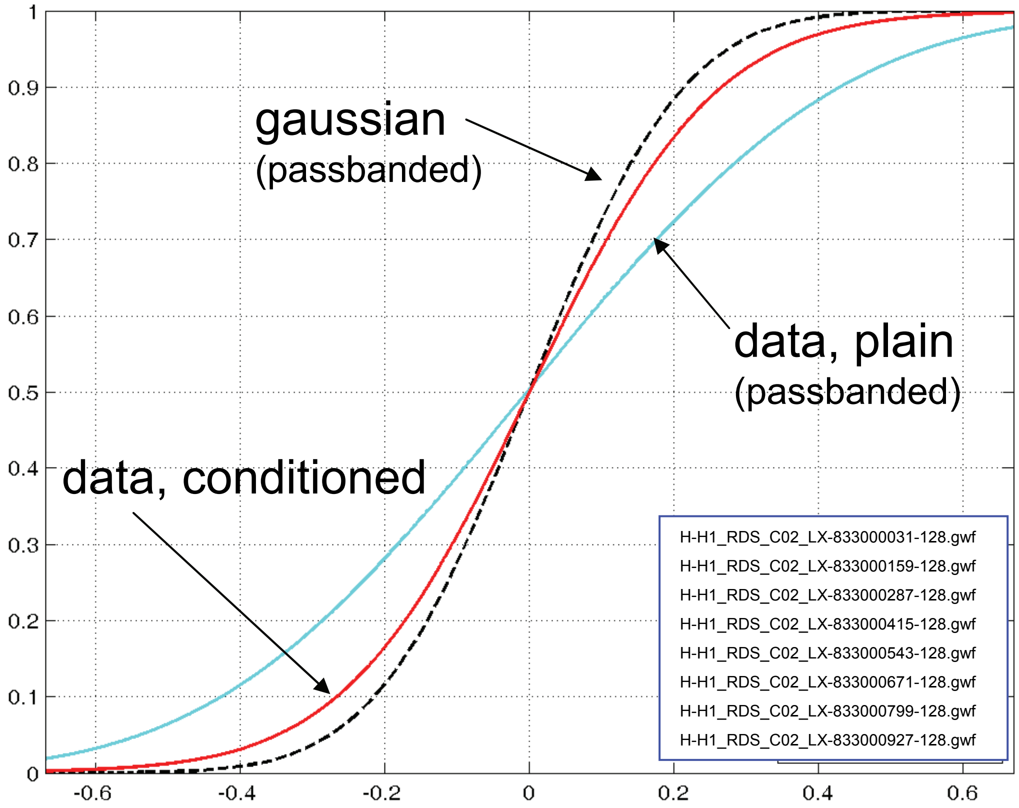
i) detector optimization;
ii) noise simulation;

Goal : building and validating a spherically invariant (aka exogenous-Gaussian) model of LIGO noise floor;

Data

H-H1_RDS_C02_LX_833000031-128.gwf
H-H1_RDS_C02_LX_833000159-128.gwf
H-H1_RDS_C02_LX_833000287-128.gwf
H-H1_RDS_C02_LX_833000415-128.gwf
H-H1_RDS_C02_LX_833000543-128.gwf
H-H1_RDS_C02_LX_833000671-128.gwf
H-H1_RDS_C02_LX_833000799-128.gwf
H-H1_RDS_C02_LX_833000927-128.gwf

LIGO (S5 sample) Noise Floor CDF



Work on IFO Noise

Median based noise-floor tracker
[Mukherjee, CQG 20 (2003) S925];

ARMA modeling of IFO noise-floor
[Mukherjee, CQG 21 (2004) S1783
Mukherjee LIGO G040361-00-Z (2004)];

Change-Point detection
[Mohanty, PRD 61 (2000) 122002
Mohanty & Mukherjee, CQG 19 (2002) 1471
McNabb et al., CQG 21 (2004) S1705
Mohanty & Jimenez, CQG 22 (2005) S1233].

Exogenous Gaussian (EG) Noise in a Nutshell

Compound (exogenous) gaussian RP

$$x(t) = s(t)\,g(t)$$

long-coherency RP

N(0,1) gaussian RP

- *"Long"* $x(t)$ time-series markedly *non - gaussian;*
- *"Short"* time-series *locally gaussian*, but with *different variance* in different time stretches (*nonstationary*)

…depending on time-scales, one may trade gaussianity for stationarity (to some extent …)

[I.M. Pinto, LIGO-G-060473-00]

Getting the Noise Floor

Passband filtering - (FIR, L=3518) chop spectra outside band of interest (40-300 Hz);

Spectral equalization -equalize spectral data using smooth-ed PSD estimate;

Narrowband features detection - Use Kay test for tones (assumes $\chi2$ distribution in each spectral bin);

Narrowband feature removal - Estimate amplitude, frequency, phase of detected narrowband features using a sliding window. Subtract estimated tones in time domain from next window;

Subsampling – Estimate correlation length, and whiten (decorrelate) data by subsampling.

Gaussianity Tests

EDF based (generic) :

Kolmogorov-Smirnov, Anderson-Darling

(*composite* hypothesis test, distribution parameters estimated *from data*)
[Dagostino & Stephens, *Goodness of Fit Techniques*, Dekker, 1986, ch.4]

Skewness/Kurtosis based (gaussian-specific)

Jarque-Bera [Intl. Stat. Rev., 55 (1987) 163.]
Urzua [Economics Lett., 53 (1996) 247]

Rank based (specific)

Shapiro & Co-workers
[Biometrika, 52 (1965) 591; J. Am. Stat. Soc., 67 (1972) 215];
Royston [The Statistician, 42 (1993) 37]

Local Gaussianity vs. Chunk Size

Chunk fraction for which Gaussianity is rejected at 5% error level				
Chunk size	**Kolmogorov.Smirnov (composite hyp.)**	**Anderson-Darling (composite hyp.)**	**Urzua (mod. Barque-Jera)**	**Shapiro-Francia**
256	0.0510	0.0488	0.0427	0.0434
1024	0.0566	0.0517	0.0889	0.0839
4096	0.0781	0.0898	0.1367	0.1328
8192	0.0547	0.1092	0.2896	0.2657
32768	0.1875	0.2812	0.5937	0.6250
131072	0.625	0.875	1.000	1.000

At $N_C = 2^{17}$ (2 min), Urzua & Shapiro-Francia (*gaussian- specific tests*) reject gaussianity for *all* chunks;

For $N_C \leq 2^8$(1/4 s), The fraction of chunks rated as non-gaussian is consistent with (less than) test significance (5%).

Residual (Short) non-Gaussian Chunks

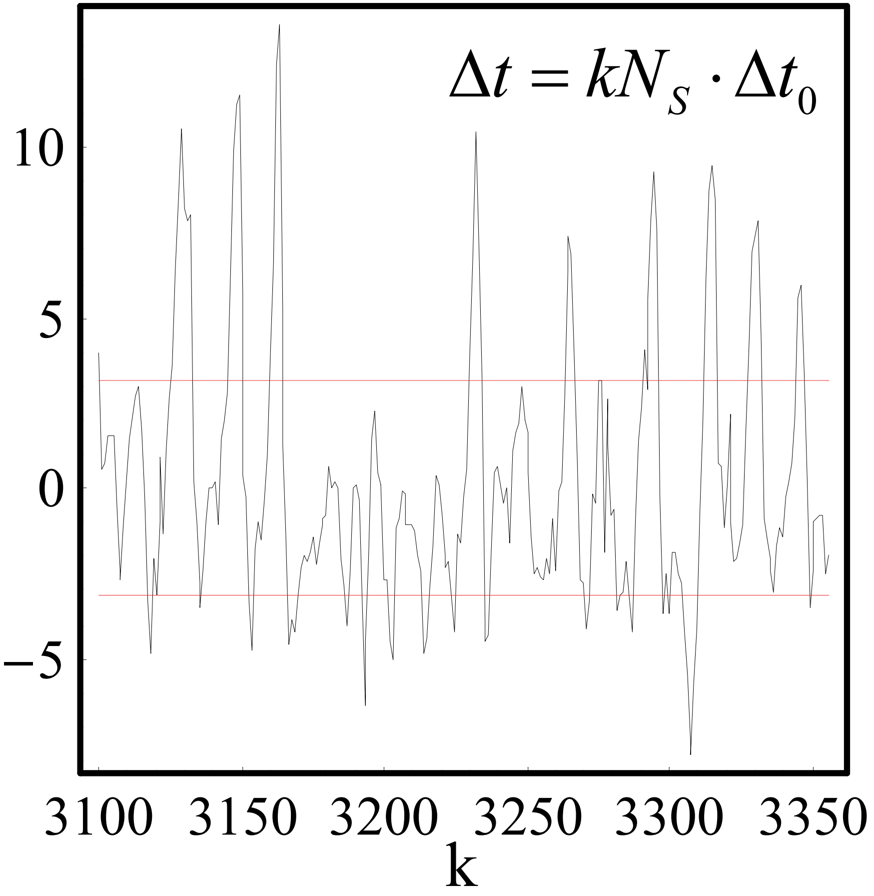
- A fraction of “short” chunks don’t pass *gaussianity* tests. They contain abrupt (*non adiabatic*) jumps in the exogenous factor;

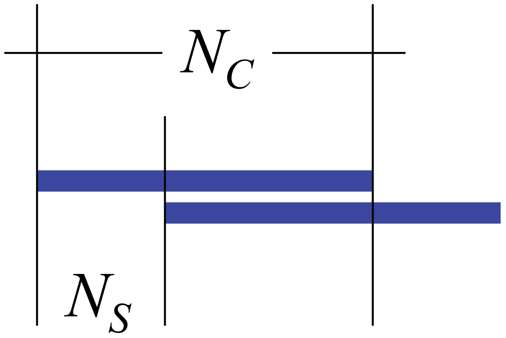
- Jumps can be finely located using *change-point* detection algorithms (e.g., blocknormal, or akin);

- It makes sense to look for *correlations* between these abrupt-changes, and data from environment-channels;

- EG models may *only* apply on the left/right of the above non-adiabatic changes.

The Exogenous Factor Robust Estimator

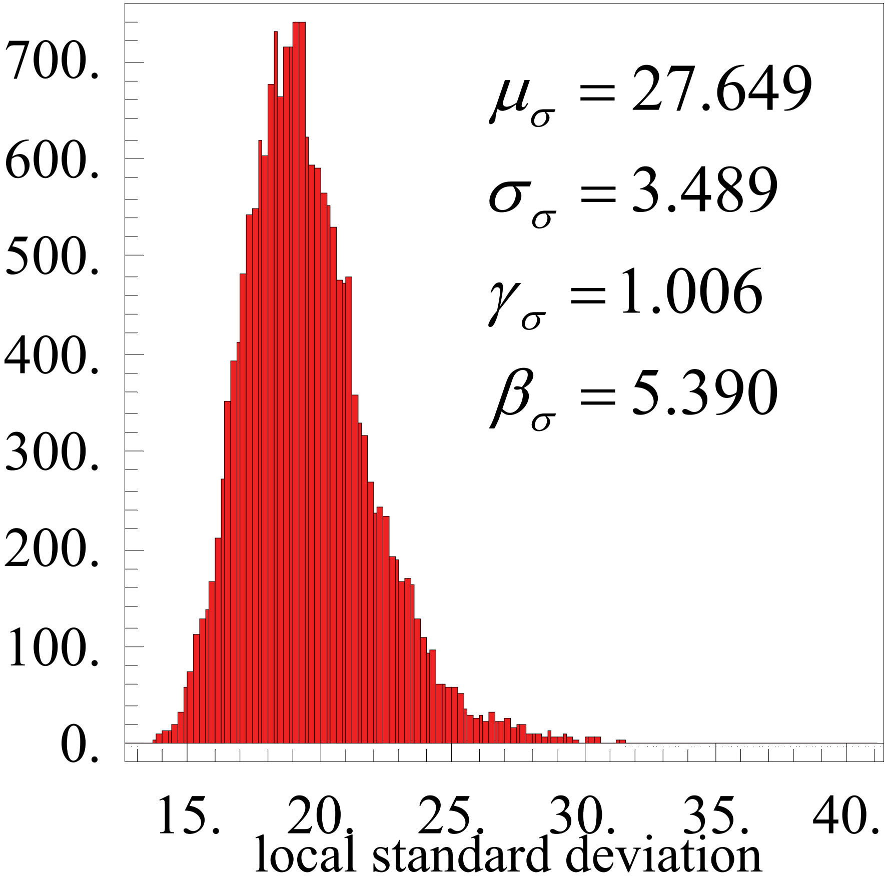




$$N_C = 256$$

$$N_S = 64$$

The Exogenous Factor Binning & Moments

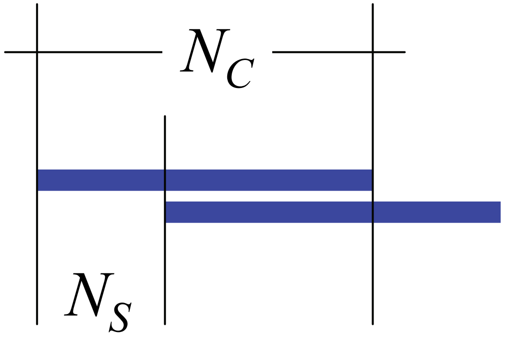


$$\mu_\sigma = 27.649$$

$$\sigma_\sigma = 3.489$$

$$\gamma_\sigma = 1.006$$

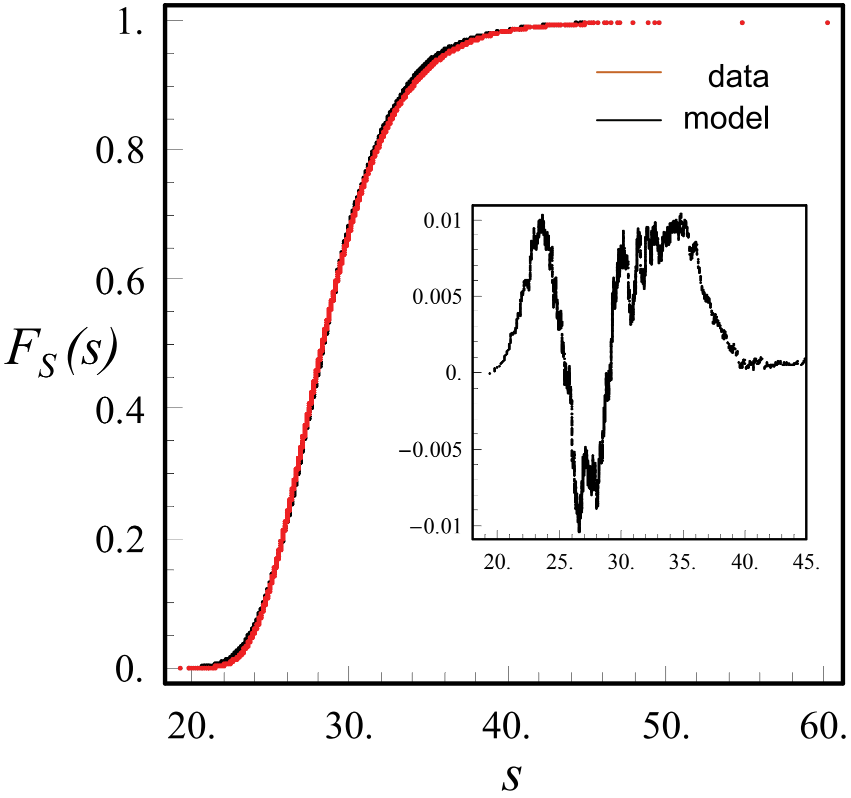
$$\beta_\sigma = 5.390$$



$$N_C = 256$$

$$N_S = 64$$

The Exogenous Factor Modeled (CDF)



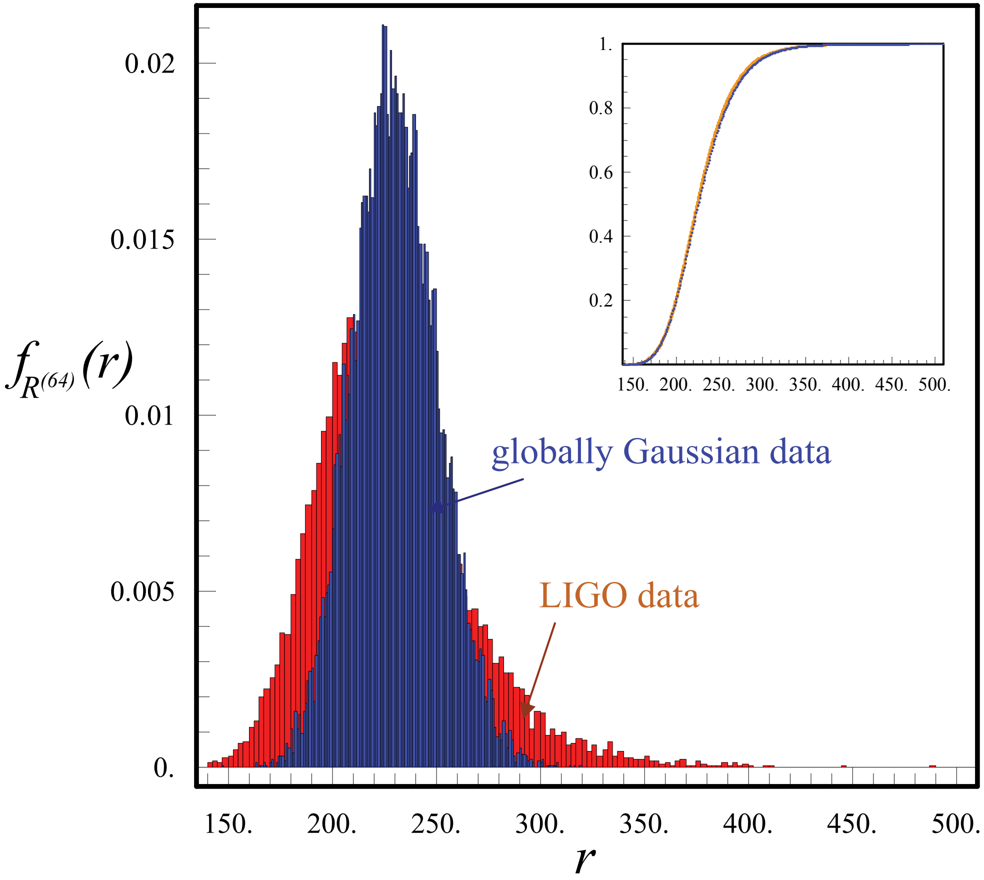
Fisher-Tippett distribution model:

$$CDF(x) = Exp\left[-Exp\left(-\frac{x-\mu}{\beta}\right)\right]$$

$$\hat{\mu} = 26.116\,, \quad \hat{\beta} = 2.803$$

(KS test passed at $\alpha \leq 0.05$)

Rangaswamy’s EG-Test



Noise N - tuples = coords. in R^N.

Distribution of radial distances in R^N related to exogenous factor distribution by :

$$f_{R^{(N)}}(r) = \frac{2^{-N/2+1}\;\; r^{N-}}{\Gamma\left(\frac{N}{2}\right)} \cdot$$

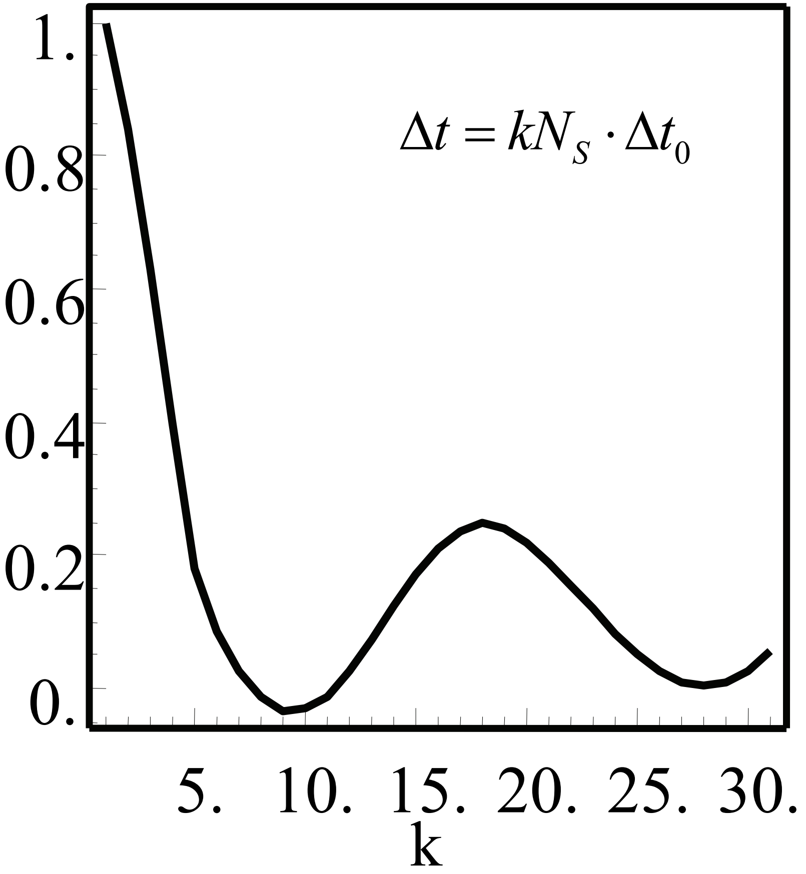
$$\cdot \int_0^\infty ds\;\; s^{-N} \exp\left(-\frac{r^2}{2s^2}\right) \widehat{f}_S(s)$$

for all N such that the noise is *locally* Gaussian …

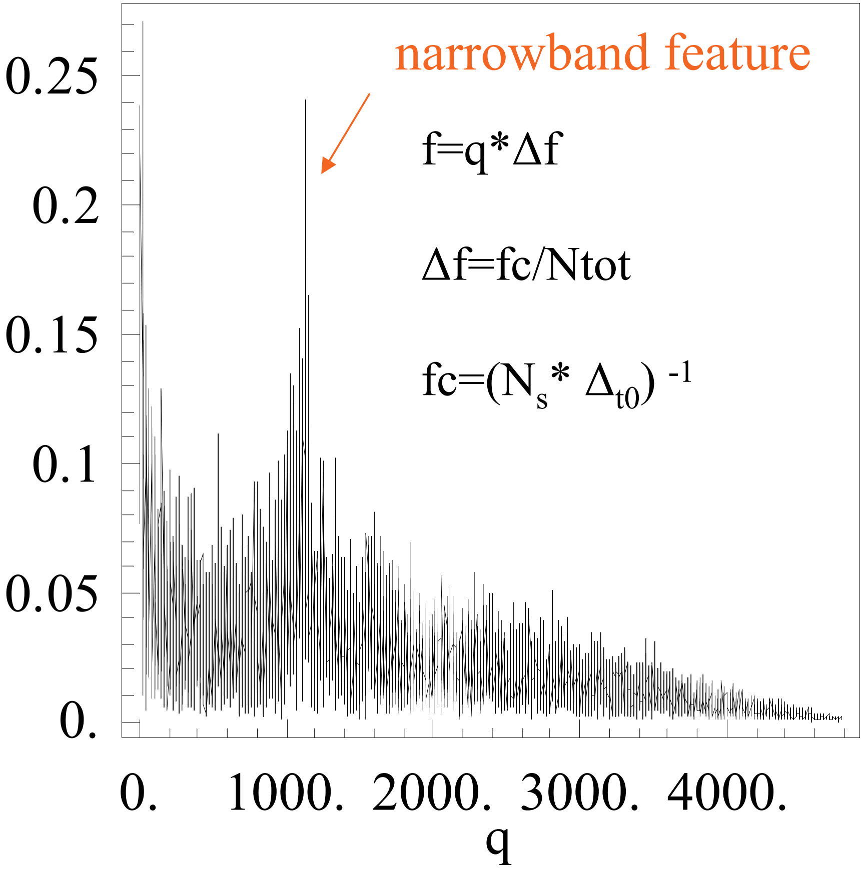
Noise Floor Simulator

- Generate a N(0,1) bandlimited AWGN time series $\{G_k, k=1,\ldots,N\}$
- Generate a vector of N/N_c exogenous process samples;
- Interpolate this latter to mimic the the spectral roll-off (and/or correlation length of the exogenous factor in the real data;
- Re-sample the resulting continuous waveform at the same rate as the Gaussian factor, ending up in a time series $\{E_k, k=1,\ldots,N\}$
- Form the time series $\{G_k E_k, k=1,\ldots,N$.

The Exogenous Factor Correlation & PS



$\Delta t = kN_S \cdot \Delta t_0$



narrowband feature

$f=q*\Delta f$

$\Delta f=fc/Ntot$

$fc=(N_s* \Delta_{t0})^{-1}$

LIGO Relevance for Detector ?

Likelihood ratio & decision rule

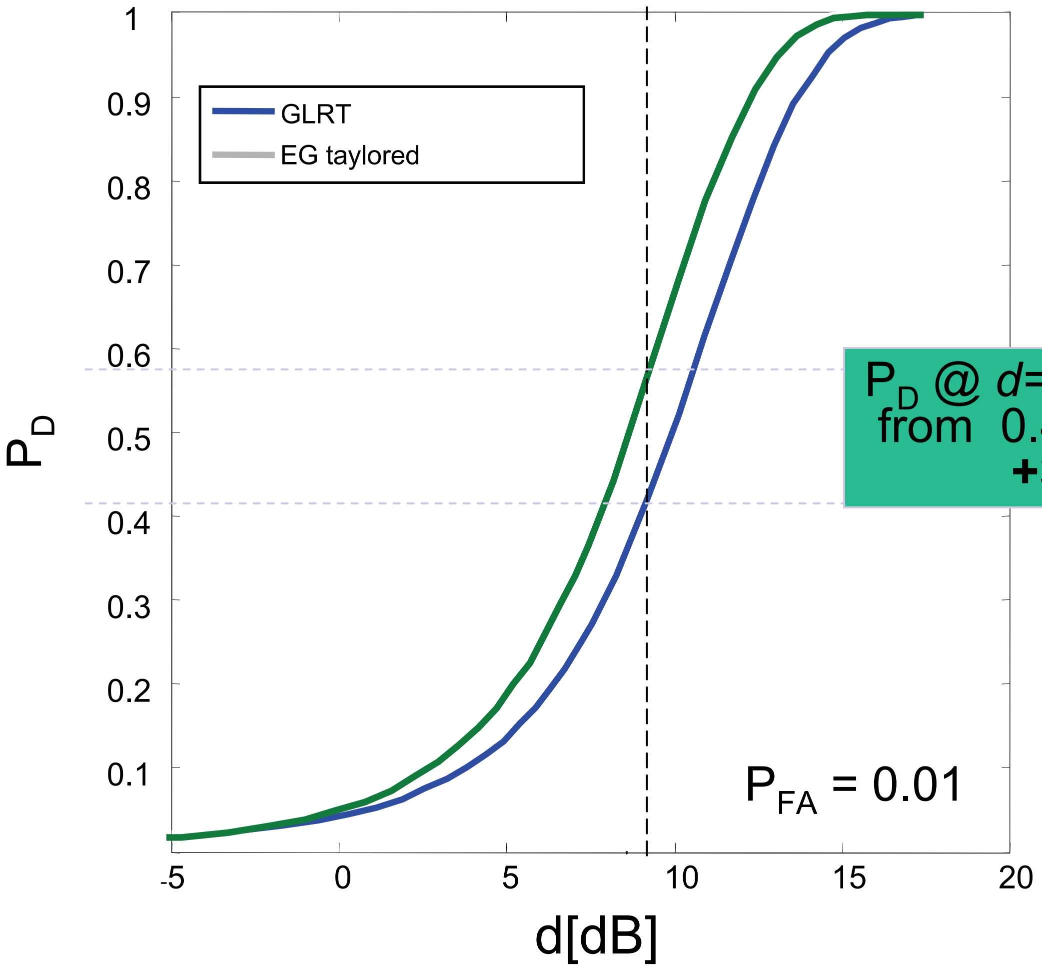
$$\Lambda = \frac{f_{(\vec{x}^{(N)}|H_1)}(\vec{x}^{(N)}_{obs})}{f_{(\vec{x}^{(N)}|H_0)}(\vec{x}^{(N)}_{obs})} :: \eta$$

Alternatives

- GLRT detector : estimate *local* variance and proceed (*heuristic* can be misleading, e.g., if there is a signal !);
- EG-detector : use known distribution of local variance to *weigh* the (conditional) statistical densities in the likelihood ratio.

GLRT vs EG Detector

Average, 10^5 CW injections equispaced in [40-300] Hz actual LIGO(S5) Noise Floor, EG-detector based on model



P_D @ *d=8* boosted from 0.42 to 0.58 **+38%**

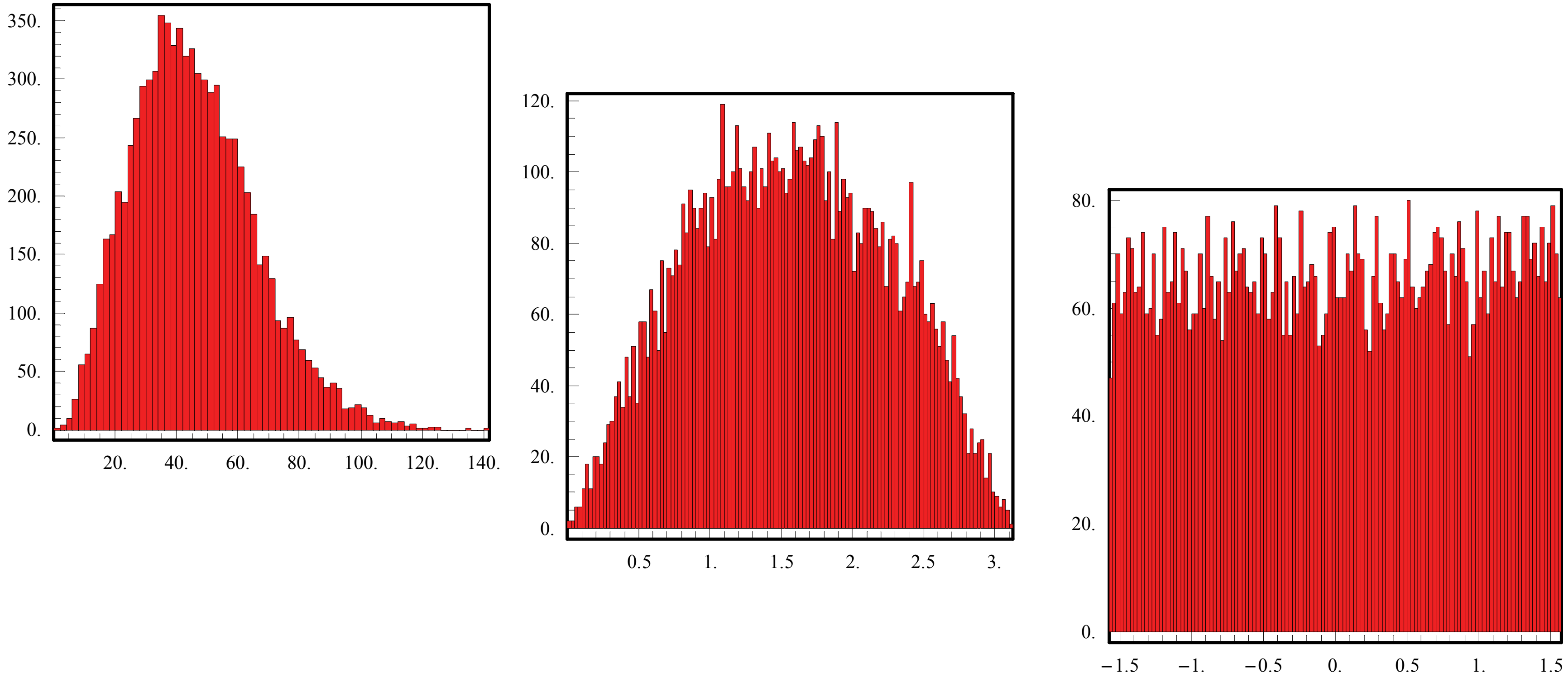
Conclusions/Directions

Exogenous Gaussian model for LIGO noise floor developed (limited amount of data available; fully functional modeling SW developed; extended tests planned);

MATLAB code implementing our exogenous gaussian LIGO noise floor *simulator* underway; extended model including glitches under investigation (Maria’s burst talk);

Performance testing of low-cost EG - noise - taylored detectors for signal classes of specific interest (e.g., chirps) in progress.

LIGO Noise Floor SIRP Behaviour in R^3



Triplets of LIGO (H1) noise-floor samples as points in R^3 distributions (histograms) of R, θ, φ.