Accidental background estimation for coherent network analyses

G.A.Prodi, *INFN and University of Trento,*
M.Drago, S.Klimenko, G.Mazzolo, G.Mitselmakher, V.Necula,
C.Pankow, V.Re, F.Salemi, G.Vedovato, I.Yakushin

UNIVERSITY OF TRENTO - Italy
Department of Physics

INFN

Scientific Motivations

Detection dilemma in "single shot" observations:

confidence has at least two sides

A. **establish confidence of *on-source* measurements against the *off-source* by frequentist statistical methods**

- ⇒ **goal is to exclude an accidental origin of the on-source result**

B. **evaluate confidence by folding in all our additional knowledge after the fact with the widest possible agreement in the community**

- ⇒ **evidence to discriminate among possible sources of the result**
- ⇒ **additional confidence on the non accidental origin ? difficult**

Must do our best on side A, life can be very controversial on side B

How to build *on-source* estimator from *off-source* measurements ?

G.A.Prodi, 27-Jan-2010, GWDAW14

Testing On-source vs Off-source

- **transient signal searches require to**
  - design the counting experiment
  - build the **reference distribution of accidental events**

    → **off-source reference**

    understand uncertainties ...
  - select test statistics (e.g. Signal-to-Noise Ratio, other)
  - find on-source results (issue of search blindness...)
  - rank on-source results against accidental reference

    → **estimate the false alarm rate**

- **standard time slides technique**

  ↻ **time shift data of detectors in the network**
  repeat the analysis

  → **reference distribution for accidental events**

- critical issues:
  - ✓ biases in off-source reference
  - ✓ uncertainties

Time slides

Common prescriptions:

✓ autocorrelation time of single detectors

⮡ *minimum time shift step O(1s)*

✓ non stationary timescale of single detectors

⮡ *maximum time shift O(1h)*

} **limit number of time slides**

✓ check for pollution by foreground or signal events in the network

- **time coincidence searches:** time-shift events

see Poster by M. Was

shift step > { **coincidence window**, **event clustering time**

same coincidences cannot repeat in *different* time slides

by construction ⇒ time slides give independent events

- **coherent searches:** time-shift data streams

same network event may repeat itself with negligible differences in *different* time slides (multiple events) ⇒ correlation among different time slides is possible even with independent detector noises

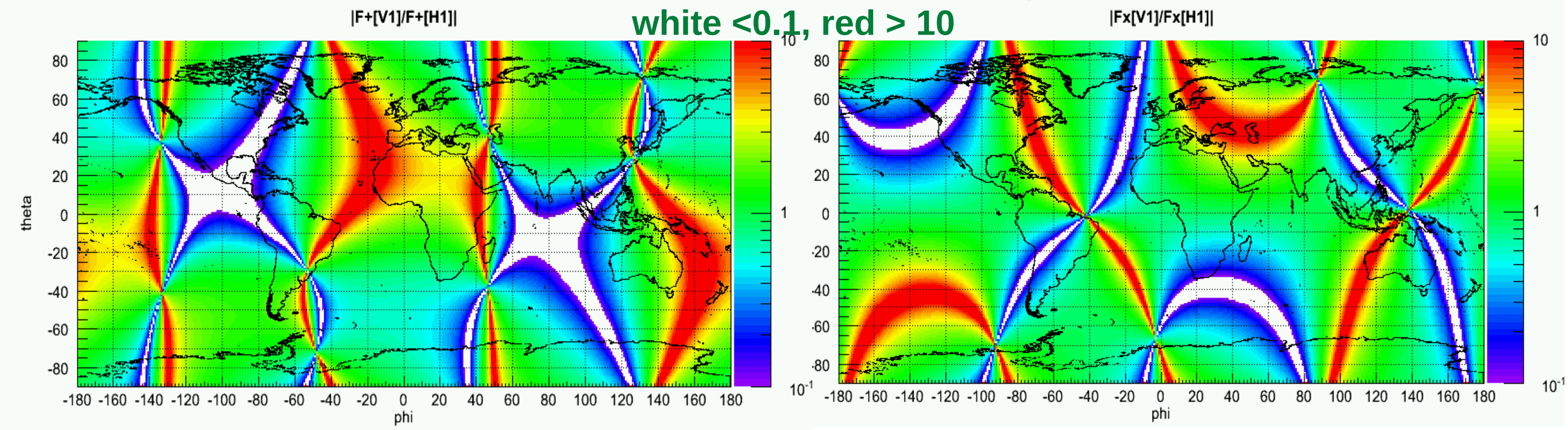
Time slides in Coherent Searches

example: all-sky searches with LSC-Virgo detectors

- sensitivity of detectors changes a lot according to direction & polarization.

RATIO of ANTENNA PATTERNS Virgo/LHO

white <0.1, red > 10



- coherent analyses **weigth each data stream** according to **directional** and **spectral sensitivity** of detectors

⇒ a full range of possibilities between two extremes

network events are not repeated in different lags:
"independent lags"
(unique events)

⟺

same network events show up in different lags:
"highly correlated lags"
(multiple events)

Multiplicity of Background

- multiple **background outliers** - **set of correlated network events produced by the same underlying event**

  e.g. for 3 detectors network: same pair of "parent" glitches in any 2 detectors may produce outliers in different time slides

- **count their multiplicities** as a function of the threshold on the chosen ranking statistic

- **best case: independent outliers min multiplicity, *m* = *1***

n_{bkg} background outliers in N_{lag} lags

⇒ expected counts for on-source:

$$\hat{n}_0 = \frac{n_{bkg}}{N_{lag}} \quad , \quad \hat{\sigma} = \frac{\sqrt{n_{bkg}}}{N_{lag}}$$

all lags are effective in improving background estimation

- **worst case: max multiplicity of outliers ⇒ *m* = N_{lag}**

$n_{bkg} = p\, m = p\, N_{lag}$ *p # of parents*

$$\hat{n}_0 = \frac{n_{bkg}}{N_{lag}} = p \quad , \quad \hat{\sigma} = \frac{m\sqrt{p}}{N_{lag}} = \sqrt{p}$$

equivalent to perform just 1 lag

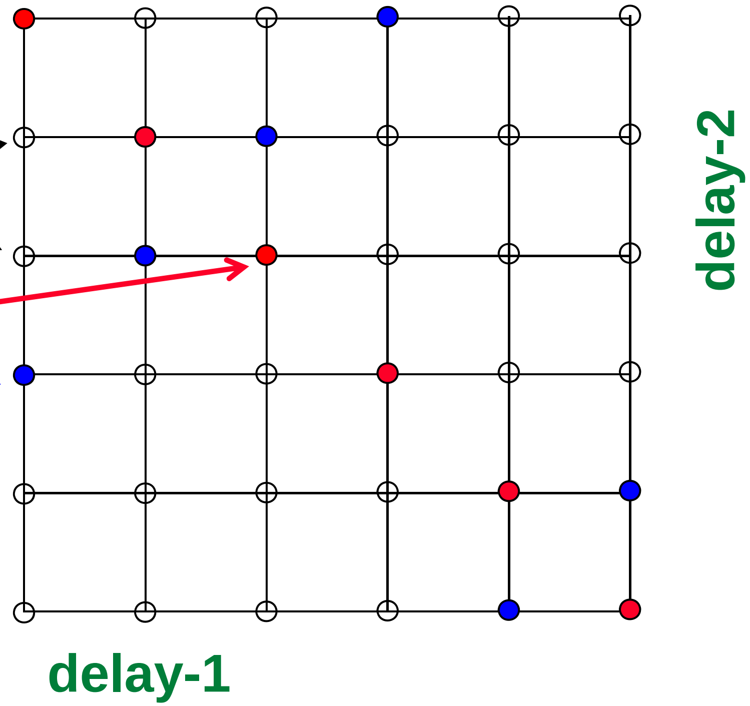
background estimation does not depend on N_{lag}

G.A.Prodi, 27-Jan-2010, GWDAW14

Best case: "unique lags"

Smarter choices of time slides: lower multiplicity ↔ smaller σ

- **set of unique lags:** never repeat relative delays between the same pair of detectors in the non zero lags ⇒ lags of the set are independent
  - PRO: no multiple network events ⇒ BEST USE of LAGS
  - CON: limited number of lags; for 3-detectors network ≈ few 1000s
  - build large background samples with low multiplicity by combining several sets of disjoint unique lags



examples of sets of unique lags:

o **grid of possible different lags for a 3 detectors network**
(2 independent time delays)

● **a set of unique lags**

● **another set of unique lags**

the two sample sets are disjoint, but resulting accidentals are in general correlated between the sets

Time Slides in Practice

intermediate cases:

n_{bkg} background outliers in N_{lag} lags

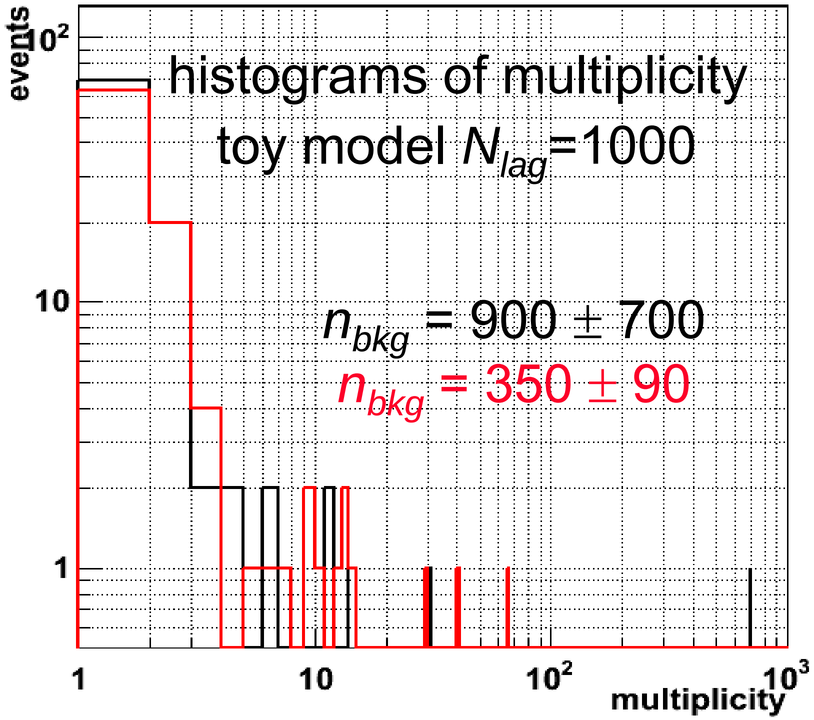
$$n_{bkg} = \sum_{j=1}^{p} m_j$$

p # of independent parents

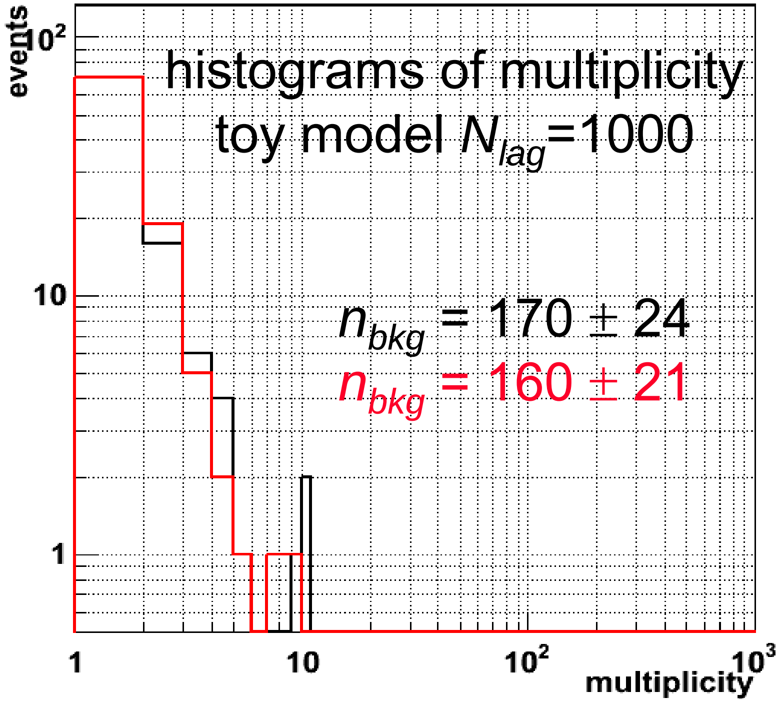
m_j multiplicity of family j

$$\boxed{\hat{n}_0 = \frac{n_{bkg}}{N_{lag}} \quad , \quad \hat{\sigma} = \frac{1}{N_{lag}} \sqrt{\sum_{j=1}^{p} m_j^2}}$$

- **shift only one “weak” detector**



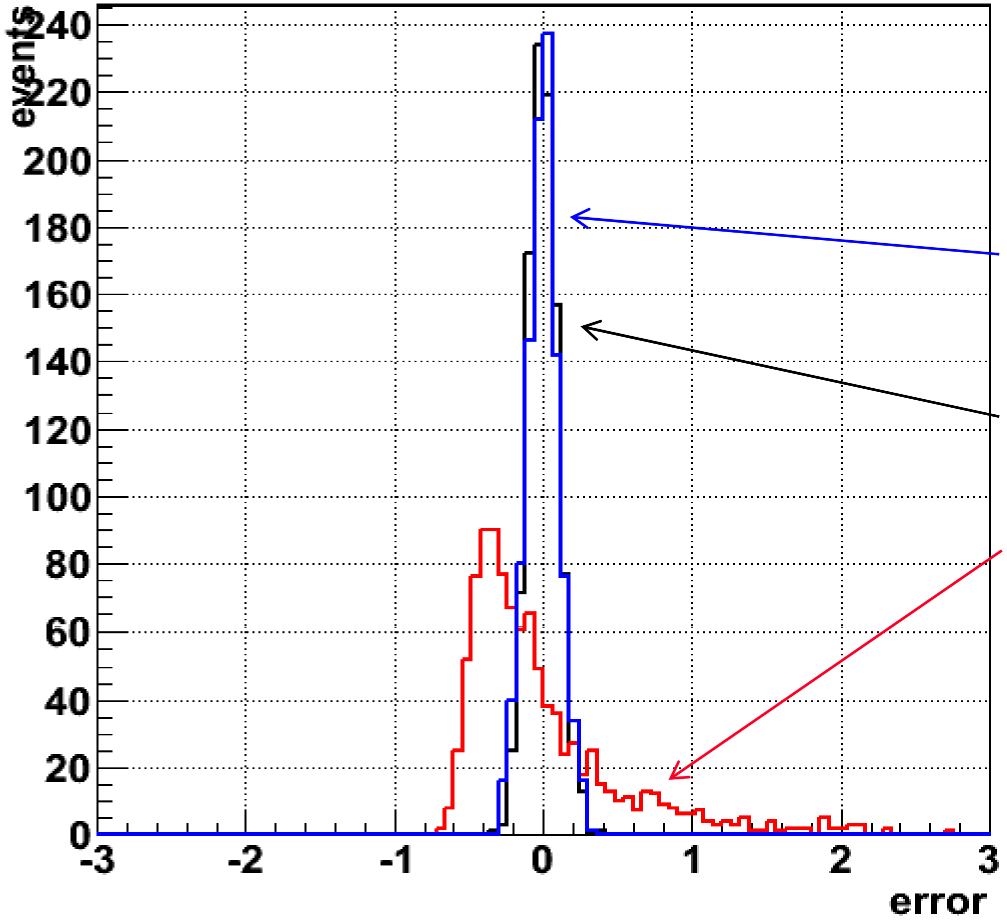
- **10 sets of 100 unique lags**



multiplicity and estimation

Smarter choices of time slides: lower multiplicity ↔ smaller σ

Histogram of rms of relative uncertainties on accidental event counts



toy model:

1000 simulations of accidental counts with mean = 100

- 1000 unique lags, m=1 : Poisson behavior
- 10 disjoint sets of 100 unique lags, m≤10 : almost Poisson
- 1000 lags with higher multiplicity, e.g. shifts of “weak” detector, larger sigma and asymmetric (right tail produced by rare events with high multiplicity)

One More Choice: Random Lags

- **select lags randomly** with uniform prob. from the set of possible lags
  - approaches the efficiency of unique lags: small multiplicities

    effective # of lags available is approx N_{lag} / (mean multiplicity)
  - allows to produce large background data samples
  - uniform sampling of the time slides space ⇒ robustness against systematics

- no bias in the background estimation because lags are selected in a blind way

- both unique lags and random lags are currently implemented in coherent WaveBurst pipeline

Summary

- estimation of the accidental background of coherent data analysis methods poses a new issue: **time slides can be correlated**
- **correlation can be measured** by counting the **multiplicity** by which the same parent events generate more network events in different time slides
- **ultiplicity increases the statistical uncertainty** of the **accidental background estimates**, without adding a new source of bias if time slides are performed in a blind way.
- **the choices of the lag set affect the background multiplicity:**
  - **unique lags**: independent but limited number
  - **more disjoint sets of unique lags**: larger statistics available
  - **set of random lags**: low multiplicity, highest statistics is possible, uniform sampling of the lag space