STATISTICAL DETECTION AND CLASSIFICATION OF TRANSIENT SIGNALS IN LOW-BIT SAMPLING TIME-DOMAIN SIGNALS

GELU M NITA*, AARD KEIMPEMA**, ZSOLT PARAGI**

*NEW JERSEY INSTITUTE OF TECHNOLOGY
**JOINT INSTITUTE FOR VLBI ERIC

2018 6th IEEE Global Conference on Signal and Information Processing (GlobalSIP 2018), Anaheim, CA, November 26–29, 2018
A well-known radio data analysis challenge
And it's not so widely-known statistical solution...
The Generalized Spectral Kurtosis Estimator
Nita and Gary 2010, MNRAS 406 L60-L64

**Theorem:** Given that, for a particular signal, the set of its power estimates \( P_k \) obeys a gamma distribution characterized by the shape parameter \( d \), the infinite series of statistical moments \( MS_2/S_1^2 \), were \( S_1 = \sum_{k=1}^{M} P_k \) and \( S_2 = \sum_{k=1}^{M} P_k^2 \) is given by

\[
E\left[\left(\frac{MS_2}{S_1^2}\right)^n\right] = \frac{M^n \Gamma(Md)}{\Gamma(d)^M \Gamma(Md+2n)} \times \left. \frac{\partial^n}{\partial t^n} \left[ \sum_{r=0}^{n} \frac{1}{r!} \Gamma(2r+d) t^r \right] \right|_{t=0}^M
\]

The Generalized Spectral Kurtosis Estimator:

\[
SK = \frac{Md + 1}{M - 1} \left(\frac{MS_2}{S_1^2} - 1\right)
\]

**Statistical properties of the SK estimator:**
- Has an **unbiased** unity expectation \( E[SK] = 1 \), independent of the integrated power \( S_1 \)
- The infinite series of statistical moments of its PDF are analytically defined only in terms of \( M \) and \( d \)

The SK estimator is well suited for detecting mixed signals not obeying the same gamma probability distribution:
Detection thresholds of deviation from unity characterized by **analytically defined probabilities of false alarm (PFA)**
Practical cases well suited for SK analysis

- Raw power estimates based on time domain real signals
  - Gamma distribution of shape factor $d=0.5$ (Chi-Square distribution)
- Raw power estimates based on time or frequency domain complex signals
  - Gamma distribution of shape factor $d=1$ (Exponential distribution)
- Accumulations of $N$ raw power estimates of shape factor $\delta$
  - Gamma distribution of shape factor $d=N\delta$
- Power estimates based on quantized time domain signals or quantized frequency domain power estimates (Nita, Gary, and Hellbourg 2017, IEEE)
  - Gamma distribution having an instrument-dependent shape factor $d$
The Spectral Kurtosis Spectrometer

Nita et al. 2007 PASP, 119, 805

The unbiased Spectral Kurtosis Estimator

\[ P_{FB} = R^2 + I^2 \]

\[ S_1 = \sum_{k=1}^{M} P_k \]

\[ S_2 = \sum_{k=1}^{M} P_k^2 \]

\[ SK \equiv \frac{M + 1}{M - 1} \left( \frac{MS_2}{S_1^2} - 1 \right) \]

\[ E(SK) = 1 \]

\[ \sigma^2(SK) \approx \frac{4}{M} \]
## Expanded Owens Valley Solar Array

World-first frequency agile interferometer equipped with an hardware embedded SK real-time computation engine

### Table 1: EOVSA Specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency range</td>
<td>2.5 – 18 GHz</td>
</tr>
<tr>
<td>Number of data channels/antenna</td>
<td>2 (dual polarization)</td>
</tr>
<tr>
<td>IF bandwidth</td>
<td>500 MHz single sideband</td>
</tr>
<tr>
<td>Frequency resolution</td>
<td>4096 spectral channels per 600 MHz band, 500 science channels variable ~1-40 MHz</td>
</tr>
<tr>
<td>Time resolution</td>
<td>Sample time: 20 ms, Full Sweep: 1 s</td>
</tr>
<tr>
<td>Polarization</td>
<td>Full Stokes (IQUV)</td>
</tr>
<tr>
<td>Number correlator inputs per poln</td>
<td>16</td>
</tr>
<tr>
<td>Number and type of antennas</td>
<td>Thirteen 2.1-m, Two 27-m equatorial (cal. only)</td>
</tr>
<tr>
<td>System Temperature</td>
<td>570 K (2 m); 35 K (27 m)</td>
</tr>
<tr>
<td>Baselines for imaging</td>
<td>78</td>
</tr>
<tr>
<td>Angular resolution</td>
<td>( \frac{56}{n_{GHz}} \times \frac{51}{n_{GHz}} \text{ arcsec} )</td>
</tr>
</tbody>
</table>

The EOVSA correlator outputs integrated power and squared power for all 15 antennas and R and L circular polarizations with 20ms-0.125MHz time-frequency resolution.
EOVSA CORRELATOR
High bit resolution POWER AND SQUARED POWER outputs
Nita, Hickish, MacMahon, and Gary 2016, J. Astronomical Instrumentation 5(4)
$S_1 = \sum_{k=1}^{M} P_k$

$S_2 = \sum_{k=1}^{M} P_k^2$

$SK = \frac{M + 1}{M - 1} \left(\frac{MS_2}{S_1^2} - 1\right)$

RFI: 2.17% (PFA=0.135%)
EOVSA Generalized SK Analysis

\[ S_1 = \sum_{k=1}^{N} P_k \]

\[ S_1 = \sum_{i=1}^{M} S_1 \]

\[ S_2 = \sum_{i=1}^{M} S_1^2 \]

\[ SK = \frac{MN + 1}{M - 1} \left( \frac{MS_2}{S_1^2} - 1 \right) \]
SK Dependence on the Integration–Relative Duty-Cycle RFI and Gaussian Transient Signals
SPECTRAL KURTOSIS: A POWERFUL SIGNAL CLASSIFICATION TOOL
The distribution of the Parkes Telescope quantized accumulated power output corresponding to a Gaussian time domain signal can be approximated by a Gamma distribution of shape parameter $d < N$, to which the Generalized Spectral Kurtosis theory may be applied.
Time Domain GSK analysis of the VLBI 2-bit sampling RCP voltage data containing the FRB 121102 signal

RFI-like statistical signature of the FRB 121102 2-bit signal
8-bit vs 2-bit Time-Domain GSK Analysis of RFI and Gaussian Signals

Distinct TDK Statistical Signatures

Undistinguishable GSK Statistical Signatures

RFI Transient Signal

Gaussian Transient Signal
8-bit vs 2-bit Spectral-Domain GSK Analysis of RFI and Gaussian Signals

Distinct GSK Statistical Signatures

RFI Transient Signal

Gaussian Transient Signal
SK and Multi-scale SK Analysis

Gaussian or 50% duty-cycle RFI Signature

PFA = 0.13%
SK > 2.342 = 0.35%
SK < 0.572 = 2.22%

PFA = 0.13%

PFA = 0.13%

PFA = 0.13%
Multi-scale SK Analysis using an adaptive starting point of integration

Unambiguous Gaussian Statistical Signature of the FRB121102 Signal

PFA = 0.13%

PFA = 0.13%
Conclusions

- Time domain Kurtosis analysis of the 2-bit quantized VLBI signals can detect both RFI and natural astronomical transients but is not capable of distinguishing them. Therefore, astronomical transients may be mistakenly flagged as RFI.

- Spectral Kurtosis analysis of the 32-bit FFT-transformed 2-bit time domain quantized VLBI signals can detect RFI while remaining blind to the presence of Gaussian transients (but may still detect sharp edges of such Gaussian signals). Therefore it is generally safe to employ spectral domain SK analysis for the purpose of automatic RFI excision.

- Multi-scale SK analysis of the 32-bit FFT-transformed 2-bit time domain quantized VLBI signals can detect both RFI and Gaussian transient signals and discriminate them based on their statistical signature. Therefore it is generally safe to employ spectral domain SK analysis for the purpose of detecting and discriminating RFI and natural transients.

- Using multiscale SK analysis we unambiguously established, for the first time, the Gaussian nature of an FRB 121102 signal detected by VLBI.