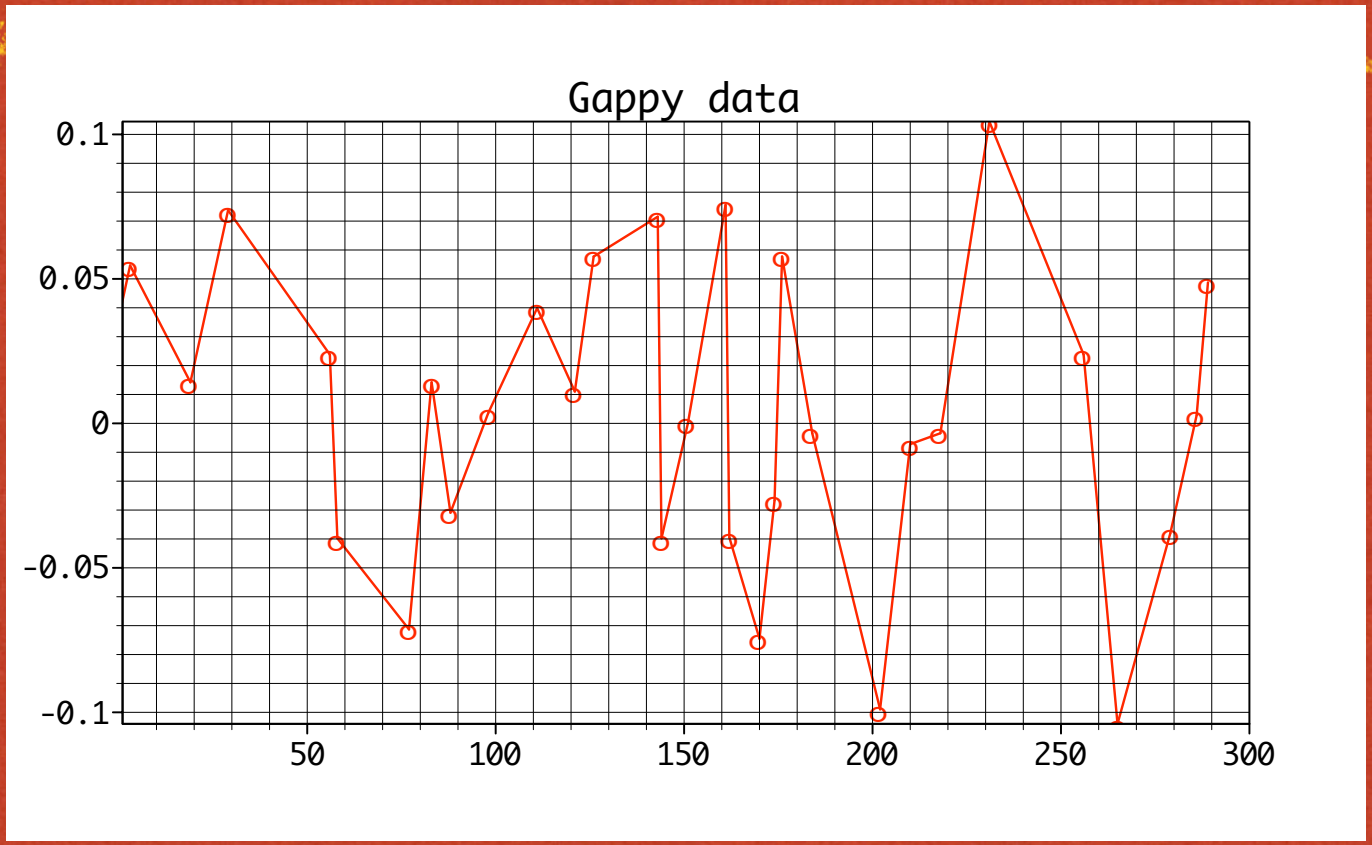
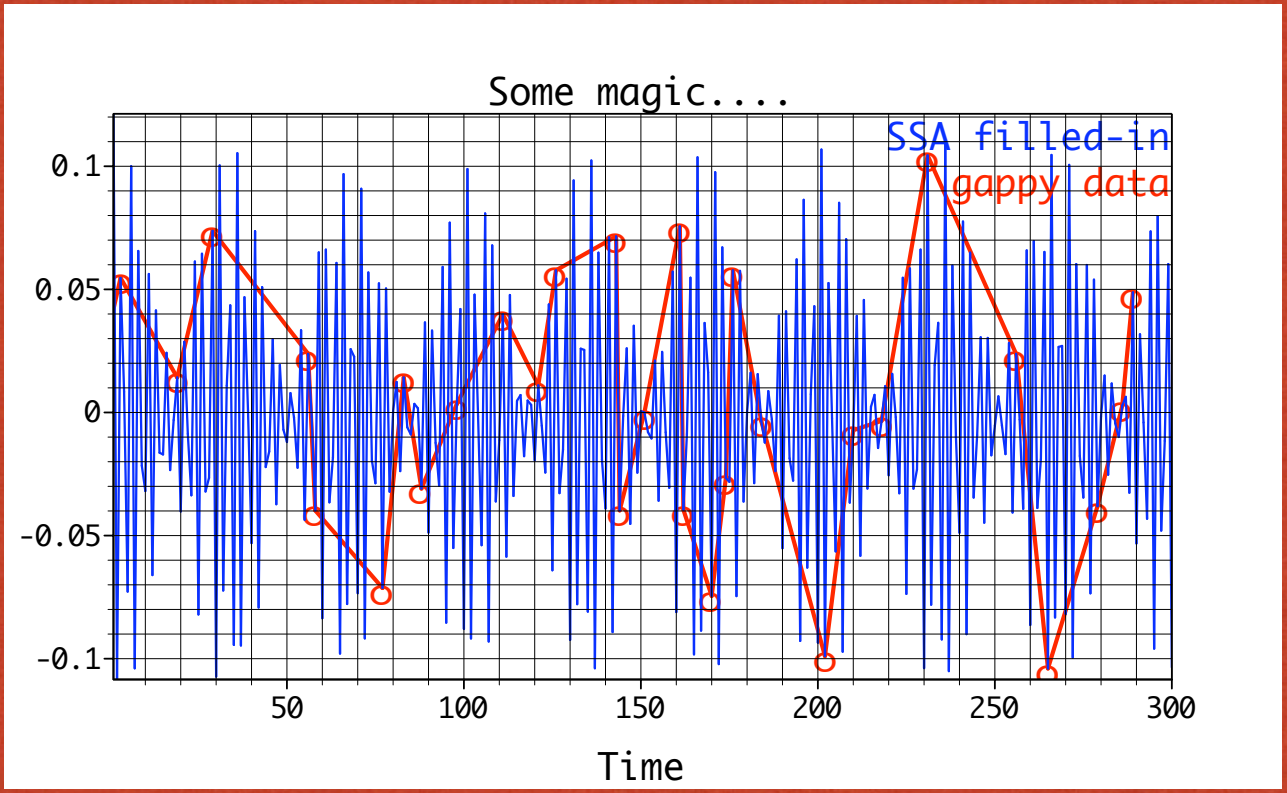
QUICK TEST FOR YOU :)



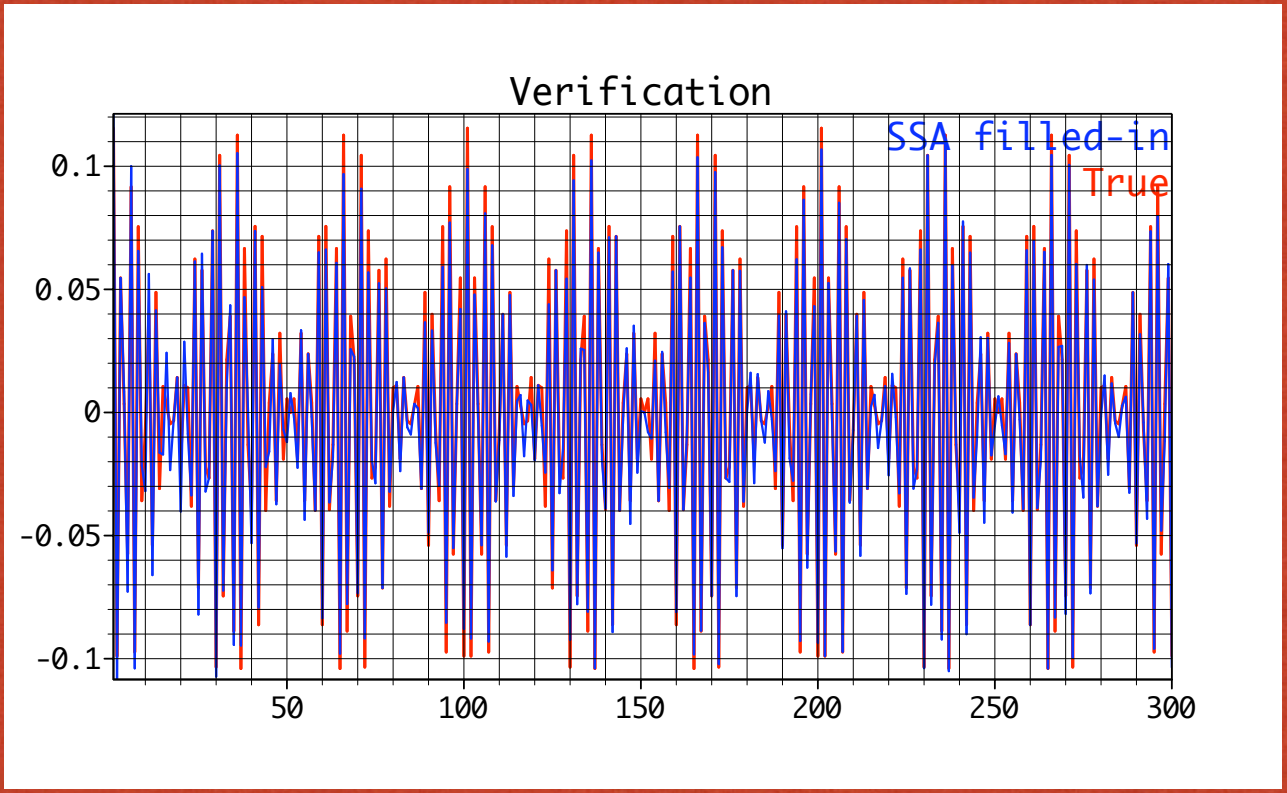
- Time series 300 data ponits long, 90% missisng. Is it periodic?

“LUCKY” GUESS :)



- Two high-freq harmonics T1=2.5, T2=2.32 resulting in a “beat” signal! Lomb-Scargle periodgram fails - false low-freq peaks!

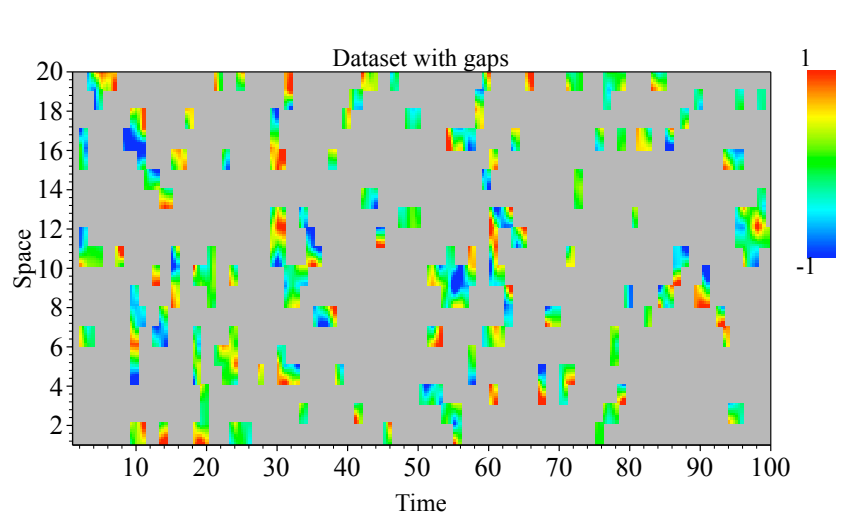
“LUCKY” GUESS :)



- Two high-freq harmonics T1=2.5, T2=2.32 resulting in a “beat” signal! Lomb-Scargle periodgram fails - false low-freq peaks!

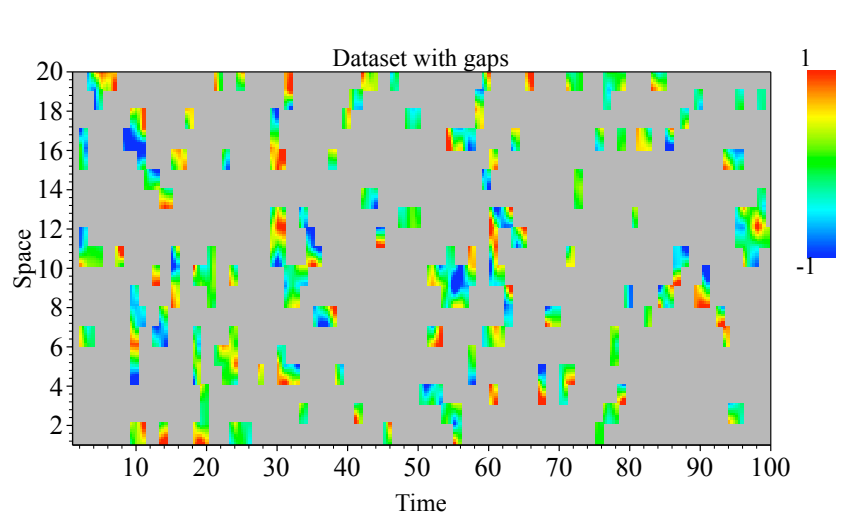
MULTIVARIATE TEST :)

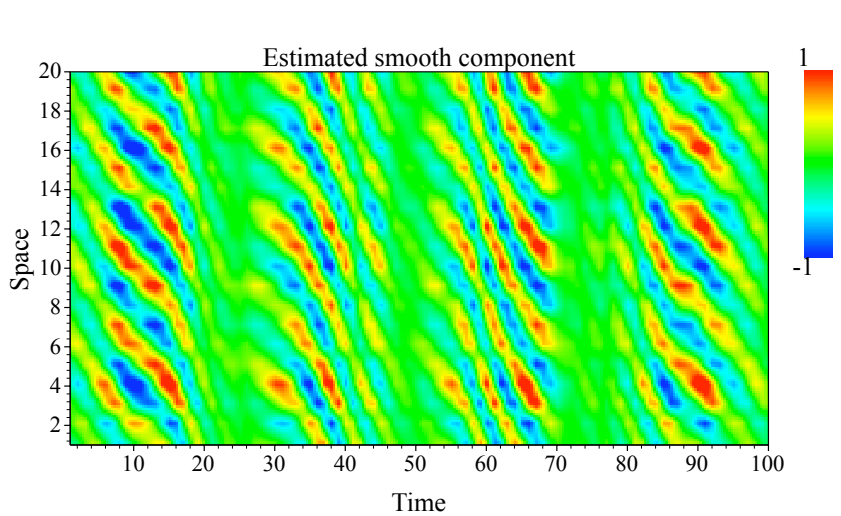
- Q: 100 (Time) x 20 (Space) points, ~70% missing. Is it periodic?



MULTIVARIATE TEST :)

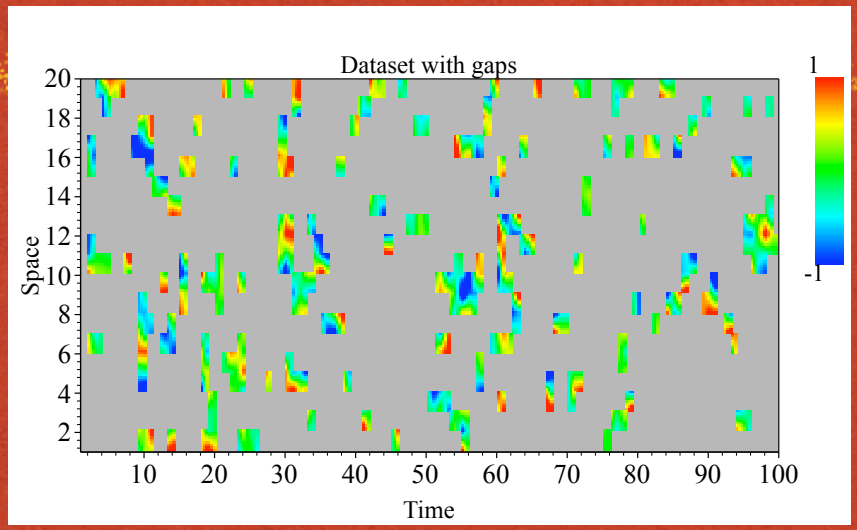
- Q: 100 (Time) x 20 (Space) points, ~70% missing. Is it periodic?

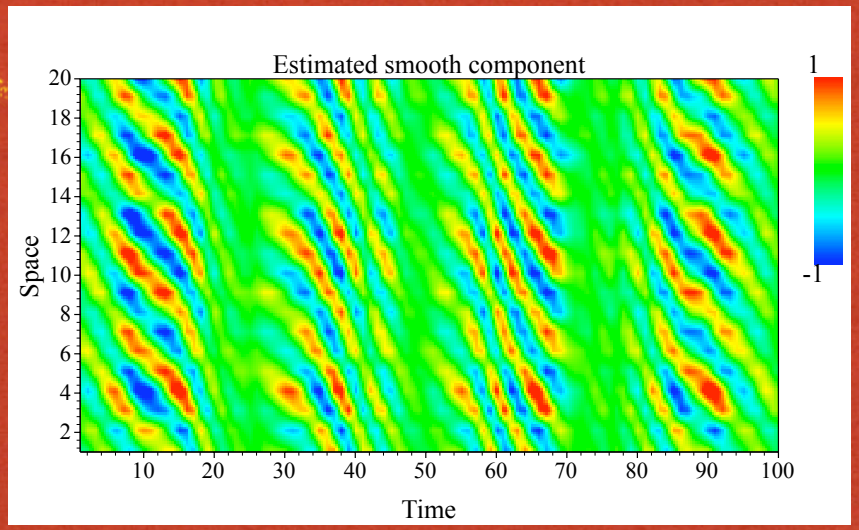


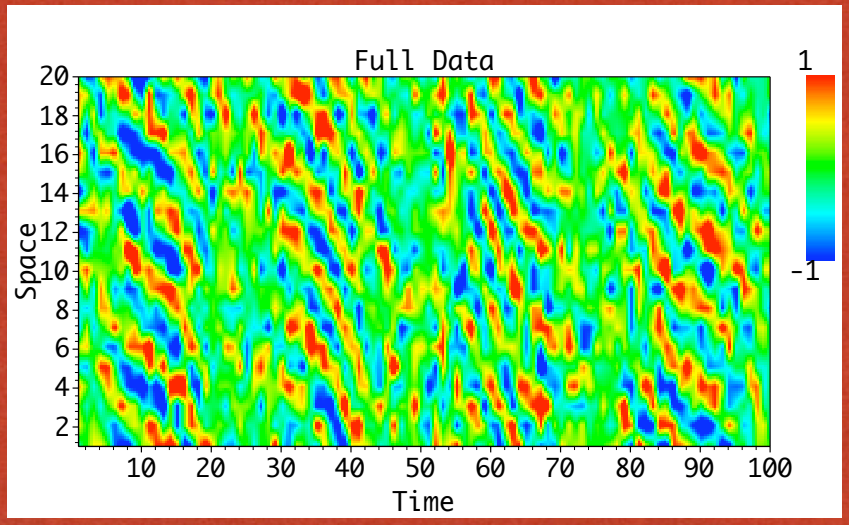


MULTIVARIATE TEST :)

- Q: 100 (Time) x 20 (Space) points, ~70% missing. Is it periodic?

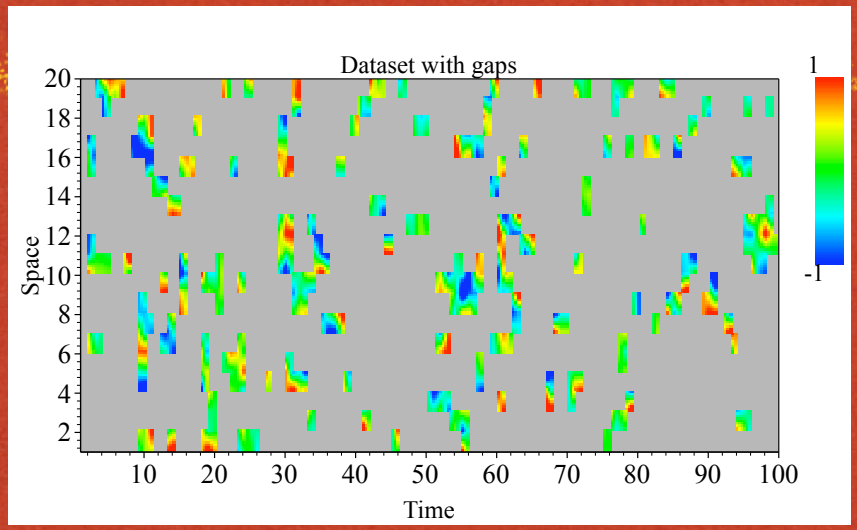


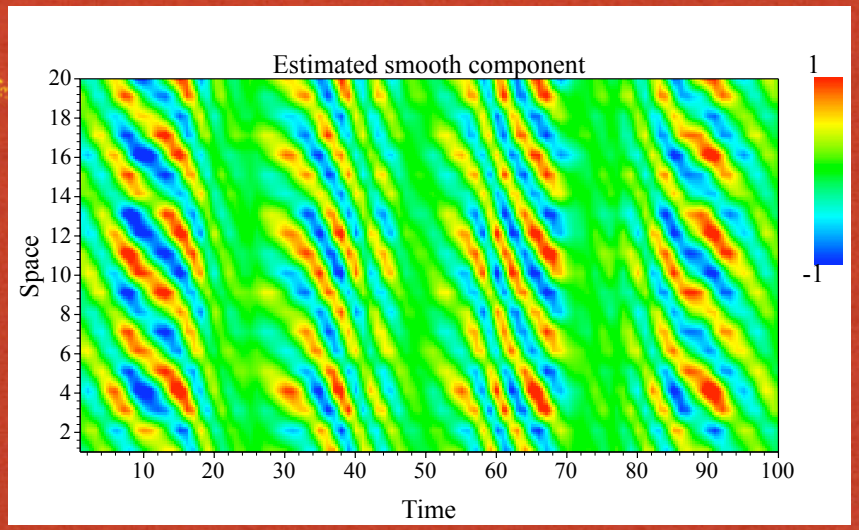


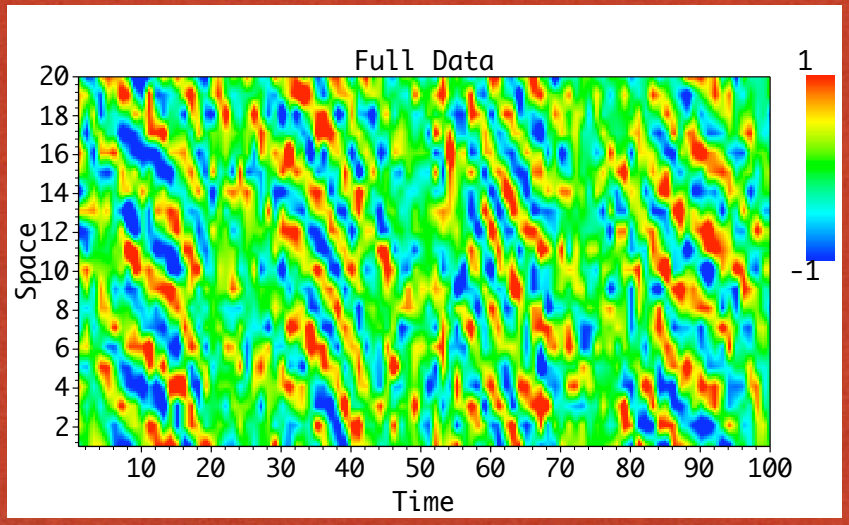


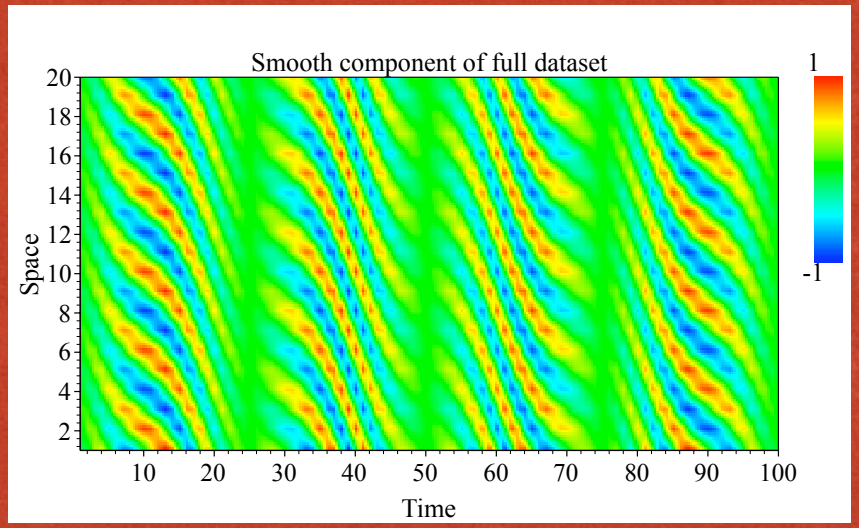
MULTIVARIATE TEST :)

- Q: 100 (Time) x 20 (Space) points, ~70% missing. Is it periodic?









Gap-Filling of Solar Wind Data by Singular Spectrum Analysis

Dmitri Kondrashov

University of California, Los Angeles

Yuri Shprits

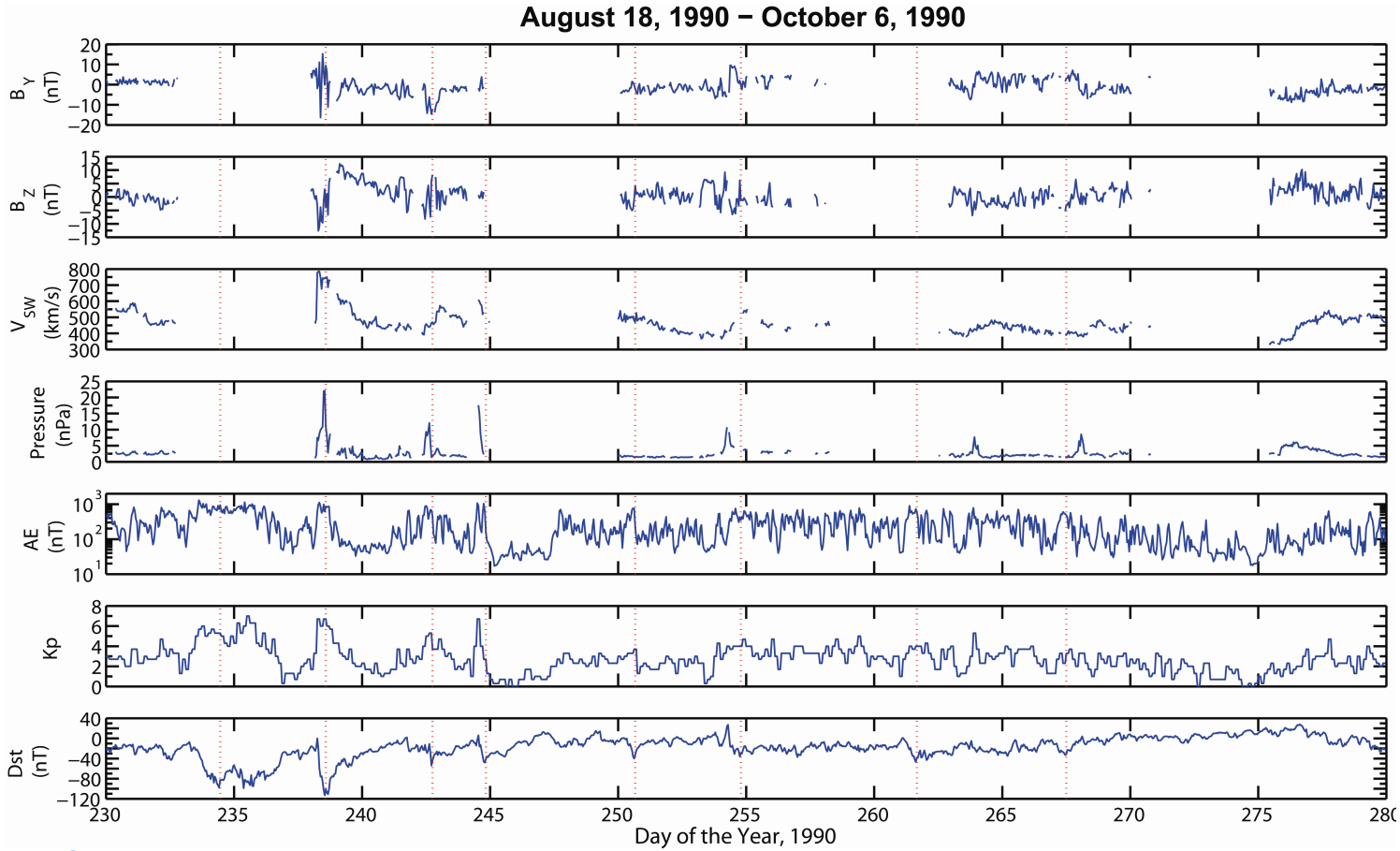
University of California, Los Angeles

Motivation

1. Observational data sets in the geosciences are often short, contain noise (errors) and are gappy: this is both an obstacle and an incentive. Continuous data is needed for modeling (boundary conditions), standard spectral estimation algorithms, etc.
2. Phenomena in the geosciences often have both regular ("cycles") and irregular ("noise") aspects. Regularities include (quasi-)periodicity ➔ - singular spectrum analysis (SSA) ➔Powerful method for spectral estimation, noise filtering and gap-filling (**producing estimates of the missing points!**)
2. Space Physics Application: **Apply SSA** to fill-in large gaps in historical solar wind and interplanetary magnetic field (IMF) data required for the state-of-the art empirical global magnetospheric magnetic field models. Important for radiation belt modeling - computing L*.

dkondras@atmos.ucla.edu

Gappy Solar Driver/Continuous Magnetospheric Response



Ni B. et al., JGR, 2009

- Gaps in solar wind satellite data before the launch of the WIND spacecraft in 1994.

- **Continuous** inner-magnetosphere indices (Kp, Dst) are ground measured time-lagged magnetic disturbances caused by interaction of Earth's magnetosphere with solar wind.

- **Main idea**: **exploit time-lagged correlations in a long record of existing data** (**Kp, Dst** and **solar driver**) **to reconstruct in the gaps of solar driver.**

Singular Spectrum Analysis (SSA)

SSA decomposes time series into ***Temporal Empirical Orthogonal Functions (EOF) and Temporal Principal Components (PC)*** based on the series' lag-covariance matrix.

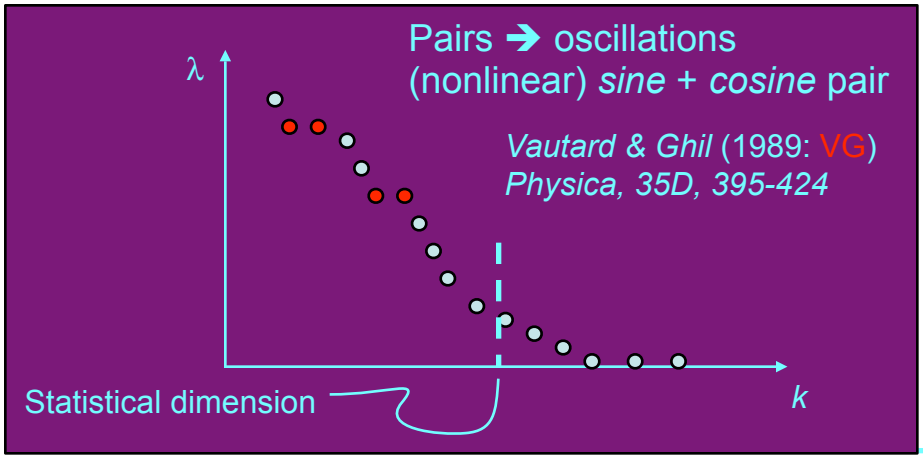
$$X(t+s) = \sum_{k=1}^{M} a_k(t) e_k(s), e_k(s) - EOF$$

e_k's are data-**adaptive "optimal" filters,**

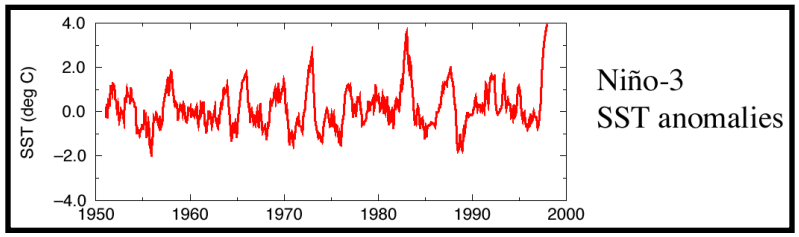
$$a_k(t) = \sum_{s=1}^{M} X(t+s) e_k(s), a_k(t) - PC$$

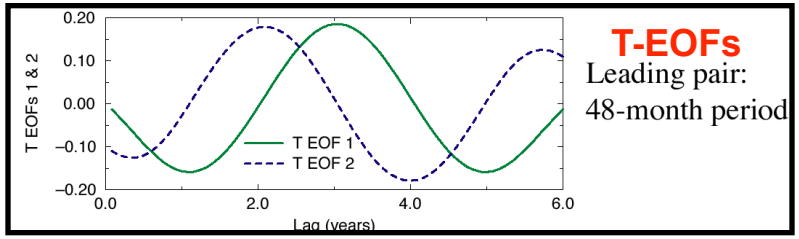
the a_k's are **principal components in time domain.**

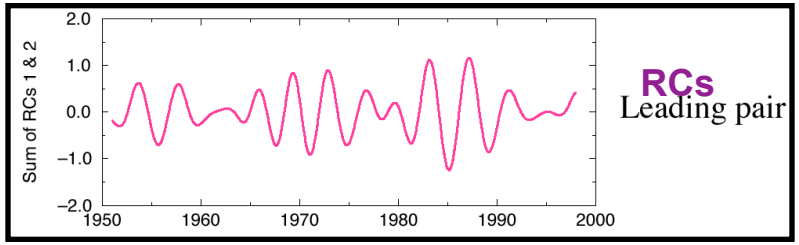
- SSA isolates oscillatory behavior via paired eigenelements.



Nino-3 index (El-Nino)







Parts of the series can reconstructed (RCs):

$$X^K(t) = \frac{1}{M} \sum_{k \in K} \sum_{s=1}^{M} a_k(t-s) e_k(s);$$

SSA Gap-filling algorithm

1. Choose window ***M*** and set ***K***=1. Flag fraction of dataset ***X(t)(t=1:N)*** as "missing" for cross-validation.

2. Update mean and lag-covariance matrix, find leading ***K*** EOFs

$$\mathbf{D} = \begin{pmatrix} X(1) & X(2) & . & . & X(M) \\ X(2) & X(3) & . & . & X(M+1) \\ . & . & . & . & . \\ X(N'-1) & . & . & . & X(N-1) \\ X(N') & X(N'+1) & . & . & X(N) \end{pmatrix}$$

$$\mathbf{C}_X = \frac{1}{N'}\mathbf{D}^{\mathrm{t}}\mathbf{D}; \mathbf{C}_X E_k = \lambda_k E_k$$

3. Reconstruct missing points using ***K*** EOFs

$$A_k(t) = \sum_{j=1}^{M} X(t+j-1)E_k(j)$$

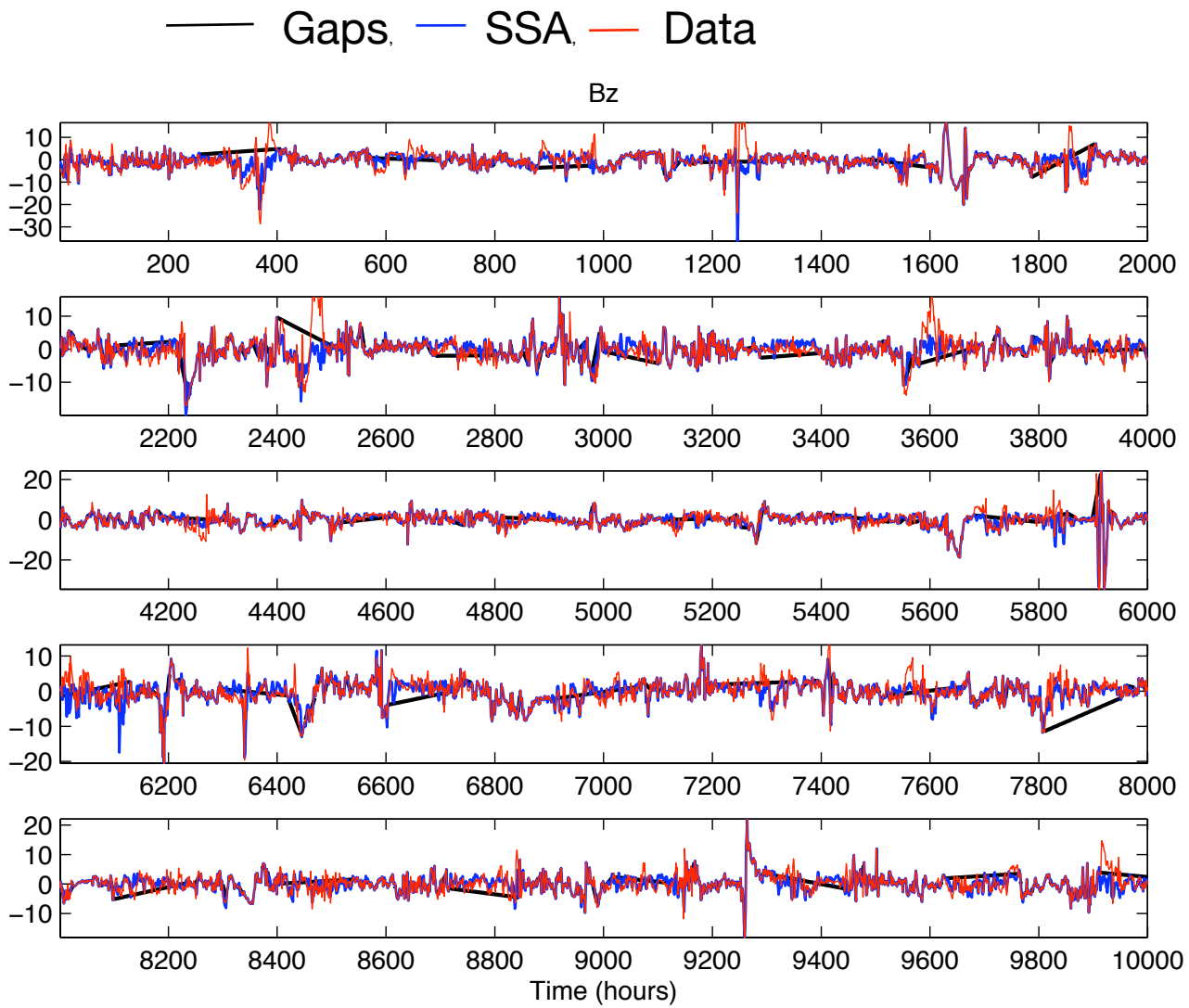
$$R_{\mathcal{K}}(t) = \frac{1}{M_t}\sum_{k\in\mathcal{K}}\sum_{j=L_t}^{U_t} A_k(t-j+1)E_k(j);$$

4. When convergence for missing points: ***K = K +1.*** Check cross-validation error, and Go to Step 2 if necessary.

- Extension of spatial EOFs for gap-filling: Beckers, J. and Rixen, M.: EOF calculations and data filling from incomplete oceanographic data sets, J. Atmos. Ocean. Technol., 20, 1839–1856, 2003.
- Utilize **spatial-temporal correlations** in **existing** data to iteratively compute maximum likelihood estimates of mean and lag-covariance matrix => can be applied to very gappy data.
- A few ***K*** leading EOFs correspond to the "**smooth**" modes, while the rest is **noise** and can be discarded.
- Cross-validation provides error estimates and optimum SSA parameters.

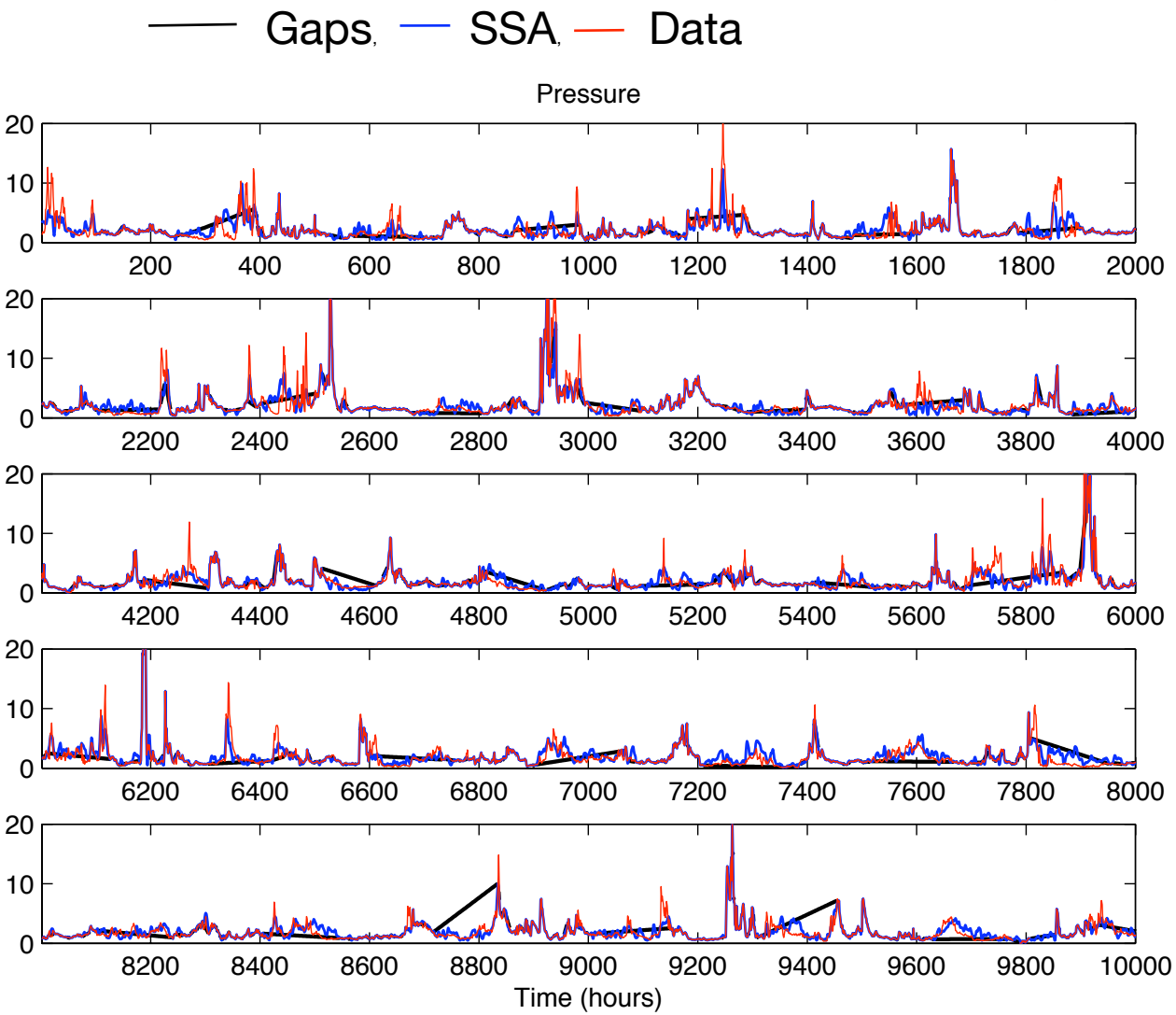
D. Kondrashov and M. Ghil, 2006: Spatio-temporal filling of missing points in geophysical data sets, Nonl. Proc. Geophys., 13, 151-159.

Filling-in 2000-2001 Bz & P with Synthetic Gaps



- Gaps of 1990-1991 are applied to hourly 2000-2001, 441 days
- SSA Window M=15hr (Bz) and M=25hr (P).

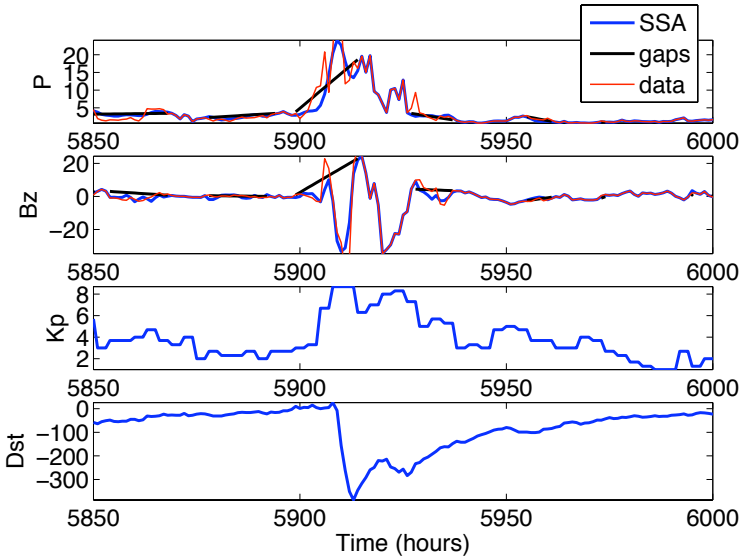
Filling-in 2000-2001 Bz & P with Synthetic Gaps

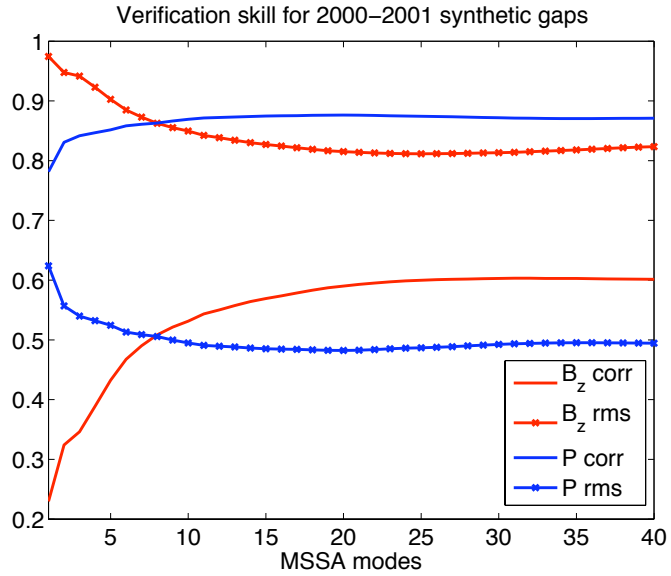


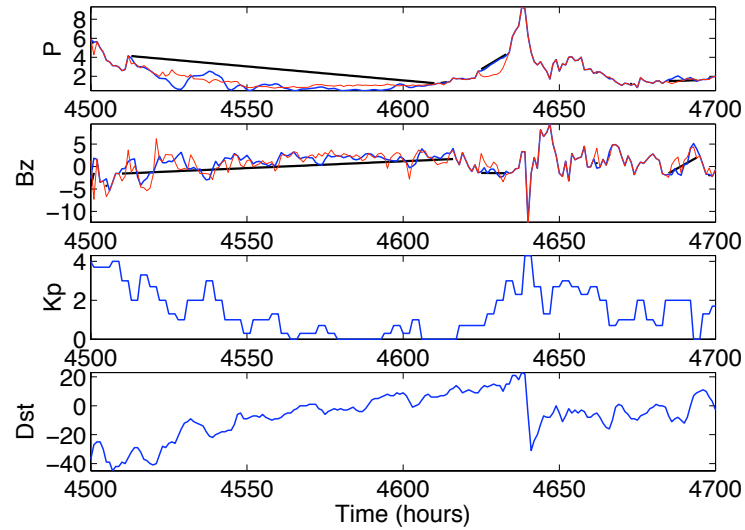
- Gaps of 1990-1991 are applied to hourly 2000-2001, 441 days
- SSA Window M=15hr (Bz) and M=25hr (P).

Filling-in 2000-2001 Bz & P with Synthetic Gaps

Gaps, SSA, Data

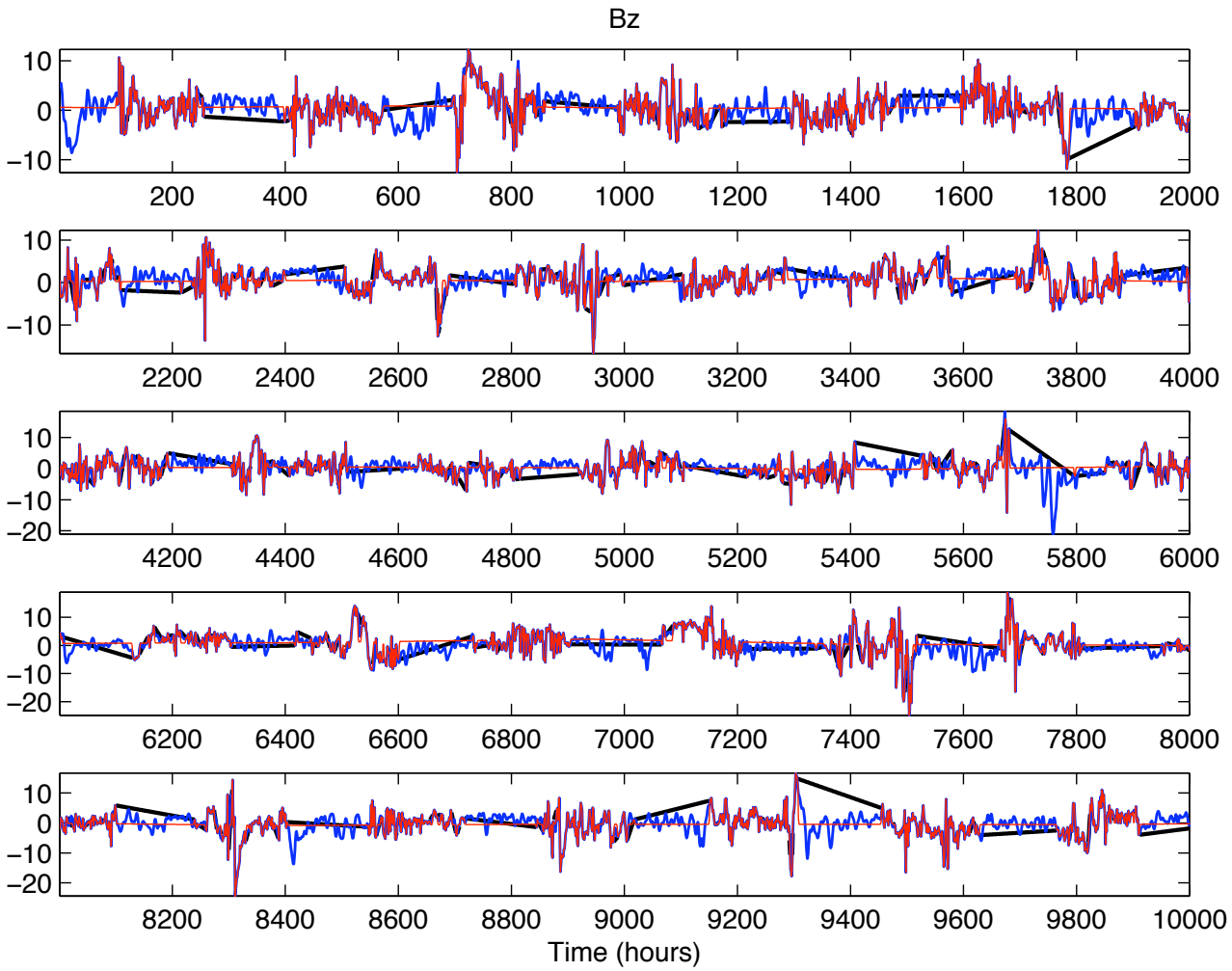






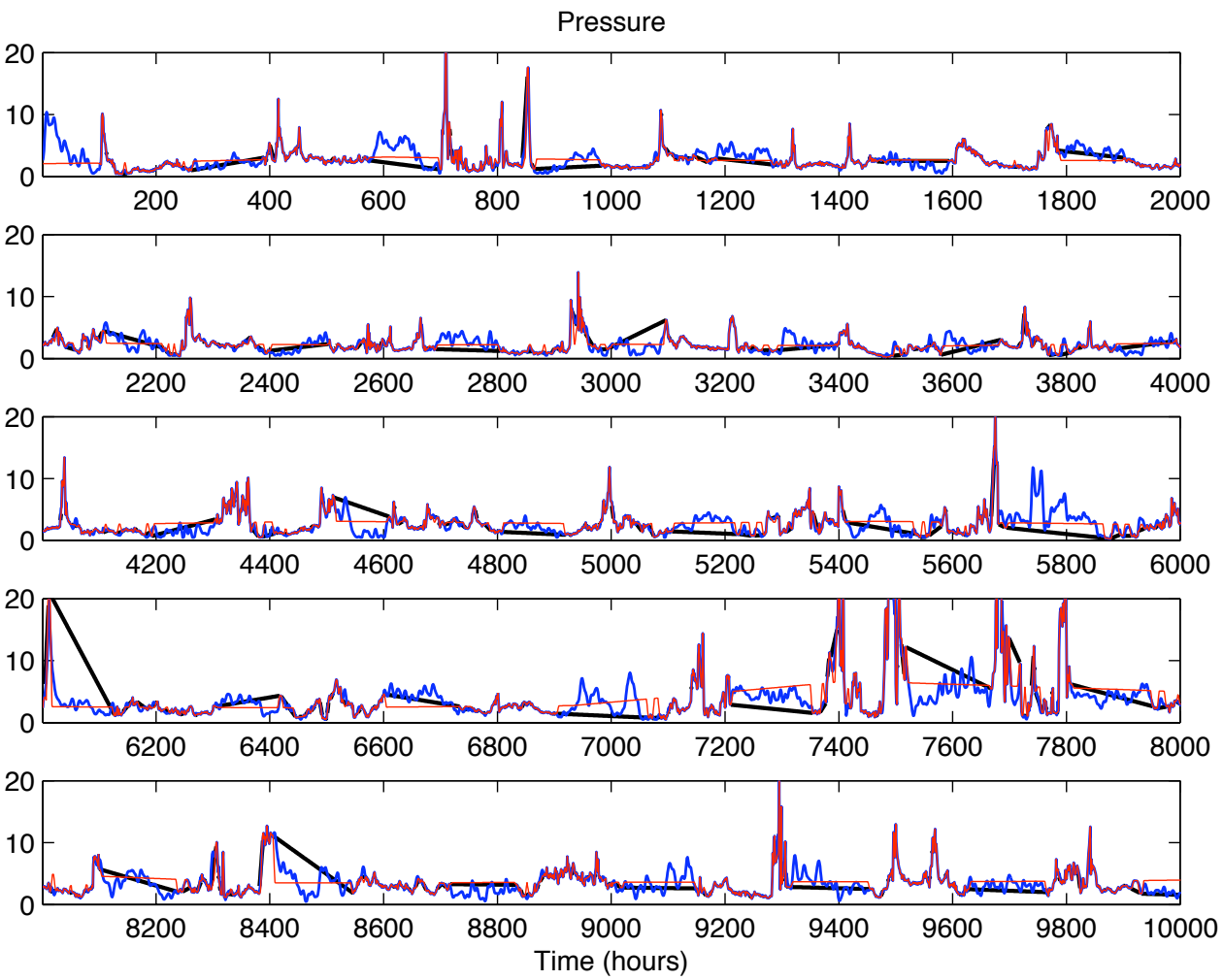
- Since both Bz & P contain noise which is discarded by SSA gap-filling, “optimum” correlation ≠1, rms≠0!
- SSA gap-filling works well for a wide range of geomagnetic conditions.

Filling-in 1990-1991 Bz & P



- SSA Window M=15hr (Bz) and M=25hr (P).

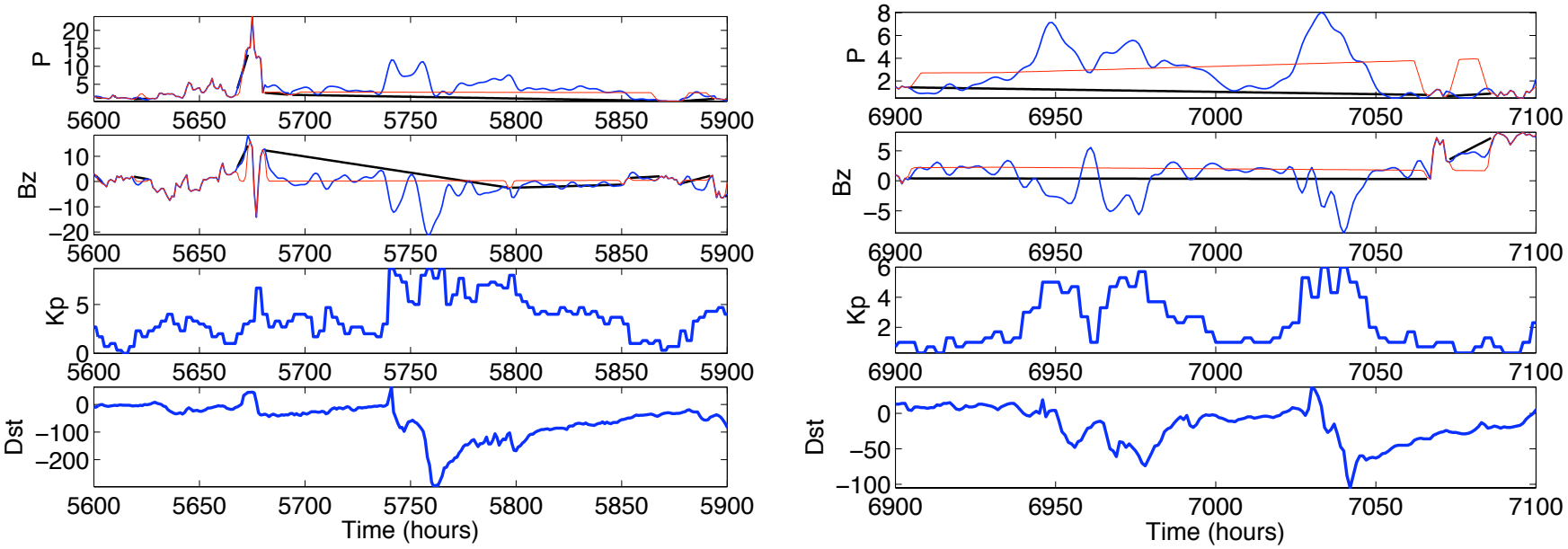
Filling-in 1990-1991 Bz & P



- SSA Window M=15hr (Bz) and M=25hr (P).

Filling-in 1990-1991 Bz & P

Gaps SSA Qin et al. 2007



- Realistic variability of SSA reconstruction!
- Error estimates are obtained through cross-validation.

Conclusions & Future Work

Promising first results of applying SSA to estimate the missing data in gappy solar wind data.

- Consider longer datasets, varying temporal scales.
- Closer look at the SSA modes responsible for successful reconstruction.
- Include more solar wind parameters: solar wind speed (V_{sw}), solar wind density (N_{sw}) and other components of magnetic field.
- Systematic search for optimum combination of inner-magnetospheric indices.

- *Ghil M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, et al., 2002: "Advanced spectral methods for climatic time series," Rev. Geophys.,40(1), pp. 3.1-3.41, 10.1029/2000RG000092.*
- *D. Kondrashov and M. Ghil, 2006: Spatio-temporal filling of missing points in geophysical data sets, Nonl. Proc. Geophys., 13, 151-159.*
- *SSA-MTM Toolkit (X11 Windows: Linux, Unix, Mac, free), http://www.atmos.ucla.edu/tcd/ssa/*
- *kSpectra Toolkit (native Mac OS X, paid, free demos), http://www.spectraworks.com*

Come to the poster!