A New Approach to Study Short-Term Nonmigrating Tidal Variability Using Information Theory and Bayesian Statistics

Komal Kumari[1], Jens Oberheide[1], Jian Du-Cains[2]

1. . Department of Physics and Astronomy, Clemson University 2. Department of Physics and Astronomy, University of Louisville

Abstract

Nonmigrating tidal diagnostics of SABER temperature observations in the mesosphere-lower thermosphere reveal a large amount of variability on time-scales of a few days to weeks. The physical reasons for the observed short-term variability are not well understood. In this paper, we present a new approach to diagnose how short-term tidal variability changes as a function of Madden-Julian Oscillation, QBO, solar cycle and others. Our approach is based on Information theory and Bayesian statistics using time dependent probability density functions, Shannon entropy and Kullback-Leibler divergence. The statistical significance of this approach is exemplified using SABER DE3 tidal diagnostics. The response of short-term DE3 variability to the QBO phase and other natural drivers will be discussed using principal component analysis of the time dependent probability density functions.

Short term tidal variability from SABER

A simple wavelet analysis of SABER DE3 amplitudes reveals a substantial amount of variability, ranging from days to years. Our objective is to better understand short-term (days to weeks) tidal variability and how it responds to the QBO, MJO, etc.

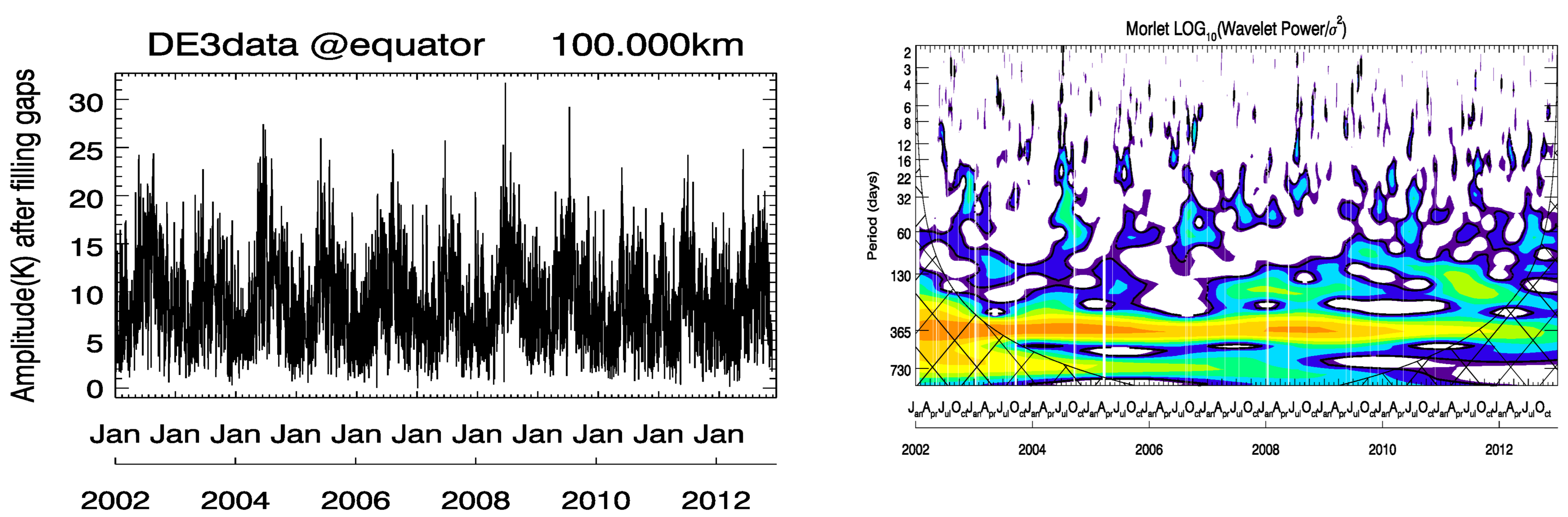

Fig 1: (Left) DE3 from SABER @equator and 100 km, (Right) Wavelet spectrum.

Time dependent Probability Density Function (TDPDF)

The first step in our approach is to estimate TDPDFs (i.e. how a PDF evolves in time), similar to contemporary studies in climate research [1]. A PDF is basically a histogram plotted using a certain bin number. The accurate estimation of the bin number is crucial as it optimizes the information content of the PDF vs noise. Here, an optimal binning scheme [2] is derived from the **Bayesian theorem**:

$$P(hypothesis|data) = \frac{P(data|hypothesis)P(hypothesis)}{P(data)}$$

$P(hypothesis)$= prior probability

$P(data|hypothesis)$= likelihood function, it quantifies the likelihood that the observed data would have been observed for a given hypothesis

$P(hypothesis|data)$= posterior probability. Starting from observed events and a model, it gives the probability of the hypotheses that may explain the observed data. It can also be written as:

$$\boldsymbol{posterior\ probability \propto likelihood\ X\ prior\ probability}:$$

$$P(M|\vec{d}, I) \propto \left(\frac{M}{V}\right)^N \frac{\Gamma(\frac{M}{2})}{\Gamma(\frac{1}{2})^M} \frac{\prod_{k=1}^{M} \Gamma(n_k + \frac{1}{2})}{\Gamma(N + \frac{M}{2})},$$

Data sample's size N, range=V, bin no.= M, n_k=number of counts in each bin

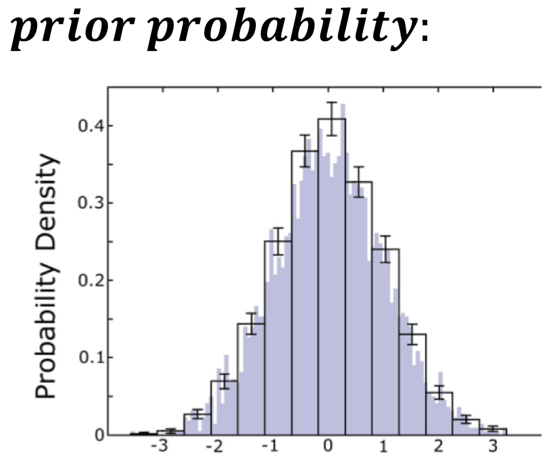

Fig 2: A PDF is basically a histogram. The optimum bin width is estimated using the Bayesian theorem [2].

Optimizing the bin width

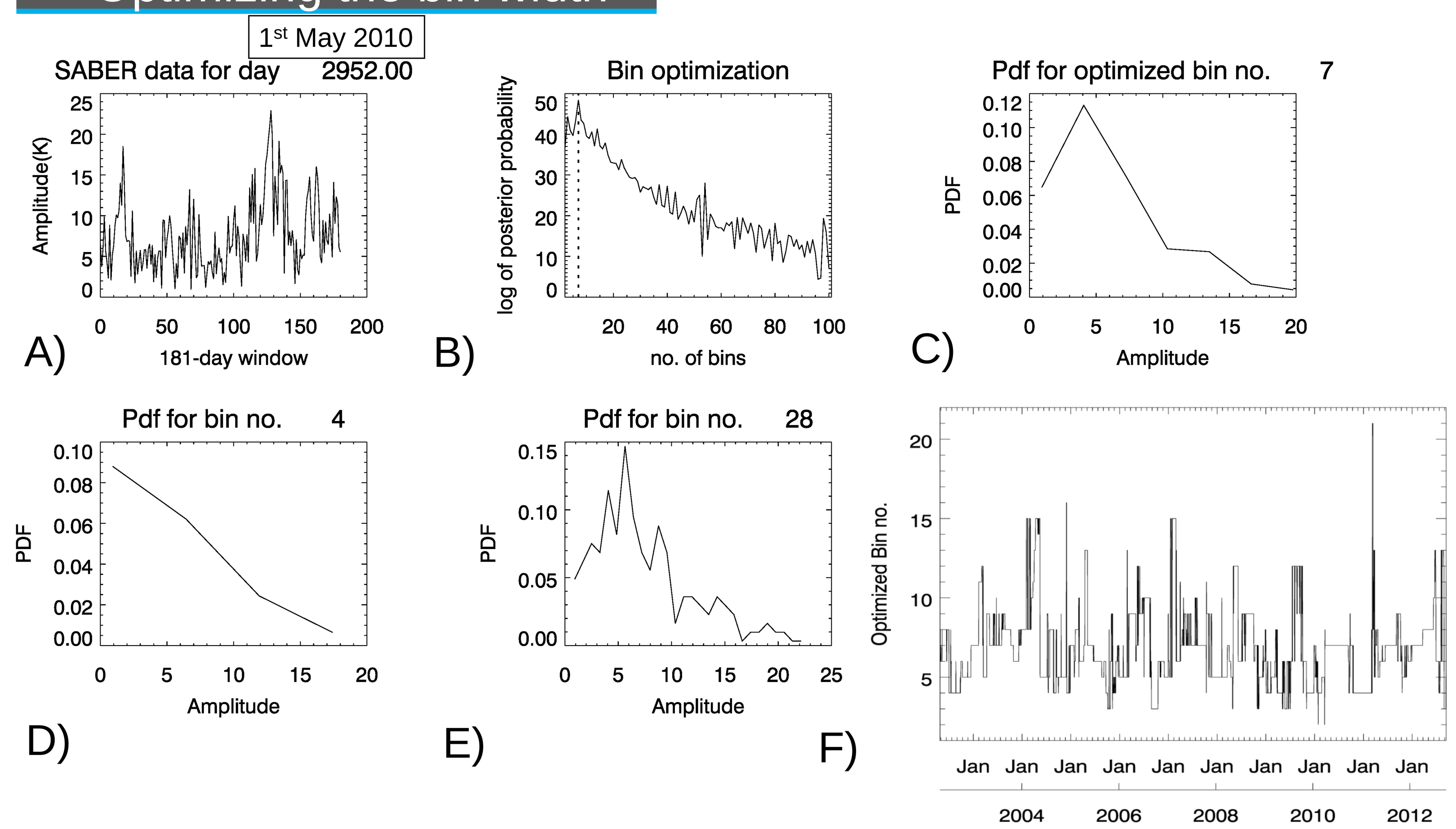

Fig 3: A) A 181-day window of DE3 data @equator and 100 km centered on the day number 2952. B) The posterior probability maximizes for optimal bin number, i.e. 7 for this day. C) PDF using optimum bin number 7 D) PDF plot using bin number 4, i.e. less than optimum bin number (loss of information) E) PDF plot using bin number 14, i.e. more than optimum bin number (too much noise) F) Optimized bin number variation for time series using 181-day windows. Here, the data range within a sample window can vary while advancing through the time series, which leads to a variation of the optimum bin number through the whole time series.

TDPDF as running mean PDF

1. A sample S (t_c-W/2, t_c+W/2) with window size W centered at t_c is chosen.
2. PDF (histogram) of the sample is estimated by using optimal binning scheme by Bayesian statistics and then PDF is assigned to the center value t_c (i.e. p(x, t_c)).
3. Advancing the window by one day each time provides the complete TDPDF.

Deseasoning

Key to studying short-term tidal variability using TDPDFs with window lengths of several days to months is the removal of seasonal variability from the original data before performing the TDPDF diagnostics: each PDF then reflects the average short-term tidal variability within a given window. The time evolution of the PDFs (TDPDF) thus reflects how the averaged short-term variability changes over time. For example, one can still diagnose short-term tidal variability using long windows such as 181 days to look into inter-annual changes of the short-term tidal variability. Diagnosing variability changes related to intra-seasonal effects (MJO) would require a shorter window.

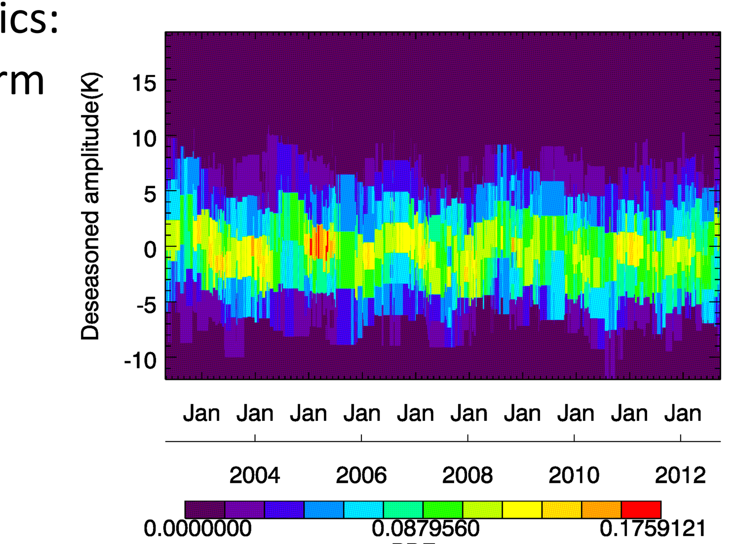

Fig 4: TDPDF for 181-day window for DE3 (equator, 100 km)

Empirical Orthogonal Function Analysis of TDPDF

Preliminary insight into physical causes of short-term tidal variability changes comes from an EOF analysis of the TDPDFs. Figure 5 exemplifies this for the 181-day window length. The first EOF (31%) reveals an QBO-like modulation and the second EOF (14%) indicates annual modulation.

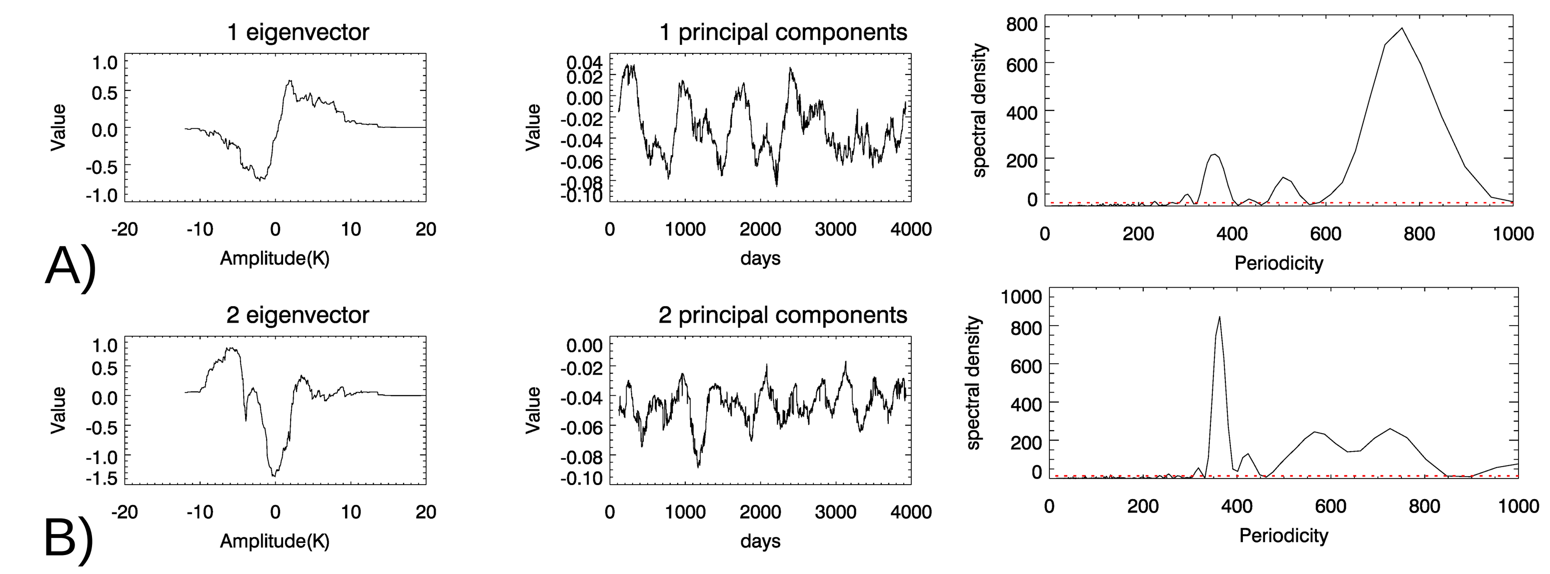

Fig 5: A) First eigenvector TDPDF using a 181-day window, corresponding principal component and its spectral analysis, the corresponding eigenvalue weight is 30.55 % and the dominant periodicity of spectral analysis is ~ 2 years (QBO), B) Similarly, second eigenvector for the TDPDF, corresponding principal component and it's spectral analysis, the corresponding eigenvalue weight 14.34%, dominant periodicity ~ 1 year (Annual time scale)

Future work

The next step is to apply information theory, i.e., Shannon entropy and Kullback-Leibler Divergence (KLD), to the TDPDFs. The structural changes in Shannon entropy will give us the information content in the PDFs and KLD will give insight into the relative tidal stability and/or rapid changes for future forecast possibilities, e.g. to construct a statistical forecast model of short-term tidal variability.

Preliminary diagnostics (not shown) point to a reasonable agreement between observed (SABER) and modeled (eCMAM30) statistical characteristics of the short-term tidal variability. This will allow us to study the physical causes and consequences of the short term variability.

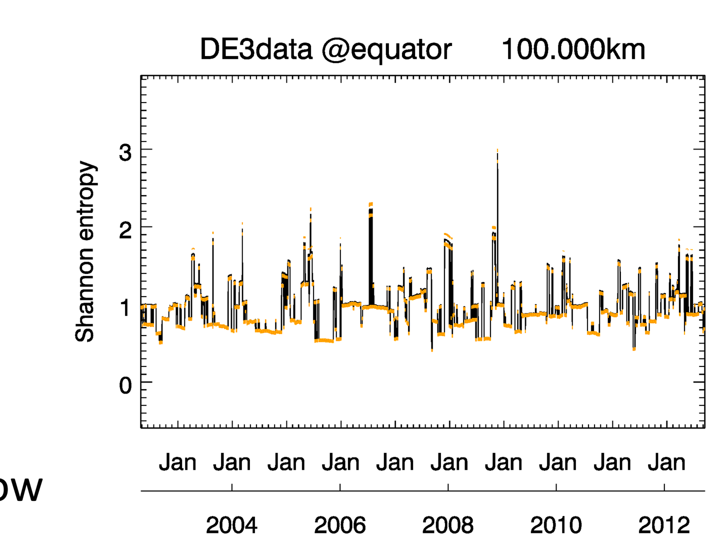

Fig 6: Shannon entropy of TDPDF using 181 days window

Conclusion

A Bayesian statistics framework for TDPDF of SABER data has been established. Preliminary results for a 181-day window using EOF analysis indicate that changes in short-term DE3 variability at the equator and 100 km occur on QBO-like and annual time scales, respectively. Initial Shannon entropy diagnostics points to several structural changes in the short-term tidal variability. Further insight will come from information theory that will also allow us to set-up a statistical forecast model in the future.

References

1. J. Walter Larson, "Visualizing Climate Variability with Time-Dependent Probability Density Functions, detecting It Using Information Theory", Procia Computer Science 9 (2012) 917-926
2. K. H. Knuth, "Optimal data-based binning for histograms, https://arxiv.org/pdf/physics/0605197.pdf (2006)
3. K. Hocke and N. Kampfer, "Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram", Atmos. Chem. Phys. 9 (2009) 4197-4206
4. N. M. Pedatella, J. Oberheide et.al, "Short-term nonmigrating tide variability in the mesosphere, thermosphere, and ionosphere", J. Geophys. Res. Space Physics (2016), 121, 3621-3633

Acknowledgement

This work has been supported by NASA grant **NNX15AJ02G**.

kkumari@g.clemson.edu