Understanding Short-Term Nonmigrating Tidal Variability in the Ionospheric Dynamo Region from SABER Using Information Theory and Bayesian Statistics

Komal Kumari[1], Jens Oberheide[1]

1. Department of Physics and Astronomy , Clemson University

Abstract

Nonmigrating tidal diagnostics of SABER temperature observations in the ionospheric dynamo region reveal a large amount of variability on time-scales of a few days to weeks. In this paper, we discuss the physical reasons for the observed short-term tidal variability using a novel approach based on Information theory and Bayesian statistics. We diagnose short-term tidal variability as a function of season, QBO, ENSO, and solar cycle and other drivers using time dependent probability density functions (TDPDFs), Empirical Orthogonal Function Analysis and Kullback-Leibler divergence. The statistical significance of the approach and its predictive capability is exemplified using SABER tidal diagnostics with emphasis on the responses to the QBO and annual cycle.

Short term tidal variability from SABER

The approach is exemplified using SABER DE3 tidal diagnostics. The uneven spatial-temporal gaps present in the data are filled using the Lomb-Scargle method [3]. After de-seasoning the data as shown in Fig. (1), the first step in our approach is to estimate TDPDFs (i.e. how a Probability Density Function (PDF) evolves in time), similar to contemporary studies in climate research [1].

Objective: To understand the causes and impact of the tidal variability on the scales ~10-30 days and how it responds to the QBO, MJO, etc

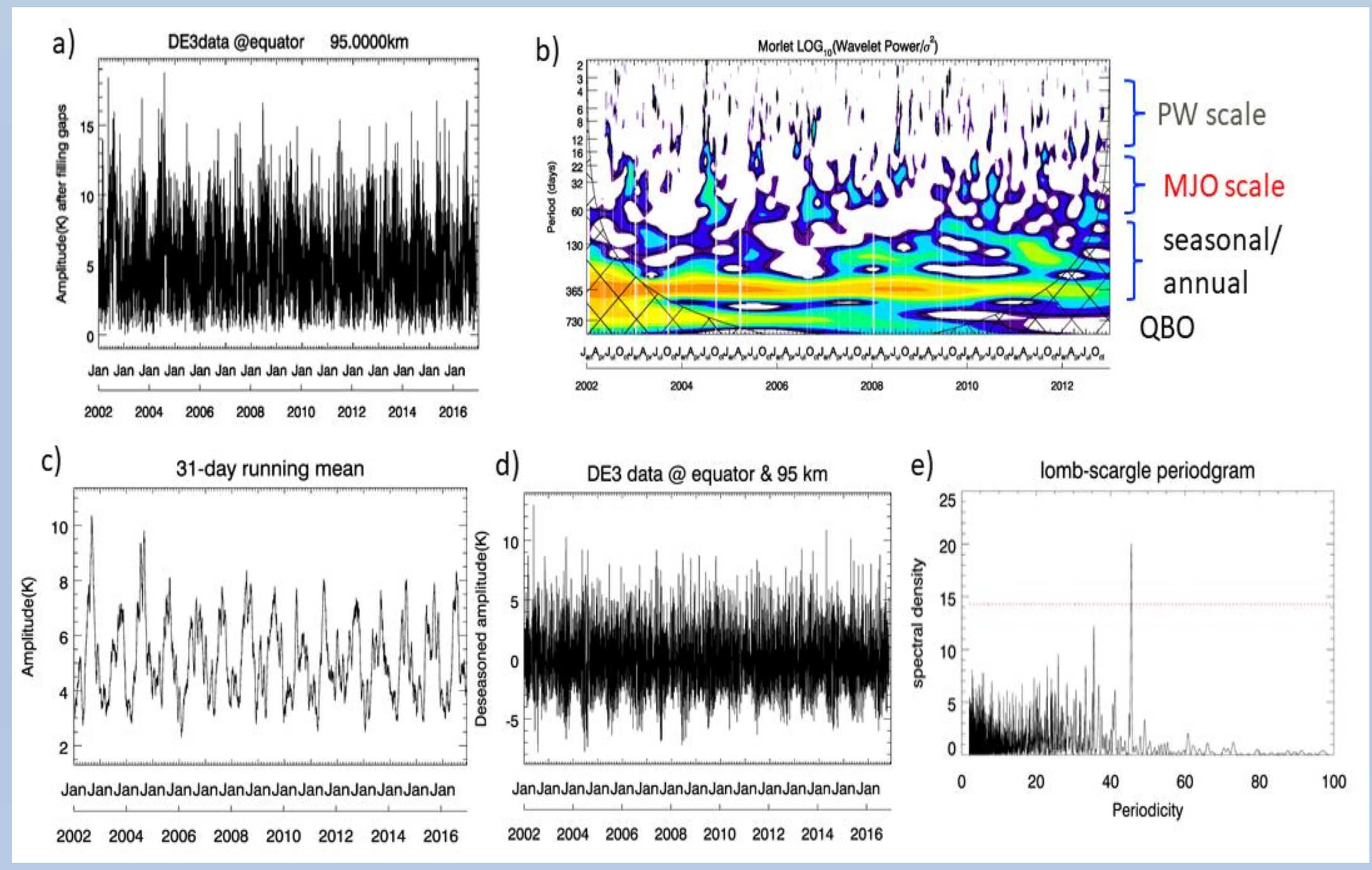


Figure 1: a) The DE3 data at 95 km and the equator after filling gaps. b) The wavelet analysis of the DE3 data at 100 km and the equator reveals periodicities on several scales. c) 31-day running mean values of the data in (a), d) the de-seasoned data after the 31-day running mean removal from the data in (a); the residual variability in the data is on time scales less than 31 days. e) the spectrum of the de-seasoned data.

Acknowledgement

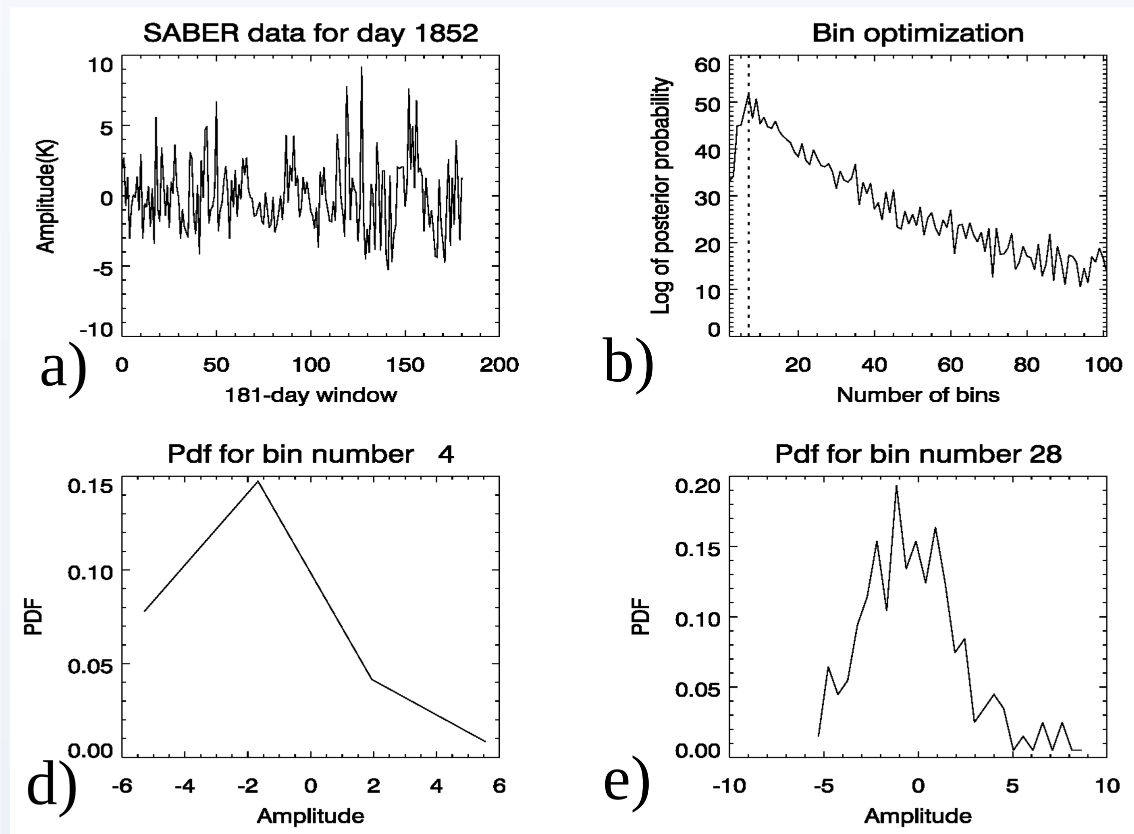
This work has been supported by NASA grant **NNX15AJ02G**.

kkumari@g.clemson.edu

Time dependent Probability density function (TDPDF)

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:**

$$\textit{posterior probability} \propto \textit{likelihood X prior probability}$$



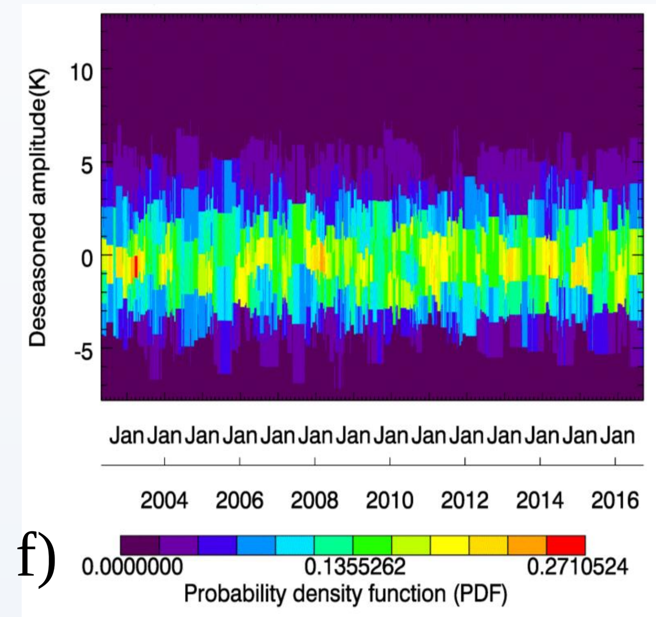


Figure 2: a) An example of the data within a 181-day window from the de-seasoned data centered on the day number 1852. b) The log of posterior probability values vs. bin numbers, which maximizes at the optimum bin number, i.e. 7. c) PDF using optimum bin number 7. d) PDF using bin number 4, i.e. less than optimum bin number (loss of information), while e) PDF plot using bin number 28, i.e. more than optimum bin number (too noisy PDF). f) The TDPDF of the de-seasoned data is obtained using a sample of 181-day window length and then advancing the window in steps of one day until the end of the time series.

How is short-term tidal variability studied using a sample of such large window length (e.g. 181 days)?
The key point is that I am considering here the ***'variability over variability'*** in TDPDF. The TDPDF using 181-day window is utilized to study inter-annual changes of the short-term tidal variability. These changes represent how short-term variability responds to other forcing factors in the atmosphere on inter-annual or larger scales. Identifying changes in short-term tidal variability related to intra-seasonal effects (such as MJO, a ~60-day periodic atmospheric event) would require a shorter window length sample selection.

Empirical Orthogonal Function Analysis of TDPDF

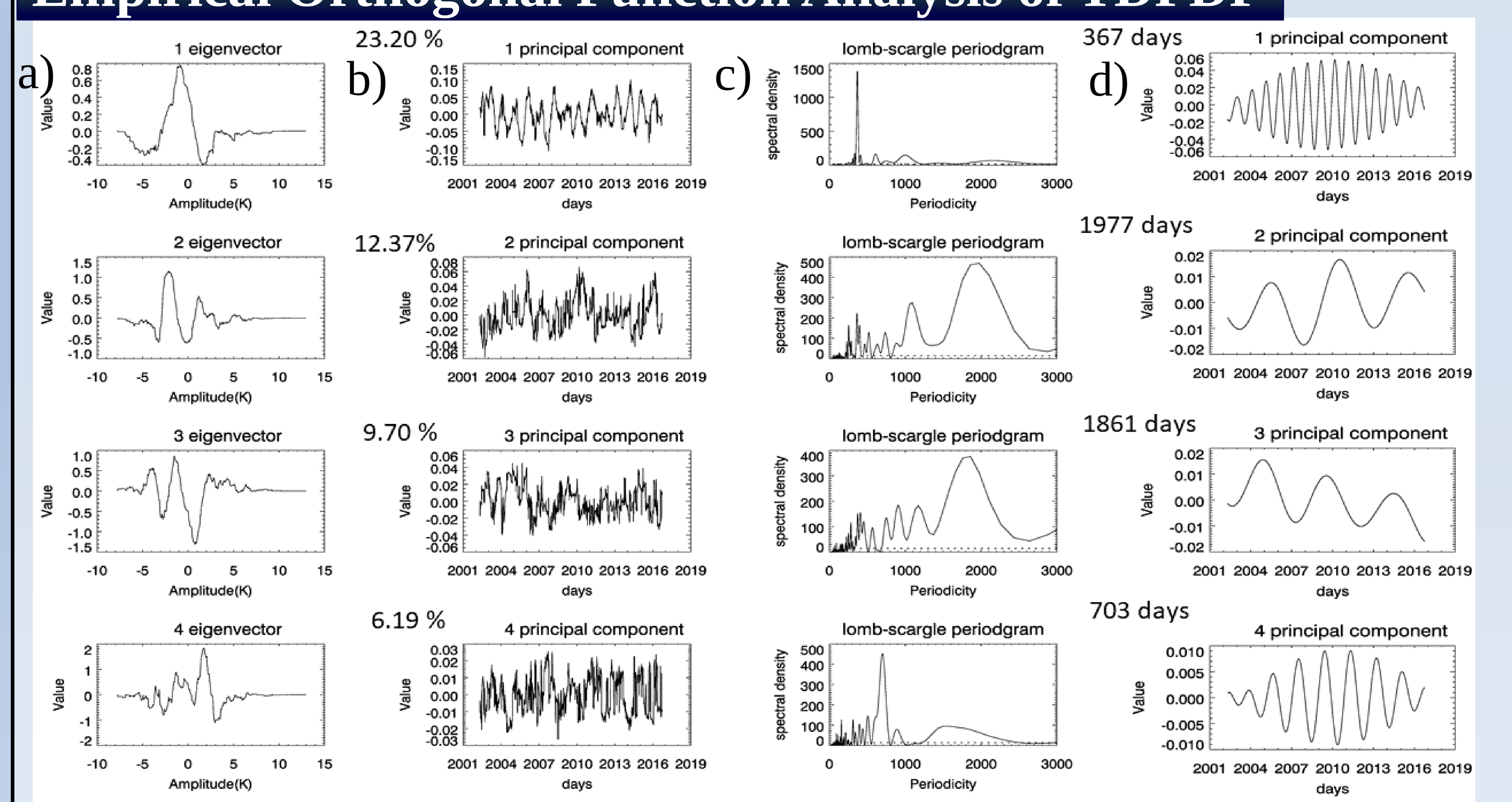
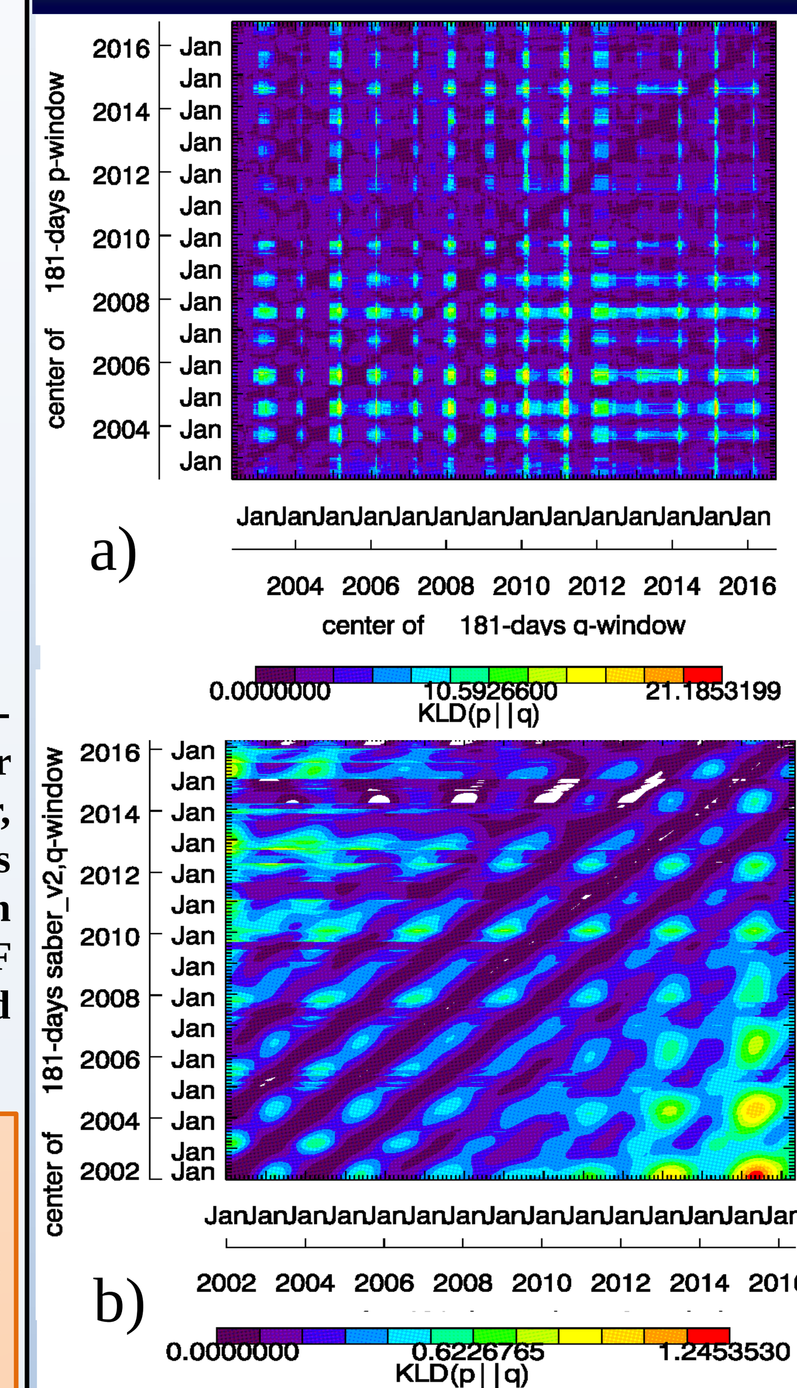


Figure 3: EOF analysis of the TDPDF, a) the plot shows the first 4 eigenmodes, b) corresponding principal components with eigenvalues (23.20, 12.37, 9.70, 6.19), c) the spectral analysis of the components with predominant periodicities of days (367 (annual), 1977, 1861, and 703 (biannual, ~QBO or ENSO)), and d) the simplified structure of the components after bandpass filtering across the most dominant frequency-width, which reveals a beat-like structure, which is yet to be understood by application of varimax rotation [5].

Information-theoretic technique on TDPDF



Kullback-Leibler Divergence (KLD) provides a tool to compare PDF $p(x, t_1)$ to another PDF $p(x, t_2)$ present in the TDPDF and it gives the measure of relative entropy or divergence between two-time shifted PDFs.

$$D(p(x)||q(x)) = \sum_x p(x) \log\frac{p(x)}{q(x)}$$

The annual variation in the KLD indicates the general seasonal variations of PDF structures from winter to summer. Moreover, for a given time window, points vertically above (below) the diagonal signify the ability of this windowed PDF to predict future (past) PDFs.

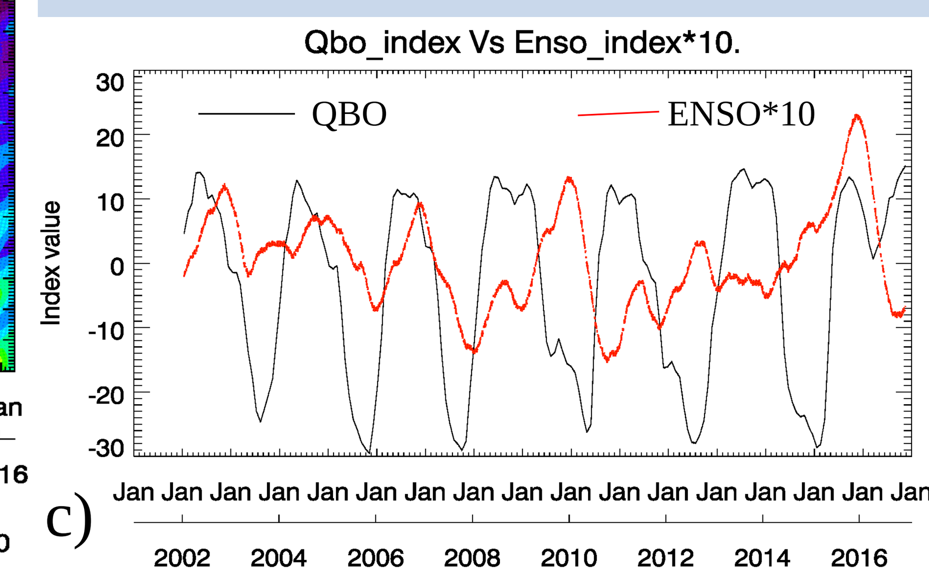


Figure 4: a) The values plotted are $D_{KL}(p||q)$ of the TDPDF from Fig. (2); this value is zero along the diagonal due to zero-time shift (i.e. no divergence or changes). A low-KLD-value region indicates less divergence in PDF structures, i.e., the period of relative stability in the tidal variability and thus good predictability. In contrast, a high-KLD-value indicates that the PDFs are structurally different, i.e. relative rapid changes in the variability and thus poor predictability. b) The KLD of the TDPDF after bandpass across (700,900) days which shows the poor prediction in 2010-2016; the probable cause is phase difference between QBO and ENSO as shown in c).

Conclusions

- ✓ An information-theoretic framework to study statistical characteristics of short-term tidal variability based on sampling periods has been established and applied to de-seasoned DE3 data.
- ✓ Preliminary results show different scales of periodicity involved in the TDPDF of short-term tidal variability in the E-region (~95 km), which will also be mapped to the F-region plasma density variability.
- ✓ The KLD plot shows that the framework will potentially contribute in the setup of a forecast model of the variability.

Future work

- ➤ To analyze features in the KLD other than annual and to develop the joint probability density function framework [4] which will be used to get mutual information between de-seasoned SABER DE3 data and QBO, ENSO effects.
- ➤ To perform Varimax rotation on the eigenmodes and the principal components of the TDPDF.
- ➤ To make the framework statistically feasible for shorter window lengths (e.g. MJO- like).

References

1. Larson, J. Walter (2012), "Visualizing Climate Variability with Time-Dependent Probability Density Functions, detecting It Using Information Theory", Procedia Computer Science 9 917-926
2. Knuth, K. H. (2006), Optimal data-based binning for histograms, https://arxiv.org/pdf/physics/0605197.pdf
3. Hocke, K. and N. Kampfer (2009), "Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram", Atmos. Chem. Phys. 9, 4197-4206
4. Batina, L., B. Gierlichs, E. Prouff, M. Rivain, F.-X. Standaert, and N. V.-Charvillon (2011), Mutual information analysis: a comprehensive study, J. Cryptol. 24, 269-291
5. Kaiser, Henry F. (1958), The varimax criterion for analytic rotation in factor analysis, Psychometrika, **23** (3)