INTERNATIONAL TRANSACTIONS ON ELECTRICAL ENERGY SYSTEMS
Int. Trans. Electr. Energ. Syst. 2015; **25**:348–358
Published online 20 December 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/etep.1847

Preprocessing synchronized phasor measurement data for spectral analysis of electromechanical oscillations in the Nordic Grid

Luigi Vanfretti[1,2*,†], Sebastian Bengtsson[1] and Jan O. Gjerde[2]

[1] *KTH Royal Institute of Technology, School of Electrical Engineering, Electric Power, Systems Department, Smart Transmission Systems Laboratory (SmarTS Lab.), Teknikringen 33, 10044 Stockholm, Sweden*
[2] *Statnett SF, Research and Development, Nydalen Allé 33, 0484 Oslo, Norway*

SUMMARY

Spectral analysis techniques have successfully been applied for near real-time monitoring of power system small-signal electromechanical oscillations using synchronized phasor data from PMUs. The methods used for this purpose commonly assume that random load variations adopt the distribution function of Gaussian noise. Hence, careful attention has to be paid so that the preprocessing of synchronized phasor measurements is capable of providing data with the characteristics expected for these methods to work properly. This can be viewed as a "conditioning" step in the data handling process that has an impact on the results from spectral analysis which consume this data. This article aims to revisit the crucial step of preprocessing of PMU data in a tutorial fashion. The goal is to share the authors' experience when dealing with *real* PMU data originating from the Nordic Grid. This article offers a systematic and detailed methodology developed by the authors which has been successfully used in studies on estimation of electromechanical modes in the Nordic Grid. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS: inter-area oscillations; mode estimation; mode meter; Nordic Grid; spectral estimation; preprocessing; PMU

1. INTRODUCTION

1.1. Background and motivation

The application of spectral analysis techniques for power spectrum estimation of power system dynamic measurements has proven useful for the near real-time monitoring of poorly damped small-signal electromechanical oscillations, especially in the system under the Western Electricity Coordinating Council (WECC) [1] and others (see references in [2,3]).

The methodologies used for spectral analysis emerge from the field of digital signal processing [4] which includes linear prediction and power spectrum estimation. In power systems, the seminal works in [5,6] introduced the basic assumptions needed to apply these techniques to power system measurements to extract electromechanical dynamics from measured data. The prevailing and underlying assumption in these techniques is that power system dynamics, under most operating conditions, are largely driven by random load variations which excite the small-signal dynamics of power systems. When this assumption holds, power system modal frequency and damping estimates from ambient data can be computed using spectral analysis techniques [6].

Random load variations result in "ambient noise" [2], a low amplitude stochastic time series. The role of loads then has an important impact on the spectrum resulting from the spectral analysis of ambient data. However, for the extraction of electromechanical dynamics for monitoring purposes,

*Correspondence to: Luigi Vanfretti, KTH Royal Institute of Technology, School of Electrical Engineering, Electric Power, Systems Department, Smart Transmission Systems Laboratory (SmarTS Lab.), Teknikringen 33, 10044 Stockholm, Sweden.
†E-mail: luigiv@kth.se, luigi.vanfretti@statnett.no

the tools designed so far generally assume that loads can be described by either Gaussian noise [6] or the integral of Gaussian noise [7].

It is then important to observe that the signal processing tools being applied for spectral analysis expect signals with the basic features upon which they have been designed for, i.e. their probability distribution highly correlates with that of a stochastic Gaussian process (that of Gaussian noise or the integral of Gaussian noise). Hence, in most of the applications based on these techniques for electromechanical mode estimation, a "preprocessing" stage prepares different signals contained in data archives [1] or synchrophasor data stream [8]. In this context, data preprocessing can be seen as a "conditioning" step, where measurements are treated so to extract the stochastic Gaussian noise embedded in PMU measurements. Therefore, the procedures involved in preprocessing have obviously an important role, and the choices made in each of them can positively or negatively affect the "quality" of the data used for spectral estimation and the estimation of electromechanical modes. A recent IEEE Task Force report (see Section 2.7 in [3]) briefly outlines some of the strategies used for preprocessing, nevertheless, a detailed account on each step used in this process is not yet available in the literature.

1.2. Previous work

Despite the wealth of work in this area (see references in [2,3]) the topic of preprocessing for spectral analysis of PMU data has not been covered explicitly nor holistically. In general, preprocessing synchrophasor data consists in removing defective data, parceling data sets, removing outliers, interpolating missing samples, removing trends and filtering unwanted dynamics.

First, missing data and outliers need to be addressed; if not, they will have a significant negative influence on mode estimation results. When no more than 6% of missing data are present in a specific data parcel [9] recommends to omit the missing data and concatenate good data; furthermore, as illustrated in this article, missing data can also be interpolated. When there exist larger amounts of missing data, concatenation will introduce artificial transients while interpolation will result in noise, both effects negatively impacting mode estimates. The primary strategy recommended in [10,11] and illustrated in this article is to create data parcels separated in segments where missing data appears.

Next, outliers need to be considered; they are data that have significantly deviated from normal measures that can result in artificial transients which degrade mode estimation results. Outliers can be detected through model prediction errors [10], residuals [12] or robust objective functions [13]; in this article, a moving median algorithm is used to detect samples deviating largely from the median within a parcel.

Finally, it has been shown in [9] that mode estimation algorithms perform best with down sampled data to approximately five samples per second. PMU data provides synchronized phasors at data rates from 30 to 50 and 60 samples per second, thus providing a bandwidth of 15 to 25 and 30 Hz, respectively. Hence, high pass filtering is used to remove unwanted low frequency effects, while a low pass filtering is utilized for down sampling. Both FIR filters, used to avoid the addition of any modes to the data, and IIR filters have been used in the literature; however, no comparison of different low-pass filtering options has been published to the author's knowledge. Each of the steps mentioned above, and additional ones, are explained in detail and illustrated in this article.

1.3. Aims, scope and contribution of this article

This article aims to revisit the crucial step of preprocessing in a tutorial fashion. The goal is to share the author's experience when dealing with *real* PMU data; to this end, PMU measurement archives originating from two substations in the high-voltage transmission grid operated by Statnett SF, the Norwegian Transmission System Operator (TSO), were obtained during 2011 (see Figure 1).

1.4. Article organization

The remainder of this article is organized as follows. Section 2 introduces a streamlined preprocessing methodology, to which the authors have arrived after several experiences dealing with *real* PMU data.

DOI: 10.1002/etep

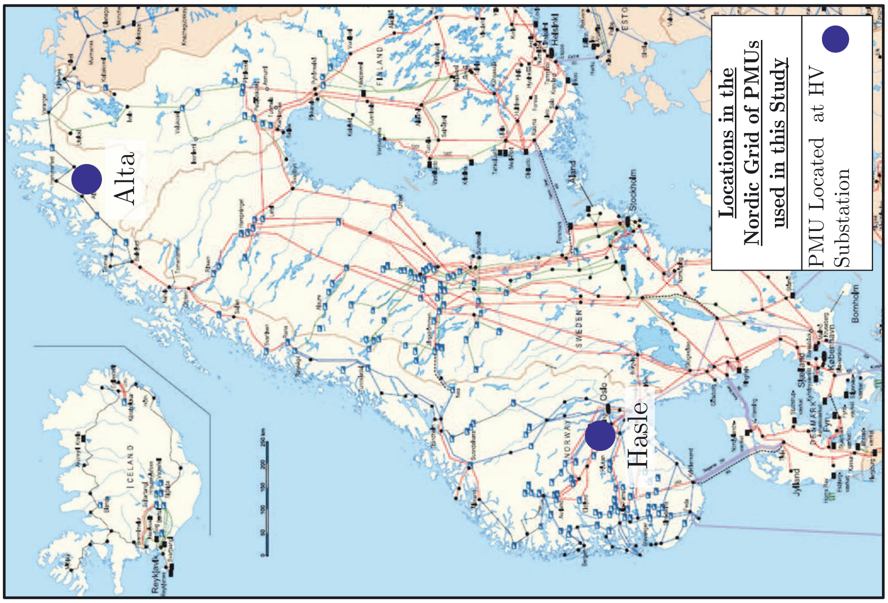


Figure 1. Nordic power system with locations of PMUs used in this study.

Data preprocessing is explained in detail by investigating each operation that takes place within the proposed process by applying to *actual* phasor measurements and giving clear illustrations on the impact of the choices made in each operation.

Section 3 illustrates the use of preprocessed data in spectral estimation; both simulated and *real* PMU data are preprocessed and power spectral densities are computed for different data archives.

Finally, in Section 4, conclusions are drawn while the Appendix gives a list and short description of MATLAB functions used in the preprocessing methodology presented. The preprocessing methodology and data presented in this article have also been used in studies on estimation of electromechanical modes of the Nordic Grid [14,15], giving satisfactory results.

2. PREPROCESSING

It is not exceptional for PMU data to have data quality issues. Prior to spectral analysis, the data should be curated to remove flawed, redundant and irrelevant data. In addition, when deterministic data errors are known, it is possible to correct these errors or to add estimates of missing samples.

System frequency control will cause power system data to exhibit deterministic characteristics. Since spectral estimators assume the input signal to be purely stochastic, deterministic components must be removed to obtain an accurate system model. To illustrate the necessity of this, the histograms in Figure 2 are computed.

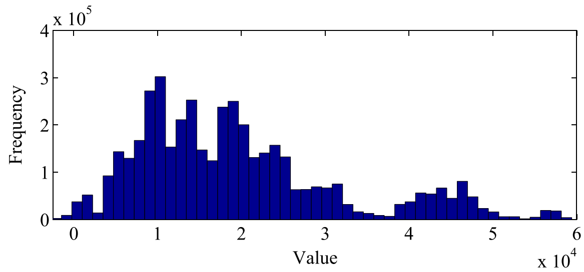
The histogram computed with raw data does not have the desired probability distribution of a stochastic process. However, after preprocessing, a histogram with the bell-shaped probability distribution of a stochastic Gaussian process is obtained. In Figure 3, the executed preprocessing steps are shown. These steps are described in detail in the following section.

2.1. Defective data removal

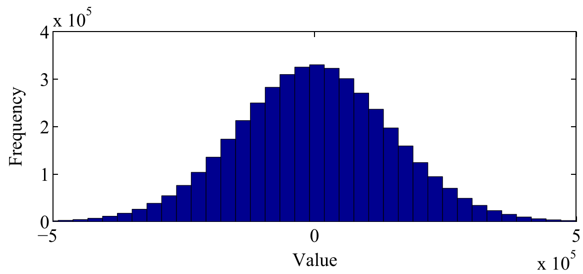
Internal PMU failures to update or retrieve sample information occur frequently and result in the following forms of defective data:

- Sequences of identical sample timestamps and/or values.
- Samples with value zero.

DOI: 10.1002/etep



(a) Histogram computed using raw frequency data from Hasle.



(b) Histogram computed using preprocessed frequency data from Hasle

Figure 2. Histograms using data from Hasle. (a) Raw data. (b) Preprocessed data.

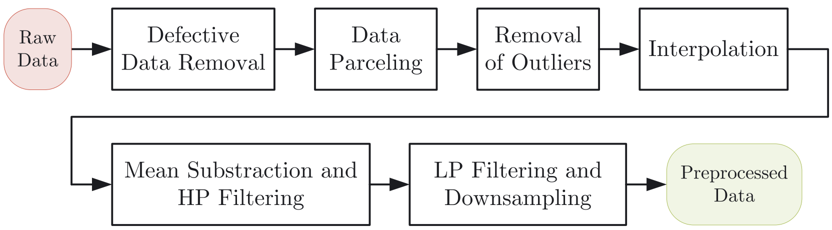


Figure 3. Preprocessing steps described in the following sections.

Errors such as these will cause a degradation in data quality and will induce outliers when computing frequency estimates from phase angle measurements. Frequency estimates can be calculated in the following manner:

$$f = f_0 + \frac{1}{2\pi}\frac{\delta\theta}{\delta t} \tag{1}$$

where f_0 is the nominal frequency of the system and θ is the phase angle.

When the samples are not correctly updated, as in Figure 4a, Equation (1) will yield a sequence of samples with value f_0 followed by an outlier, as in Figure 4b.

To prevent this degradation in data quality, imperfect data must be detected and removed.

2.2. *Data parceling*

Filters are highly sensitive to gaps due to missing data. Data located in the vicinity of such a gap will be corrupted when filtered. This corruption is very similar to that one illustrated in Figure 5, showing the effect of a filter's output when data containing an outlier is filtered. To mitigate this corruption, the data is split into parcels without any missing samples,[1] and then the individual parcels are processed separately.

[1]Estimates will later on be computed for smaller sections of missing samples; parcels are therefore allowed to contain such sections.

DOI: 10.1002/etep

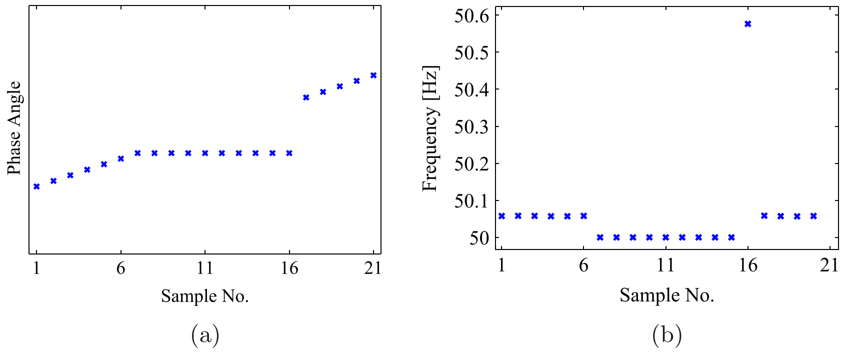


Figure 4. Data from Hasle. (a) Sample value is not correctly updated. (b) Corresponding frequency estimates.

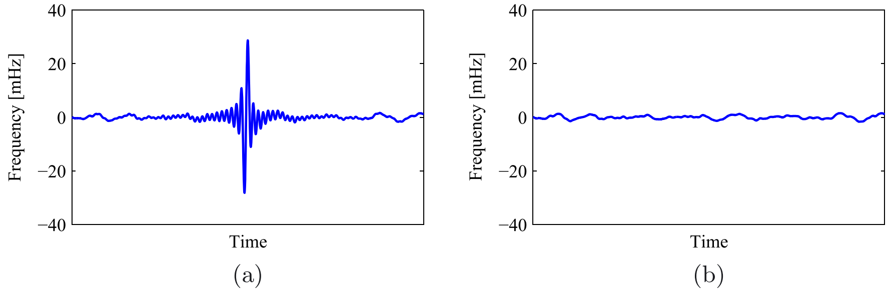


Figure 5. Data from Hasle. (a) Filtered data containing an outlier. (b) Filtered data, outlier removed prior to filtering.

2.3. Removal of outliers

An outlier is a sampled value that is subject to measurement error and therefore deviates significantly from the majority of the other samples. Outliers have an undesirable impact on the performance of filters, as illustrated in Figure 5 where the effects of a single outlier are shown. As can be seen from the figure, outliers induce artificial transients when the data is filtered. To avoid this sort of data corruption, outliers in the data must be removed.

Data comprised by a large amount of oscillations and with moving mean puts high requirements on the algorithms that are to detect outliers. An algorithm that is able to determine the moving median of the signal is suitable for this task.

2.3.1. Moving median algorithm. The moving median is computed (using a median filter of order 20) for the signal and samples that deviate more than k standard deviations from the moving median are removed. Results from applying this algorithm are shown in Figure 6, where $k = 3$ is used.

Note that since data characteristics for different PMUs tend to vary, it is necessary to select a value of k that fits each particular PMU.

2.4. Interpolation

Estimates replacing missing samples as well as removed data and outliers can be computed using interpolation. Because electromechanical modes can be observed in the frequency region up to 2 Hz [16]; the fastest dynamics that needs to be taken into consideration is a 2 Hz oscillation. In Figure 7, such an oscillation is depicted.

For the sampling rate of 50 Hz, there are 25 samples distributed over one period of the oscillation. Because of high frequency components and measurement noise, interpolation methods that make use of the slope of the data are inappropriate. The goal is to capture the general trend of the signal; hence, linear interpolation is a viable option for this purpose.

DOI: 10.1002/etep

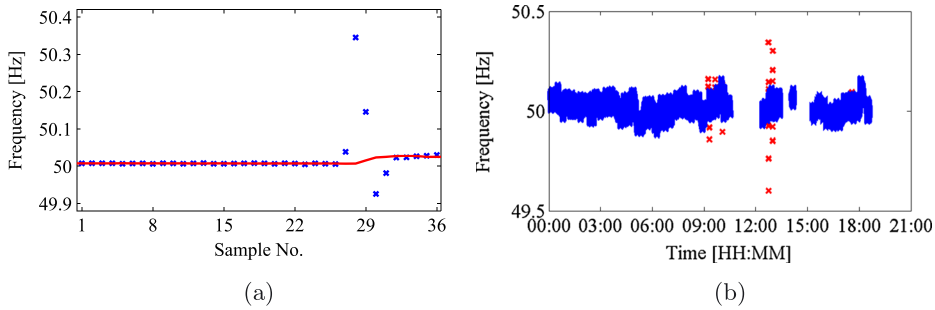


Figure 6. Data from Hasle. (a) Showing moving median (red line) and frequency (blue crosses). (b) Showing removed outliers (red crosses) and frequency (blue crosses).

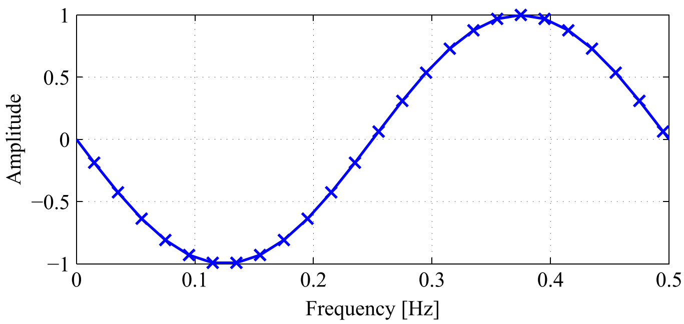


Figure 7. 2 Hz oscillations, 50 Hz sampling rate.

For a 2 Hz oscillation, one may obtain good estimates for up to 10 missing samples. However, if the peak of the oscillation is missing, accurate estimates can only be obtained about three missing samples [14].

2.5. *Removal of mean and high pass filtering*

System frequency control causes slow gradual changes in the measured signals. Such characteristic is a signature of governing-control actions, and hence the mean value of the frequency will change over time. These deterministic components must be removed to obtain the stochastic signal required for spectral analysis. However, if they are not removed, the relatively high amplitude of this governing mode, illustrated in Figure 8, will result in spectral leakage into the frequency range of the electromechanical modes.

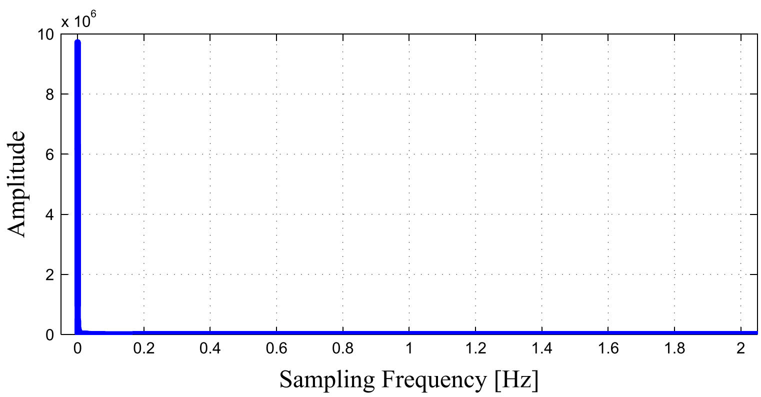
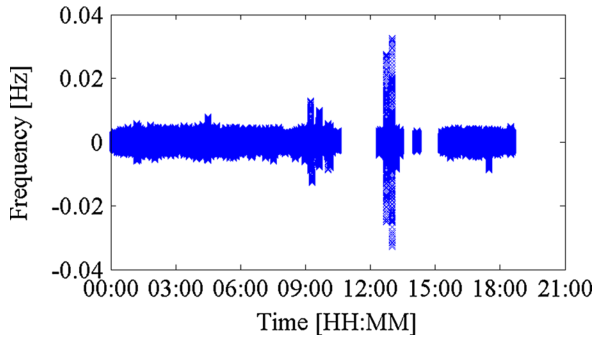


Figure 8. DFT using 24 h of data from Hasle showing the electromechanical frequency range and the governing mode located close to 0 Hz.

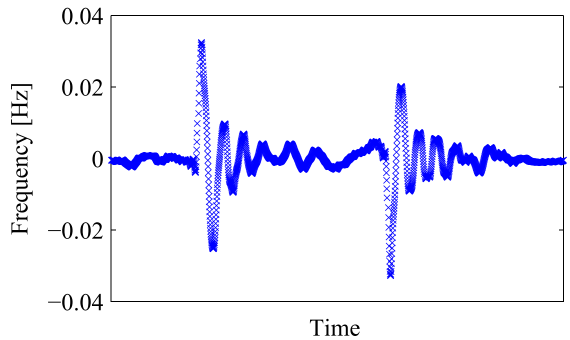
Since the frequency components related to the systems control actions are of relatively low frequency compared to electromechanical modes, they can be attenuated using a high pass filter. In Figure 9, the signal after subtraction of mean and high pass filtering is shown.

2.6. *Low pass filtering and downsampling*

Sampling rates that are significantly faster than the power systems electromechanical dynamics will result in redundant data that will slow down computation and give additional high requirements on processing power. Furthermore, when parametric methods are used, the autocovariance matrix that these methods rely on will become ill-conditioned for high sampling frequencies, resulting in poor numerical sensitivity. Consider an example of the conditioning of the Yule–Walker autocovariance matrix shown in Figure 10.



(a)



(b)

Figure 9. Data from Hasle. (a) After mean subtraction and high pass filtering. (b) Closeup of two transients from (a).

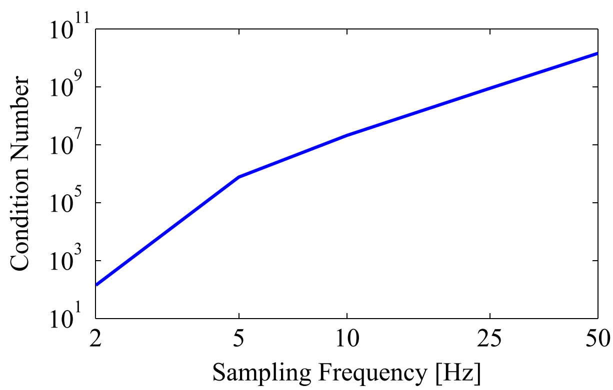
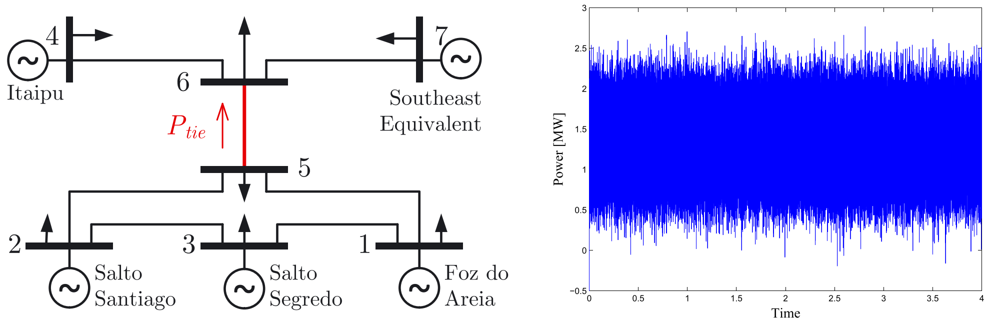


Figure 10. Condition number of the Yule–Walker autocovariance matrix for different sampling frequencies.

DOI: 10.1002/etep



(a) (b)

Figure 11. (a) Single-line diagram of the simulated power system. (b) 4 h of simulated data.

Table I. Mode frequencies and damping ratios of the simulated system.

Mode frequency [Hz]	Damping [%]
0.8308	6.9316
1.7869	14.5385

The amount of redundant data that can be safely removed without losing valuable information is given by the Nyquist–Shannon sampling theorem [17]. This theorem states that, for a sampled signal, perfect reconstruction can be assured if the sampling frequency is at least two times the highest frequency component of the signal. Since electromechanical modes are located in the frequency range up to 2 Hz [16], perfect reconstruction is guaranteed for sampling frequencies equal to or higher than 4 Hz.

To avoid errors due to aliasing, the signal must be passed through a low pass filter prior to being downsampled [17]. The filter should remove all frequency content above the Nyquist frequency, i.e. half the sampling frequency.

3. SPECTRAL ANALYSIS

3.1. Validation using simulated synthetic PMU data

Once the data have been properly preprocessed, it can be used as input to spectral estimators. To show the validity of the preprocessing approach developed above, synthetic PMU data is generated from a power system simulation.

The simplified seven bus model of the Southern/Southeastern Brazilian power system in Figure 11 (a) and used for analyses of forced oscillation effects on mode meter algorithms in [15] is used here. The model includes the Itaipu hydro power plant, which is connected by a 765 kV line to the load area. Four 500 kV buses (Buses 1–5), which form a ring with three generators, are connected to the Itaipu power station by 765kV tie-line (line between Bus 5 and Bus 6). All generators are modeled with a fifth-order model, and include controls. All model parameter data can be found in [18].

A classical small signal stability study reveals one inter-area mode with a frequency of 0.8308 Hz and a damping ratio of 6.9316%, and one local mode with a frequency of 1.7869 Hz and a damping ratio of 14.5385%, as shown in Table I. These values will be used as references in determining the validity of the data preprocessing approach.

Ambient data is simulated by adding uniformly distributed pseudo-random values (white noise[2]) in all load buses in the system. A signal of the tie-line active power, P_{tie}, which is sampled with frequency

[2]Uniformly distributed pseudo-random values are drawn from the standard uniform distribution using MATLAB's rand command.

DOI: 10.1002/etep

of 50 Hz, is shown in Figure 11b and used as input to mode meter algorithm. Note that the choice of this measurement will have an impact on the estimates computed because it will bear a more significant content of the inter-area mode oscillatory frequency.

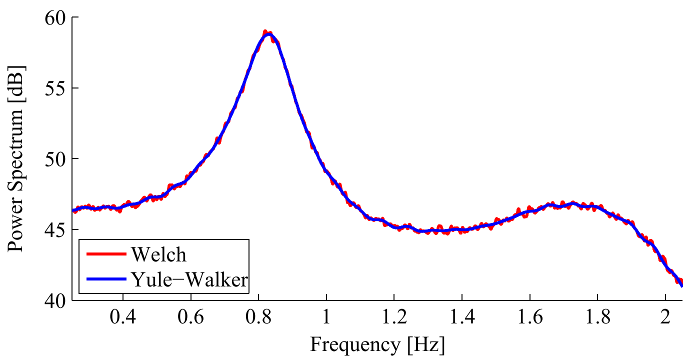
Spectral estimates of the tie-line active power, computed by Welch's and Yule–Walker methods, are shown in Figure 12. The spectral estimates are computed with parcels of both 24 h and 10 min of simulated data. The estimates show two peaks corresponding to the modes in Table I. Note that the variance in the PSD in Figure 12 (a) is larger to that of Figure 12 (b) because the PSD in Figure 12 (b) is computed with a much smaller parcel of data; nevertheless, the estimation of the inter-area mode at 0.8308 Hz is clear since the signal used has mostly content from this mode.

3.2. Application to real PMU data from the Nordic Grid

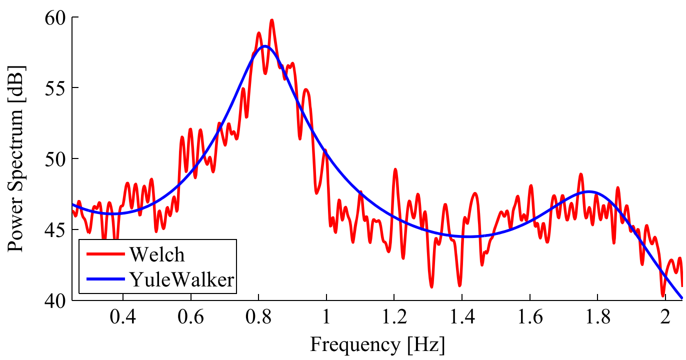
Spectral estimates are obtained for measurements from the Nordic Grid. Note that PMU data from power systems often exhibits far more oscillations, not all of them being "true" system modes; this can be seen in Figure 13 where two estimates using data from Alta are shown. Note that Figure 2 uses 24 h of data to compute the PSDs, while Figure 2 only uses 10 min. The model order of the YW estimator has not been modified for illustration purposes; however, note that when using 10 min of data, the model order should be decreased for better spectral estimation accuracy.

Figure 13 shows electromechanical oscillations within the ranges: 0.3457 – 0.3535 Hz, 0.5 Hz, 0.7695 – 0.7891 Hz and 1.328 – 1.332 Hz. These results are consistent with studies of electromechanical oscillations in the Nordic Grid both from measurement data [19–21,14,15] and models [22,23].

The PSDs also show forced oscillations [24,25]. A more detailed analysis of forced and inter-area oscillations is out of the scope of this paper, and the reader is referred to [14,15] for a more detailed analysis using recent PMU measurements from the Nordic Grid including analysis of forced oscillations.



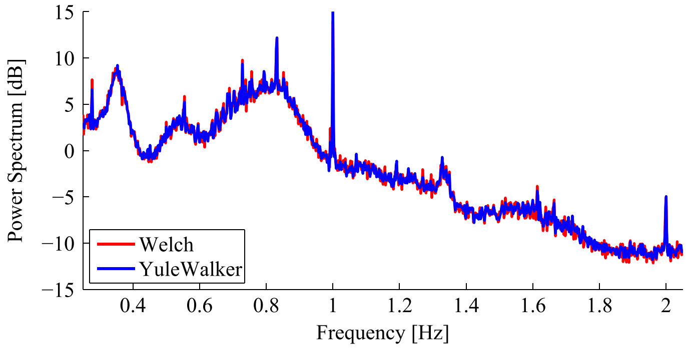
(a)



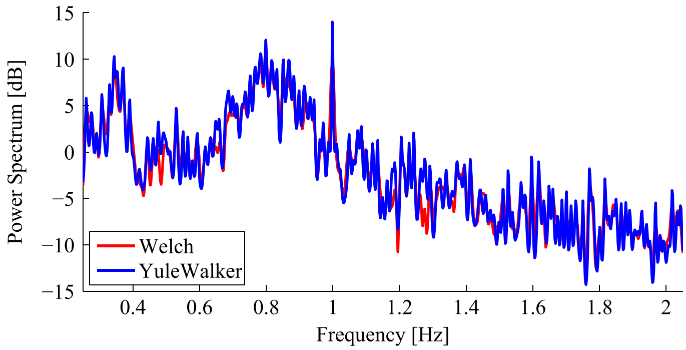
(b)

Figure 12. PSDs computed using (a) 24 h and (b) 10 min of simulated data.

DOI: 10.1002/etep



(a)



(b)

Figure 13. PSDs computed using (a) 24 h and (b) 10 min of data from Alta.

4. CONCLUSIONS

Preprocessing of PMU data has to be tailored to each individual application that will exploit it. In this case, the estimation of electromechanical oscillations requires the "conditioning" of the measurement data to expose key characteristics on which spectral analysis techniques operate.

This article has revisited the crucial step of preprocessing of PMU data for spectral analysis applications used in electromechanical mode estimation. The presentation has been carried out in a tutorial fashion, with the goal of sharing the methodology developed by the authors. Note that such a systematic methodology focusing on preprocessing is not available in the literature; therefore, the authors believe that the one presented in this article can be useful for other researchers, especially those starting to carry out investigations in this field.

The method developed by the authors has been validated using synthetic data from power system simulations, and *real* PMU data from the Nordic Grid. Moreover, the methodology has also been used in studies on estimation of electromechanical modes of the Nordic Grid [14,15], giving satisfactory results.

APPENDIX

Useful MATLAB functions used for preprocessing of PMU data

`linspace`: Linearly spaced vector.
`interp1`: 1-D interpolation.
`medfilt1`: One-dimensional median filter.
`fir1`: FIR filter design using the window method.
`filtfilt`: Zero-phase forward and reverse digital filtering.
`downsample`: Downsample input signal.

DOI: 10.1002/etep

REFERENCES

1. Hauer A, Trudnowski D, DeSteese J. A Perspective on WAMS Analysis Tools for Tracking of Oscillatory Dynamics. *IEEE Power Engineering Society General Meeting* June 2007; **2007**:1–10.
2. Trudnowski D, Pierre JW, in Inter-area Oscillations in Power Systems: A Nonlinear and Nonstationary Perspective, ser. Power Electronics and Power Systems, A. R. Messina, Ed.Springer, ch. Signal Processing Methods for Estimating Small-Signal Dynamic Properties from Measured Responses, 2009; 1–36.
3. IEEE PES Task Force on Modal Identification of Electromechanical Modes, "Identification of Electromechanical Modes in Power Systems," June 2012, [Online]. Available: http://www.pes-store.org/p-13616.htm.
4. Proakis J, Manolakis D. *Digital Signal Processing Principles, Algorithms, and Applications*, 4th ed. Prentice Hall: Upper Saddle River, New Jersey, 2007.
5. Hauer J, Cresap R. Measurement and Modeling of Pacific AC Intertie Response to Random Load Switching. *IEEE Transactions on Power Apparatus and Systems* Jan. 1981; **PAS-100**(1):353–359.
6. Pierre J, Trudnowski D, Donnelly M. Initial Results in Electromechanical Mode Identification from Ambient Data. *IEEE Transactions on Power Systems* Aug. 1997; **12**(3):1245–1251.
7. Vowles D, Gibbard M. Illustration of an analytical method for selecting signals and locations for power system modal-estimators. in *2010 IEEE Power and Energy Society General Meeting*, Jul. 2010; 1–7.
8. Parashar M, Mo J. Real Time Dynamics Monitoring System (RTDMS): Phasor Applications for the Control Room, in *42nd Hawaii International Conference on System Sciences*, *2009* (*HICSS '09*), Jan. 2009; 1–11.
9. Trudnowski D, Pierre J, Zhou N, Hauer J, Parashar M. Performance of three mode-meter block-processing algorithms for automated dynamic stability assessment. *IEEE Transactions on Power Systems* May 2008; **23**(2):680–690.
10. Ljung L. *System Identification Theory for the User*, 2nd ed. Prentice Hall: Upper Saddle River, New Jersey, 1999.
11. Zhou N, Pierre JW, Trudnowski D, Guttromson R. Robust RLS Methods for On-line Estimation of Power System Electromechanical Modes. *IEEE Transactions on Power Systems* May 2007; **22**(3):1240–1249.
12. Van Overschee P, De Moor B. *Subspace Identification for Linear Systems: Theory-Implementation-Applications*. Kluwer Academic Publishers, London, 1996.
13. Kovacevic B, Milosavljevic M, Veinovic M. Robust Recursive AR Speech Analysis. *Signal Processing* 1995; **44**:125–138.
14. Vanfretti L, Bengtsson S, Aarstrand V, Gjerde JO. Applications of spectral analysis techniques for estimating the nordic grid's low frequency electromechanical oscillations, in *IFAC Symposium on System Identification Brussels*, *2012*, invited Paper: Special Session on "Development of System ID Methods for Power System Dynamics, 2012.
15. Vanfretti L, Bengtsson S, Peric V, Gjerde JO. Effects of forced oscillations in power system damping estimation, in *2012 IEEE International Workshop on Applied Measurements for Power Systems*, 2012; 1–6.
16. Klein M, Rogers G, Kundur P. A fundamental study of inter-area oscillations in power systems. *IEEE Transactions on Power Systems* aug 1991; **6**(3):914–921.
17. Glad T, Ljung L. *Control Theory: Multivariable and Nonlinear Methods*. Taylor and Francis, London, 2000, ISBN 0-7484-0878-9.
18. Martins N. Eigenvalue and frequency domain analysis of small-signal electromechanical stability problems, in *IEEE Symposium on Application of Eigenanalysis and Frequency Domain Method for System Dynamic Performance*, 1989; 914–921.
19. Uhlen K, Warland L, Gjerde J, Breidablik O, Uusitalo M, Leirbukt A, Korba P. Monitoring amplitude, frequency and damping of power system oscillations with PMU measurements. *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, July 2008; 1–7.
20. Hemmingsson M, Samujelsson O, Pedersen K, Nielsen A. Estimation of electro-mechanical mode parameters using frequency measurements. *IEEE Power Engineering Society Winter Meeting* 2001; **3**:1172–1177.
21. Vanfretti L, Dosiek L, Pierre JW, Trudnowski D, Chow JH, Garca-Valle R, Aliyu U. Application of ambient analysis techniques for the estimation of electromechanical oscillations from measured pmu data in four different power systems. *European Transactions on Electrical Power* 2011; **21**(4):1640–1656. [Online]. Available: http://dx.doi.org/10.1002/etep.507.
22. Uhlen K, Elenius S, Norheim I, Jyrinsalo J, Elovaara J, Lakervi E. Application of linear analysis for stability improvements in the Nordic power transmission system. *IEEE Power Engineering Society General Meeting* 2003; **4**:–2103.
23. Vanfretti L, García-Valle R, Uhlen K, Johansson E, Trudnowski D, Pierre J, Chow J, Samuelsson O, Østergaard J, Martin K. Estimation of eastern denmark's electromechanical modes from ambient phasor measurement data, in *2010 IEEE Power and Energy Society General Meeting*, july 2010; 1–8.
24. Trudnowski D. Oscillation use cases, in *NASPI Meeting*, *Orlando*, *FL.*, *Feb. 29-Mar.1*, 2012.
25. Van Ness J. Response of large power systems to cyclic load variations. *IEEE Transactions on Power Apparatus and Systems* july 1966; **PAS-85**(7):723–727.