Data processing of electric probe data acquired under locally close-to-coherent noise conditions

F M Dias
Instituto de Plasmas e Fusão Nuclear – Instituto Superior Técnico, 1049-001 Lisboa, Portugal
E-mail: francisco.dias@ist.utl.pt

Abstract. A numerical procedure able to effectively eliminate close-to-coherent noise from measured data is presented. Although the data processing of probe measurements in Tokamak machines was the initial motivation, the current procedure is quite general concerning the kind and the source of data it can be applied to. As a result, the procedure can be useful for decreasing drastically the acquisition-time that otherwise is required to achieve acceptable signal-to-noise ratios in a very wide range of measurements.

1. Introduction
Removing high-level noise from probe characteristics is always a complex task due to the non-linearity of the current-voltage (I, V) function. Indeed, though in some cases noise removal may be effective enough so that “look nice” characteristics are achieved, distortions are unavoidable whenever fluctuations, whatever the physical processes involved, are detected as voltage-noise or as electron energy distribution function (EEDF) modulations. As a result, noise itself becomes “information” in the sense that its evaluation is as required as that of what is intended to be measured because actual fluctuations must be known in order to correct their effects in non-linear systems.

The simplest filtering procedures can be applied when the noise and the signal have spectral contents in completely different frequency regions, e.g. a low-frequency signal having a high-frequency noise. Conversely, filtering is more and more challenging as the above spectral regions merge. As a result, the instrumentation must be a guardian against noise by imposing a bandwidth not wider than that required for the signal being measured. In case of data acquisition systems, the above means a wise choice of the sampling rate and of the matching anti-aliasing filter, and eventually the need for external triggering in low signal-to-noise ratio situations.

Considering that the spectra of digitised data are commonly evaluated using the standard Fourier series, i.e., based on $n^{th}$-order harmonics where $n$ is an integer, filtering based on the above approach becomes extremely ineffective when the total acquisition-time is not an integer-multiple of the fundamental period of the noise. Then, a purely sinusoidal function and the $\sin(nt + \phi)$ terms of the Fourier series are not, in general, orthogonal functions, which means that the above series converges very slowly, and that spectra becomes too broaden. As a result, common filtering techniques may remove also a non-negligible part of the signal spectrum.

Unfortunately, spectral broadening is quite common in many kinds of measurements, either due to mains’ induced fluctuations in case of non-triggered data-acquisitions based on modern, non-real-time operating systems, or in case of measurements in magnetised plasmas. In the former case, it cannot be
exactly predicted when data is actually sampled (there are exceptions but they are expensive and not too common). In the latter case, fluctuations do not feature a fixed, fundamental frequency during the whole acquisition-time. In all cases, and as a final result, the acquired data seems to have random- or variable-frequency superimposed disturbances. To our knowledge, the previous, simple solution for handling non-coherent disturbances was based on an increased total acquisition-time so that disturbances could be considered as random variables, which obviously implied a waste of precious time.

Although the above is already a challenging noise-problem, noise can be too complex to be described accurately as a simple, variable-frequency disturbance because its amplitude is usually also time-modulated. Yet, in most cases it is acceptable to assume that amplitudes and frequencies are slow-varying functions of time. This is what we hereby define as “locally close-to-coherent noise” and the aim of the current work is precisely to provide a way on how such type of noise can be eliminated from measured data without an acquisition-time waste.

As a logical continuation of our previous works on the same subject [1,2], here we present an improved, adaptive, self-tuning, numerical procedure able to identify, to evaluate, and to eliminate locally close-to-coherent noise without the need for a useless waste of the acquisition-time. In addition, the pertinent information concerning noise is not lost, which means that a posteriori deconvolution and reconstruction procedures can be applied.

2. A glance into the real world
In order to avoid possible misunderstandings, we consider as noise whatever is present in the measured data other than what was not initially intended to be measured.

Unfortunately, the idea that noise is just a minor, random, high-frequency disturbance is too widely spread. As a result, too many people still believe that noise can always be readily eliminated using some simple, low-pass filter. Also unfortunately, but now under an educational point of view, manufacturers of lock-in amplifiers, boxcars, etc., use to “sell” the idea that such devices readily eliminate whatever noise problems users may have. The real problem is that new users “buy” such idea without paying attention for what kind of noise those equipments are actually effective as filters.

Here we show the difficulties that automated procedures based on the standard Fourier series have while trying to identify a purely sinusoidal function, $F$, which is the simplest kind of disturbance. Let us consider the case when the frequency, $f$, of the sinusoidal function is an odd-multiple integer of one-half of the total acquisition-time, $T$, i.e., $f = (2n + 1)/(2T)$, and let us calculate how many terms of the series are required to recover $F$ within a given error. Such results are shown in figure 1 for five values of $n$.

![Figure 1. RMS relative error as a function of the highest order of the Fourier series terms taken into consideration, for five values of the $n$ parameter (see text).](image-url)
As can be seen in figure 1, actual frequency values are somewhat difficult to be identified for small values of \( n \). Conversely, a huge number of the Fourier series terms need to be taken into consideration for high values of \( n \). Yet, an inexperienced observer can readily and accurately identify amplitude and frequency values in a plotted sequence of half-periods, which means that a standard Fourier series is not the general, best approach. Conversely, one single term of a modified Fourier series (based on a non-integer order of the harmonics) is all what is needed to evaluate the amplitude if the frequency is known \textit{a priori}. Although such means an \textit{eigenvalue} approach, the overall numerical procedure may run much faster.

Unfortunately, single-frequency functions do not belong to the real world of instrumentation. So, let us be realistic: what if a non-sinusoidal disturbance, whose amplitudes or/and frequencies are time-modulated, is superimposed over a signal? Well, then the fluctuation spectrum is broader and filtering operations are problematic because they will noticeably deplete information from the signal in the regions where the signal and the fluctuation spectra merge.

The very short analysis in this Section was solely intended to aware the reader that, within the limitations imposed by particular constraints, the spectral contents of a disturbance (or its time-variation) should be known as accurately as possible in order to choose the best filtering approach and to minimise the inherent errors.

3. Filtering procedures

It should be kept in mind that filtering and fitting are essentially equivalent procedures, which means that what are the sound mathematical guidelines for a fitting \cite{3} should also be followed for a filtering operation. Namely, the user must have an educated guess on what actual signal and noise functions are before attempting to separate them. Otherwise, filtering is not too different from “witchcraft”.

A common educated guessing is when measurements are intended to validate a theory other than the one that supports the measuring method. For instance, if, and only if, probe measurements are intended to prove that EEDFs are Maxwellian, then fittings are the best filtering procedures. Yet, and by definition, then the experimentalist should not be able to achieve anything else than Maxwellian EEDFs. Fortunately, the number of degrees of freedom is often large enough for the experimentalist to have good grounds to suggest that EEDFs are something else, \textit{e.g.} bi-Maxwellians \cite{2,4--5}.

Another educated guessing is the so called optimal (Wiener) filtering \cite{6}. Although it does not rely on a completely objective approach, its name clearly suggests that even applied mathematicians consider it as “the best one may get”. The idea is simple: considering that each kind of signal and noise have spectral \textit{signatures} of their own, which experienced users are expected to be readily able to identify, whatever unexpectedly deviates from a signal-trend is noise, and vice-versa; optimal filtering is the procedure that corrects the above deviations so that the expected trends are kept. This approach is well-suited for \textit{white} noise and for coherent noise because they feature precise spectral \textit{signatures}.

As the reader may have already understood, filtering becomes as problematic as the \textit{a priori}, required information about the signal and/or the noise functions are missing. Concerning probe measurements in Tokamak machines, though it has been assumed that EEDFs could be considered as Maxwellian, it was shown later that such was not the case (see references above). In addition, this author prefers to evaluate EEDFs without any \textit{a priori} assumptions concerning the EEDF. Fortunately, an analysis of several probe characteristics has shown that disturbances and signals were not correlated (note: there is an apparent correlation between the signal and the amplitude of the noise, which we interpret as inherent to how the plasma fluctuations are detected by a probe, and which the reader should not misunderstand as an actual signal-noise time-correlation). As a result, the soundest educated guess is the peculiar locally close-to-coherent noise feature, which makes our approach somewhat non-orthodox. Indeed, instead of identifying the signal and then the noise is “what else”, which is the most trivial approach, we start by identifying the noise and the signal is what remains after noise is removed. Hence, it is not strange that our procedure becomes more accurate for higher noise amplitudes (see below) because then it is easier to identify what actually is the noise component.
We wish to emphasise a few more filtering basics before starting a more comprehensive description of our current procedure.

1) Filtering operations carried out in the time-domain typically use finite impulse response filters [7] due to practical reasons, i.e., finite sets of data.
2) Coherent (or close-to-coherent) noise typically asks for filter functions whose duration [7] must keep a close relation with the period (or the characteristic period).
3) The larger the filter function duration the lower the frequency of the signal components that will be removed by the filtering operation, hence the higher the distortion in those particular cases where the spectral density of the signal is a decreasing function of the frequency, which is typically the case for probe measurements.

In simpler words, the above means that an effective filtering can only be achieved if the filter-size and the characteristic period values are similar in value and that the signal will be unavoidably distorted for large values of the characteristic period unless corrective measures are applied (see [8] for possible solutions).

Readers more interested in the theoretical grounds of less-common, time-frequency transforms are hereby suggested to look for wavelet, in general, and for Short-Time Fourier Transform (STFT), in particular, in general mathematical text books. Due to the possible similarity between our procedure and a STFT, we wish to aware the readers that we are not calculating “local” spectra. Instead, what we do is to locally fit the noise by a function that satisfies our locally close-to-coherent noise definition.

3.1. The locally close-to-coherent noise filtering procedure

Here we describe how our numerical procedure handles locally close-to-coherent noise while processing the 1st derivative of probe data, which is the starting point for EEDF evaluation under collisional conditions, e.g. in magnetised plasmas. Nevertheless, and considering that the actual noise processing only starts at step 3) below, readers are free to “input” whatever data they may like to be filtered at that step. There are other reasons why we use the 1st derivative instead of the original data: derivatives enhance the amplitudes of high-frequency components, our procedure works better for higher noise levels, and we expect that noise will have a high-frequency content higher than that of the signal.

Our procedure works in the following way:

1) The \((I, V)\) measured data is sorted, and interpolated in order to provide 1024 uniformly \(V\)-spaced bins as a way to speedup the whole processing (explanation: considering that sampled voltages are usually less noisy than those of the currents, this step avoids the need for a more complex \(V\)-filtering; we currently follow a linear interpolation based on the 3, closest measured \(V\)-points but the reader is free to use any other sorting procedure).
2) A raw, first derivative, \(dI/dV\), is numerically calculated from the convolution of the above data with the first derivative of an instrument function (IF), whose full with at half maximum (FWHM) is adaptively chosen aiming for locally equal noise and error values. Noise is deduced from 2nd derivatives using 3 adjacent data points [8], and error from a one-step algebraic reconstruction technique (ART) based on a deconvoluted form of Hayden procedure [9] as mentioned in [10]. As in previous works (see e.g. [1]), our basic IF, \(g(j)\), is a Hann-like window

\[
g(j) = \frac{1 + \cos(\pi j/n)}{2n},
\]

where \(n\) is the initial FWHM value. Anyway, each convolution is followed by an one-step ART, hence the actual IF is \(2g - g \ast g\), where “\(\ast\)” stands for convolution.
3) The characteristic period of the disturbance is evaluated as the lowest value of \(\tau\) \((0 < \tau < T/2)\) that maximizes the function:

\[
\tau^{1/2} \int_0^T I(t)^2 dt \sqrt{\int_0^{T/2} I^2(t)dt} \int_0^{T/2} I^2(t + \tau)dt.
\]
We found that the above autocorrelation-like procedure is more reliable than all the other methods that we have tested. The so calculated characteristic period is currently only used as a seed for the first point of the iterative process carried out as described in steps 5)–9) below.

4) An approximated value of the local period, which is $V$-dependent, is deduced at each point, i.e., for each window, as being the window size that maximizes the amplitude of the fundamental under a Fourier series approach and approximated values of its COS and SIN components are calculated.

5) The first derivatives of the disturbing signal amplitude, phase, and frequency are calculated (note: in order to avoid an increased processing-time, which seems to be useless for Tokamak data, we are not yet accounting for the 1st derivative of the frequency variation).

6) The period of a dummy, amplitude- and frequency-modulated sinusoidal function matching the apparent period found in step 4) is calculated accounting for the above first derivatives, and a more accurate amplitude value is calculated accounting for the 2nd derivative, $d^2 I/dV^2$, which is evaluated at step 8) during the previous loop (explanation: we wish to emphasize our non-orthodox approach as mentioned in section 3; indeed, what we are actually evaluating is noise, and higher derivatives of the signal are being used to achieve a higher level of accuracy concerning noise evaluation). Note that it is in this step that the actual close-to-coherent fitting takes place.

7) $dI/dV$ is filtered using the same FWHM that leads to a notch-filter for the dummy function (explanation: the noise will be absolutely made equal to zero if it satisfies our initial assumptions).

8) $d^2 I/dV^2$ is calculated from the filtered $dI/dV$ achieved in the previous step.

9) The procedure loops back to step 5) until convergence is reached.

10) An additional ART is carried out to guess the error of the whole procedure.

In a former version of our numerical procedure, and as shown in [1], the error values evaluated at the above step 10) were in good agreement with the actual error under model test conditions. Although in the past we were pleased by such result because it provided a means to have a likelihood parameter, latter we realised that values achieved at step 10) and those of the initial disturbance were too closely correlated to be considered as an error, hence they should be considered as information instead. As a result, in the current work we decided to use the results achieved at step 10) as a further improvement to the filtering ability of our procedure. The drawback from such decision as well as the directions for further improvements will be discussed below at the appropriated places.

4. Model tests

To our best knowledge, the current numerical procedure is the only that effectively filters data measured under locally close-to-coherent noise conditions, e.g., that achieved in Tokamak machines, without imposing unnecessary, restrictive fittings. Although our results have been confirmed by those obtained using less-detailed approaches [2,4–5], we think that a model test is the best way to prove the correctness of our procedure. Considering that the procedure was initially intended to process probe data measured in collisional conditions, we have chosen a test probe characteristic whose 1st derivative of the signal component is (see e.g. [5] and references therein):

$$\left(\frac{dI}{dV}\right)_{\text{signal}} = \begin{cases} a(V_p - V)e^{\frac{V-V_p}{b}} & V \leq V_p \\ 0 & V > V_p \end{cases} \quad (3)$$

The value of the plasma potential, $V_p$, was set to 70 V, and the parameter $b$, which provides information on the electron temperature, was set to 20 V. The 1st derivative of the noise component (see solid curves on the left side of the appropriate figures 2–4 below) is given by:

$$\left(\frac{dI}{dV}\right)_{\text{noise}} = f(j)\sin\left(\frac{2\pi j}{c(1 - 5 \times 10^{-7} j^2)}\right), \quad (4)$$

where $j$ is the bin order ($0 - 1023$), $c$ is a scaling parameter for the local period, and $f(j)$ provides an amplitude variation similar to that found in Tokamak data and scales amplitudes. Scaling is hereby
used to analyse how much the characteristic period and amplitude of the noise have an influence on the accuracy of our procedure. As reference conditions we shall consider $c = 120$ and a maximum amplitude value equal to 5.93 a.u..

Figure 2. Model test results for the reference conditions (see text). Left: 1st derivative of the noise component (solid), filtered 1st derivative of the signal (dash), error (dot), and the associated probe characteristic (dash dot). Right: error (solid) plotted in an expanded scale.

The model test results for the reference conditions are depicted in figure 2, while figures 3 and 4 show those achieved, respectively, when the amplitude and the local period was made twice as large.

The error, which we consider the most pertinent result, can be seen in an enlarged scale on the right side of figures 2–4, and its analysis shows that: i) the initial noise is virtually absent in the filtered data; ii) error values are rather small (<1%) except in the vicinity of the plasma potential and close to the boundaries of the data-interval; iii) the filtering efficiency increases as noise amplitude increases (note that the span of the error scale is the same in figures 2 and 3) but, conversely, error values in the vicinity of the plasma potential are very sensitive to that of the noise period.

Figure 3. Model test results when noise amplitude is doubled. Left: 1st derivative of the noise component (solid), filtered 1st derivative of the signal (dash), error (dot), and the associated probe characteristic (dash dot). Right: error (solid) plotted in an expanded scale.

A more precise analysis can be achieved checking the accuracy as the current procedure identifies the value of the plasma potential, which is a key parameter in probe diagnostics. The resulting errors are 2.5, 1.4, and 7 V, respectively, in the reference, the doubled-amplitude, and the doubled-period case. Under the same conditions, the FWHM of the Hann-like filter are, respectively, 7, 7, and 13 V.
Concerning the doubled-amplitude case, and taking into account that the error value is only 1/5 of that of the FWHM, we consider the above result as very good. Conversely, the similar, relative result for the reference case is 1.75 times worst, which suggests that the current numerical procedure still has some limitations for lower-level amplitudes, whose explanation we attribute to the fact that the 1st derivatives of the period were not taken into account. Concerning the doubled-period case, the error was expected to increase due to the unavoidable increase of the FWHM values, which the current results confirmed. Nevertheless, we wish to emphasise that (3) leads to a discontinuity of \( \frac{d^2I}{dV^2} \) at \( V_p \), which suggests that the expected error values should be smaller under a real condition situation. The reader should keep in mind that the electron temperature is accounted in (3) through the \( b \) parameter, whose value was set to 20 V, which means that all the above errors concerning the plasma potential evaluation would be fully acceptable as a first, uncorrected probe diagnostic.

Cross-correlations between the actual noise amplitudes and those evaluated by our numerical procedure lead to an analysis similar to that achieved above concerning the plasma potential. Indeed, cross-correlation values are 0.999245, 0.999634, and 0.991431, respectively, in the reference, the doubled-amplitude, and the doubled-period case (note: in the above calculations the data-interval was shortened by “1” FWHM at each boundary to minimise end-effects; see explanation below). Under the same model test conditions, local periods are evaluated with 2.02, 1.42, and 6.58 % RMS averaged errors.

The increased error close to the boundaries of the data-interval is trivial as a result and always arises in case of a finite set of data, which means that the size of the input data-interval should be always larger than that of the data to be actually evaluated. As a guideline, each extended region should be larger than the respective FWHM value at each boundary.

5. Discussion and Conclusions

We have improved a previous version of our numerical procedure intended to filter locally close-to-coherent noise, e.g. that found in Tokamak probe data, and we have analysed how much the characteristic values of the noise amplitude and of the period have an influence on the filtering efficiency. The improvement was made possible by considering as “information” what previously was assumed as just a measure of the error. This approach is justified by the fact that we found that the above evaluated “error” and the noise were closely correlated. Although this new approach made us lose our previous likelihood parameter, we think that such is just an intermediate phase that, in a near future, will lead to further improvements and to a more accurate error evaluation because an increased filtering efficiency is the basic condition so that an increased number of deconvolution iterations are possible without retrieving back a noticeable amount of noise.
Considering that filtering in the time-domain, i.e., a convolution, is a reversible operation, readers may wonder about the validity of the last statement in the above paragraph (note: we shall skip here the practical limitations when discrete convolution operations are applied to finite sets of data). Yet, when deconvolutions are carried out using iterative procedures, which is the scheme we use, the recovering speed is spectrum-dependent, which means that it is not the same for the signal and for the noise components. The most striking example is when a DC signal having a superimposed, constant amplitude and period, sinusoidal noise is filtered by a rectangular window whose width is precisely equal to the period. Then, the result of deconvolutions intended to sharpen the IF is errorless and noiseless independently of the number of iterations carried out. Our wishful approach is precisely to build a filtering procedure able to eliminate completely the noise superimposed on data.

Following the same line of thought, we expect that further improvements will be possible when the 1st derivative of the period is taken into account because then the filtering efficiency should be even greater hence also the possibility to increase the number of deconvolution iterations without a noticeable increase of the noise level.

Concerning the influence of the noise amplitude on the filtering efficiency, we have shown that the latter increases as amplitudes increase, which was explained as a direct result from the so increased ability of the numerical procedure to identify what noise is. As a corollary, errors decrease under the same conditions. Conversely, an increased noise period leads to a similar error variation, which was the expected result due to the inherent FWHM value increase of the IFs and due to the merging of the signal and the noise spectra.

The current numerical procedure keeps, as a heritage from its previous versions, features that are not the most appropriated for high-accuracy, numerical processing. For instance, some of the variable arrays are of the integer-type. Yet, the wishful idea of the author was (and still is) that the filtering program may be included in the firmware of a front-end, digital data processor, in which case speed constraints need to be considered very seriously whenever data-processing operations are to be carried out during the time between consecutive data-acquisitions. Conversely, we have good reasons to expect that the filtering efficiency will be further increased as soon as all numerical variables are handled as floating-point entities.

As a final remark, we wish to emphasise that our filtering procedure is able to accurately retrieve noise amplitudes and frequencies, which are basic requirements for EEDF reconstruction, and which is the starting point to achieve accurate values of the electron density and of the plasma potential. We have the knowledge on how to reconstruct EEDFs from averaged probe data provided that additional information concerning EEDF modulation is made available, which is not possible from one-probe measurements. As a result, the filtered results that we can achieve are bound to be close-to-useless concerning EEDF evaluation unless information achieved from other simultaneous diagnostics, or any sound theoretical guidelines, are provided so that the appropriated EEDF reconstruction can be applied. Otherwise, many thousands of probe measurements, which have never been processed for “having some superimposed weird disturbance”, will be forever lost.

Acknowledgments
This work was partially supported by FCT/FEDER in the framework of the project “Ecological Engineering Plasma Laboratory” POCI/FIS/61679/2004. The author wishes to thank all those who have provided noisy probe characteristics so that the reliability of the numerical procedure could be tested in real situations. The author also wishes to thank Dr. Tsv. Popov for the fruitful discussions concerning EEDF evaluation of probe measurements in magnetised plasmas.

References
[1] Dias F M and Popov Tsv K 2007 Proc. 34th EPS Conf. on Plasma Physics (Warsaw) ECA 31F P-5.080
[2] Popov Tsv K, Ivanova P, Stöckel J, Dejarnac R and Dias F M 2007 Proc. 34th EPS Conf. on Plasma Physics (Warsaw) ECA 31F P-5.104
[3] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge: Cambridge University Press) chapter 15.0 pp 656–7

[4] Popov Tsv K, Ivanova P, Benova E, Bogdanov T, Dias F M, Stöckel J and Dejarnac R 2008 *Proc. 35th EPS Conf. on Plasma Physics (Hersonissos)* ECA 32 P-4.081

[5] Popov Tsv K, Ivanova P, Dimitrova M, Stöckel J and Dejarnac R 2008 *J. Phys.: Conf. Ser.* 113 012004

[6] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge: Cambridge University Press) chapter 13.3 pp 547–9

[7] Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge: Cambridge University Press) chapter 13.1 p 538

[8] Dias F M and Popov Tsv K 2007 *J. Phys.: Conf. Ser.* 63 012005

[9] Hayden H C 1987 *Comput. Phys.* 1 74–5

[10] Dias F M and Popov Tsv K 2002 *Vacuum* 69 159–63