



SCOUT: Signal Correction and Uncertainty Quantification Toolbox in MATLAB

Richard Semaan

Technische Universität Braunschweig

Vikas Yadav

Technische Universität Braunschweig

Abstract

This manuscript describes the software package **SCOUT**, which analyzes, characterizes, and corrects one-dimensional signals. Specifically, it allows to check and correct for stationarity, detect spurious samples, check for normality, check for periodicity, filter, perform spectral analysis, determine the integral time scale, and perform uncertainty analysis on individual and on propagated signals through a data reduction equation. The novelty of **SCOUT** lies in combining these various methods into one compact and easy-to-use toolbox, which enables students and professionals alike to analyze, characterize, and correct for signals without expert knowledge. The program is oriented towards time traces, but an easy adaptation to spatial distributions can be performed by the user. **SCOUT** is available in two variants: a graphical user interface (GUI) and a script-based version. A key motivation of having two variants is to offer maximum flexibility to adaptively and visually adjust the analysis settings using the GUI version and to enable large batch processing capabilities and own code-integration using the script-based version. The package includes both variants as well as three example scripts with their corresponding signals.

Keywords: stationarity, spurious sample detection, filtering, uncertainty analysis, MATLAB.

1. Introduction

Signal processing and uncertainty quantification are two very broad yet related research areas. They have applications in many fields ranging from engineering to mathematics, to music. With the proper tools, signal analysis allows us to discover valuable traits and characteristics in signals, whereas uncertainties quantification informs us about the origin, propagation, and interplay of different sources of errors. Due to their relevance, many open-source and commercial software packages have been developed, such as the **SciPy** (Jones, Oliphant, Peterson *et al.* 2001–2019) and the **Signal Processing Toolbox** (Mathworks 2019) packages for signal analysis, and **UQLab** (Marelli and Sudret 2014) and **propagate** (Spiess 2018) packages for

uncertainty propagation. These packages, and others, offer an impressive range of capabilities and options but are mainly geared toward expert users in either field. This limits their usability in many engineering fields.

In this manuscript, we present **SCOUT**, an easy-to-use signal processing and uncertainty quantification **MATLAB** package that is well suited to today's students and professionals alike. It offers the main tools necessary to analyze, categorize, and quantify the uncertainty of acquired signals. Specifically, it allows to check and correct for stationarity, detect spurious samples, check for normality, check for periodicity, filter, perform spectral analysis, determine the integral time scale, and perform uncertainty analysis on individual and on propagated signals through a data reduction equation. The uncertainty analysis yields uncertainties on central moment up to the fourth moment. **SCOUT** is available in two variants: a graphical user interface (GUI) version we label **SCOUT GUI**, and a script-based version we call **SCOUT Script**. The GUI version offers maximum flexibility to adaptively and visually adjust the analysis settings, whereas the script version enables large batch processing capabilities and own code-integration.

SCOUT is available as free software and can be downloaded from <https://www.richardsemaan.com/home/software/>, with all supplementary materials which include both variants and three example scripts with their corresponding data.

2. General Characteristics

The two **SCOUT** versions, the graphical user interface (GUI) and the script versions are intended to be complimentary. **SCOUT GUI** provides visual as well as numerical output at every step. It allows dynamic and adaptive analysis with a range of options and parameter tuning. On the other hand, **SCOUT Script** is command-based designed for integrated and batch processing. In other words, it allows all of **SCOUT**'s capabilities to be integrated within the user's own analysis code. In addition, the script version enables batch processing of large data. An envisioned workflow scenario would be to perform initial analysis using **SCOUT GUI**, where the settings are fine-tuned, followed by the **SCOUT Script** for integrated or batch processing. In the following, we list general characteristics common to both versions:

- Each signal to be imported should *only* include the time trace without the time vector, i.e., a one-dimensional array. **SCOUT** generates the time vector from the specified sampling frequency and the inferred record length. Including more than one signal in the file will result in only one signal imported; the first one that is read.
- All signals *must* be sampled at a constant sampling frequency f_s (i.e., equal time interval Δt between the samples). Moreover, if uncertainty propagation is sought, all propagated signals should be sampled at (or decimated to) the same rate.
- Up to 5 signals can be simultaneously analyzed, and, if desired, their uncertainty propagated.
- It is recommended that the user performs the analysis in the presented order, i.e., 'Stationarity Analysis', followed by 'Spurious Samples', and so on.

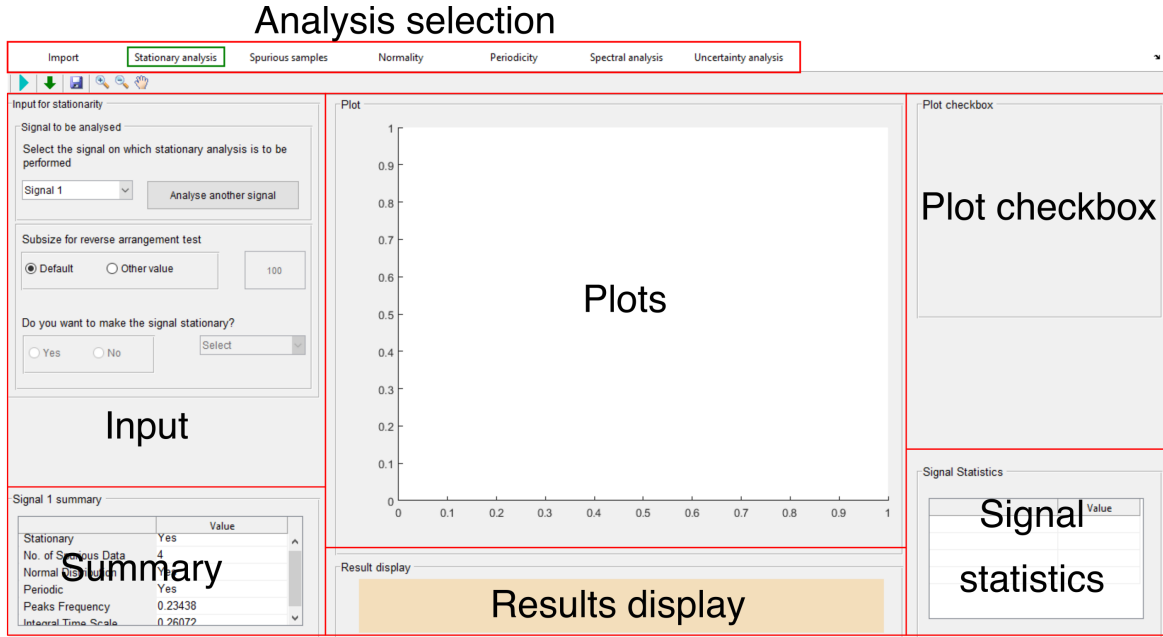






Figure 1: The general layout of **SCOUT GUI**. The red boxes highlight the various sections of the interface.

3. Interface and Code Execution

The two **SCOUT** versions have two very different interfaces. While the GUI version is adaptive, the script version is static and requires a configuration file. This section details both interfaces.

3.1. SCOUT GUI Interface

A screenshot of **SCOUT GUI**'s interface is presented in Figure 1. The red boxes highlight the various regions of the layout, which are detailed below. The layout structure is general and applies to most of the analysis steps. It has the following general characteristics:

- The user is required to press the green  button to execute an analysis step.
- Hovering over the  buttons provides information and tips.
- Each figure can be manipulated and saved using these buttons  on the top left.
- The user can generate and download a summary file using the  button.
- Each signal is constantly updated and logged in the workspace with each analysis step. This enables the user to select any previously-analyzed signal version for further examination.

For the remainder of this section, we detail the various interface sections.

Analysis selection section

The analysis selection section consists of seven tabs for the various analyses. It is recommended the user performs the analyses in the suggested order from left to right. In other words, after importing the signal(s), the next recommended analysis step is to check for stationarity, followed by checking for spurious samples, and so on. It is worth to note that the user is not *obligated* to perform all suggested steps. Each processing step can be independently conducted.

Input section

The input section requires the user to input the information necessary for each process. The input type ranges from numerical values to checkboxes, to drop-down menus. The input section changes with each analysis tab.

Summary section

The summary section displays the latest general summary of the signal analyzed. This section is constantly updated after each analysis step. The summary section displays the following information about *each* signal:

- Signal stationarity status
- Number of spurious samples
- Whether the signal is normally distributed
- Whether a periodic component is removed
- The detected peak frequency
- The integral time scale

It is worth to note that this section distinguishes itself from the ‘Signal statistics’ in that it does not display any statistical information about the signals.

Plot and plot checkbox sections

The plot section in the middle is reserved for plotting the various results. Plots range from time traces to histograms, to spectra. The user can toggle between various signals using the checkboxes on the right side.

Results display

The purpose of the results display section is to ‘announce’ important results after the completion of a certain analysis step. This includes the following:

- Whether the signal is stationary with 95% confidence.
- Whether the signal is normally distributed with the specified level of confidence.
- The frequency of the highest detected peak in the spectrum.

Signal statistics

The signal statistics section displays the latest statistical information of the signal being analyzed. This section is constantly updated after each analysis step. The signal statistics section displays the following statistical information about *each* signal:

- Mean
- Standard deviation
- z-score (for stationary analysis)
- Kurtosis
- Skewness

3.2. SCOUT Script Interface

SCOUT Script requires one post-processing configuration file for each imported signal, and a single optional uncertainty analysis configuration file. The configuration files contain the various analysis settings. Hence, unlike **SCOUT GUI**, all analysis settings are fixed during the execution.

The code execution is performed by calling on the main function `SCOUT_Script` followed by the various configuration files:

```
>> SCOUT_Script(ConfigFile1('U','Workspace','u',10000),UncertaintyConfigFile)
```

This code snippet, taken from `Example1.m` of the package, shows how `SCOUT_Script` handles the two types of configuration files:

1. Configuration file type 1 (e.g., `ConfigFile1`): is required for the entire analysis sequence *except* for the uncertainty analysis (the last tab in **SCOUT GUI**).
2. Configuration file type 2 (e.g., `UncertaintyConfigFile`): to specify the uncertainty analysis settings.

In case uncertainty analysis is not sought, the user can simply withhold the type 2 configuration file. Each type 1 configuration file requires four direct inputs:

1. The name of the output summary file.
2. The location of the signal (`Workspace` or `Directory`).
3. The name of the signal in case the file is loaded from the workspace, or the name of the file including its path in case the signal is loaded from a directory.
4. The sampling frequency.

These input options allow batch processing of different signals with different names, saved at different locations, and sampled at different rates. The reader is referred to **Example3** in the package for a batch processing example script.

Similar to the GUI version, the script version can simultaneously process five signals. A code execution example for simultaneously analyzing two signals with two different configuration files without uncertainty analysis would be (refer to **Example2.m**):

```
>> SCOUT_Script(ConfigFile1('U','Workspace','u',10000), ...
ConfigFile2('V','Directory','Data/Sig2.mat',10000))
```

Besides the three direct input variables, the script version requires a list of other settings in the configuration files. The various analyses offered and their corresponding settings are detailed in section 4. The user is also referred to the example scripts in the package and the therein-included comments for further details.

It is worth to note that unlike **SCOUT GUI**, **SCOUT Script** does not offer the option to select different signal variants at different stages of the analysis. Specifically, **SCOUT Script** sequentially performs all analysis steps activated in the configuration file always using the most up-to-date signal from the previous step. For example, if the users wishes to analyze a signal *without* removing any possible spurious samples, they are simply required to deactivate the relevant variables in the configuration file type 1 by setting: `I.Spurious.Chauvenet.n = 'NA'` and `I.Spurious.Postprocessing = 'NA'`. The variables in the aforementioned scenario are detailed in section 4.2.

4. Analyses

All analyses begin with importing signals. Up to 5 signals can be simultaneously imported. In **SCOUT GUI**, the total number of signals to be imported is provided as an option at the top of the input section. Each signal can be either imported from the drive or MATLAB's workspace. The user is required to provide the sampling frequency for each signal. It is important to note that when a file containing several signals, *only* the first read signal will be imported.

After importing the signal(s), the user is offered a range of analysis possibilities. This section briefly presents the theory behind each analysis step and details the corresponding options.

4.1. Stationarity analysis

This analysis examines and optionally renders signals stationary. The stationarity check is based on the reverse arrangement test (Siegel and Castellan 1988), which we will briefly review. For more details about the method, the reader is referred to the ample literature on the subject (Siegel and Castellan 1988; Bendat and Piersol 2011).

The reverse arrangement test first requires the sample record to be specifically prepared. The signal is split into M equal time intervals, where the data in each interval may be considered independent. For each interval, the mean square value is then computed as

$$\overline{x_1^2}, \overline{x_2^2}, \overline{x_3^2}, \dots, \overline{x_M^2}.$$

The reverse arrangement test for stationarity is then performed on the mean square values. The process involves counting the number of times that $\overline{x_i^2} > \overline{x_j^2}$ for $i < j$. Each such inequality

is called a reverse arrangement. The total number of reverse arrangements is denoted by A , where A is defined from the set of observations $\overline{x_1^2}, \overline{x_2^2}, \dots, \overline{x_M^2}$ as follows,

$$h_{ij} = \begin{cases} 1 & \text{if } \overline{x_i^2} > \overline{x_j^2} \\ 0 & \text{if otherwise} \end{cases}$$

Then

$$A = \sum_{i=1}^{M-1} A_i,$$

where

$$A_i = \sum_{j=i+1}^M h_{ij}.$$

A is then compared to the z -score, which is defined as,


$$z = \frac{A - \left\lfloor \frac{M(M-1)}{4} \right\rfloor}{\sqrt{\frac{2M^3 + 3M^2 - 5M}{72}}}.$$

The null hypothesis for this test is that the data points in the signal are independent observations from a stationary random variable. To reject the null hypothesis at 95% confidence level requires a z -score of $z \geq 1.96$ or $z \leq -1.96$.

Stationary analysis requires only one input; The sample size for the reverse arrangement test, which should be a positive number $1 < M < \text{length}(\text{signal})$. The default sample size is 100. The sample size should be chosen such that it is neither too short nor too long with respect to the fundamental period of the signal.

If the signal is not stationary, the options for rendering it stationary become active. Signals can be made stationary using two methods: the MATLAB inbuilt `detrend` function, or the proposed `Polynomial fit` method. MATLAB's `detrend` function simply detects and removes linear trends in the data. The user is referred to MATLAB user manual for details. The polynomial fit method uses sequentially higher polynomial orders up to third order¹ to fit the data, and then automatically chooses the best polynomial order based on a compromise between accuracy (fit error) and complexity (polynomial order). The steps are summarized in Algorithm 1.

In the last step of Algorithm 1, an explicit bias of -0.2 is introduced to favor simplicity. The algorithm is tested on two non-stationary signals of polynomial order 1 and 2 with large superimposed noise. As Figure 2 shows, the resulting polynomial orders are correctly identified. After the detrending process, the reverse arrangement test is again repeated to verify stationarity.

In **SCOUT GUI**, the process can be separately performed on all imported signals, with every signal individually selected from the drop-down menu. Each time a new signal or a new detrending method is selected the user must press the  button to execute the analysis. When a new signal is selected, all settings return to default. The main steps of the detrending process are registered and updated in the 'Signal summary' section. Similarly, the 'Signal statistics' section gets updated with the new statistical values from the detrended signal.

¹Higher order polynomials than the third are not recommended (Bendat and Piersol 2011).

Algorithm 1 Polynomial order identification for stationarity analysis**Input:** non-stationary signal

```

1: Fit three polynomial curves with order 1–3 to the signal
2: if  $E_1 < E_2, E_3$  then                                ▷  $E_i$  is the fit standard error
3:   Select  $\mathcal{O} = 1$                                        ▷  $\mathcal{O}$  is the polynomial order
4: else
5:   Fit a spline curve through  $E_i/E_1$  and  $\mathcal{O}_i$ , for  $i = 1, 2, 3$ 
6:   Compute the gradient of the spline curve  $\mathcal{G}(E)$ 
7:   Infer  $\mathcal{O}_{min} = \min(\mathcal{G}(E))$ 
8:    $\mathcal{O} = \lfloor \mathcal{O}_{min} \rfloor - 0.2$                         ▷  $\lfloor \cdot \rfloor$  denotes the floor rounding function
9: end if
10: return Polynomial order

```

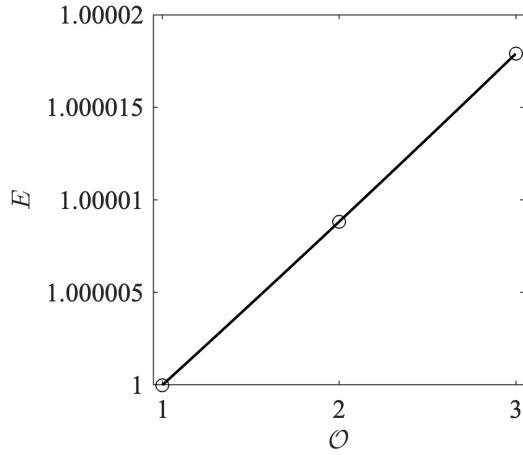
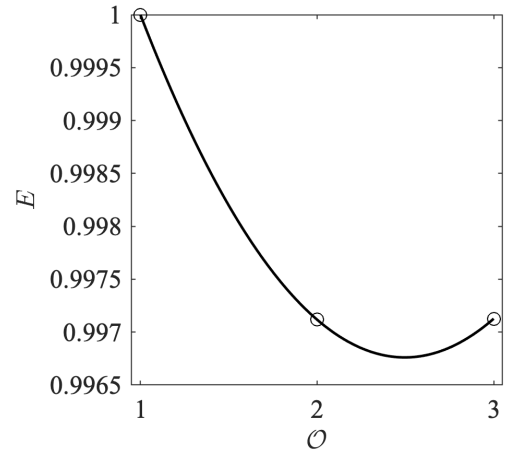
(a) Selected polynomial order: $\mathcal{O} = 1$.(b) Selected polynomial order: $\mathcal{O} = 2$.

Figure 2: The polynomial order selection for detrending based on the error distribution fit. Two scenarios showing the selection outcome for (a) a first order $\mathcal{O} = 1$ and for (b) a second order $\mathcal{O} = 2$ polynomial.

In **SCOUT Script**, the stationarity process is controlled through 3 variables in the configuration file:

- **I.Stationarity.Subsize**: the sample size for the reverse arrangement test.
- **I.Stationarity.MakeStationary**: a binary selection (**Yes/No**) on whether to render the signal stationary.
- **I.Stationarity.StationaryMethod**: the choice of the detrending method.

4.2. Spurious Samples

This section describes the detection method and the handling of spurious samples. Two possible algorithms for detecting spurious samples are offered: the Chauvenet criterion, and

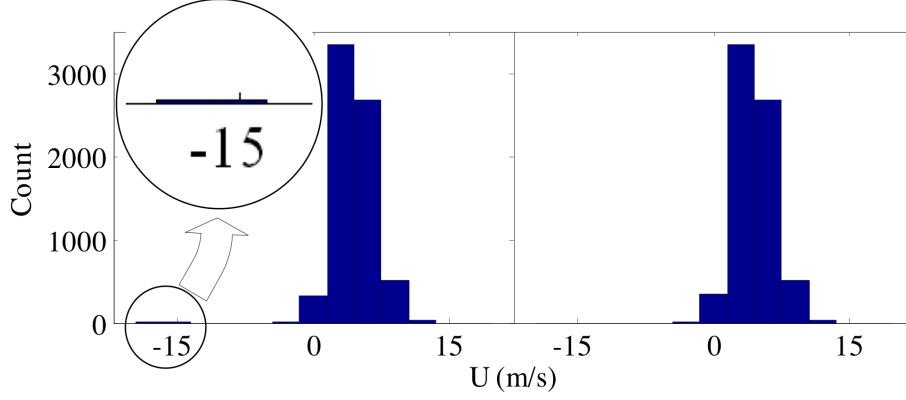


Figure 3: Illustration of the detection and the elimination of spurious points using the histogram method (Semaan and Naughton 2013).

the histogram method. The Chauvenet criterion (Chauvenet 1960; Coleman and Steele 2009) is a widely accepted method for spurious sample detection. It specifies that all points that fall within $1 - 1/(2N)$ probability band around the mean value should be retained, where N is the sample record length. The Chauvenet criterion can be used on most random data, except when the measured probability density function is skewed or multi-modal, which causes a rejection bias, where ‘real’ data get clipped. Under this scenario, the so-called histogram approach should be used instead. The histogram approach (Semaan and Naughton 2013) simply constructs a coarse histogram and detects outliers as samples that are separate from the main histogram body. This approach is illustrated in figure 3, where the spurious point in the original data on the left side has been removed, as shown on the right side.

If spurious samples are detected, the options to deal with them become active. In this case, the user has two options: removal, or replacement. As the name suggests, the removal option simply deletes the spurious samples. Removal is the preferred option when no subsequent spectral analysis is planned since any Fourier transform requires data sampled at equal time intervals. Specifically, this will exclude performing ‘Periodicity’, ‘Spectral analysis’ and ‘Uncertainty analysis’. If any of the above-mentioned analysis is desired, the replacement option should be selected. Here, spurious samples are replaced by their local average values.

In **SCOUT GUI**, the main steps of the spurious sample detection are visualized in red in the time trace and the histogram plots, and are reported and updated in the ‘Signal summary’ and the ‘Signal statistics’ sections.

In **SCOUT Script**, this analysis is controlled with 4 input variables in the configuration file:

- `I.Spurious.DetectionMethod`: the choice of the detection method.
- `I.Spurious.Chauvenet.n`: the extent of the sample retention region away from the mean as a number of standard deviations. The default value is the one computed from the Chauvenet criteria.

- `I.Spurious.Histogram.Bins`: the number of bins for the histogram method.
- `I.Spurious.Postprocessing`: selection between the removal and the replacement options.

4.3. Normality

This section describes the method to identify normally-distributed signals to a pre-selected confidence level. Among many motivations, knowing whether a signal is Gaussian yields significant simplifications in the uncertainty analysis. The method to check for normality is the χ^2 goodness-of-fit test (Cochran 1952), which is used in a wide range of applications. In the following, we shall briefly describe it.

Consider a sample record of N independent observations ² from a random variable x with a probability density function $p(x)$. Let the N observations be grouped into K bins, which together form a frequency histogram. The number of observations falling within the i th bin is denoted by f_i . The number of observations that would be expected to fall within the i th bin for the true normally-distributed probability density function $p(x_0)$ are denoted by F_i . The discrepancy between the observed frequency and the expected frequency within each bin is given by $f_i - F_i$. The measure of the total discrepancy for all bins is the sum of the normalized squares of the discrepancies in each bin,

$$X^2 = \sum_{i=1}^K \frac{(f_i - F_i)^2}{F_i}. \quad (1)$$

The X^2 distribution in equation (1) is similar to that of the χ_n^2 . To compare against χ_n^2 , one needs to know the degrees of freedom n , which for a normal distribution is $n = K - 3$. The hypothesis is accepted if $X^2 \leq \chi_{n,\alpha}^2$, where α is the level of significance. The method is summarized in Algorithm 2.

Algorithm 2 χ^2 goodness-of-fit test

Input: confidence level, number of bins

- 1: Group the sampled observations into K bins
 - 2: Compute the expected frequency for each interval f_i
 - 3: Compute X^2 as in equation (1)
 - 4: **if** $X^2 > \chi_{n,\alpha}^2$ **then**
 - 5: The hypothesis that $p(x) = p(x_0)$ is rejected at the α level of significance
 - 6: **else**
 - 7: The hypothesis is accepted at the α level of significance
 - 8: **end if**
 - 9: **return** Test result
-

The outcome of Algorithm 2 is the test result on whether the distribution is normal. In **SCOUT GUI**, the normality test result is displayed in the ‘Result display’ section for the

²**SCOUT** offers the possibility to compute the integral time scale, and hence to identify the independent samples. However, such capabilities are not integrated in the normality test. It is the user’s responsibility to make sure the observations are independent.

selected confidence level. As a visual guide, a normally-distributed reference probability density function $p(x_0)$ with the same mean and standard deviation as the signal is displayed in red in the histogram plot. As before, the main steps of the normality test are registered and updated in the ‘Signal summary’ section.

In **SCOUT Script**, the normality test is controlled via 2 variables in the configuration file:

- **I.Normality.ConfidenceLevel**: the desired confidence level (to remain consistent and avoid confusion, the confidence level is used instead of the level of significance employed in Algorithm~2).
- **I.Normality.NumberOfBins**: the number of bins. The default recommended bin size is $\Delta x \approx 0.4 s$, where s is the standard deviation.

4.4. Periodicity

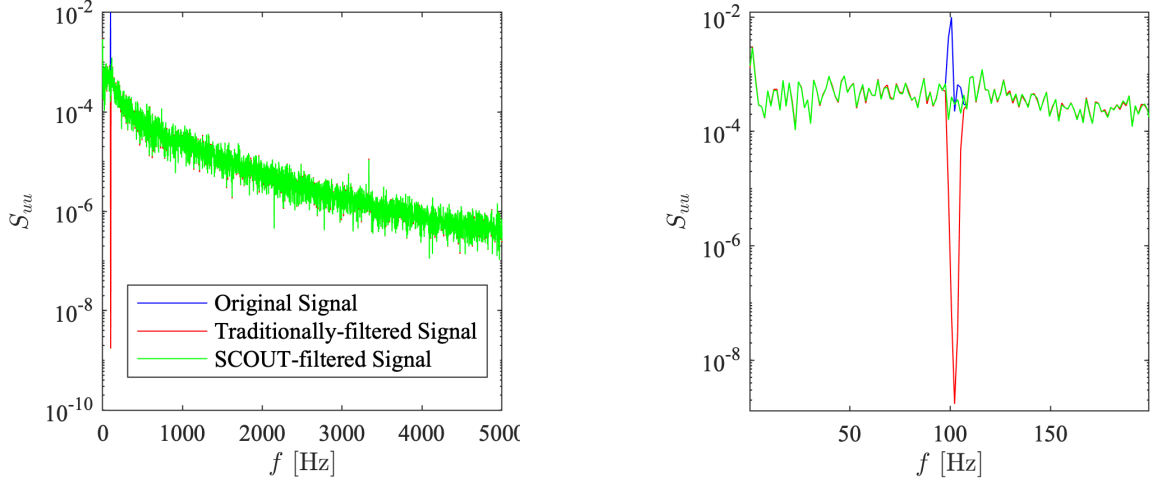
This section describes methods to detect periodic components and to optionally filter them. Removing deterministic (e.g. periodic) components from signals is a necessary prerequisite before uncertainty quantification. The periodicity analysis starts with computing the autospectral density function (Welch 1967), which is a standard function in MATLAB. The magnitude of the highest spectral peak, as well as its frequency, are then detected.

The user is subsequently offered the option to filter the detected (or any other) frequency band using a specially-modified Fourier filter. Unlike typical filters, such as Butterworth or Chebychev, which require a lot of tweaking and tend to over-attenuate the spectrum in the target frequency band yielding a distorted signal in the time domain, the modified Fourier filter is very easy to set and yield a smooth spectrum. The filter simply attenuates the desired frequency range by replacing the Fourier coefficients at those frequencies with a uniformly-distributed *random noise* that have a user-selected standard deviation. The addition of random noise acts as an alternative to windowing (Kemaio 2015), and minimizes ripple effects. The attenuation magnitude ranges between level 0, where the Fourier coefficients are replaced with random noise with low standard deviation, and level 1, where the Fourier coefficients are replaced with random noise whose standard deviation is similar to the detected peak magnitude. A comparison between a traditional and the proposed filtering technique is presented in Figure 4, where the spectra of the original signal (blue), of the MATLAB-filtered signal (red), and of the filtered signal using **SCOUT** (green) are shown. A zoom-in on the filtered frequency range in Figure 4 (b) clearly shows the over-attenuation when using MATLAB’s **bandstop** despite setting the steepness and the attenuation to the very low levels of 0.5 and 10, respectively.

Despite windowing and overlapping, spectral edge-effects are sometimes unavoidable. This issue becomes particularly clear when transforming the signal back to the time domain. **SCOUT** addresses this issue, by offering the user to append points at both ends of the sample record *before* filtering is initiated (Mayhew 1985). These appended points are mirror samples, such that for K appended points on each end, the signal becomes

$$\mathbf{x} = \{x_K, x_{K-1}, \dots, x_1, x_1, x_2, \dots, x_{N-1}, x_N, x_{N-1}, \dots, x_{N-K+1}, x_{N-K}\},$$

where x_1 and x_N are the original and first and final sample, respectively. After time-domain reconstruction, the appended samples are chopped and the sample record is returned to its original length.



(a) Entire frequency range.

(b) Zoom-in on the filtered frequency range.

Figure 4: Comparison between traditional filtering using MATLAB's `bandstop` and **SCOUT**'s filtering method. The spectra of the original signal (blue), of the MATLAB-filtered signal (red), and of the filtered signal using **SCOUT** (green) are shown. Also shown (b), is a zoom-in on the filtered frequency range.

In **SCOUT GUI**, the detected peak frequency is registered in the 'Signal summary' section. When applicable, the 'Signal statistics' section gets also updated with the new statistics of the filtered signal.

In **SCOUT Script** the spectrum computation and the filtering process is controlled by the following variables:

- `I.Periodicity.Window`: the window type used for the power spectrum.
- `I.Periodicity.NumberOfBlocks`: the number of blocks to divide the sample record.
- `I.Periodicity.Overlap`: the percent amount of overlap between the blocks.
- `I.Periodicity.DiscreteFFT`: the number of discrete FFT points.
- `I.Periodicity.RemovePeriodic`: a binary selection Yes/No on whether to band-stop filter the signal.
- `I.Periodicity.MinLimit`: the lower frequency of the stopband frequency range.
- `I.Periodicity.MaxLimit`: the upper frequency of the stopband frequency range.
- `I.Periodicity.Attenuation`: the filter attenuation level.
- `I.Periodicity.Points`: the number of appended points. The default number is 0.03 N .

4.5. Spectral Analysis

In this section, we detail the spectral analysis that consists of computing the autospectral density function and the autocorrelation coefficient, and estimating the integral time scale.

The integral time scale is relevant for physical insights, and for identifying the number of independent samples, which are necessary for the uncertainty analysis. The integral time scale is defined as

$$I = \int_0^{\tau_{\max}} \rho(\tau) d\tau, \quad (2)$$

where τ_{\max} is the maximum time lag, and ρ is the autocorrelation coefficient,

$$\rho = \frac{R_{xx}}{s^2}, \quad (3)$$

where R_{xx} is the autocorrelation, and s is the standard deviation. In **SCOUT**, R_{xx} is computed with an indirect approach from the inverse Fourier transform of the autospectral density function (Bendat and Piersol 2011). This approach is computationally faster and makes use of block averaging, delivering a smoother autocorrelation distribution.

Executing the analysis in **SCOUT GUI** computes and visualizes the auto-spectral density and the autocorrelation coefficient function. The integral time scale value is displayed on the right of the autocorrelation coefficient plot alongside the maximum time lag τ_{\max} . The default τ_{\max} is set to the nearest point around the third zero-crossing of ρ . The integrated area is colored turquoise. If the third zero-crossing is not satisfactory, the user can change the integration range manually by shifting the slider next to the autocorrelation coefficient plot. The final computed integral time scale is registered in the ‘Signal summary’ section.

In **SCOUT Script**, the options for computing the autospectral density function are the same as those listed for the ‘Periodicity’ analysis’ in section 4.4 (`I.Spectral.Window`, `I.Spectral.NumberOfBlocks`, `I.Spectral.Overlap`, `I.Spectral.DiscreteFFT`). Two additional options are required for computing the autocorrelation coefficient and the integral time scale:

- `I.Spectral.TimeShift`: the total lag number L ($L = \tau_{\max}/\Delta t$).
- `I.Spectral.AutocorrNumOfBlocks`: the total number of blocks for computing the autocorrelation.

5. Uncertainty Analysis

This section describes the uncertainty analysis of individual and of propagated signals. The analysis yields uncertainty quantification of the mean result and of higher-order central moments up to the fourth. Due to its ease-of-use and direct interpretability, the propagation of uncertainty is performed using the first-order Taylor series method (Tukey 1957; Helton and Davis 2003). The uncertainty is propagated through a data reduction equation (DRE) of the form

$$r = r(\text{Signal1}, \text{Signal2}, \dots, \text{Signal5}),$$

where r is the result, and `Signal1`, `Signal2`, `Signal5` are the uncertain depend variables. To keep the analysis simple and approachable, the uncertainty propagation analysis is limited to the most practical and recommended approaches:

1. The random uncertainty is computed directly on the result r , as recommended (Bendat and Piersol 2011). In other words, individual random uncertainties are not propagated

through the DRE, thus bypassing the need to estimate possible correlations among them. This, however, requires that all dependent signals are sampled at (or decimated to) the same sample rate and record length.

2. The systematic uncertainty is only propagated through one accepted functional form of the DRE,

$$r = C(\text{Signal1})^{n1}(\text{Signal2})^{n2}(\text{Signal3})^{n3},$$

where C and the n 's are user-defined constants. For example, the drag coefficient equation can be expressed as $C_D = 2F_D\rho^{-1}V^{-2}A^{-1} = (2\rho^{-1}A^{-1})F_DV^{-2}$, where F_D is the measured force from a balance, ρ is the measured flow density, V is the free stream velocity, and A is the projected area of the model. Assuming a constant flow density and a constant model surface area, this yields

- The result $r = C_D$, which is the variable into which we want to propagate the uncertainty
- $C = 2\rho^{-1}A^{-1}$, which is constant
- $\text{Signal1} = F_D$, $n1 = 1$
- $\text{Signal2} = V$, $n2 = -2$

It is worth to note that when the uncertainty of an individual (non-propagated) signal is desired (e.g. only that of Signal1), the user can simply set the DRE constants to unity, i.e. $C = 1$ and $n1 = 1$.

3. A large sample record is assumed, which yields a coverage factor $t = 2$, i.e., an expanded uncertainty $U_r = 2u_r$, where u_r is the combined uncertainty of the result.
4. All reported uncertainties are for 95% confidence level, which is typical for engineering applications.

In **SCOUT**, the random uncertainty of higher-order central moments can be computed using two approaches. The first one employs simplified equations for the second and fourth central moments,

$$u_{x^2} = \sqrt{\frac{2}{N_{eff}}}, \quad (4a)$$

$$u_{x^4} = \sqrt{\frac{11}{N_{eff}}}, \quad (4b)$$

which assumes a normally-distributed signal. N_{eff} refers to the effective *statistically independent* number of samples, which is estimated as

$$N_{eff} = \begin{cases} \frac{T}{2I} = \frac{N\Delta t}{2I} & \text{if } \Delta t < 2I \\ N & \text{if } \Delta t \geq 2I \end{cases}$$

where T is the total sampling time. For a Gaussian process, the odd moments are identically zero. Alternatively, the uncertainties of all central moments can be estimated directly with

$$u_{x^m} = \sqrt{\frac{1}{N_{eff}} \frac{\langle x^{2m} \rangle - \langle x^m \rangle^2}{\langle x^m \rangle^2}}, \quad (5)$$

where $\langle \rangle$ is the time-averaging operator. Direct equation (5) is only recommended when the signal(s) is (are) sampled for a sufficiently long time.

Executing the uncertainty analysis in **SCOUT GUI** yields two plots and various outputs:

- The time trace of the result r in the center top.
- The autocorrelation coefficient distribution of the result in the center bottom. Similar to the ‘Spectral analysis’ tab, the user can adjust the integration range using the slider next to the figure.
- The data reduction equation in a readable simplified format in the top right corner.
- The random uncertainty of the individual signals as well as of the result.
- The systematic uncertainty of the result.
- The combined and expanded uncertainty of the result mean \bar{r} , which is also used to formally present the mean with its uncertainty band $\bar{r} = r_{mean} \pm U_{\bar{r}}$.
- The integral time scale for the result.
- Uncertainties of the second, third and fourth central moments using either method.

The entire uncertainty analysis in **SCOUT Script** is set with one type 2 configuration file. In the following, we list the setting options:

- **U.Uncertainty.Coefficient**: the constant C in the data reduction equation.
- **U.Uncertainty.InputSystematic**: a binary selection (Yes/No) on whether to include the systematic uncertainties.
- **U.Uncertainty.TimeShift**: the total lag number L .
- **U.Uncertainty.Exponent1** (and other similar variables): the corresponding signal exponent in the data reduction equation.
- **U.Uncertainty.Systematic1** (and other similar variables): the relative systematic uncertainty of the corresponding signal.
- **U.Uncertainty.AutocorrNumOfBlocks**: the total number of blocks for the autospectral density function to compute the autocorrelation of the result.
- **U.SecondCentral.Normal** (and other similar variables): binary selection (Yes/No) on whether to compute the second order central moment considering normal distribution.
- **U.SecondCentral.Formula** (and other similar variables): binary selection (Yes/No) on whether to compute the second order central moment using a direct equation.

6. Conclusions

Signal processing and uncertainty quantification and propagation are fundamental fields for the engineering sciences. However, the wealth of information and the complexity of methods are hindering broad adaptation. Thus, a compact, interactive, and easy-to-use toolbox provides an attractive alternative to existing expansive packages, which typically require expert knowledge. The price to pay is the limited offered options for signal processing and for uncertainty analysis, which can be easily remedied through own-code integration. With its two flavors, **SCOUT** offers unique capabilities for both interactive and integrated analysis. We believe **SCOUT** could well be a suitable first choice for data analysis and uncertainty quantification.

SCOUT builds on years of experience and best practices in processing experimental data. It covers a range of analyses typically encountered when post-processing measured signals. These include stationarity analysis, spurious samples detection, normality check, periodicity check, filtering, spectral analysis, and uncertainty analysis of individual and of multiple propagated signals through a data reduction equation. Numerical checks show the consistency and validity of the results.

Future developments on **SCOUT** include more spectral analysis options, such as wavelet transform, and expanding the uncertainty propagation analysis to include more general functional forms and direct computation of the gradient for the Taylor series expansion method.

References

- Bendat JS, Piersol AG (2011). *Random Data: Analysis and Measurement Procedures*. 4th edition. John Wiley & Sons. ISBN 978-0-470-24877-5.
- Chauvenet W (1960). *A Manual of Spherical and Practical Astronomy*, volume II. 5th edition. Reprinted in Dover Publication, New York.
- Cochran WG (1952). “The χ^2 Test of Goodness of Fit.” *The Annals of Mathematical Statistics*, **23**(3), 315–345. ISSN 00034851.
- Coleman HW, Steele GW (2009). *Experimentation, Validation, and Uncertainty Analysis for Engineers*. 3rd edition. John Wiley & Sons. ISBN 978-0-470-48568-2.
- Helton JC, Davis FJ (2003). “Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems.” *Reliability Engineering & System Safety*, **81**(1), 23–69. doi:10.1016/S0951-8320(03)00058-9.
- Jones E, Oliphant T, Peterson P, et al. (2001–2019). **SciPy**: Open source scientific tools for Python. URL <http://www.scipy.org/>.
- Kemao Q (2015). “Applications of Windowed Fourier Fringe Analysis in Optical Measurement: a Review.” *Optics and Lasers in Engineering*, **66**, 67–73. doi:10.1016/j.optlaseng.2014.08.012.
- Marelli S, Sudret B (2014). “**UQLab**: A Framework for Uncertainty Quantification in MATLAB.” In *Proceedings in the 2nd International Conference on Vulnerability, Risk Analysis and Management (ICVRAM2014)*, Liverpool, United Kingdom, pp. 2554–2563. doi:10.1061/9780784413609.257.

- Mathworks (2019). *MATLAB Signal Processing Toolbox*. URL <http://www.mathworks.com/products/signal.html>.
- Mayhew MA (1985). “Curie Isotherm Surfaces Inferred from High-Altitude Magnetic Anomaly Data.” *Journal of Geophysical Research: Solid Earth*, **90**(B3), 2647–2654. doi: [10.1029/JB090iB03p02647](https://doi.org/10.1029/JB090iB03p02647).
- Semaan R, Naughton JW (2013). “Three-Component Laser-Doppler-Anemometry Measurements in Turbulent Swirling Jets.” *AIAA Journal*, **51**(9), 2098–2113. doi: [10.2514/1.J051783](https://doi.org/10.2514/1.J051783).
- Siegel S, Castellan NJ (1988). *Nonparametric Statistics for the Behavioral Sciences*. 2nd edition. McGraw-Hill. ISBN 978-0070573574.
- Spiess AN (2018). *propagate: Propagation of Uncertainty in R*. URL <https://CRAN.R-project.org/package=propagate>.
- Tukey JW (1957). “The Propagation of Errors, Fluctuations, and Tolerances Basic Generalized Formulas.” *Technical Report 10*, Statistical Techniques Research Group, Princeton University, N. J.
- Welch P (1967). “The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms.” *IEEE Transactions on Audio and Electroacoustics*, **15**(2), 70–73.

Affiliation:

Richard Semaan
 Institute of Fluid Mechanics
 Technische Universität Braunschweig
 Hermann-Blenk-Str. 37
 38108 Braunschweig, Germany
 E-mail: r.semaan@tu-bs.de
 URL: <https://www.richardsemaan.com/home/software/>