The INSTANT Neuronal Spike-Train ANalysis Toolbox

Martin Nilsson, RISE Research Institutes of Sweden (E-mail: `instant@drnil.com`)

December 16, 2020

Overview

The INSTANT Neuronal Spike-Train ANalysis Toolbox is a program that inputs a spike train and estimates the corresponding neuronal parameters. This program also estimates errors and performs statistical tests of stationarity and independence. The program is implemented in MATLAB (MathWorks Inc., Natick, MA, USA).

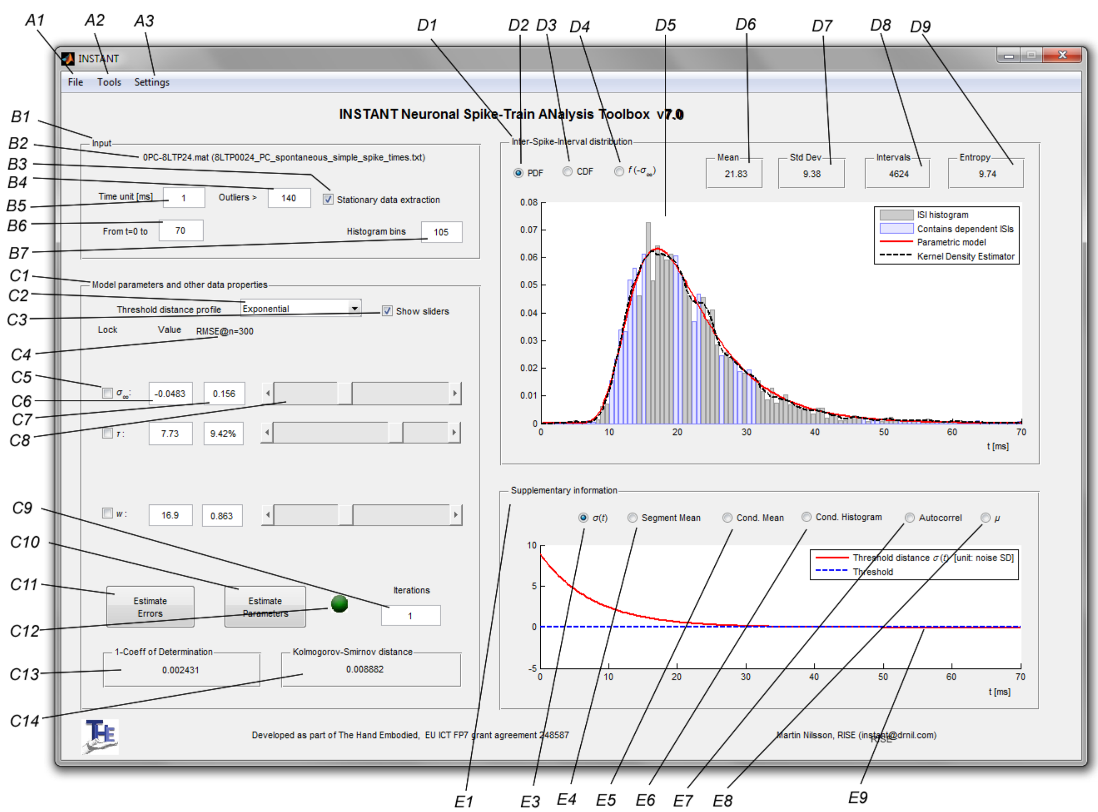

Figure 1: The INSTANT user interface. Labels within parentheses in the text refer to this figure.

Workflow

The basic workflow proceeds as follows:

1. Load a raw data file of spike times or spike intervals using the **Open command in the File menu** (A1). The raw ISI data should be in text format, one item per row. INSTANT will autodetect whether the data are given as spike times or interval lengths.

2. Declare the **time unit** (B5) of your data in the **Input pane** (B1). The default is 1 ms. The other parameters are computed automatically and will usually be accurate, but they can be adjusted if necessary.

3. Choose a **Threshold distance (membrane potential) profile** (C2) in the **Model Parameters pane** (C1). Normally, `Exponential` is a good initial choice.

4. Adjust the **sliders** (C8) or set the **editable parameter fields** (C7) manually in such a way that the model (red curve) roughly fits the histogram in the **main diagram** (D5).

5. Click the **Estimate Parameters button** (C10). This action starts the search for the parameters. Usually, the program finds a solution almost instantly (thus the program's name). The estimated parameter values will be shown in the **parameter fields** (C6). If the initial values are far from optimal, the execution could take a long time. The **indicator light** (C12) will switch to a red square during a search.

6. Save the results using the **Save command in the File menu** (A1). This command saves only the parameters for the data in a parameter (`.mat`) file and will never modify the original raw data file. The next time that the user wants to use the data, it is sufficient to load the parameter file. This will load the raw data automatically.

7. Experiment with the various features of INSTANT, e.g.:

   - To perform statistical tests, click the **radio buttons** E4–E7 in the **Supplementary Information pane** (E1).
   - To compute an error estimate using bootstrap, set the field **Multistarts** (C9) to the number of samples and click the **Estimate Errors button** (C11). This process is a time-consuming operation, but the computation can be accelerated using a multicore computer (see below).
   - To perform a global search, set the field **Multistarts** (C9) to the number of random initial points and click the **Estimate Parameters button** (C10). This operation is slow but is also amenable to acceleration by parallel processing.

Installation

If the version of INSTANT is packed into a Base64-encoded archive `instant.txt`, the file can be decoded under Microsoft Windows using the command

```
certutil -decode instant.txt instant.zip
```

The integrity of the archive can be checked using the command

```
certutil -hashfile instant.txt SHA512
```

The result should be

```
SHA512 hash of instant.txt:
55bb50e7365d85af6ec0c593b3c8e327dff72b7dc0d054c7241c431dc4b81aa5
66ad5e67218c1daaf9bb578c87e487330066f835b7ac9cc0c8876a32305542d4
CertUtil: -hashfile command completed successfully.
```

The program can be executed directly in MATLAB. It has been tested in MATLAB release 2013a under Windows 7 and Windows 10.

INSTANT inherits some quirks from MATLAB. The most annoying of these is perhaps the difficulty to interrupt ongoing computations. Therefore, before starting a long computation, test a smaller scale version. A process can always be stopped with the Windows task manager, but those results that have not been saved will then be lost.

INSTANT writes a log file named `instant.log`, but this name can be changed by the command **Set logfile...** in the Settings menu (A2). It contains only INSTANT-specific output and *appends* new information.

Menu commands

In the `File` roll-down menu (A1):

New clears the parameter settings after the previously loaded data.

Open loads either a text file of ISI data or a binary parameter (`.mat`) file, which contains the raw ISI data file name and other relevant parameters, saved by INSTANT on a previous occasion.

Print screen saves the screen content as a PNG-format non-lossy raster image file.

Print... prints the screen using MATLAB's standard printing routine.

Save parameters saves the current parameters in a parameter file.

Save parameters as... saves the current parameters in a parameter file and prompts for a file name.

Save diagram saves the main diagram (D5) as a table of data.

The `Tools` roll-down menu (A2):

Batch process is a submenu for the batch commands, i.e., they will apply some procedure on all of the parameter files that are specified in a batch file. Execution continues until all of the files have been processed. The batch file should be a text file, giving one parameter file name per line. Blank lines are ignored. Comments can be included if preceded by a hash symbol (#), semicolon (;), or percent (%). The following batch commands are available:

- **Estimate parameters** performs a global search to estimate the parameters for every spike train. The results are printed in the logfile. The number of random starting points should be given beforehand in the **Multistarts field** (C9).
- **Estimate parameters and save back** performs the same operation as **Estimate parameters** but it also saves the results in the corresponding parameter file.
- **Estimate errors** estimates the root-mean-square-error (RMSE) for the estimated parameters using "0.632" bootstrapping. The results are printed in the log file. The number of resamplings should be given beforehand in the **Multistarts field** (C9).

Estimate errors and save back performs the same operation as **Estimate errors**, but also saves the results in the corresponding parameter file.

Generate documents applies the **Print screen** command to every parameter file to generate a set of screen shots.

Save diagram saves the **main diagram** (D5) for every parameter file.

Print parameters prints estimated parameters on the log file.

CRLB and Fisher information computes the Fisher information matrix and the Cramér-Rao Lower Bound (CRLB) for the estimated parameters. This information is output into the log file only.

Synthesize generates a synthetic set of ISI intervals that has the same parameters as the current raw data. This set can be used for testing parameter estimation. Analyzing these synthetic data should produce the same parameters that were used for generating the data.

The `Settings` roll-down menu (A3):

The Settings roll-down menu controls the choice of the ordinary differential equation (ODE) integration algorithm, the name of the log file, and parallel execution.

ODE integrator: Three different ODE integrators are available: the fixed-step `odeRK4` and the variable-step `ode23` and `ode45`. The fixed-step integrator `odeRK4` is the most robust but also the slowest; `ode45` is the fastest but most fragile, and `ode23` is in-between. For local search in the neighborhood of an estimate, it is better to use `ode45`, whereas for global search with random starting points, it is better to use `odeRK4` or `ode23`. The choice `auto` (default) automatically selects the integrator depending on the task at hand.

Log file: The name and location of the log file can be set here.

Single core or multicore: This setting determines whether to use a single core or multiple cores (i.e., parallel processing) if available. A four-core computer can sometimes speed up execution by a factor of approximately three. This speed-up can be helpful when performing slow operations such as global search and error estimation. A disadvantage of the parallel mode is that the progress indicator, which is shown in the **Multistart field** (C9), is only updated on every 300th iteration.

Input pane (B1)

The **top line** (B2) shows the name of the input file. Input data can be given as spike times or as spike intervals. Specifying which is unnecessary because this choice is determined automatically. The default unit is 1 millisecond. For different units, a scale factor can be given in the box **Time unit** (B5). Data less than zero or larger than the value **Outliers** (B4) will be dropped. Intervals below "**to**" (B6) values will be processed, and the others will be placed in an overflow bin. The integer parameter **Histogram bins** (B7) determines the number and size of the histogram bins and kernel density estimator bandwidths. Ticking the box **Stationary data extraction** (B3) means that the program will extract only stationary sections of data and discard other data (Johnson, 1996).

Model Parameters and other data properties pane (C1)

The threshold distance (membrane potential) profile can be chosen in the **Threshold distance profile roll-down menu** (C2). Depending on this choice, up to five parameters become visible. These can be set either by the sliders to the right (C8) or by the editable parameter boxes to the left (C6). The possible ranges for the parameters are the following:

1) **Refractory period threshold limit ($\sigma_{\max}$):** $0 \leq \sigma_{\max} \leq 9$. For the double exponential profile, this parameter instead means the amplitude of the second exponential component.

2) **Limiting threshold distance (σ_∞):** $-4 \leq \sigma_\infty \leq 6$.

3) **Mean ion channel open duration** ($\tau = 1/\mu$): $0.01 \leq \tau \leq 100$.

4) **Time shift** ($w = \tau \ln \sigma_0$, where σ_0 is the projected hyperpolarisation depth): $-5 \leq w \leq 50$.

5) **Re/hyperpolarization phase (fast K-channel) time constant (θ):** $0.1 \leq \theta \leq 10$.

The different threshold distance profiles are the following (see Fig. 2):

a) **Constant:**

$$\sigma_{\mathrm{Const}}(t) = \sigma_\infty$$

b) **Single Exponential:**

$$\sigma_{\mathrm{Exp}}(t) = \sigma_\infty + (\sigma_0 - \sigma_\infty)\, e^{-t/\tau}$$

c) **Truncated Exponential:**

$$\sigma_{\mathrm{TruncExp}}(t) = \min\left[\sigma_{\max}, \sigma_{\mathrm{Exp}}(t)\right]$$

d) **Smooth Truncated Exponential:**

$$\sigma_{\mathrm{SmoothTruncExp}}(t) = -\ln\left[e^{-\sigma_{\max}} + e^{-\sigma_{\mathrm{Exp}}(t)}\right]$$

e) **Sigmoid:**

$$\sigma_{\mathrm{Sigmoid}}(t) = \sigma_\infty + \frac{\sigma_{\max} - \sigma_\infty}{\exp(t/\tau)/(\sigma_0 - \sigma_\infty) + 1}$$

f) **Double Exponential:**

$$\sigma_{\mathrm{DoubleExp}}(t) = \sigma_{\mathrm{Exp}}(t) - \sigma_{\max}\, e^{-t/\theta}$$

By clicking the **Estimate Parameters button** (C10), the model parameters are estimated by either fitting the model PDF to the kernel density estimator (KDE) of the data or fitting the model CDF to the sorted ISI data. The choice between the PDF, where one minus the coefficient of determination ($1 - R^2$) is minimized, or the CDF, where the Kolmogorov-Smirnov distance is minimized, is selected with the radio buttons (D2-D3) in the **Inter-Spike-Interval Distribution pane** (D1). During optimization, the green round busy indicator (C12) turns into a red square. A yellow triangle indicates that an error occurred. INSTANT employs a global optimizer using several simultaneous optimization processes if a number larger than one is entered into the **Multistarts field** (C9). This option is available only for PDF fitting. For optimal results, the initial values are generated in a Sobol sequence.

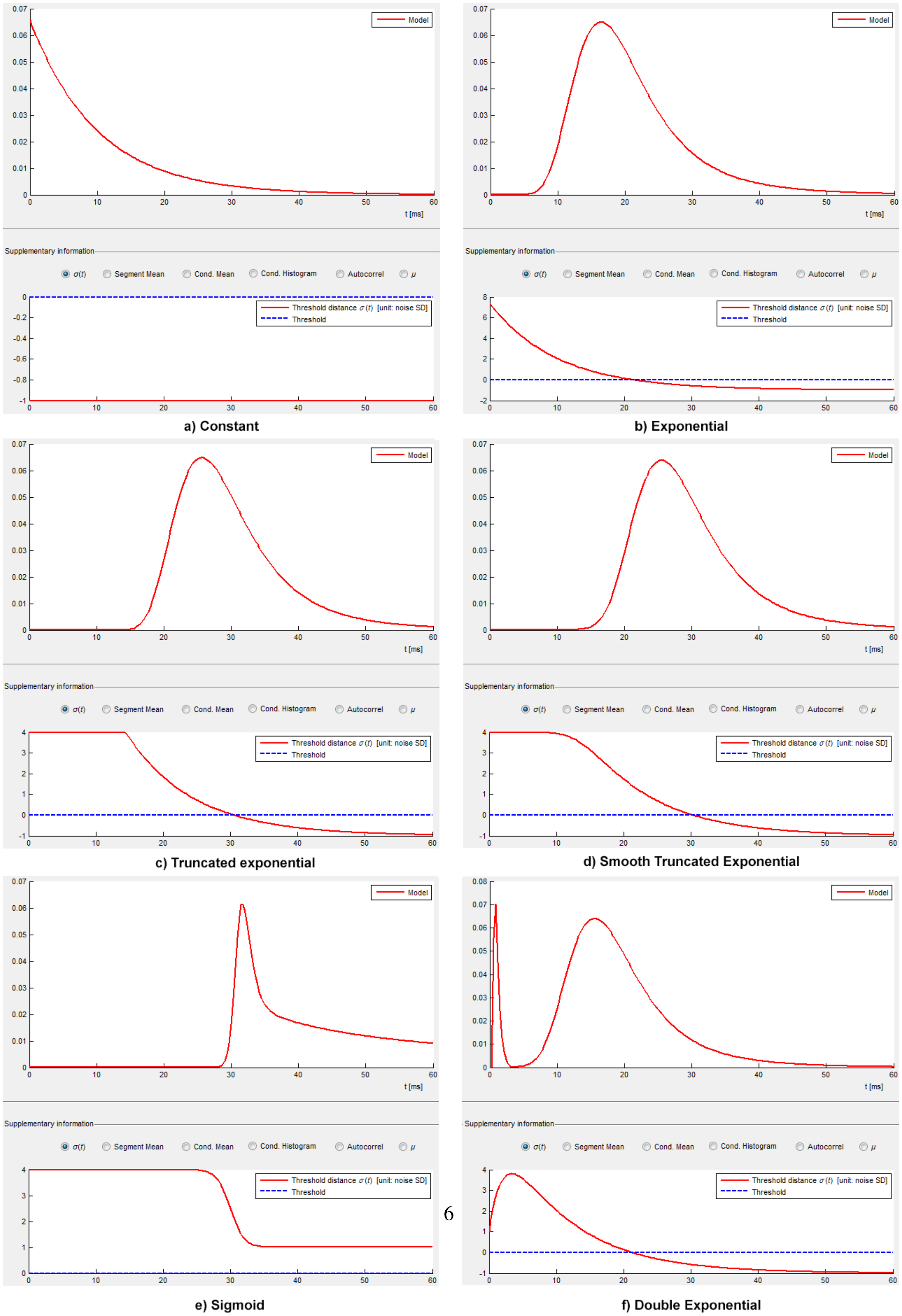

Figure 2: Threshold distance profiles and their corresponding ISI distributions.

The **Estimate Errors button** (C11) randomizes the parameters and uses "0.632" statistical bootstrapping on the ISI data to estimate the root-mean-square-error (RMSE) of the accuracy of the estimated parameters. In this case, the number in the multistarts field describes the number of resamplings. The results are displayed in the RMSE fields (C7) and appended to the log file. The RMSE for the τ parameter is given as a percentage because it is logarithmically represented internally. The **RMSE header** (C4) will show the number of resamplings at which the current RMSE was computed.

The **Lock tick boxes** (C5) tells the estimation not to change the values of the corresponding parameters. Here, it is important to verify that MATLAB has really noted the change. When the **Show sliders tick box** (C3) is clear, the statistical integrity measures shown recently in the **Supplementary information window** (E9) will be shown instead of the sliders. The editable parameter field of the μ parameter is always ignored.

The fit of the model to the data is displayed as **one minus the coefficient of determination** $(1-R^2)$ (C13) and the **Kolmogorov-Smirnov distance** (C14). For both indicators, the value of zero indicates a perfect match.

Inter-Spike-Interval distribution pane (D1)

The three radio buttons, **PDF** (D2), **CDF** (D3), and $f(-\sigma_\infty)$ **transfer** (D4) select between the PDF, the CDF, or the output frequency in hertz as a function of the threshold distance, respectively, to be shown in the **main diagram** (D5).

The units on the x-axis are time in milliseconds for the PDF, time in milliseconds for the CDF, and distance between the limiting potential and threshold in units of noise standard deviations. The units on the y-axis are probability intensity (i.e., relative statistical frequency), probability, and spikes per second (Hz).

There are four information fields (D6-D9) that show the statistics of the current inter-spike-interval (ISI) distribution: the **mean** (D6), **standard deviation** (D7), **number of intervals** (D8), and **entropy** (D9) in bits per spike interval. The unit for mean and standard deviation is determined by the time unit (B5).

The **main diagram** (D5) shows the model as a solid red curve and the ISI data as the kernel density estimator (KDE). The ISI distribution is additionally shown as a traditional histogram, but the KDE is a more accurate representation that does not exclude as much information. The blue stripes indicate that there are ISI outliers with respect to the conditional mean independence test. Such data are not included in the parameter search, but they are always included when computing the **final fitness values** (C13-C14).

Supplementary Information pane (E1)

There are six radio buttons (E3-E8) for selecting auxiliary graphic displays, mostly for established statistical tests for stationarity and independence (Johnson, 1996). The default shows the model's **Threshold distance profile** $\sigma(t)$ (E3) in units of noise standard deviations. The **Segment Mean** (E4) divides the input data into the minimum of 100 and $\sqrt{n}$ segments, where n is the number of spike intervals, and computes the mean of each interval. This computation is a statistical test for investigating the stationarity of the spike-train. The mean (the red line) should not protrude outside of the blue dashed lines. The **Conditional Mean** (E5) computes the mean which is conditioned on the length of the preceding interval, and is a test of the first-order independence of the intervals. Again, the red line should not protrude significantly beyond the dashed blue lines. The **Conditional Histogram** (E6) displays ISI histograms that are conditioned on the preceding interval, and is a test of second-order independence. The alternative **Autocorrelation** (E7) tests the autocorrelation of the spike train, which should give a narrow spike at zero and be nearly zero elsewhere. The μ **option** (E8) shows an alternate estimate of the $\mu = 1/\tau$ parameter from the tail of the ISI distribution.

References

Johnson, D. H. (1996). Point process models of single-neuron discharges. *J. Comput. Neurosci.*, 3:275–299.