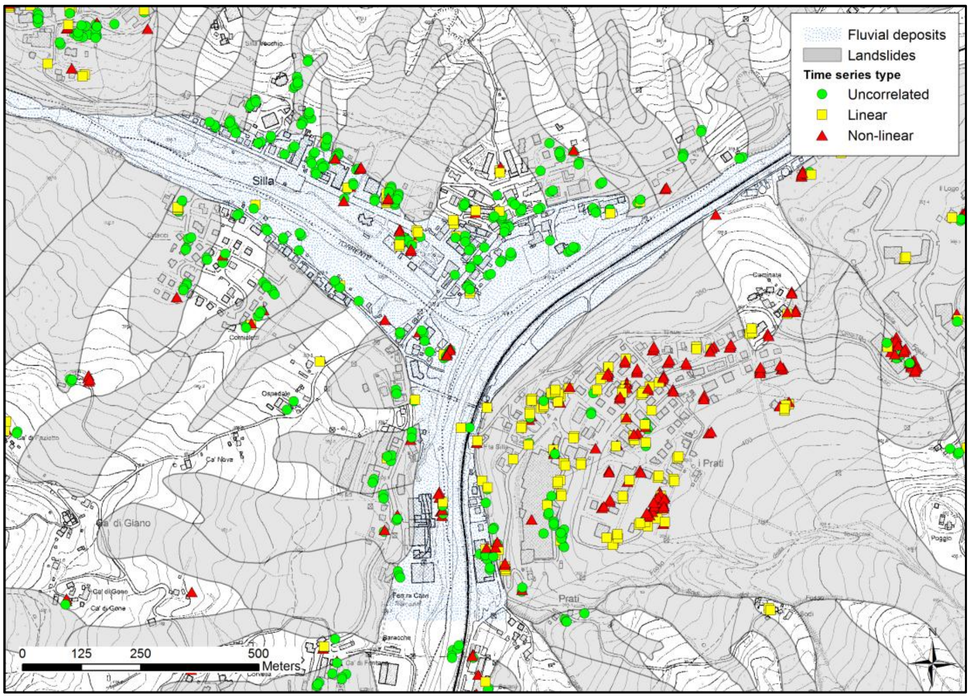
Dipartimento di Scienze Biologiche, Geologiche e Ambientali
Università di Bologna - Italy

PSTIME

A free program for the automatic classification of Permanent Scatterers time series

By Matteo Berti (matteo.berti@unibo.it)



Version 1.1- April 2013

<u>Working Group</u>

Matteo Berti, University of Bologna (Italy)
Silvia Franceschini, University of Bologna (Italy)
Alessandro Corsini, University of Modena and Reggio Emilia (Italy)
Jean Pascal Iannacone, University of Modena and Reggio Emilia (Italy)

Table of contents

To download PSTime and the electronic version of this user's manual,

please go to www. bigea.unibo.it/it/ricerca/pstime

If you have questions about PSTime or its use, please contact:

Matteo Berti: matteo.berti@unibo.it

1 Introduction

PSTime is a free application that provides an automatic classification of Persistent Scatters Interferometry (PS) time series on the basis on a conditional sequence of statistical tests. Time series are classified into distinctive predefined target trends (such as uncorrelated, linear, quadratic, bilinear and discontinuous) related to different styles of ground deformation.

Beside the classification tests, PSTime computes several descriptive parameters of the time series useful to evaluate the data scatter around the mean trend and the presence of periodic fluctuations in the data. All the relevant information obtained from the statistical analyses can be imported into a geo-referenced GIS software to support the interpretation of PS data.

This user manual describes how to use PSTime and the key features of the program. Please refer to the original paper from Berti et al. (2013) and to the cited references for details on the theory behind PSTime as well as for the discussion of important issues such as the calibration of statistical thresholds, the interpretation of the results, and the limitations of the method.

Reference paper

Berti M., Corsini A., Franceschini S., Iannacone J.P. (2013). Automated classification of Persistent Scatterers Interferometry time-series. *Natural Hazards and Earth System Sciences (NHESS)*, in press.

Disclaimer

PSTime was developed for scientific purpose and it is free for non-commercial users. You are the sole responsible for your use of the program. The author and the University of Bologna are not responsible for any damage however caused which results, directly or indirectly, from your use of PSTime.

Acknowledgments

This work was supported by the Civil Protection Agency of the Emilia-Romagna Region under the framework agreement “Special activities on support to the forecast and emergency planning of Civil Protection with respect to hydrogeological risk” (ASPER-RER, 2011-2015) and by the Italian Ministry of University and Scientific Research (PRIN 2010-2011, Ref. 2010E89BPY_005, “Time-space prediction of high impact landslides under the changing precipitation regimes”).

2 Getting started

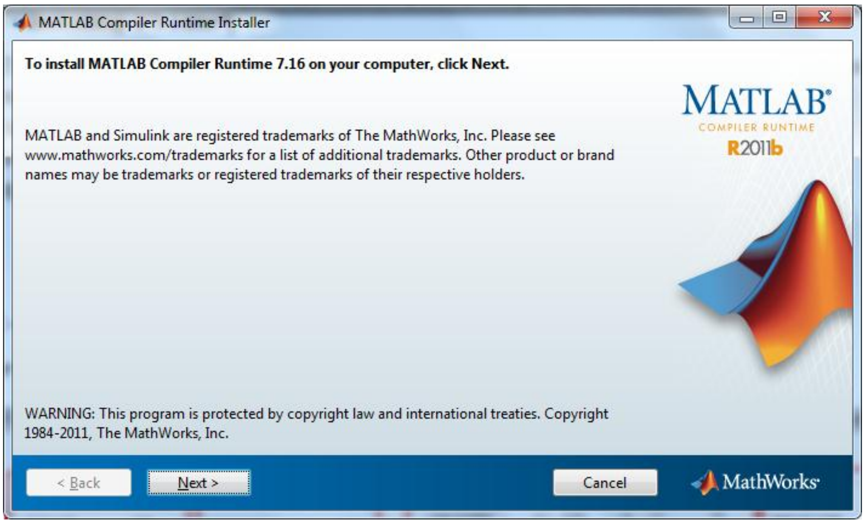
2.1 Installing PSTime

PSTime was developed in Matlab and compiled as a standalone application. The application can run on any windows system without having Matlab or its auxiliary toolbox.

If you do not have Matlab installed on your computer (release R2011b or higher), download the "Installation package with Matlab libraries" (423 MB) and save it on your computer in a separate folder. The installation package contains:

- the PSTime executable
- two sample data files
- the shared Matlab libraries

Run the self-extracting executable and follow the instructions displayed by the setup program. The program will extract the PSTime executable and the sample data files on the installation folder, and it will run the Matlab Compiler Runtime (MCR)



The MATLAB Compiler Runtime (MCR) is a standalone set of shared libraries that enables the execution of compiled MATLAB applications or components on computers that do not have MATLAB installed. Once the installation is completed, you can run the PSTime executable.

If you have Matlab installed on your computer (release R2011b or higher with the Statistics Toolbox), just download the zip file (433 kB) that contains the PSTime executable and the two sample data files. Unzip the file and run the application.

2.2 The Graphical User Interface

The Graphical User Interface (GUI) of PSTime is shown in Fig. 1.

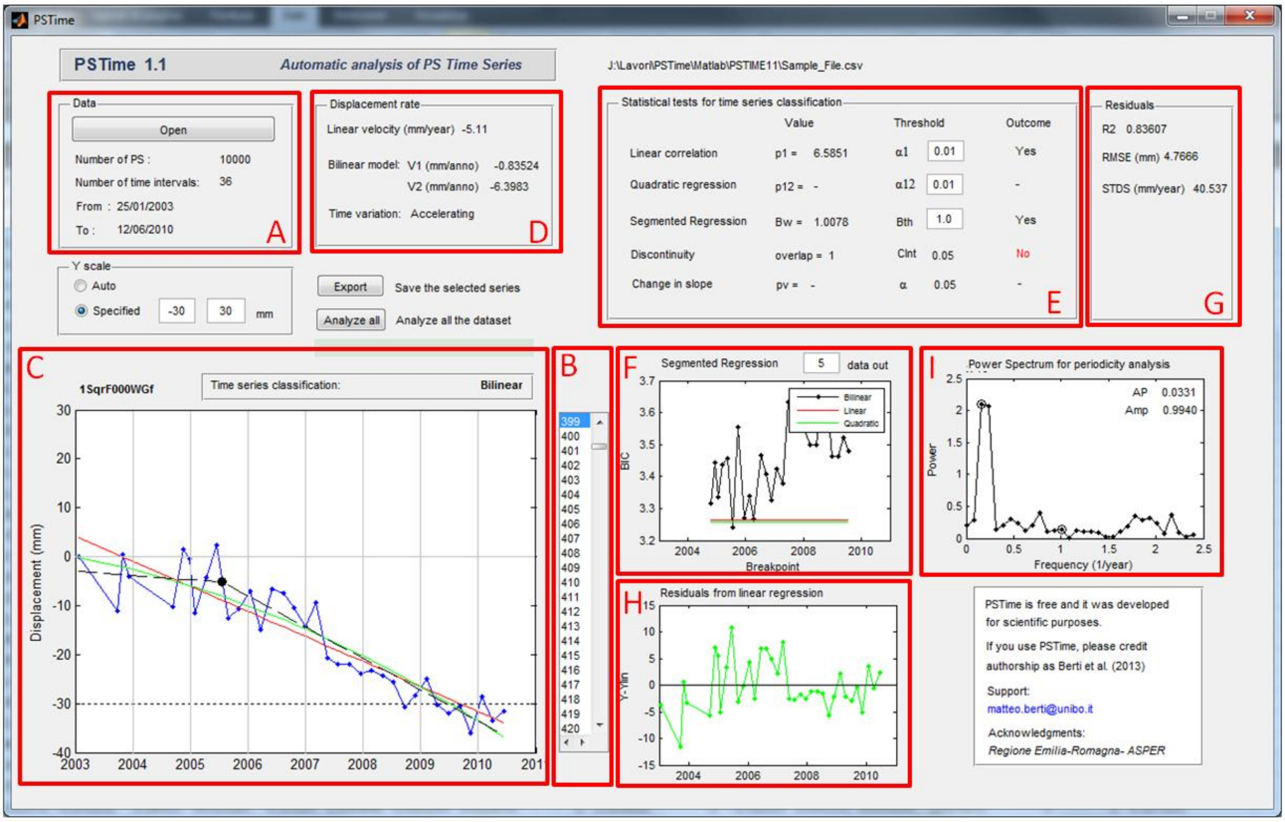


Fig. 1. The PSTime graphical user interface

A short description of the GUI panes is given below:

Pane	Description
A	Load the PS input file (in CSV format)
B	List of the PS contained in the dataset
C	Displacement time series fitted by a linear (green), quadratic (red), and bilinear (black) regression model. The time series classification is shown on the top.
D	Mean displacement rate obtained by a linear model and, for bilinear time series, linear velocities of the two segments.
E	Results of the statistical tests used for the time series classification and corresponding values of the statistical thresholds
F	Segmented regression analysis used to detect bilinear time series
G	Statistical parameters describing the scatter of the time series.
H	Scatter plot of the residuals from linear regression
I	Spectral analysis used to evaluate the Annual Periodicity index (AP) and the amplitude (Amp) of the annual component.

2.3 How to run a PSTime analysis

1. Input file

- Open your PS dataset, and use Excel or any other spreadsheet program to create a data table like the one below:

Column header (text)

Survey dates (dYYYYMMDD)

code	d20030125	d20030927	d20031101	d20031206	d20040911	d20041120
1SqrF000VCQ	0	0.7	-0.1	1.9	-0.7	-3.6
1SqrF000VBw	0	-2.7	-4	0	-4	4
1SqrF000V3A	0	1	3	4.1	-5.1	-2.6
1SqrF000V6b	0	2	2.5	5.2	9.7	-2.5
1SqrF000V2q	0	2.1	2.9	4.9	-3	-3.5
1SqrF000V8O	0	12.8	11.3	14.5	11.4	10.9
1SqrF000V8j	0	8.4	4.5	4.9	3.6	-0.9
1SqrF000V8P	0	9.6	6.1	4.2	5.6	0.2
1SqrF000UhH	0	-3.4	1.9	3.9	-0.4	5.6
1SqrF000VCA	0	-1	-6	-5.4	-4.2	3.4
1SqrF000Uwp	0	7.1	3.1	1.2	0.7	4.8
1SqrF000UgR	0	8.9	5.7	6.4	3.6	2.3
1SqrF000Ugr	0	5.2	2	2.9	1.5	4
1SqrF000V2N	0	10.8	-0.2	6	0.7	-7.1
1SqrF000WIZ	0	3.2	3.5	6.2	2.6	3.4
1SqrF000WK3	0	7.9	9.4	4.4	-1.5	5.2
1SqrF000WPD	0	5	9.2	3.5	0.9	2.6

Permanent Scatterers ID

Fig. 2. Structure of the input file (to be saved as comma-separated values)

- The first column must list the PS identification codes and the column header must be a text (e.g. "code" or "ID").
- The subsequent columns must contain the measured ground displacements with respect the first scene (the first column of zeros). The header of each column indicates the acquisition time in the format: dYYYYMMDD (d=displacement; Y=Year; M=Month; D=Day).
- Export the data table as a comma-separated values file (CSV extension). A sample input file (sample_file.csv) can be found in the installation folder.

2. Start the analysis

Run PSTime and click the "Open" button to browse the CSV input file (pane A in Fig.1). PSTime loads the data and shows the classification (pane C) and the descriptive parameters (panes D-E-F) of the first PS listed in the dataset.

The results are refreshed every time the user click on a different item on the PS list (pane B). Click on the "Export" button to save the selected displacement time series in Excel format together with the values predicted by the linear, quadratic, and bilinear regression models. The export file is named export.xls and saved in the same folder as the input file.

3. Analyze all the dataset

Click the “Analyze all” button to perform a sequential analysis of all the dataset. PSTime will apply the statistical classification tests and the descriptive analysis to all the time series and will store the results in an output table. The output table is saved as a CSV file (filename: InputFileName_Proc.csv) in the same folder as the input file. A sample output file (sample_file_Proc.csv) can be found in the installation folder.

The output table can be join to the original PS shapefile through the field “Code” and used for display with appropriate symbology to improve the spatial analysis of PS.

3 Features and specifications

3.1 Statistical analyses performed by PSTime

3.1.1 Classification of Permanent Scatterers time series

PSTime applies a sequence of statistical tests to classify a PS time series into one of the six pre-defined target trends shown in Fig. 3. These trends were recognized as “typical” pattern of PS data by the visual analysis of 1000 time series collected by ERS and ENVISAT satellites (Berti et al., 2013).

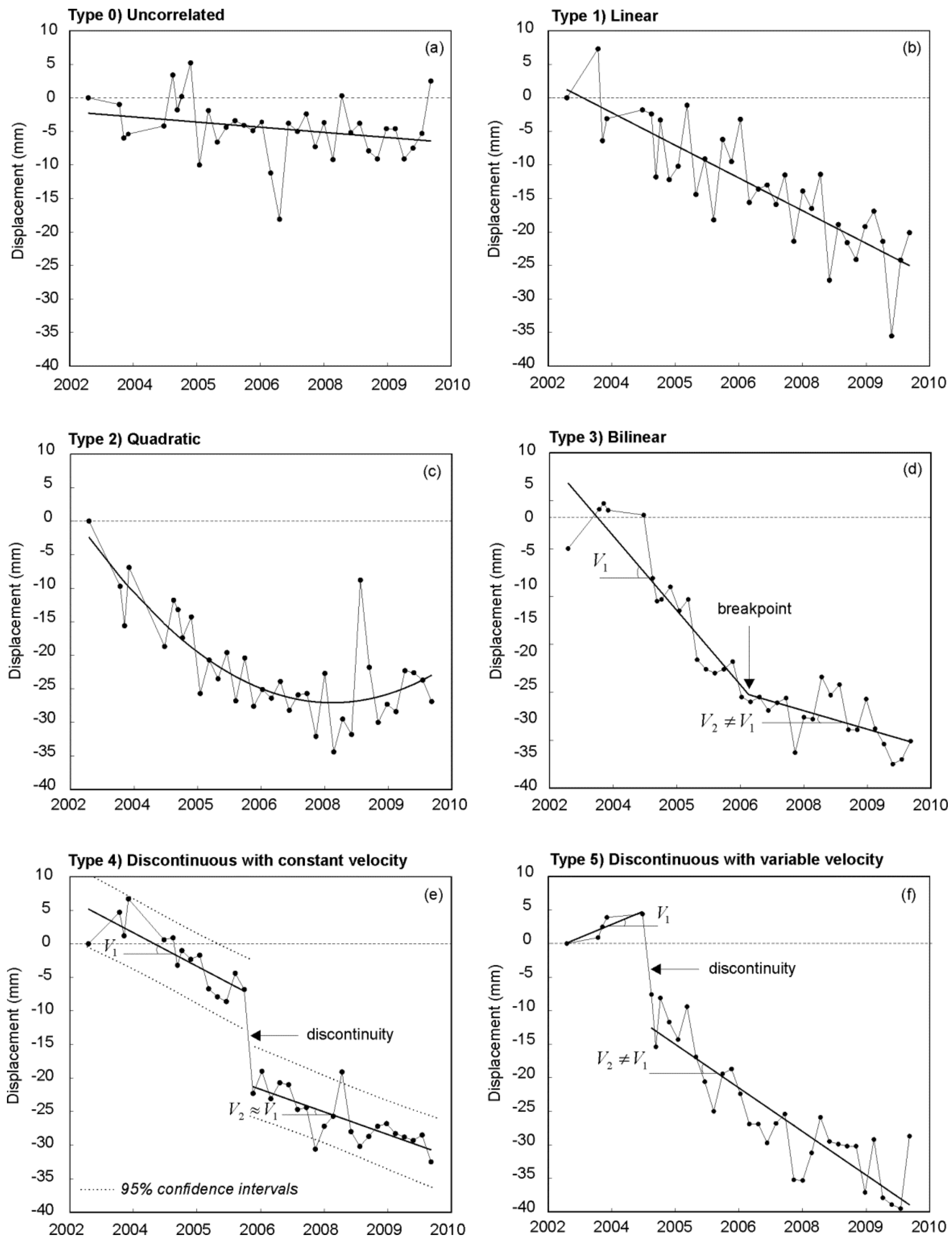


Fig. 3. Typical ground displacement trends identified by visual inspection of 1000 Permanent Scatterers time series.

The conditional sequence of statistical tests is shown in Fig. 4. All tests are conventional statistical tests that are commonly used to check the validity of the regression assumptions (A, B, C in Fig.4) or to determine whether there are significant differences between two series of data (D, E). Refer to Berti et al. (2013) and to the literature (Schwarz, 1978; Weisberg, 1985; Davies, 1986; Main et al., 1999; Steven, 2001; Quinn and Keough, 2002;Wagenmakers and Farrell, 2004) for details.

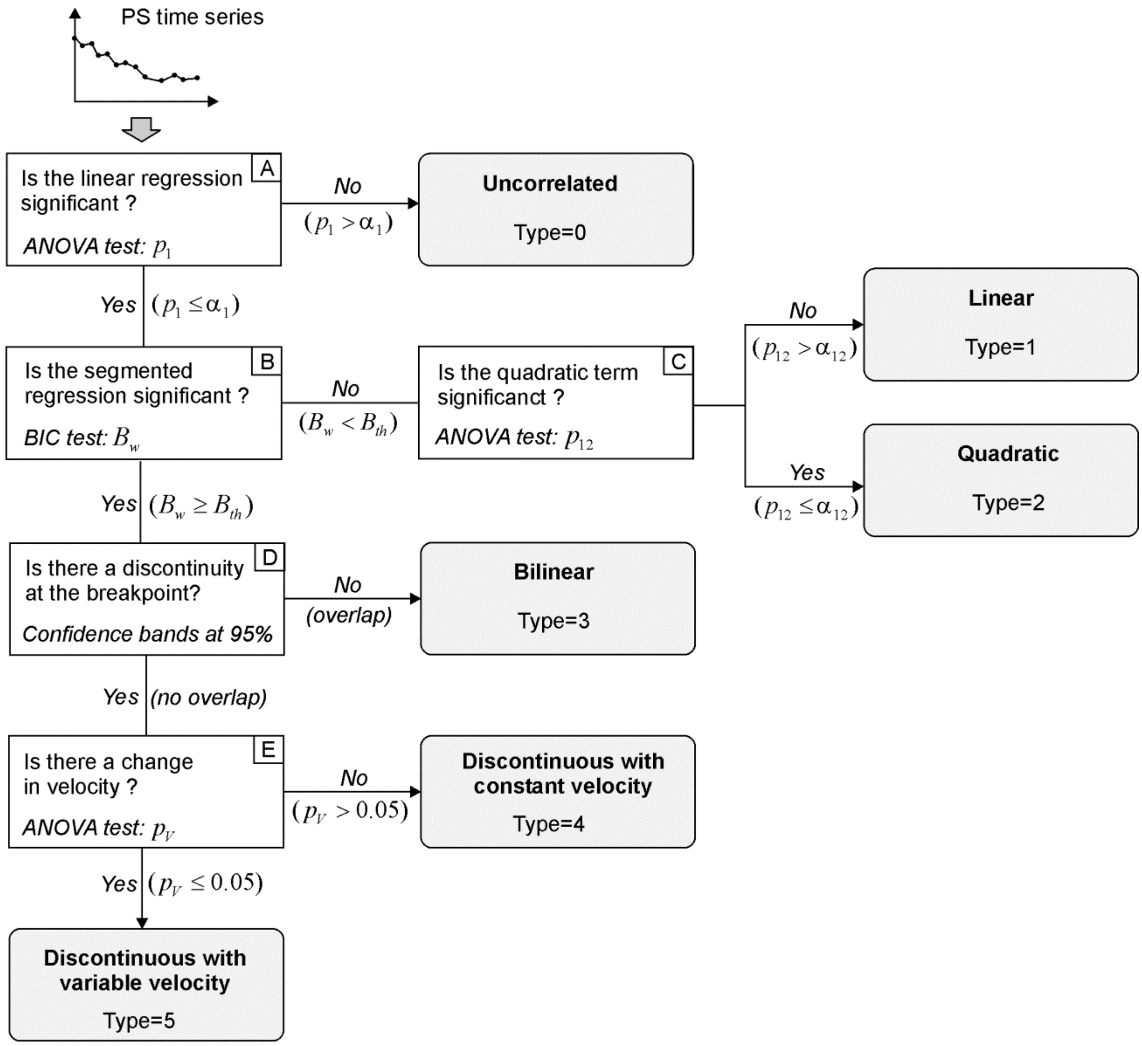


Fig. 4. Workflow of the automatic classification procedure.

The result of the time series classification is reported in pane C (Fig. 1). The chart shows the time series of the selected PS together with the linear (green), quadratic (red), and bilinear (black) regression curves. Pane E (Fig. 1) summarizes the results of the statistical tests and the corresponding statistical thresholds used to accept or reject the null hypothesis (see section 3.2).

The chart in pane F shows the result of the segmented regression, which is applied to detect a bilinear trend in the time series. The chart compares the goodness of fit of a linear (red line), quadratic (green line), or bilinear (black line) model using the Bayesian Information Criterion (*BIC*) (Main et al., 1999). The smaller the *BIC,* the better the fit of the regression model. The *BIC* values of the bilinear model are plotted against the position of the breakpoint along the time series. If the minimum value of *BIC* for the bilinear model is lower than the *BIC* of both the linear and quadratic model, there is statistical evidence that a breakpoint exists in the time series. Although this condition is sufficient to classify the time series as “bilinear”, PSTime uses a more restrictive criterion based on the so called "evidence ratio" that also allows the user to calibrate the test outcome (Berti et al., 2013).

The pane D (Fig. 1) reports the mean displacement rate obtained by the linear model and, for segmented time series (type 3-4-5), the mean velocities in the two linear reaches (V_1 and V_2). In this latter case PSTime compares the absolute value of V_1 and V_2 to evaluate if the displacement rate is increasing (acceleration: $|V_2| > |V_1|$) or decreasing (deceleration: $|V_2| < |V_1|$).

3.1.2 Scatter of the time series

The scatter of the time series is measured by three statistical parameters (panes G and H, Fig. 1):

- the coefficient of determination, r^2
- the Root Mean Square Error, $RMSE$
- the Standard Deviation of Slope, SDS

The coefficient of determination of a time series made of n values is computed as:

$$r^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \qquad (1)$$

where y_i are the measured displacement values, $\bar{y}$ is the mean displacement, and $\hat{y}_i$ are the estimated values predicted by a linear regression model.

The Root Mean Square Error is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{(n-2)}} \qquad (2)$$

High values of r^2 (close to 1) or small values of $RMSE$ (close to 0) indicate a smooth time series with low scatter with respect the average linear trend.

The Standard Deviation of Slope is a statistical index proposed by Frankel and Dolan (2007) to measure the roughness of natural surfaces:

$$SDS = \left[\frac{1}{n^2} \sum_{i=1}^{n^2} (m_i - \bar{m})^2 \right]^{0.5} \qquad (3)$$

where m_i indicates the slope between two successive points of the time series and $\bar{m}$ the mean slope. Low values of SDS (close to 0) indicate low data scatter.

3.1.3 Annual periodicity

The Annual Periodicity index (AP) was introduced to detect periodic fluctuations in the displacement time series with a wavelength of approximately one year. The index is based on a Fast Fourier Transform analysis (Priestley, 1981) and it is defined as:

$$AP = \begin{cases} 0.5\dfrac{P_1}{P_0} & (if\ P_0 \geq P_1) \\ \\ 1 - 0.5\dfrac{P_0}{P_1} & (if\ P_0 < P_1) \end{cases}$$

where P_0 and P_1 are the spectral peaks in the frequency bands $f_0 = 0 \div 0.5$ yr^{-1} (fundamental frequency) and $f_1 = 0.8 \div 1.2$ yr^{-1} (annual frequency) respectively. AP provides a measure of the relative importance of annual periodicity compared to the long-term fluctuation (Berti et al., 2013). The index ranges from 0 (no annual periodicity) to 1 (very strong annual periodicity).

The results of the Annual Periodicity analysis are reported in the pane I of the GUI (Fig. 1). The chart shows the power spectral density of the selected time series, while the parameter "Amp" indicates the amplitude (in mm) of the periodic annual component.

3.2 Choice of the statistical thresholds

It is important to realize that the results of the statistical classification tests strongly depend on the selected level of significance.

As an example, let's take the ANOVA test for the significance of the linear regression (test A in Fig. 4). The null hypothesis H_0 for this test is that the slope of the regression line is zero (no relationship between time and displacement) while the alternative hypothesis is that a significant linear relationship exists between the two variables. The null hypothesis can be rejected if the computed probability value p_1 is less than the selected level of significance α_1. The significance level α_1 represents the probability of a Type I error, that is the probability of rejecting H_0 when it is actually true.

The choice of the level of significant is arbitrary, and tough it is advisable to choose the significance level before taking the data to give more objectivity to the decision, it is common to adjust the selected value to fit the specific needs. For instance, if the PS dataset contains noisy time series with large data scatter, we may choose a low level of significance (e.g $\alpha_1 < 0.01$) in order to reject the null hypothesis only when a strong evidence of linearity exists (low probability to commit a Type I error). On the contrary, a less stringent criterion (e.g. $\alpha_1 > 0.05$) can be adopted when analyzing high-quality data characterized by low noise.

The default values of the statistical thresholds used by PSTime to test the hypothesis of a linear, quadratic, or bilinear model are reported in pane E (Fig. 1). These values were obtained by calibration of a sample dataset collected in the Northern Apennines of Italy (Berti et al., 2013). However, they should be adjusted to the specific needs.

Our recommendation is to run the analysis with the default values and to check a number of PS time series to see if the results are reasonable. If the outcome of the automatic classification does not agree with your expert (subjective) judgment, the statistical thresholds reported in box D should be changed until an appropriate classification is obtained.

Beside this visual verification, one can also analyze all the dataset by clicking on the "Analyze all" button, and plot the frequency distribution of the classified PS time series from the output table. On a large area, we expected that most of the PS is stable (uncorrelated time series, Type 0) or characterized by (slow) steady movements (linear time series, Type 1), while non-linear temporal trends (Type 2 to 5) should represent a smaller percent of the dataset. If this is not the case, the statistical thresholds are probably inappropriate, or the dataset is affected by a systematic error as discussed by Berti et al. (2013).

3.3 PSTime output file

PSTime creates an output file which contains the classification index and the descriptive parameters for all the PS time series. Table fields are described below:

Field	Description
Code	PS code used to join the output table to a georeferenced shapefile
VLin	Mean velocity by linear regression (mm/year)
R2	Coefficient of determination of the linear regression
RMSE	Root Mean Squared Error of the linear regression (mm).
STDS	Standard Deviation of Slope (mm/year)
AP	Annual Periodicity index
P1	F-statistic for the significance of the linear regression (to be compared with the selected level of significance α_1)
P2	F-statistic for the significance of the quadratic regression (not used in the classification procedure)
P12	F-statistic for the significance of a quadratic term added to a linear regression (to be compared with the selected level of significance α_{12})
BL	Result of the bilinear regression analysis: 0= the bilinear model does not provide a better fit to the data than the linear and quadratic models; 1=the bilinear model provides a better fit
BICW	Evidence ratio of the breakpoint B_w (see Berti et al., 2013)
Type	Time series classification: 0=uncorrelated; 1=linear; 2=quadratic; 3=bilinear; 4= discontinuous with constant velocity; 5= discontinuous with variable velocity.
V1	Mean velocity before the breakpoint (V_1, mm/year) computed by bilinear regression for non-linear time series (types 2 to 5)

V2	Mean velocity after the breakpoint (V_2, mm/year) computed by bilinear regression for non-linear time series (types 2 to 5)
Break	Breakpoint date (mm/dd/yyy) computed by bilinear regression for non-linear time series (types 2 to 5)
dV	Change in velocity before and after the breakpoint (mm/year): dV=$\|V_1 - V\|_2$
Acc	Sign of the change in velocity: -1=deceleration; 0=constant velocity; 1=acceleration
Type3	Time series classification with non-linear trends (types 2 to 5) grouped in a single class: 0=uncorrelated; 1=linear; 6=non-linear
Amp	Amplitude (mm) of the annual periodic component

3.4 Maximum size of the input array

The largest size of the input array (number of PS*number of time frames) depends on the operating system and configuration.

On a 32-bit Windows XP system, the number of elements in largest real double array is about 150e6 (≈1.2 GB), while on a 64-bit Windows XP system is about 2000e6 (≈16 GB). See the MathWorks web site (http://www.mathworks.it/support/solutions/en/data/1-IHYHFZ/index.html) for further information.

To prevent "Out of memory" errors with very large dataset, you can split the input file into smaller files and perform the analysis separately.

4 Release history

Do not hesitate to contact the author (matteo.berti@unibo.it) if you find any error and bug, or if you have comments or suggestions to improve PSTime. We need your feedback!

Release	Date	Description
1.1	April 2013	First release of PSTime

5 References

Berti M., Corsini A., Franceschini S., Iannacone J.P. (2013). Automated classification of Persistent Scatterers Interferometry time-series. *Natural Hazards and Earth System Sciences (NHESS),* in press.

Davis J. C. (1986), *Statistics and Data Analysis in Geology*, John Wiley and Sons, New York, USA.

Frankel K. L., and J. F. Dolan (2007), Characterizing arid-region alluvial fans with airborne laser swath mapping digital topographic data, *J. Geophys. Res.*, *112*, F02025.

Main I. G., T. Leonard, O. Papasouliotis, C. G. Hatton and P. G. Meredith (1999), One slope or two? Detecting statistically significant breaks of slope in geophysical data with application to fracture scaling relationships, *Geophysical Research Letter, 26*, 2801-2804.

Priestley M. B. (1981), *Spectral Analysis and Time Series*, Academic Press, New York, USA.

Quinn G. P., and M. J. Keough (2002), *Experimental Design and Data Analysis for Biologists*, Cambridge University Press, Cambridge.

Schwarz G. E. (1978), Estimating the dimension of a model, *Annals of Statistics*, 6, 461–464.

Steven A. J. (2001), Inference and estimation in a changepoint regression problem, *The Statistician*, 50, 51-61.

Wagenmakers E. J., and S. Farrell (2004), AIC model selection using Akaike weights, *Psychonomic Bulletin & Review*, 11, 192-196.

Weisberg, S. (1985), *Applied Linear Regression*, John Wiley & Sons, New York, USA.