

## SSA Gap-filling

---

Here we demonstrate Toolkit capabilities of analyzing univariate dataset with gaps using iterative form of SSA. Gaps are filled-in by utilizing spatio-temporal correlations estimated from the available data. Missing values in the datafile should be marked with "NaN". Unfortunately(!), **Grace** (default graphics package) does not support plotting NaNs, so dataobjects with missing values should be plotted **outside** the Toolkit.

### Steps for **Gap-filling** in the Toolkit:

Gap-filling feature is available in **Reconstruction** option of **SSA (MSSA)** panels, for univariate (multivariate) data, respectively.

The user needs:

1. to select the data from the data pop-up menu of SSA (MSSA) tool,
2. specify the SSA (MSSA) window size (large enough to cover longest temporal correlations; for a guidance it can be the largest period in the dataset if it is known),
3. set the number of SSA (MSSA) components **on the main(!)** SSA (MSSA) panel in **Components** field, and the method for estimating the covariance matrix that is decomposed (diagonalized) by SSA (MSSA).

4.

Then gap-filling is done then by clicking **Fill** in the **Reconstruction** panel with the result stored in **RC-sum Vector**. (**DO NOT(!!!) USE Compute** on the main SSA panel or **Reconstruct** in **Reconstruction** panel, the program will crash violently!) When **Fill&Smooth** box is checked, then the result will be equal to estimated smooth component everywhere, i.e. both the points where the original data is missing and is available. Otherwise, original (potentially noisy data) is retained in the points where data is available.

5. To visualize the filled-in data please use **Plot->Vector** in the Main menu, and not **Plot** in **Reconstruction!!!** (the latter will put both original (with missing values) & the filled-in data on the same plot, & since Grace can not handle NaNs nothing will be plotted!!!)

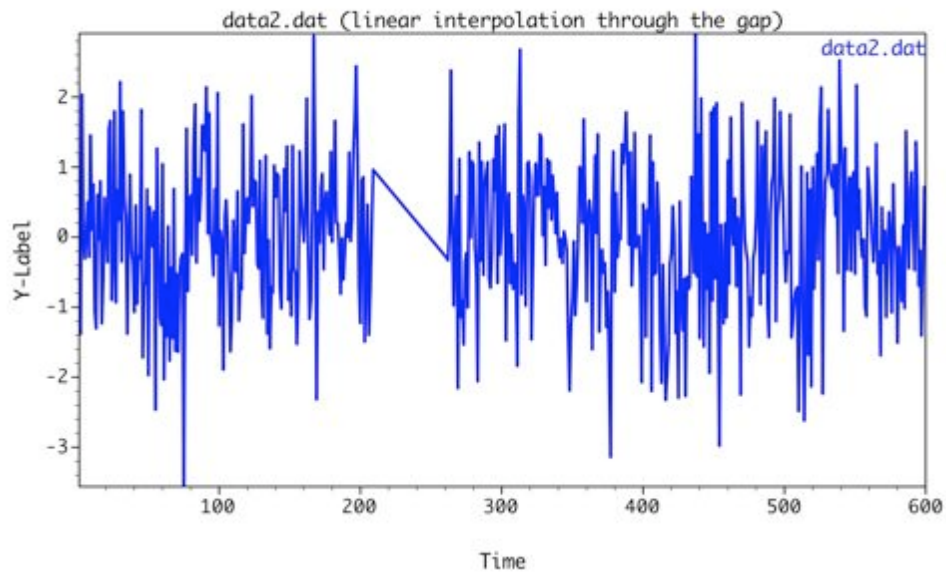
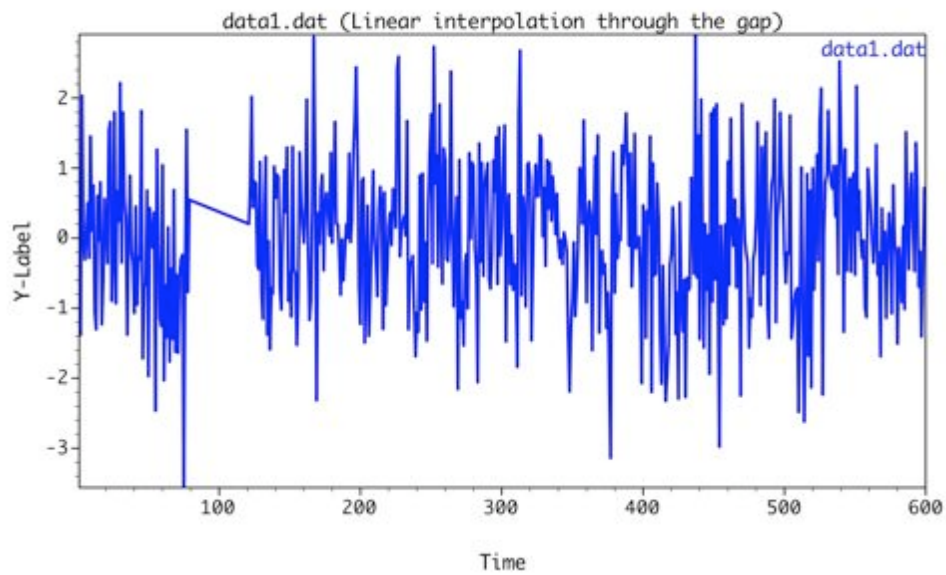
The number of SSA (MSSA) components one has to use depends on the dataset, and in particular on the amount of noise present. The main idea is to discard higher-ranked components corresponding to noise, and use only "smooth" components of "signal".

---

Here we demonstrate gap filling on univariate synthetic time series. It is 600 data points long, and consists of the oscillatory carrier signal with a period of  $T=40$  units. This oscillation is modulated in phase with period of  $T=120$ , and in amplitude with period of  $T=200$ ; it is also contaminated by large amplitude white noise (**fulldata** file).

### 1. Filling-in gappy data:

Double click **Spectra**, go to **File/Data->Read Vector**; navigate to the folder **SUMMER\_SCHOOL/GAPS** to select file **data1.dat** in **Files** panel, change name in "**Store data in vector**" field to **data1**, and click **Read file**. Repeat the same procedure to read the data from **data2.dat** and **full.data.dat**, **data2** and **fulldata** vectors, respectively. Two gappy data sets (**data1** & **data2**) have large continuous gaps in different locations (**data1** ([80 120]) and **data2** ([210 260])) and have been plotted outside the Toolkit with a 3rd party software:



Go to SSA in **Analysis Tools**, select **data1** from Data pop-up menu, set Covariance option to **Broomhead-King**, change the **Window** value to 200, set number of **Components** to **7** that contain the oscillatory signal only. Go to **Reconstruction**, uncheck the **Fill&Smooth** box to estimate the smooth component of the **data1** only in the gap, set the name of filled-in vector to **ssarcvec1** and hit the **Fill** button (not **Reconstruct**!).



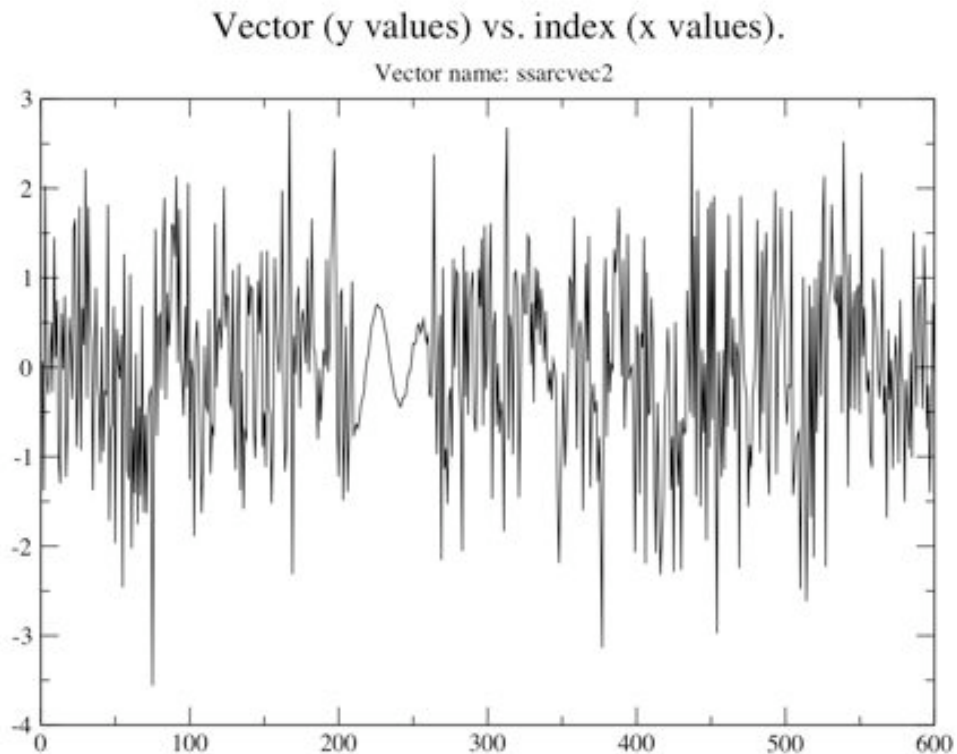
Repeat above procedure with the **data2** to obtain filled-in vector **ssarcvec2**.

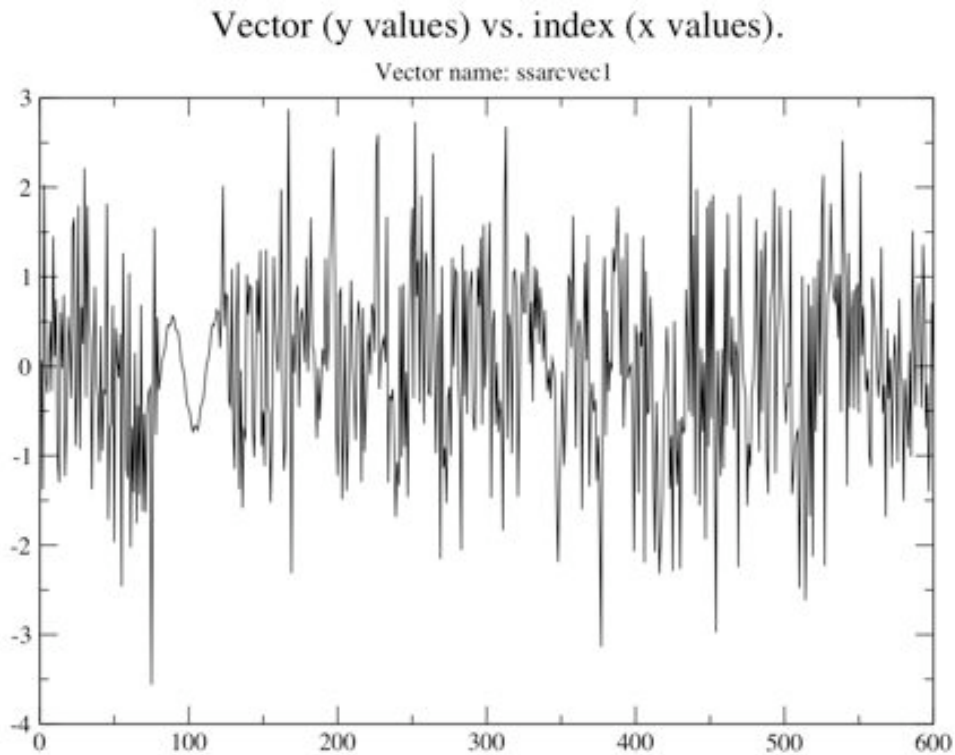
---

**Important Note: Grace can not handle NaNs when plotting (unfortunately). Use Plot->Vector to plot SSA filled-in data (ssarcvec1 & ssarcvec2), not Plot in SSA Reconstruction!!**

---

Note oscillatory mode estimated in the gaps:





To obtain the smooth components of data1 & data2 repeat the above analysis with the **Fill&Smooth** box checked.

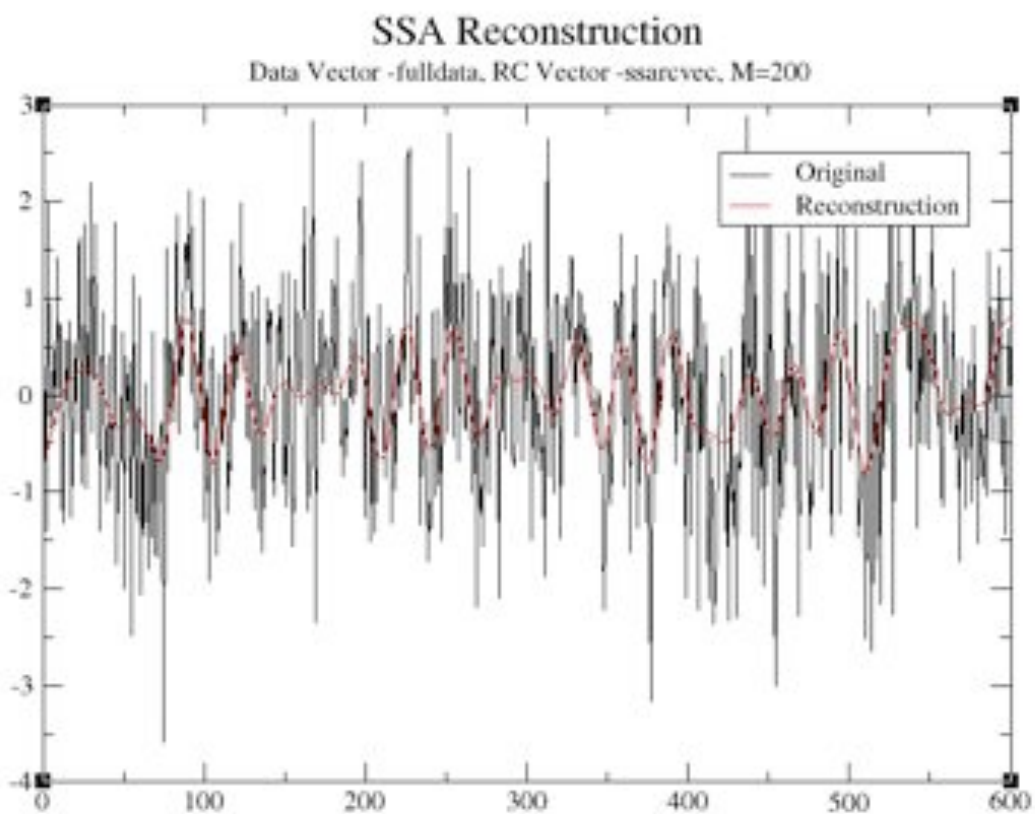
## 2. Compare filled-in "signal" with a smooth component estimated from the full data:

Go to SSA in **Analysis Tools**, select **fulldata1** from Data pop-up menu, set Covariance option to **Broomhead-King**, change the **Window** value to 200, set number of **Components** to 7 that contain the oscillatory signal only, click **Compute**. Click **Reconstruction** on the main SSA panel, and set "1 2 3 4 5 6 7" in **Specify Components**. Set the name of RC-Vector to "ssarcvec"



Then click **Reconstruct** and **Plot** to see the smooth component of the full dataset.

**Figure 1 (print to a file)**

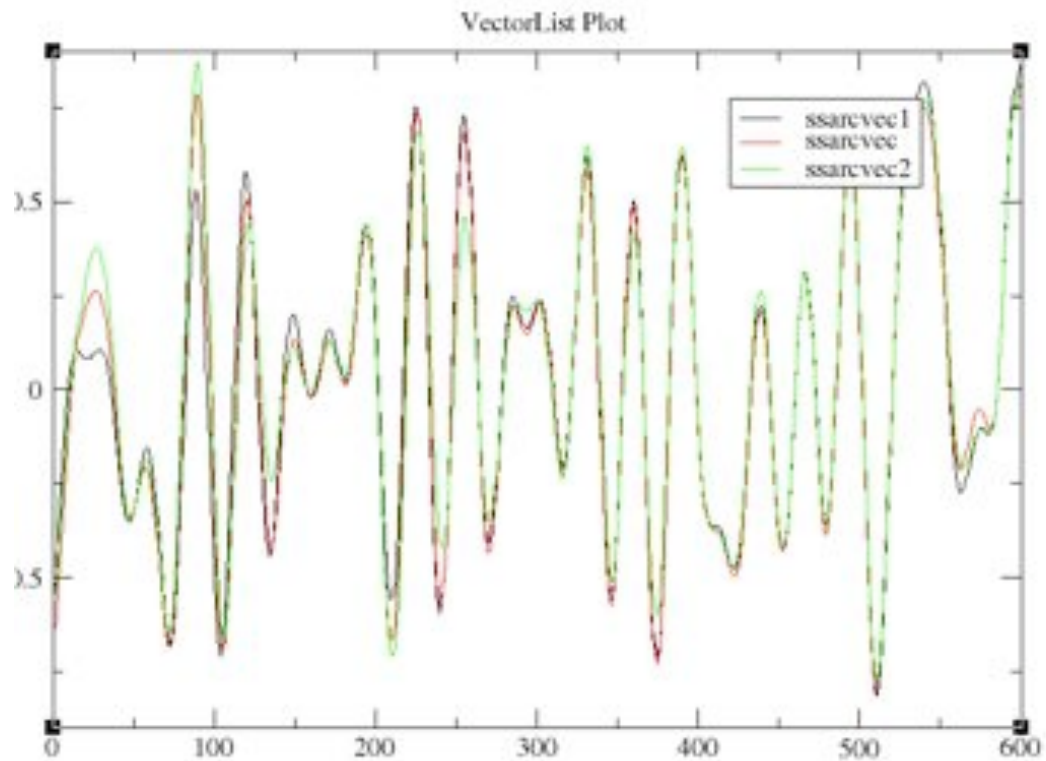


To compare with the smooth component (**ssarvec**) of the full dataset, go to the **Plot/Vector List** option and **Add** vectors **ssarvec**, **ssarvec1** (**ssarvec2**) to the **List of Vectors**,

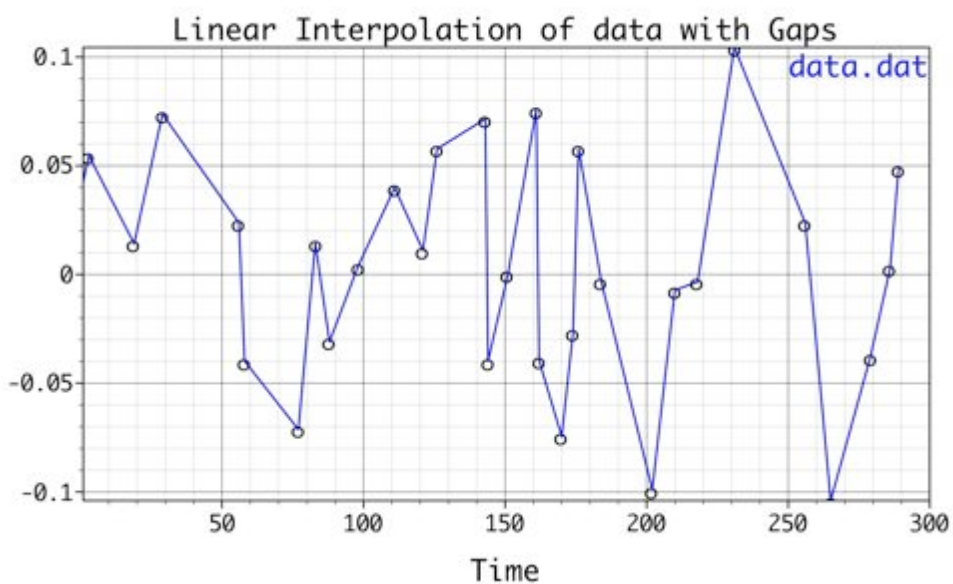


then click **Plot**.

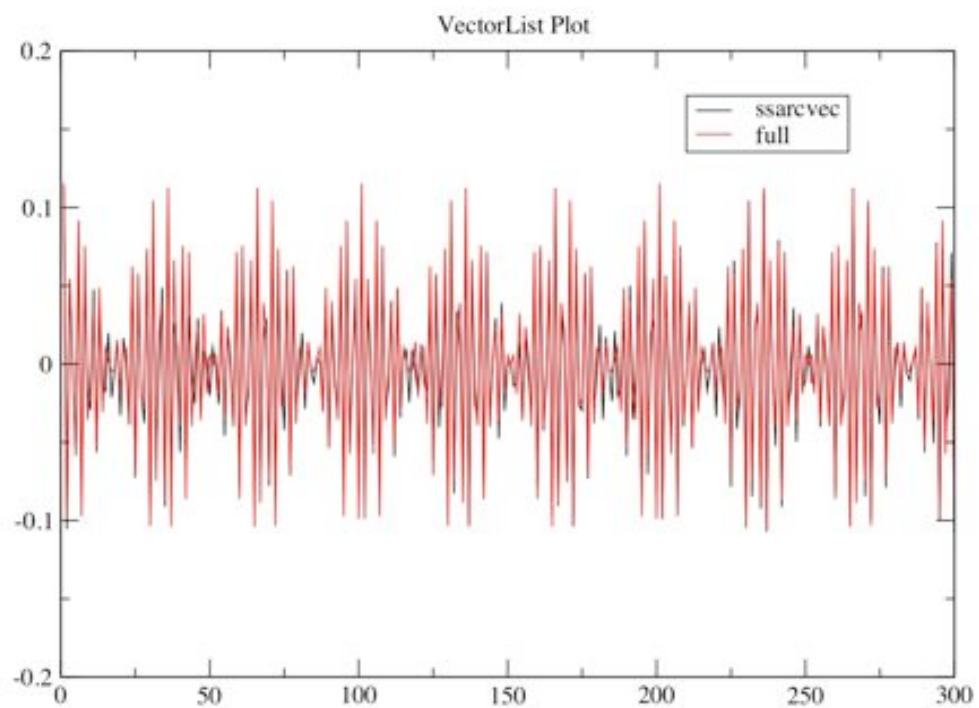
**Figure 2 (print to a file)**



3. Apply SSA gap-filling to mysterydata.dat (300 data points, 90% is missing)



& try to obtain the result close to a right answer (mysteryanswer.dat):



**Figure 3 (print to a file)**

**Hint: several gap-filling attempts & playing with parameters may be necessary.**