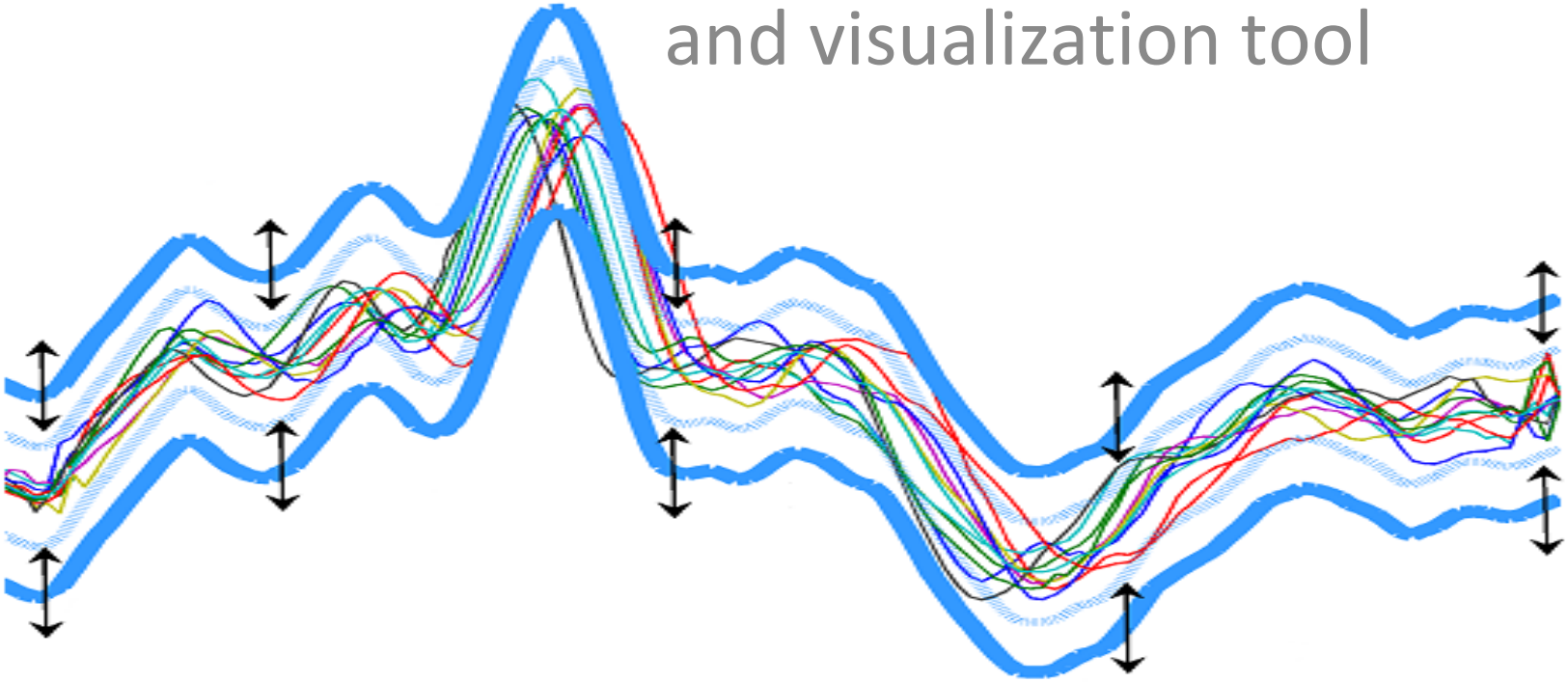
iMotifs

an interactive *multiresolution*

time series motif discovery

and visualization tool



Nuno C. Castro and Paulo J. Azevedo

iMotifs - Interactive Motif Discovery Tool

Time Series Database Directory

A:\data\10 ...

Run

database version
subsequence version

Motif Length 1000
Overlap% 20

Min Cardinality 4
Max Cardinality 32

MinThreshold 2

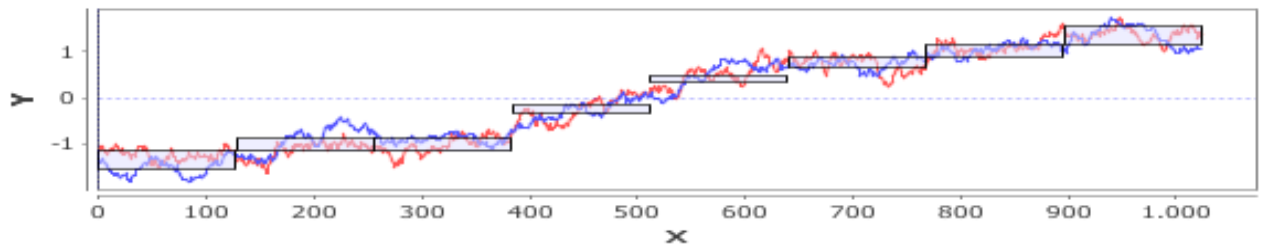
Motif Hierarchy Tree

- 0,0,0,1,2,3,3,3^{4}(163)
  - 0,1,1,3,5,6,6,7^{8}(8)
    - 1,2,2,6,10,12,13,14^{16}(2)
    - 1,2,2,7,11,13,12,14^{16}(1)
    - 1,2,3,6,11,12,12,14^{16}(1)
    - 1,2,2,6,11,12,12,14^{16}(1)
    - 1,2,3,7,11,12,13,14^{16}(1)
    - 1,2,2,7,11,12,13,14^{16}(1)
    - 1,2,2,6,10,13,13,14^{16}(1)
  - 0,1,1,2,4,6,6,7^{8}(6)
    - 1,2,3,4,9,12,12,15^{16}(1)
    - 1,3,3,5,9,13,13,14^{16}(1)
    - 1,2,3,5,9,12,13,14^{16}(1)

Invert Tree

show all

Original Time Series



Draw Motifs

Send to back

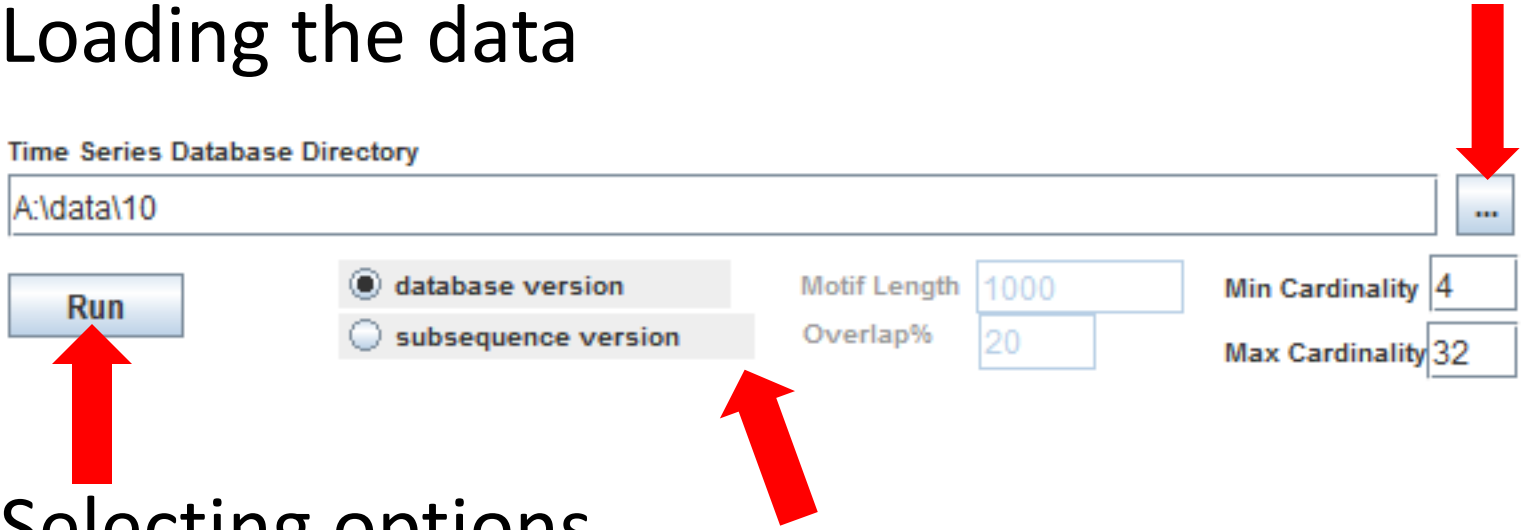
Maximize

Finished. Idle.

Ok

Usage

- Loading the data



- Selecting options
- Running the MrMotif algorithm

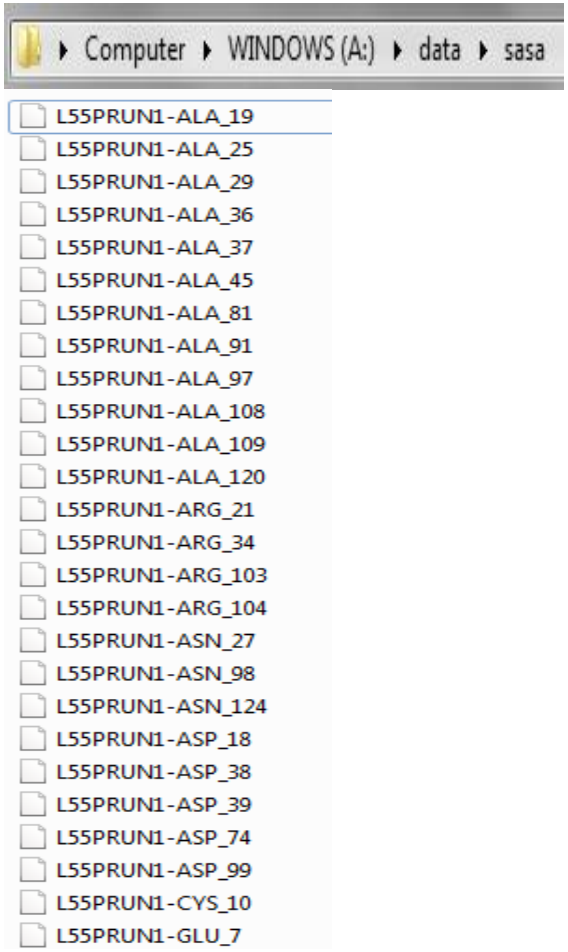
Loading the data

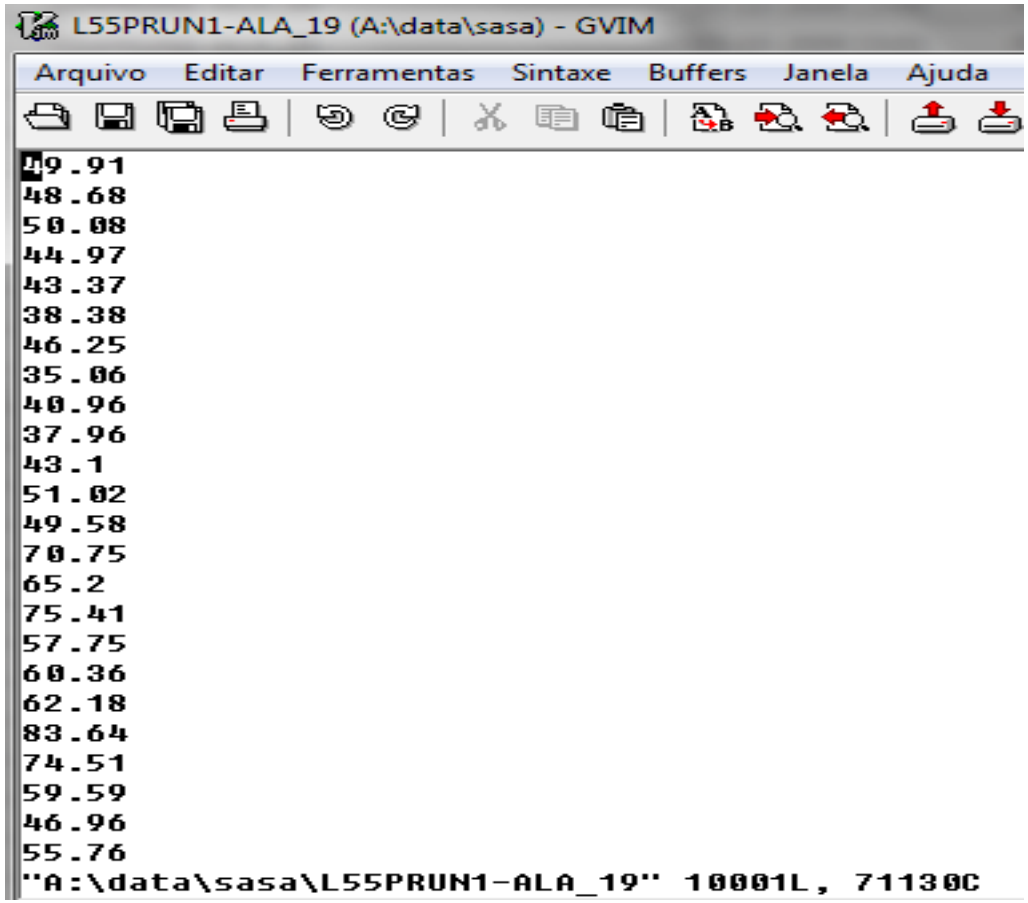
- Select a directory with your time series files
- Each file can be:
  1. A time series (one point per line);
  2. A set of time series (one time series per line)

  This is detected automatically (no parameter necessary)
- Each time series point is separated by whitespace (not tab)

Loading the data (1)

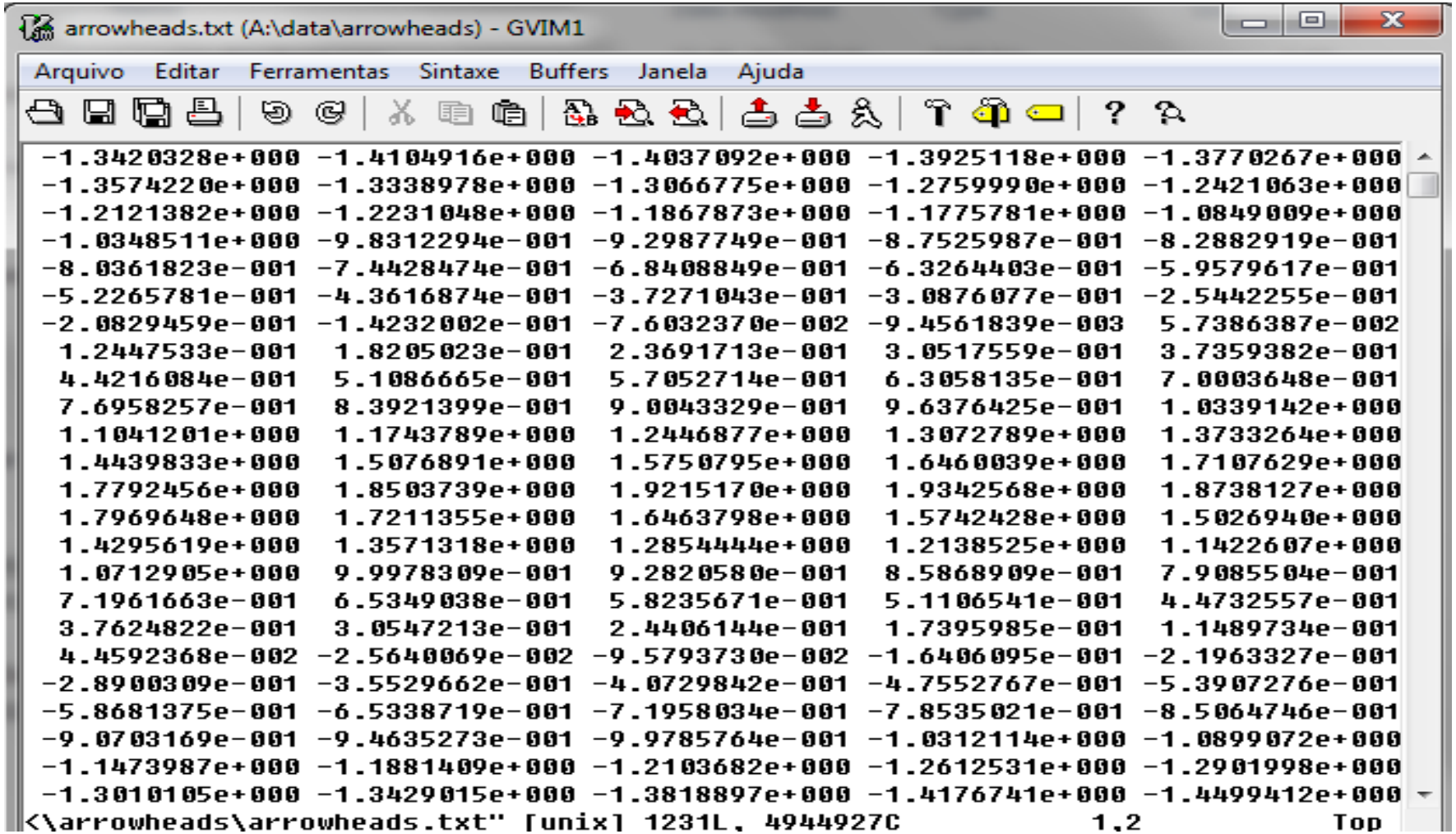
- Example (one time series per file):





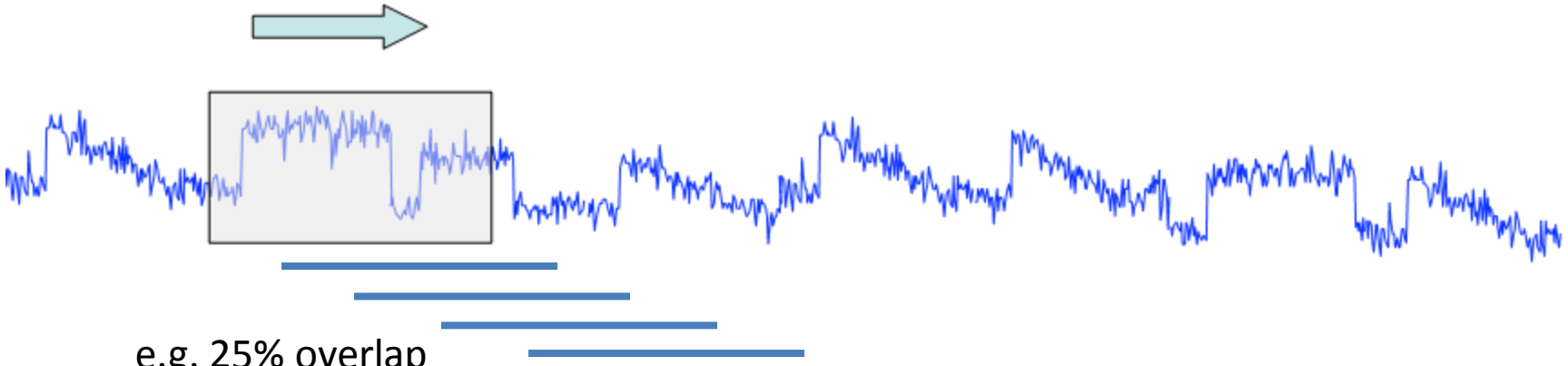
Loading the data (2)

- Example (multiple time series per file, one per line):



Options

- Subsequence Version
  - Motif are extracted using a sliding window of length *Motif Length*.
  - Each sliding window overlaps *overlap%* with the previous one (overlap% helps controlling trivial matches)



e.g. 25% overlap

image from U. Rebbapragada. Introduction to machine learning: Research on time series. Available at the URL *http : //timemachine.iic.harvard.edu/site_media/i/umaatutorial.pdf*, 2007.

Options (cont.)

- Database version
  - There is no sliding window
  - We have a time series database, where each time series is a different object in the DB, e.g. Database of shape time series, where each time series represents a shape. Motifs are similar frequent shapes.

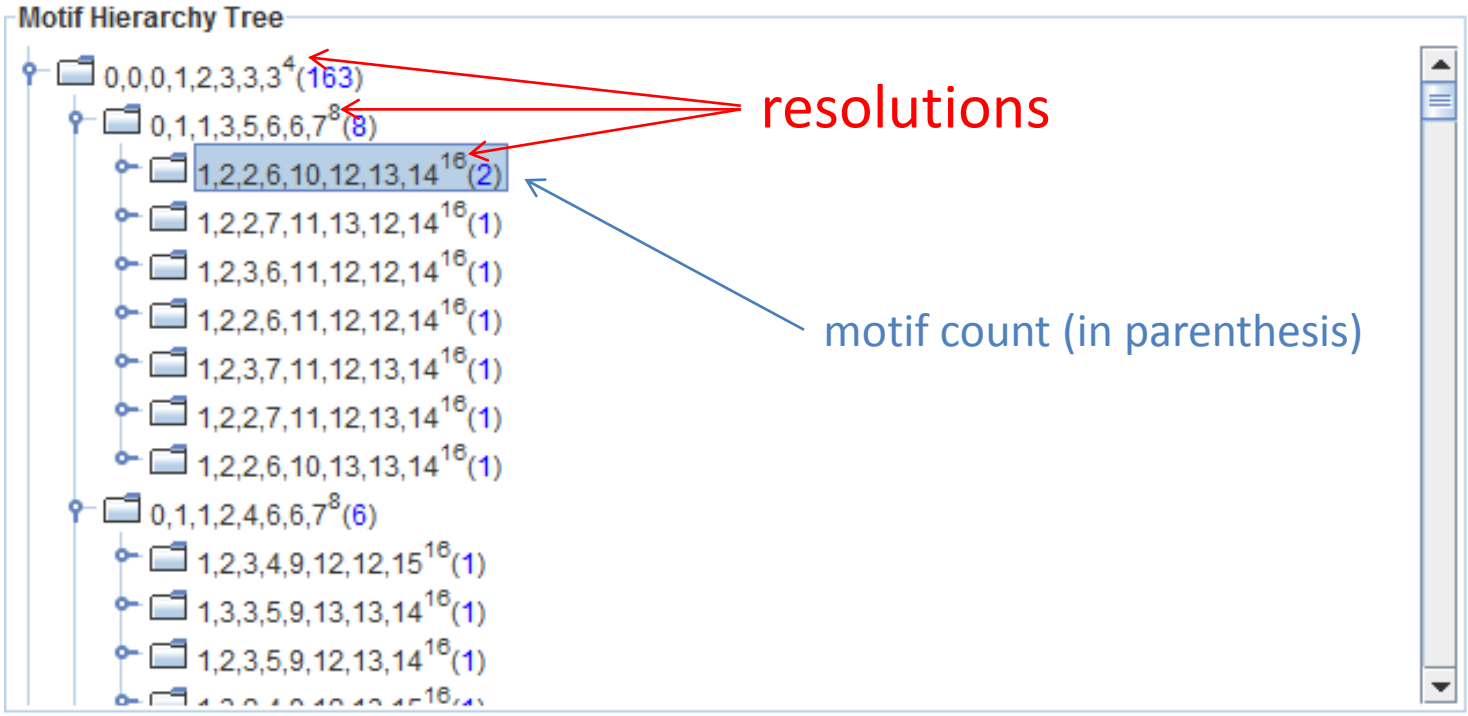
Options (cont.)

- *Minimum* and *maximum* resolution (AKA alphabet size or cardinality) can be set to 4 and 32, unless you want to play with the algorithm
- MrMotif finds motifs for all powers of 2 between *min* and *max*, e.g. 4,8,16,32
- The larger the resolution, more difficult it is to find motifs. So, motifs that persist as resolution increases are best.

→ Main idea behind the MrMotif algorithm

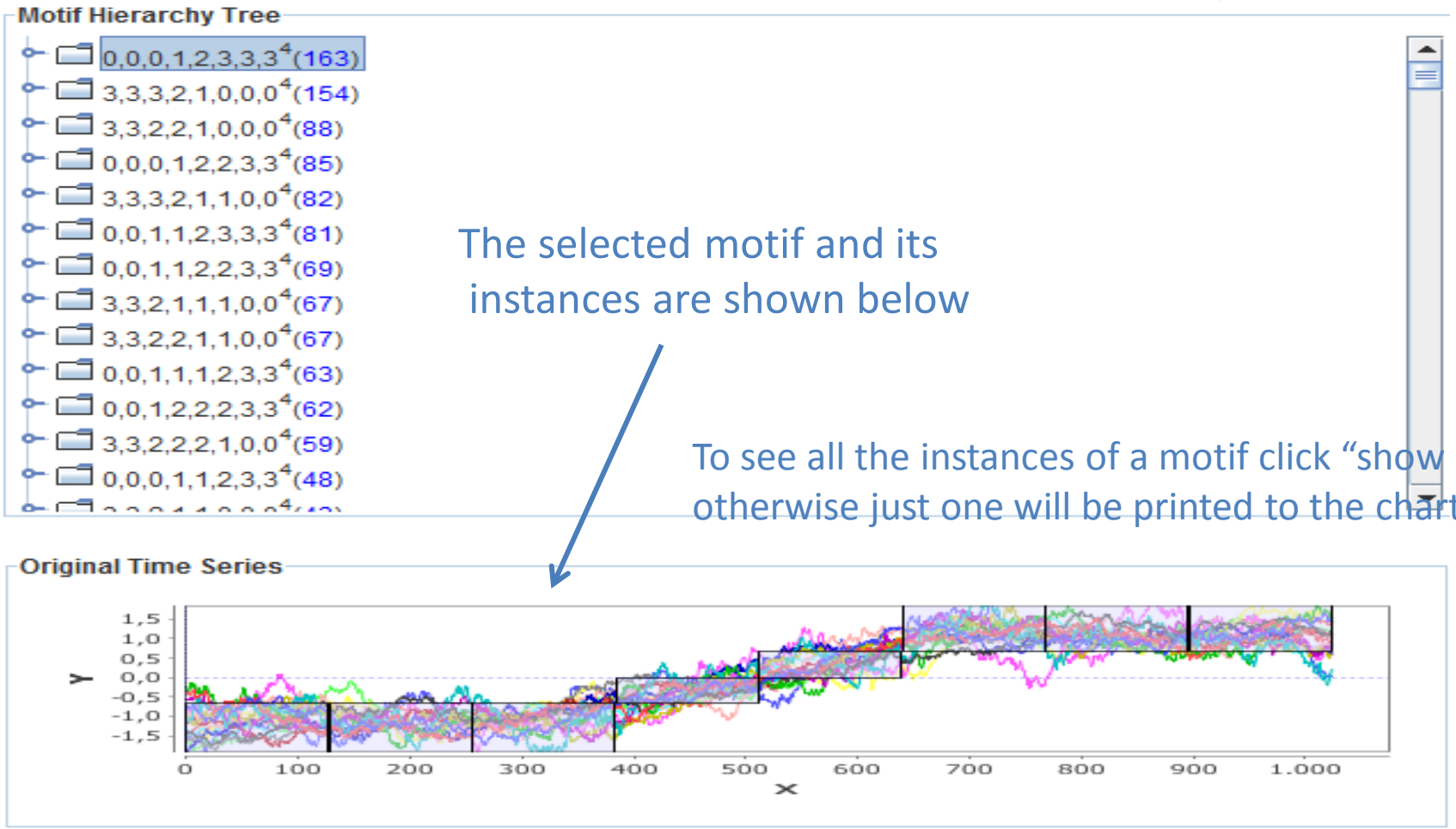
Running MrMotif and visualizing

- As MrMotif searches the DB for motifs in all resolutions between *min* and *max*, we can draw a motif hierarchy tree for each motif:



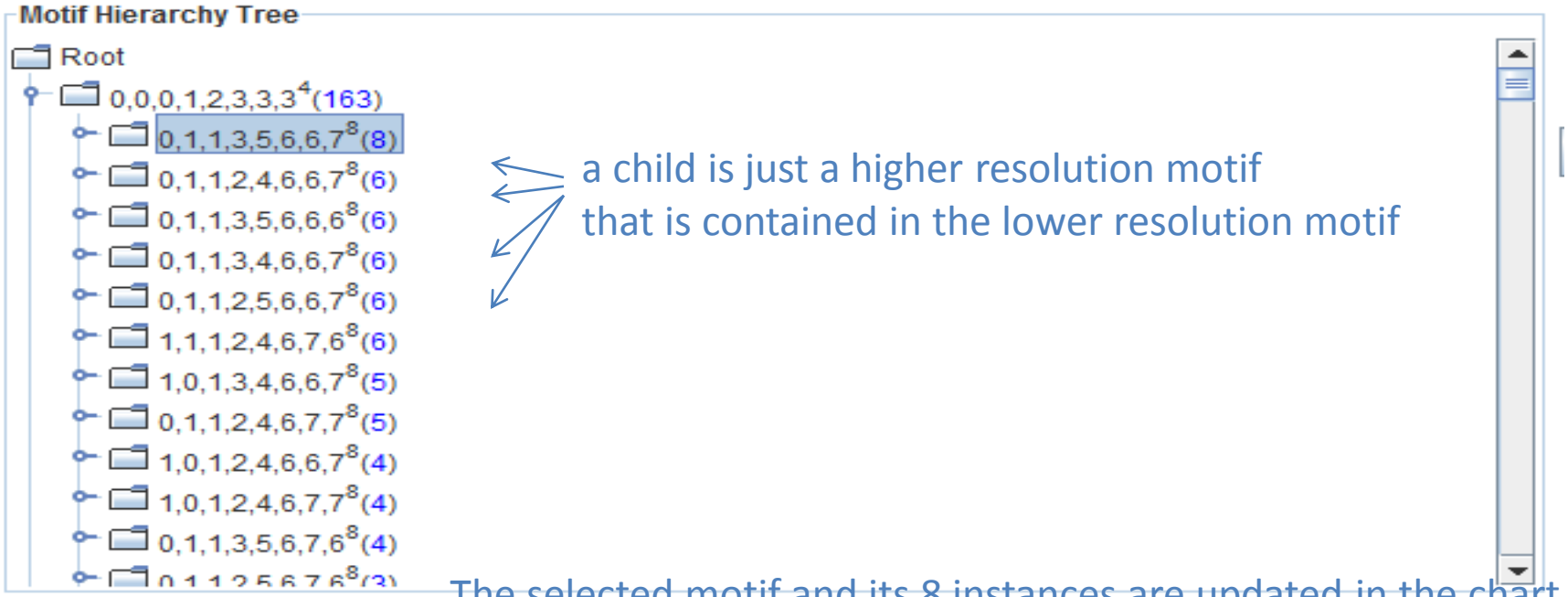
Motif Hierarchy

- First, all motifs are shown for *min* resolution (4)



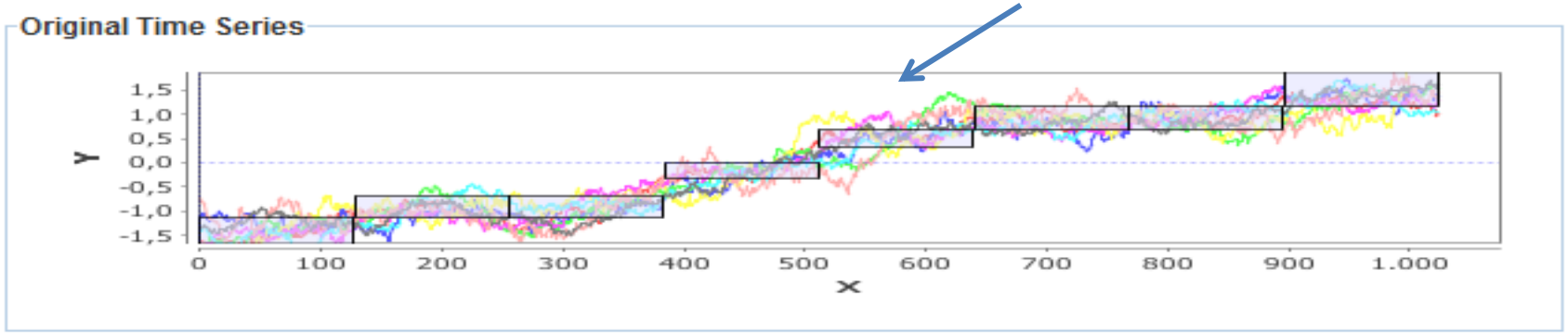
Motif Hierarchy

- By clicking a motif, we “drill-down” to its children in the next resolution (8)



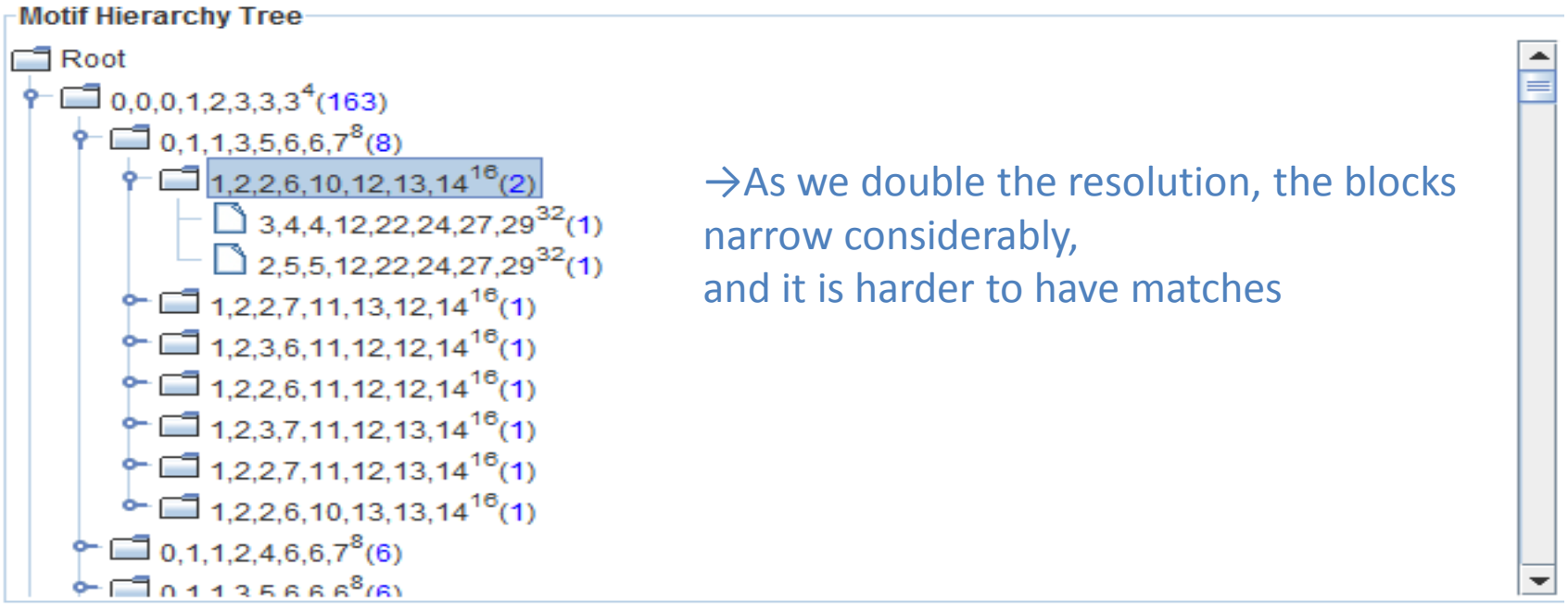
a child is just a higher resolution motif
that is contained in the lower resolution motif

The selected motif and its 8 instances are updated in the chart

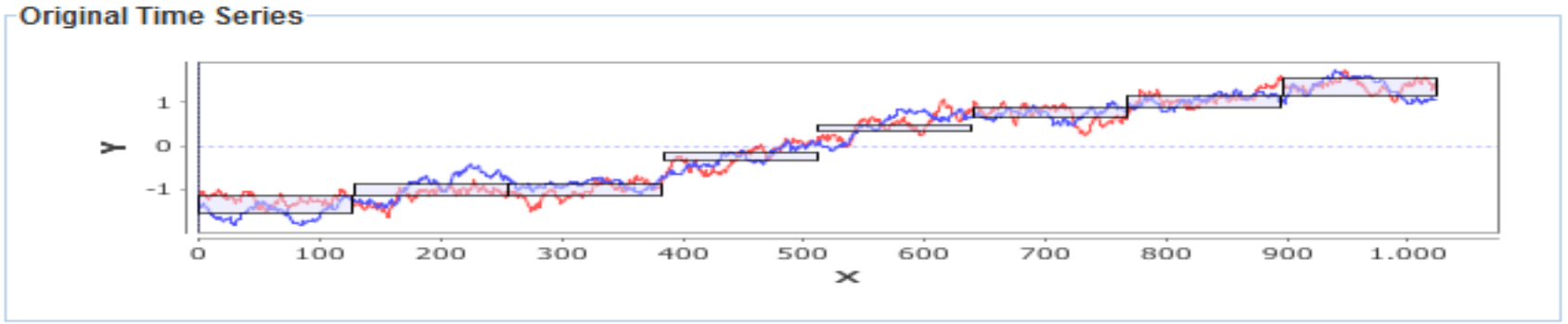


Motif Hierarchy

- And so forth:



→As we double the resolution, the blocks narrow considerably,
and it is harder to have matches

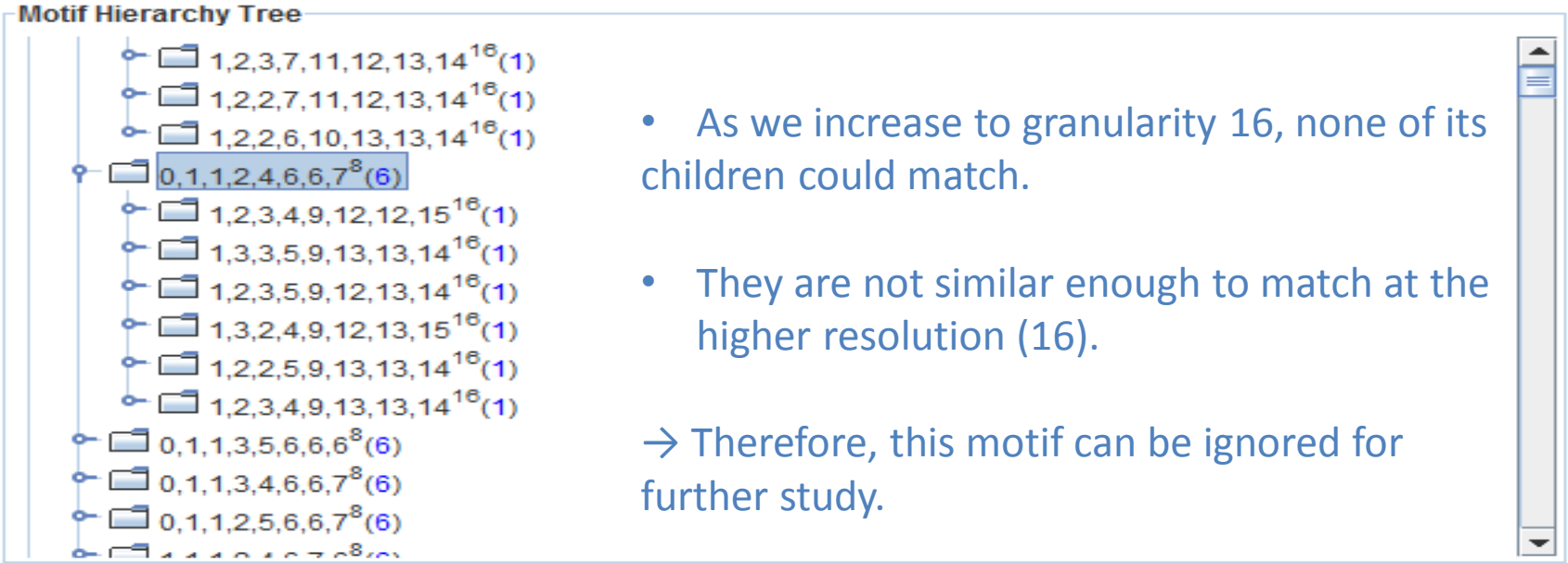


Multiresolution

- At the larger resolutions (16 and 32) it is considerably harder to have a match:
  - The time series need to be really similar to match those tiny blocks
  - Therefore we can skip expensive distance (e.g. Euclidean) calculations when we do have a match
- This is the main intuition behind the MrMotif algorithm
  - it allows the algorithm to execute faster

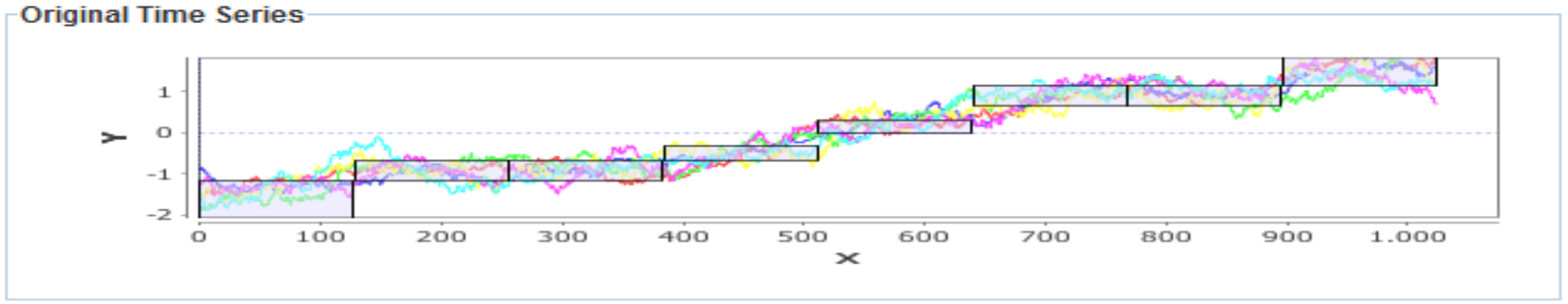
Example

- Take for example, motif {0,1,1,2,4,6,6,7} at resolution 8, having 6 repetitions in the database:



- As we increase to granularity 16, none of its children could match.
- They are not similar enough to match at the higher resolution (16).

→ Therefore, this motif can be ignored for further study.



Multiresolution

- This permits the user to gain useful insight over its time series database
  - He can study how the most frequent motifs behave over the several resolutions
  - He can play with the motif hierarchy and find interesting patterns that he wasn’t expecting

Other features

- Invert tree
- Location and Euclidean distance calculation

Invert tree feature

- This feature allows to invert the hierarchy tree, i.e. Show the most frequent motifs at the highest resolution (32).
- When using this feature, the *max* resolution should be set to 16 or 8, since it may be hard to find motifs at resolution 32.

Location and Euclidean dist.

- By clicking with the second mouse button on a motif in the tree, there is the possibility to:
  - Get the location of the motif instances in disk
  - Calculate the Euclidean distance among all the instances that compose a motif.