LNCC

Jornada de Ciência de Dados

Motifs identifications in Spatial-Time Series

CEFET/RJ

Eduardo Ogasawara
http://eic.cefet-rj.br/~eogasawara

Collaborators

CEFET/RJ: **Murillo Dutra, Riccardo Campisano (MSc.students)**

LNCC: **Fabio Porto**

INRIA: **Florent Masseglia, Esther Pacitti**

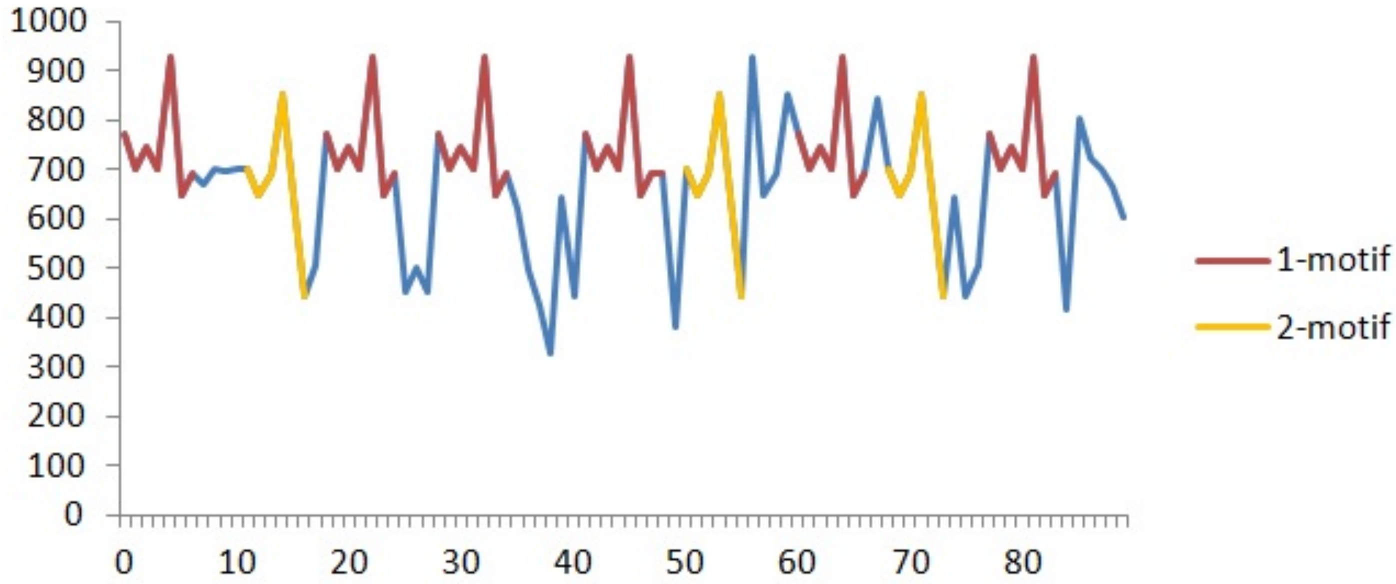
CEFET/RJ

National Laboratory Scientific Computing

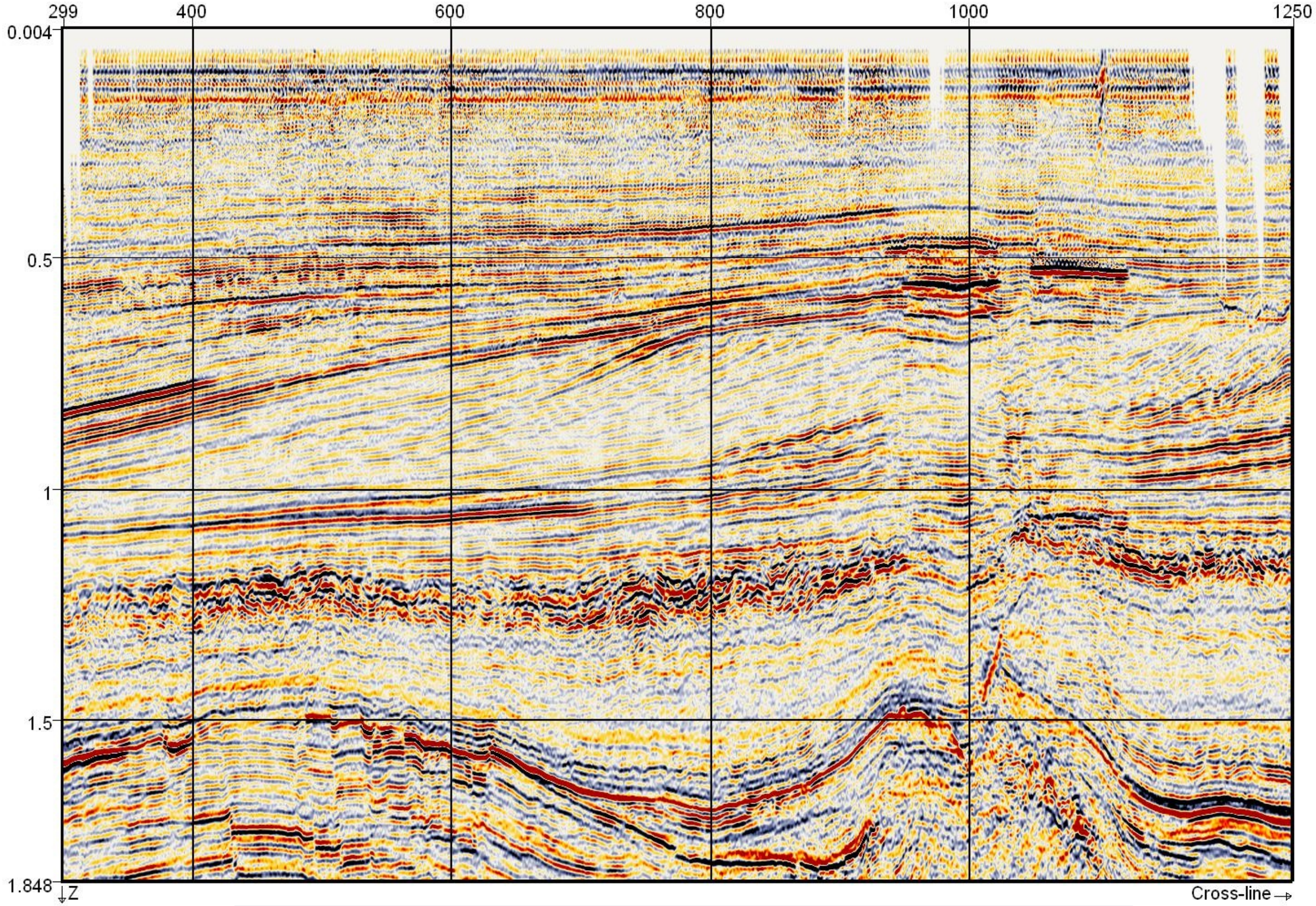
informatics mathematics

Motifs in Time Series

- Time series express phenomenon of interest
- Identifying motifs (patterns) in time series brings knowledge and enables predictions



Seismic Traces Analysis



Crossline: 100 (951 time series with 462 values)
Netherlands dataset

Analysis of Delays in Airports

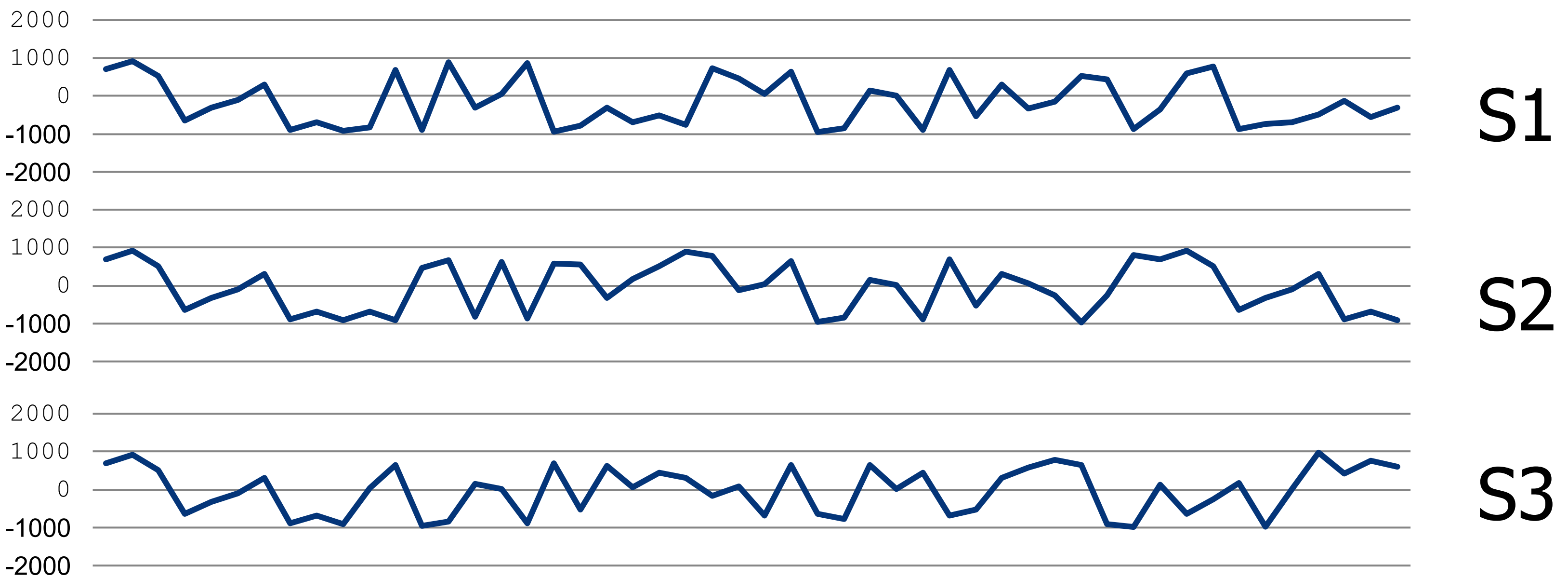
Analysis of delays in airports according to time

Buses Stops Analysis

Buses are sensors: Trajectory Data
Spatial-time aggregation of buses according to buses stops
Buses Stops becomes Spatial-Time derived sensors

Motifs in Spatial-Time Series

- Motifs identification and analysis in time series is being studied during last decade



Absence of studies of motifs identification and analysis in spatial-time series

Time Series and Sequences

Definition 1. *A **time series** t is an ordered sequence of values in time [1], where each t_i is a value, $|t| = m$ is the number of elements in t, and t_m is the most recent value in t.*

$$t =< t_1, t_2, \ldots, t_m >, t_i \in \mathbb{R}$$

Definition 2. *The p-th **sub sequence** [2] of size n in a time series t, represented as $t^{p,n}$, is an ordered sequence of values $< t_p, t_{p+1}, \ldots, t_{p+n-1} >$, where $|t^{p,n}| = n$ and $1 \leq p \leq |t| - n$.*

$$t^{p,n} = subseq(t, p, n)$$

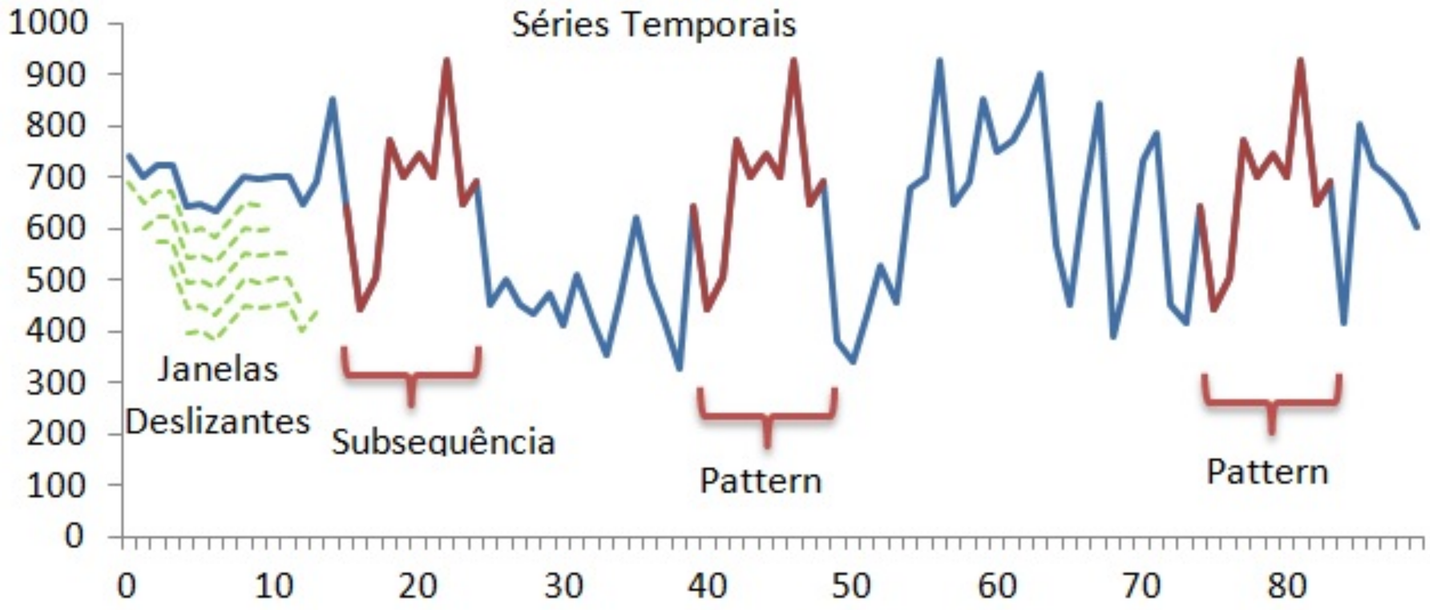
Sliding Window

Definition 3. *A* ***sliding window*** *[3] is a function* $sw(t, n)$ *with arguments* t *and* n *that produces a matrix* W *of size* $(|t| - n + 1)$ *by* n *that contains all sub sequences of size* n *of time series* t*. Each line in* W *is a sub sequence of* t *of size* n*. Given* $W = sw(t, n)$, $\forall w_i \in W$, $w_i = t^{i,n}$.

Definition 4. *Let* $q =< q_1, q_2, \ldots, q_n >$ *and* $t =< t_1, t_2, \ldots, t_m >$ *be two time series, such that* $|q| = n$, $|t| = m$, *and* $m > n$. q *is* ***included*** *in* t $(q \prec t)$ *iff* $\exists\, w_i \in W, W = sw(t, n) \mid q = w_i$.

Motif in Time Series

Definition 5. *Given two time series q and t, q is a* ***motif*** *[4] with support σ, iff q is included in t at least σ times. Formally, given time series q and t such that* $W = sw(t, |q|)$, $motif(q, t, \sigma) \leftrightarrow \exists\, R \subseteq W$, *such that* $\forall w_i \in R,\ w_i = q \wedge |R| \geq \sigma$.



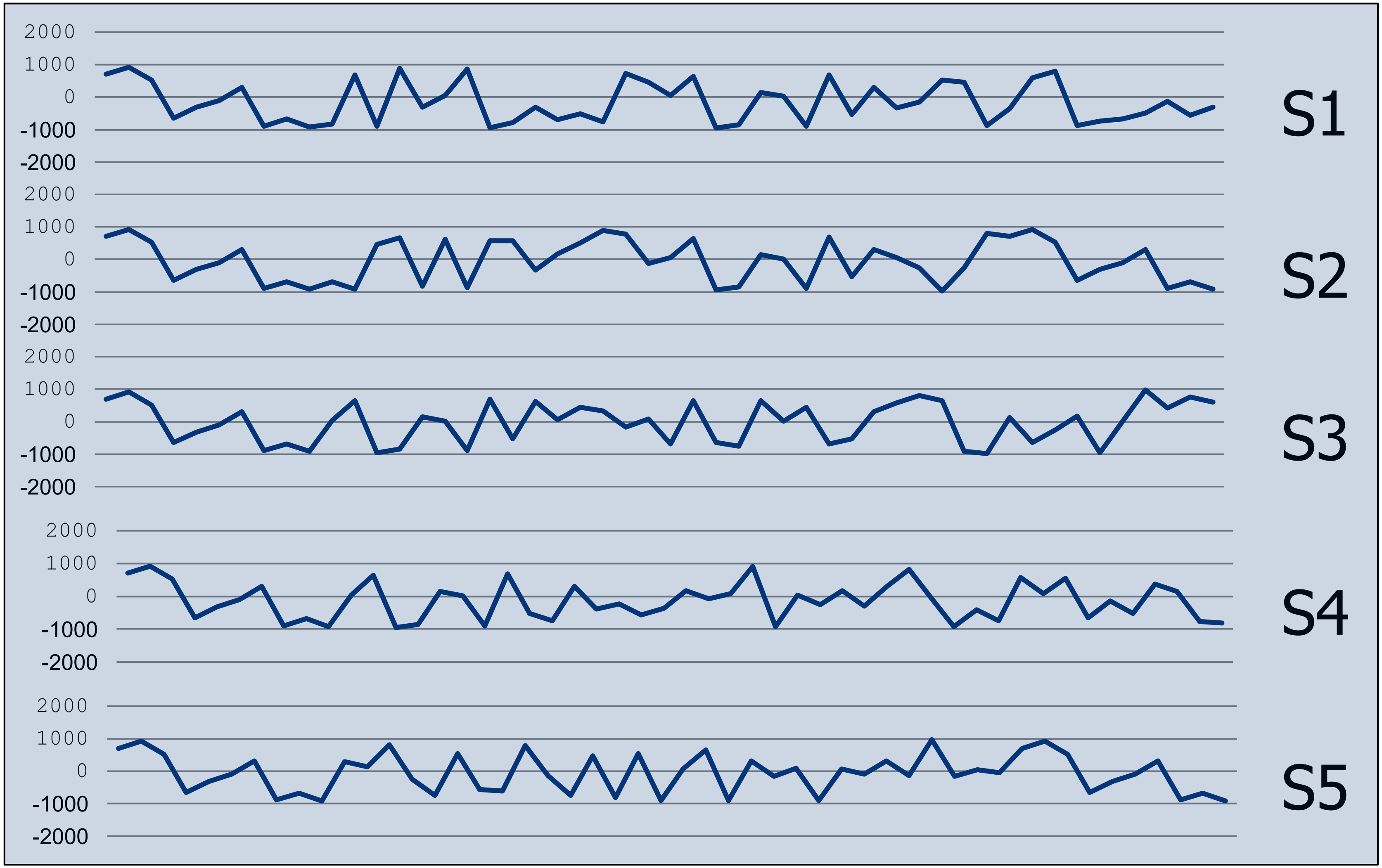
Spatial-Time Series

Definition 6. *A **spatial-time series** s is an abstraction for a time series with an associated position in space. Formally, a spatial-time series s is a time series composed of coordinates x and y and time series* $t =< t_1, t_2, \ldots, t_m >$. *s.x and s.y are coordinates of s, and s.t is time series for s.*

Definition 7. *A **spatial-time series dataset** (for short, dataset) S is a set of spatial time series* $\{s_z\}$. *We define* $t_{max}(S)$ *as the maximum number of observations for all spatial series* s_z *inside dataset S. Formally,* $t_{max}(S) = max(\{|s_z.t|\}), \forall s_z \in S$.

Assuming that spatial-time series inside a dataset are uniformly distributed in space, it is possible to define a bounding-box for them $BB(x_{min}(S), y_{min}(S), x_{max}(S), y_{max}(S))$. Through simple coordinates translation, we can assume that $x_{min}(S) = y_{min}(S) = 0$ and $x_{max}(S) = sw$ and $y_{max}(S) = sh$. In this way, we have a bounding-box $BB(0, 0, sw, sh)$ with size $sw \times sh$ for dataset S.

Spatial-Time Series Example

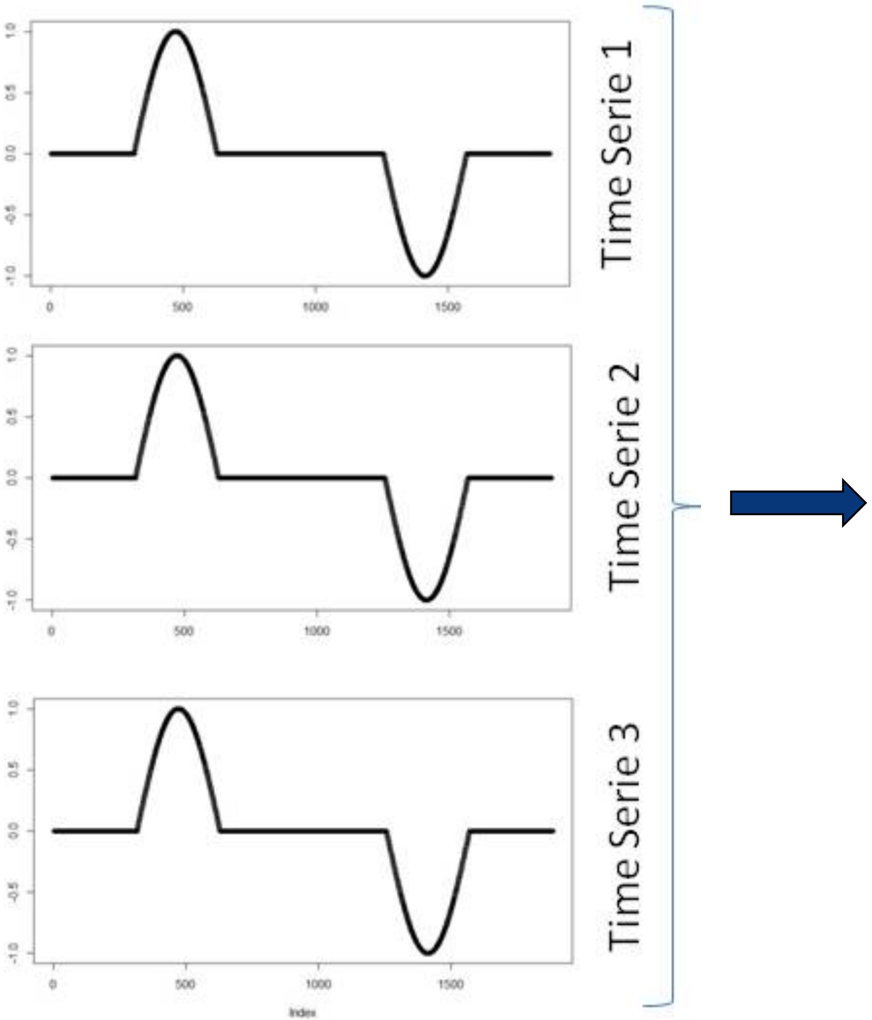


Bounding Box

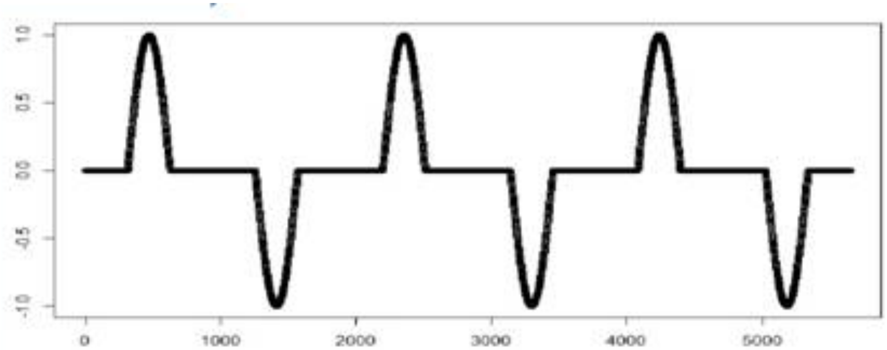
Spatial-Time Motif

Definition 10. *Let σ and κ be two support values such that $\sigma \geq \kappa$. A time series q is a **spatial-time motif** in a parallelepiped $pt_{i,j}^{k,n} \in S$ iff q is included at least σ times in $pt_{i,j}^{k,n}$ and $\forall s_z.t^{kn,n} \in \overline{pt}_{i,j}^{k,n}$, q is included in $s_z.t^{kn,n}$, $\overline{pt}_{i,j}^{k,n} \subseteq pt_{i,j}^{k,n}$ and $|\overline{pt}_{i,j}^{k,n}| \geq \kappa$.*

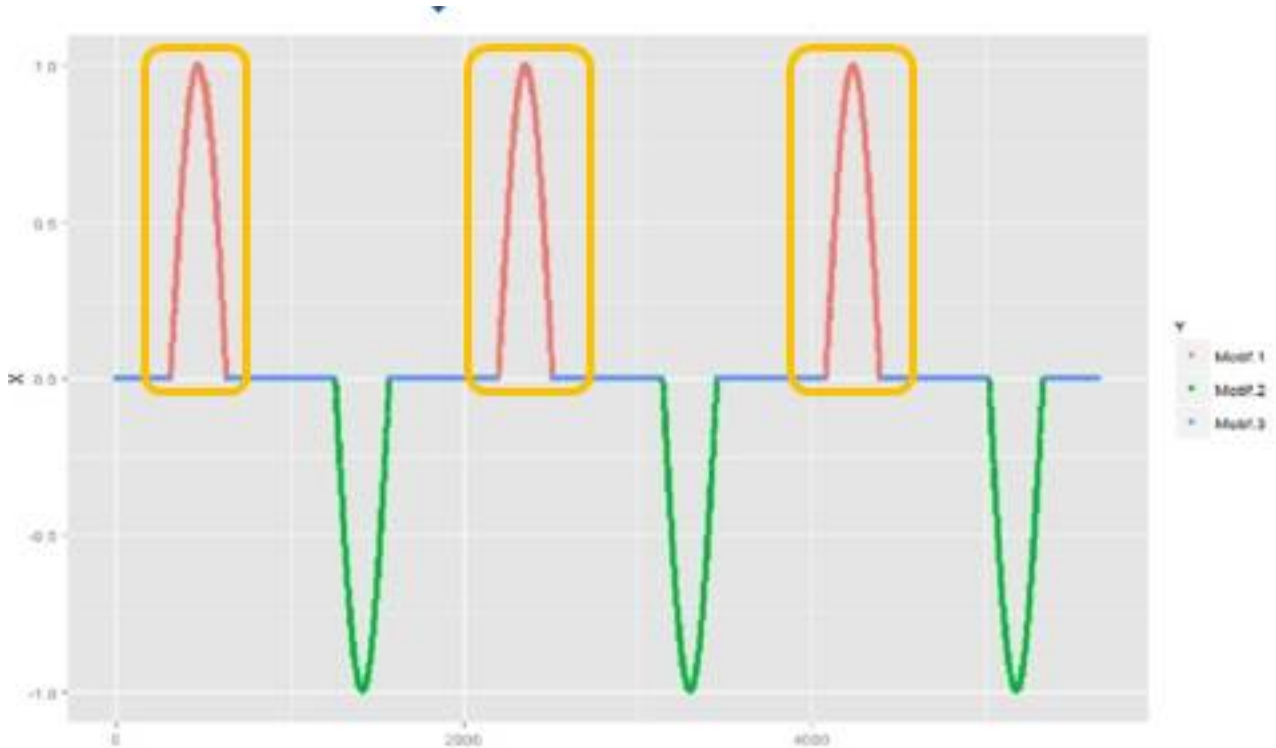
Combined Spatial-Time Series



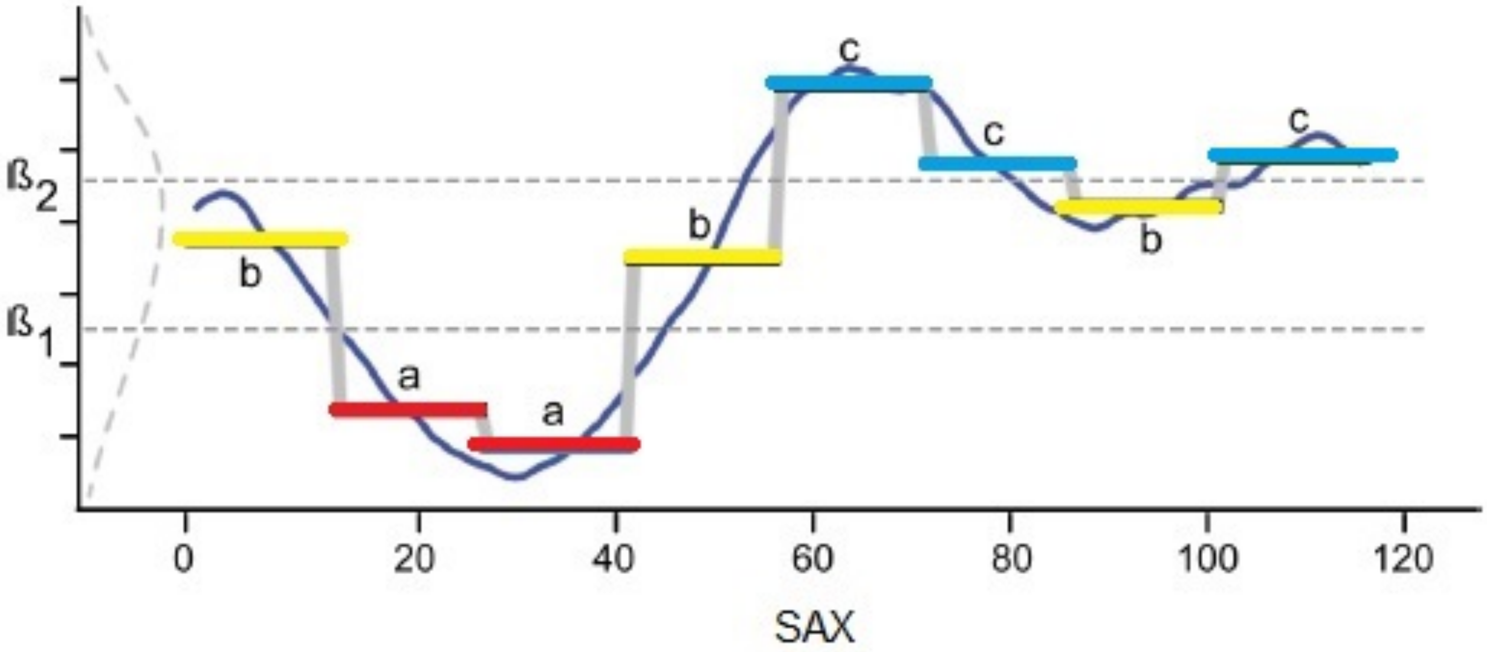
Combined time series



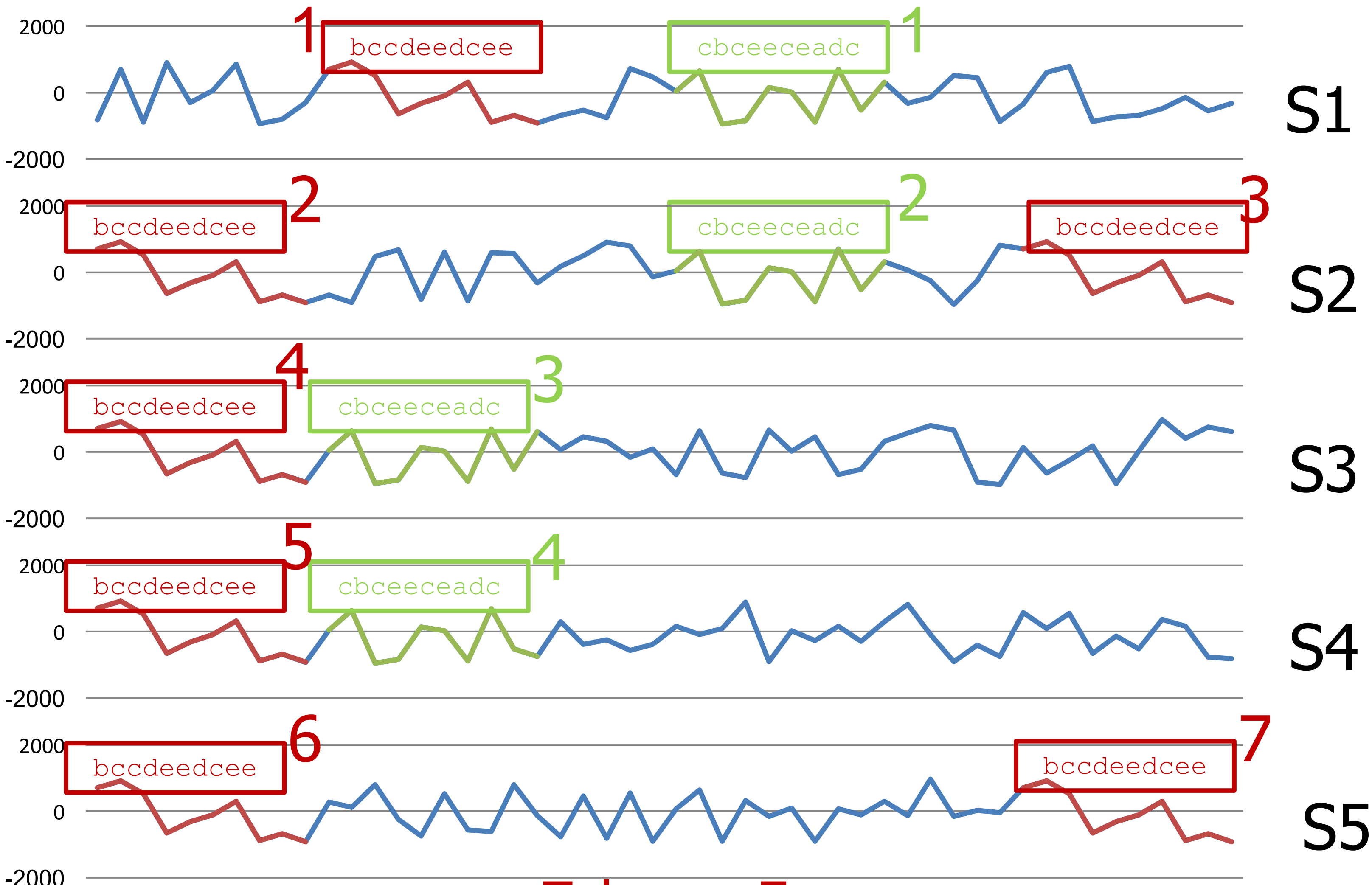
Motifs in Combined Spatial-Time Series



SAX Indexing



Identified Motifs in Original Spatial-Time Series



$\sigma = 7 \mid \kappa = 5$

$\sigma = 4 \mid \kappa = 4$

Spatial-Time Motif Ranking

- Rank identified spatial-time motifs

Motif	Word	σ	κ	Spatial-Time Motif
Motif 1	bccdeedcee	7	5	Yes
Motif 2	cbceeceadc	4	4	No

σ: total motif occurrences in block

κ: number of series that occurs the identified motif

Restriction Parameters:

$\sigma \geq 5$

$\kappa \geq 3$

Algorithm

function STMotif(b, sw, w, a, bs, bt)
    $b_i \leftarrow partition(b, bs, bt)$
    for each $b_i \in b$ **do**
        $t \leftarrow combine(b_i)$
        $CSTM \leftarrow identify(t)$
        $STM \leftarrow STM \cup constraintST(CSTM)$
    end for
    $rankSTM = aggregate(STM)$
    return $rankSTM$
end function

Jornada de Ciência de Dados

LNCC

Motifs identifications in Spatial-Time Series

CEFET/RJ

Eduardo Ogasawara

http://eic.cefet-rj.br/~eogasawara