# Multi-level Association Rules and Directed Graphs for the Lagrangian Analysis of the Mediterranean Ocean Forecasting System (MFS)

B. Petelin (1), V. Malacic (1), A. Malej (1), M. Kukar (2), and I. Kononenko (2)

(1) National Institute of Biology, Marine Biology Station, Ljubljana, Slovenia (petelin@mbss.org)
(2) University of Ljubljana, Faculty of computer and information science, Ljubljana, Slovenia

NATIONAL INSTITUTE OF BIOLOGY

This poster participates in OSP

Outstanding Student Poster Contest

## 1) Introduction

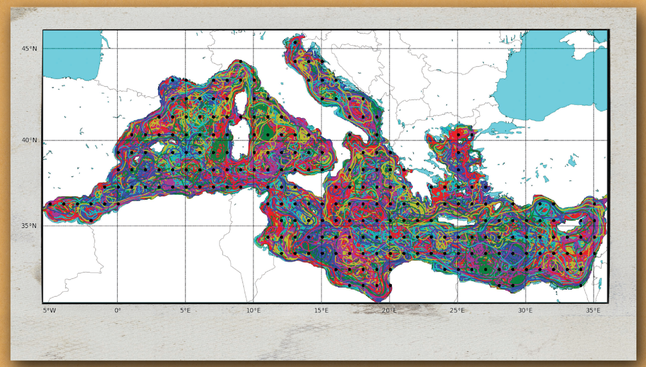
Lagrangian particle tracking is one of basic methods for modeling the transport of water parcels and the dispersion of biological species. To avoid the difficulty of the analysis of a large number of trajectories and its visual presentation, we propose a new methodology which includes data mining and different visualization techniques, namely, association rules and directed graphs.

Oceanographic data exhibit strong spatial and temporal dependencies, so we extend the basic association rules discovery to spatial and temporal association rules mining. In addition, we suggest a novel visualization method which uses multi-level directed graphs with different levels of space and time granularity. Moreover, we can intertwine the knowledge from various disciplines related to oceanography, e.g. marine ecology.

The proposed methodology consists of the following steps. First, we generate a vast number of Lagrangian trajectories in the domain of the numerical model, or use many observations of various floats, which must cover the area of interest. Second, we derive the spatial association rules resulting on these trajectories. Then we construct a multilevel directed graph based on these rules and finally include additional attributes of various types, which represent the oceanographic knowledge.

## 2) Methodology

### 2a) Lagrangian particle tracking

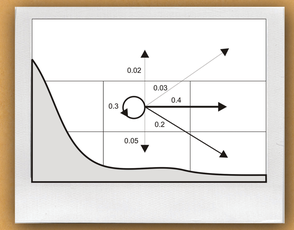



- velocity fields of ocean currents from the numerical model of Mediterranean Ocean Forecasting System - MFS [2] for the period 1999-2011
- Lagrangian trajectories calculated using Ariane (http://stockage.univ-brest.fr/~grima/Ariane/).
- starting positions (black dots) in a depth from 0 to 40 m at the beginning of each month
- total number of trajectories was 96280 with the length of 365 days each.

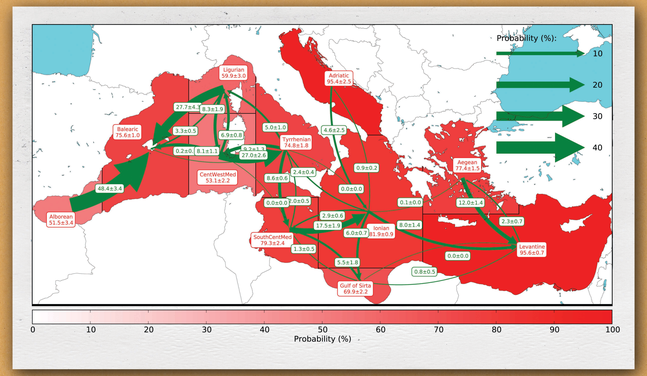
### 2b) Spatial Association Rules

The problem can be formulated as follows: "If the particle in the current time period is located in the area A of the given spatial domain, then the same particle in the next time period will be located in the area B with a certain probability".

First we discretize the given spatial domain in smaller areas and choose an appropriate time interval. Then we determine the probability of displacement of particles from one area to another using the association rules [1]. The measure of confidence of these rules denotes that probability.




### 2c) Multi-level directed graphs

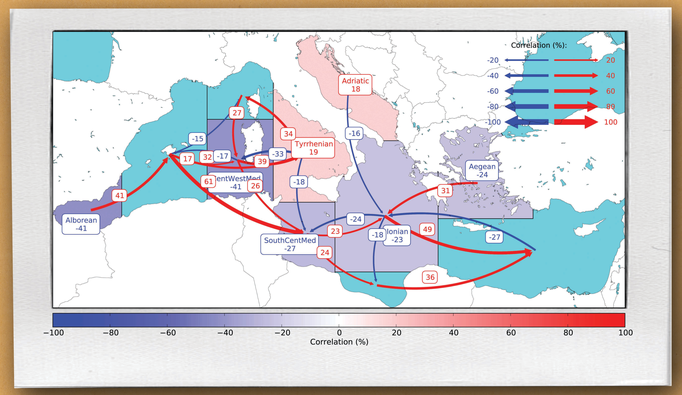


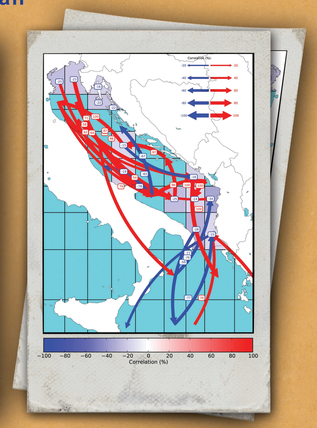

Spatial association rules in the whole domain of study form a multi-level directed graph with the sea areas as nodes. The weights of the edges denote the probability of transition of particles between areas in a given time interval.

Above graph was obtained by Fourier analysis of time series of edge weights in the period of 1999-2011. The values indicate the mean transition probability and the amplitude of fluctuations (period of 12 months) around this mean, respectively.

## 3) Experiments and Results

### 3a) Correlation with the wind power transmitted to the ocean

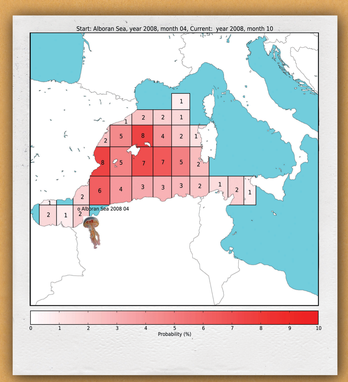


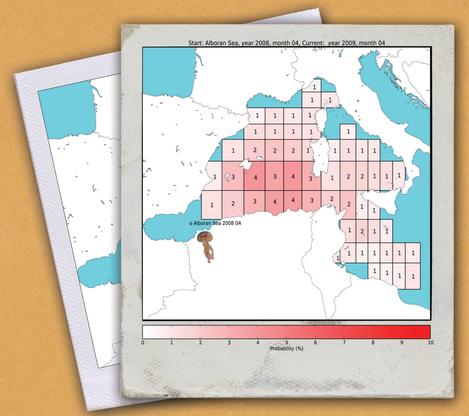



We calculated the correlations between the wind power transmitted to the ocean in the sea areas and the probabilities of transitions of the particles between these areas within the interval of one month.

Higher average wind power in an area results in a higher probability of moving particles to the neighboring areas (left figure). The exceptions are Adriatic Sea and Tyrrhenian Sea, which behave more like closed areas where the wind is causing the recirculation of surface particles within these areas. The movements of particles inside Adriatic Sea are highly correlated with the wind power in contrast to particles which cross the boundary between the Adriatic and Ionian Sea (right figure).

### 3b) The long term distribution of biological species




Each digraph from the period of 1999-2011 covers one month and has an associated transition matrix of discrete time non-homogenous Markov process. Graph nodes (sea areas) represent the states and the elements of matrix denote the probabilities of movements of particles (biological species) between the areas. We can obtain the long term probabilities of movements by multiplying the matrices sequentially in time.

The figures above show the hypothetical distribution (in %) of Pelagia noctiluca (Scyphozoa), which was observed in the Alboran Sea in April 2008, after six (left) and 12 months (right).

## 4) Conclusions

Association rules and multi-level directed graphs form a novel and promising methodology in the spatial data-mining in oceanography:

1) Takes into account the integration of various oceanographic data and applications on these data
2) Implies the use of graph data mining [3] on resulting graphs

## 5) References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile. pp. 487-499. Morgan Kaufmann (1994)

2. Tonani, M., Pinardi, N., Dobricic, S., Pujol, I., Fratianni, C.: A high-resolution free-surface model of the Mediterranean
3. Washio et al., "State of the art of graph-based data mining," SIGKDD Explor. Newsl. vol. 5, pp. 59-68., Jul. 2003