*International Journal of Geographical Information Science*
Vol. 00, No. 00, July 2004, 1–15

# Highlighting Space-Time Patterns: Effective Visual Encodings for Interactive Decision Making

Mike Sips, Jörn Schneidewind, Daniel A. Keim

*(August 2006)*

Dynamic space-time pattern and potential interesting events in space and time have in practice a much higher complexity than available visual encodings can handle. Challenges arise because newly available data sets are from heterogeneous data sources and often have millions of records or even far more. The data is in general defined over a geo-spatial context with some associated attributes such as numerical statistical parameters, text, images, GPS-data, network logs etc. The analysis involves a wide variety of objects, with varying attributes in time; it is often hard to see what is emphasized. This paper focuses on developing a geo-visual analytic approach to highlight patterns that are defined over a geo-spatial context and change in time. The aim is to discover interesting and unknown complex patterns, and present them in an easy-to-understand visual encoding. Highlighting expressly supports human interpretation, analytical reasoning, and decision-making and it applies human perceptual abilities to the analysis of real-world data sets.

## 1 Introduction

In many application scenarios data is collected and referenced by it's geo-spatial location at a certain point in time (time-stamp). The analysis of such data sets is an important task, since for decision makers, analysts or emergency response teams it is often essential to rapidly extract relevant information from the flood of data.

In emergency management, for example, GPS-navigation plays an important role to provide effective coordination for various organizations to enable more efficient emergency assistant. Therefore, the GPS information from selected vehicles is collected which results in large and complex databases. A research challenge on the one hand is to find efficient methods to handle such massive information flows and on the other hand to provide visualizations that fuse this information together. The aim is to allow analysts to identify space and time patterns and to provide efficient visual awareness and coordinated help in emergency cases.

University of Konstanz, Germany, e-mail: sips@dinfovis.uni-konstanz.de
University of Konstanz, Germany, e-mail: schneide@infovis.uni-konstanz.de
University of Konstanz, Germany, e-mail: keim@infovis.uni-konstanz.de

Another example is credit card fraud protection where the geographic information of credit card transactions at certain points in time can help to prevent fraud. Credit card companies may verify customer authorizations for those transactions which (a) show a great difference in the distance of transactions in a very short time or (b) transactions that have been processed in "high risk" countries within a short time frame (countries that are well known sources for credit card fraud).
In the described examples the collected data typically results in large and heterogeneous data sets with geo-spatial information, time stamps, pictures, text and other information. The visual analysis of these massive volumes of data is a challenge for existing visualization techniques. Space-time-patterns can be seen as a series of multivariate profiles. The difficulty is to provide effective visual awareness in multi-dimensional heterogeneous data spaces that represents the available recourses or products as well as different kinds of events and alerts. Effective visual awareness is based on the perception of objects in an environment with a volume of space and time, the comprehension of their meaning, and the projection of their status in the near future. Interactive decision therefore includes three levels:

- Highlighting space-time patterns
- Comprehension and their changes over time
- Projection of their future status

We focus on the combination of automated data analysis methods and smart visual encodings. to face this problem as proposed in the context of Visual Analytics (see Thomas and K.A.Cook (2005)). The data analyst typically specifies some parameters to restrict the search space, then automated data analysis is performed by an algorithm and finally the detected patterns are presented to the data analyst. Many existing geo-visualization techniques analyze geospatial information along 1 or 2 single dimensions, typically time or space. But due to the high number of patterns generated by an automatic algorithm, it is almost impossible to interpret and to evaluate each of them. Therefore an effective combination with effective visual encodings to support interactive decision making is needed. It is important to show the state of information and their interconnectedness at a certain point in time for an efficient emergency management handling, like a large industrial accident, or a sudden epidemic growth of diseases. Therefore an effective combination with effective visual encodings to support interactive decision making is needed. Highlighting expressly supports human interpretation, analytical reasoning, and decision-making and it applies human perceptual abilities to the analysis of real-world data sets.
This work provides an approach that visually analyzes space-time patterns and their changes over time to take these demands into account. We focus on data that is stored in Data Warehouses, a typical way to organize and inte-

grate large data volumes from multiple data sources. Data Warehouses provide efficient access to the data sources via Online Analytical Processing (OLAP). OLAP allows an effective drill-down / roll-up functionality to the data based on data cubes. In real world scenarios, the data cube is usually defined by three components: place (geo-spatial context in 2-D or 3-D), time (with a continuous direction) and a set of attributes (see the work proposed by Stolte et al. (2002) for more details with focus on relational databases). More general information on data cubes with focus on geo-spatial data is presented in Chen et al. (2005), Chen et al. (2006) and Guo et al. (in press).
We demonstrate our technique based on an application to company sales data, but the technique is of course adaptable to other scenarios such as health and disease analysis, networks and fraud detection. Our data set contains sales data from an Italian company over 5 years. Each cell in our cube is defined by a spatial object (Italian Regions like Lombardi, Emilia-Romagna), a sales period (year, month, day), and customer- and product attributes.

## 2 Highlighting Space-Time Patterns – An interactive approach

Our approach provides a suite of easy-to-understand visual encodings that highlight changes of geo-patterns over time stored in a Data Warehouse environment. It allows the interactive exploration of the data by providing slices of space-time-attributes. A slice of space-time attribute includes the place and time, and some potential interesting statistical parameter.

### 2.1 *Access to the data*

A slice of space-time-attributes is the result of an SQL-Query to the Data Warehouse and is often an aggregated view of a more detailed dataset. We aggregate the data by grouping it into different geo-spatial objects at different geographic scales such as region, province, etc. We compute individual views to the data for each geo-spatial object, because it enables an effective highlighting of space-time patterns.

### 2.2 *Interface*

Our visual data exploration follows a three step process: analyze first, zoom and filter, further analysis and details on demand (called Visual Analytics Mantra proposed by Keim (2005)). The interface allows the data analyst to select a slice of space-time-attributes. The slice is visualized in the main window to the user. The right upper sub window shows some additional statistical measures about the statistical parameters, the events, alerts and patterns de-

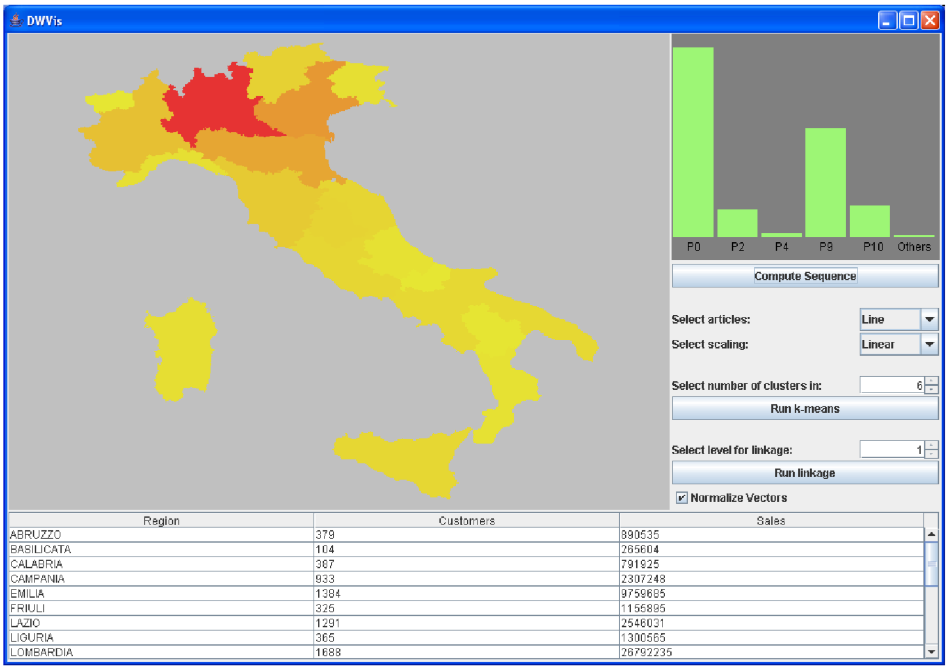



Figure 1. *Warehouse Interface* – Selection of space-time-attribute slices from the Data Warehouse based on SQL-queries. The data is aggregated by grouping it into different geo-spatial objects at different geographic scales

fined within each individual time step. Figure 1 shows the interface working on our example dataset.

In addition to the core visualization technique, it is necessary to provide interaction techniques to enable effective data exploration. Our interface allows the data analyst to directly create and select different views to the data. The data analyst can then guide the highlighting process, where the highlighting algorithm is based on a time separation and a geo-space aware clustering at each time step followed by a pair wise comparison of trees. The trees are the result of the hierarchical geospace-aware clustering for each separate time moments in the data (see section 3 for details). Our framework provides navigation methods to explore pattern defined in the different tree levels and to judge the interestingness of patterns. The patterns can be explored at their coarsest level (root of the tree is the aggregated view of all pattern) to individual details. The navigation can be performed in both ways bottom-up and top-down.

### 2.3 *Highlighting Space-Time Pattern*

The objective of this research is to develop new methods and techniques to discover the spatial inter-connectedness of information in time over a

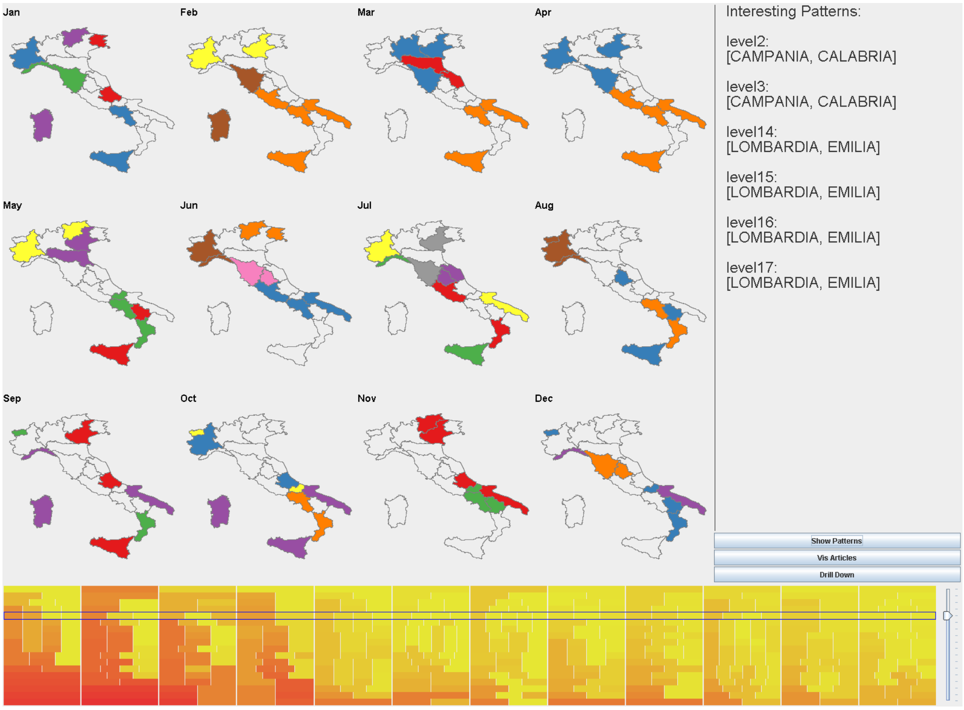



Figure 2. *Analysis Interface* – Highlighting the evolution of geo-spatial pattern in time. The regions Sicilia, Puglia, Campania and Lazio are grouped together in a common cluster *(orange cluster)* that is defined from February until April (some little changes in March).

geo-spatial context.

The complexity of the visual analysis boosts with the time dimension, and the idea is to appropriately combine smart visualizations with automatic analysis methods. We visualize the inter-connectedness of information in time over a geo-spatial context by highlighting and tracking events, alerts and pattern at every single time step. The visualization is tightly coupled with the highlighting algorithm to achieve an overview of statistical parameters at individual places and times. The ordering of the different time steps goes from left to right (intuitive direction of time) with line breaks at the edges of the available screen. The individual patterns are uniquely encoded by color. If an event, alert or pattern occurs over a certain time frame the color remains constant (to enable an efficient visual awareness).

The overview visualization is dynamically coupled with the highlighting algorithm to show space-time patterns in detail.

*Algorithm 1* **Procedure** Visualize Highlighting(cluster A, cluster B, time t, time t+1)

// hl contains all at time step $t$ active clusters
hashlist hl();
// $\lambda$ is the current color of all members in hl
array $\lambda$();

**// Step 1: Test if (A,B) are already clusters in time sequence** $t-1$
$id = hl\{(A,B)\}$;
**If** $id \neq undef$ **do**

    **color** cl = $\lambda[id]$;

**else do**

    **color** cl = generate_color(); // $cl \in RGB \setminus \lambda$
    $hl\{(A,B)\} = (A,B)$;

**endif**

**// Step 2: Visualize** $(A,B)$ **in time steps** $t$ **and** $t+1$
**Image** Vis1 = vis_array[t];
**Image** Vis2 = vis_array[t+1];
**Paint**(Vis1, A, cl);
**Paint**(Vis2, B, cl);

Dynamically coupled means, if events, alerts or patterns are occurring in neighboring spatial regions than the highlighting algorithm continuous recursively into the next lower level (e.g. from state into county level) to visualize more details. This is important to understand the evolution of spatial pattern in time. In this situation we highlight the evolution of such events, alerts and patterns showing the current status over individual time steps at the lowest level on which they occur. This approach is a synthesis of the degree of interest proposed by Furnas (1988) and multi-resolution approaches by Keim and Schneidewind (2005), with extension of some ideas from scalability approaches proposed by Sips et al. (2006).
An important issue in our framework is to determine the probability of events and their projection to their near future, for example in case of emergency management scenarios the probability that a chemical fire expands to other chemical facilities in the factory.

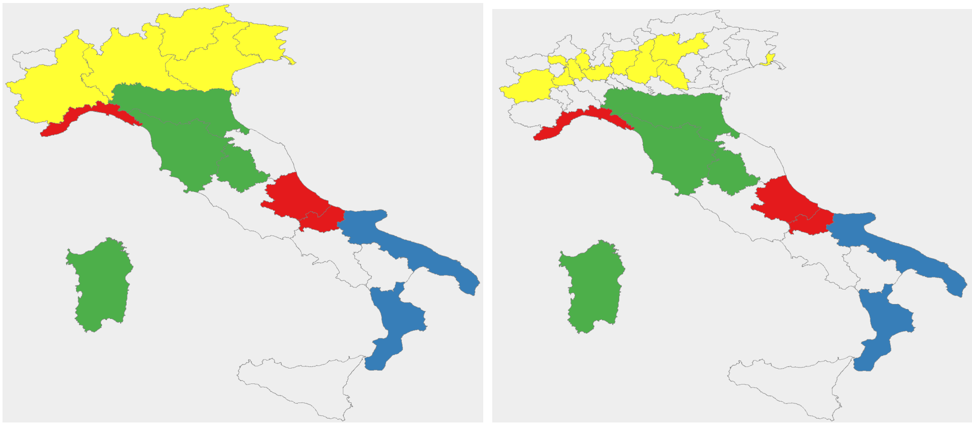


Figure 3. *Enhance the interesting cluster automatically* – if events, alerts or patterns are occur in neighboring spatial objects than the highlighting algorithm continuous recursively into the next lower level (from state into county level) to visualize more details to enable visual awareness

## 3 Analysis and Highlighting Algorithm

Now we can give a detailed definition of the proposed algorithm. The overall approach is to map the space-time slices into a feature space, then highlight clusters and their changes over time using an adapted single linkage algorithm (see Jain and Dubes (1988) and Han and Kamber (2001) for details) with some further extensions.

### 3.1 *Step 1: Feature Space Transformation*

We map in a first step the chosen slices of space-time-attributes into a feature space defined by two discrete dimensions (place and time) and continuous dimensions (statistical parameters). Because real-world data set distributions are often non-uniform, the data points form readily identifiable point clouds.

### 3.2 *Step 2: Identify Geo-Pattern*

Next we use our extended single-linkage schema to group geo-spatial objects with similar statistical parameter at each time step together. Basically, the extended single-linkage schema includes three basic steps:

(1) Initialize each data point as it own cluster
(2) Merge the most two closest cluster into a new parent cluster and repeat this step until all clusters are merged in one root cluster
(3) Cut off the root and leafs in the resulting tree

Note, the root and leafs are just the whole database or elementary information and do not represents a useful highlighting.
We measure the similarity of two clusters by taking the neighborhood of the geo-spatial object into account. More precisely, we use a weighted similarity measure based on the distance between clusters in the geographic space and the similarity in the statistical parameters. The weight $\omega$ corresponds to the distance between the centers of gravity of the clusters $\| \cdot \|$.
Let $A$ and $B$ two clusters and $\|\cdot\|$ the measure for the distance in the geographic space and $|\cdot|$ the measure for the similarity in the statistical parameter. Let $C_A$ and $C_B$ the center of gravity of both clusters and $S(A), S(B)$ the associated statistical parameters. We set $\omega = \|C_A, C_B\|$ and then compute the similarity $dist(A, B)$ of $A$ and $B$ in the following way:

$$dist(A, B) = \omega \cdot \big(1 + |S(A), S(B)|\big)$$

where we use the Euclidean distance as distance measures in both cases.

### 3.3 *Step 3: Highlight Changes in Time*

The geospace-aware clustering follows a pair wise comparison of trees. The trees are the result from the previous step. Let $H(t)$ and $H(t+1)$ two trees computed at the time steps $t$ and $t+1$. Let $l$ the current level in the exploration process and $\big\{H(t)[l]\big\}$, $\big\{H(t+1)[l]\big\}$ the set of all nodes defined on level $l$. The pair wise comparison follows the two basic steps:

(1) Compute the pair wise set intersection of all nodes at level $t$ with

$$\forall v \in \big\{H(t)[l]\big\}, w \in \big\{H(t+1)[l]\big\} : v \cap w = \big\{u : u \in v \wedge u \in w\big\}$$

(2) If $\big|v \cap w\big| > \frac{|u|}{\lambda \cdot (|w| + |v|)}$ than compute detail resolution and output to the user

Starting in the top parent level we compute the pair wise set intersection for all nodes for each level $l$. If the intersection set of two nodes contains more than $\lambda$ percent common elements than we highlight this cluster in both time steps. The threshold $\lambda$ is given by the user; the default value is 2.

### 3.4 *Step 4: Compute Detail Resolution*

It is crucial to automatically enhance important features in the clusters. The idea is to take the raw clusters and compute the detail resolution in order to be able to quickly detect the important aspects. The cluster detail depends

on the one hand on the number of cluster members and on the other hand on the connectivity of the geo-spatial objects.
Let $O = \{o_1, \ldots, o_n\}$ a set of spatial objects. Let $r$ the radius of the query ball $Q$. The radius $r$ is given by the user and can be adjust to different application scenarios. We define $Q(o_i, r) = \{o : \|center(o), center(o_i)\| \leq r\}$ with $center(o)$, $center(o_i)$ denotes the center of gravity of the geo-spatial objects $o, o_i$. Then the neighborhood function is defined as follows:

$$nbh(o_i, o_j) = \begin{cases} true & o_i \in Q(o_j, r) \wedge o_j \in Q(o_i, r) \\ false & else \end{cases}$$

The analysis and highlighting algorithm goes recursively into the next lower geographic scale in which the the geo-spatial objects $o \in v \cup w$ are embedded if the number of cluster members $|v \cup w|$ is greater than a given threshold $\beta$ and all geo-spatial objects $o$ are connected $\bigwedge_{o_i \in v, o_j \in w} nbh(o_i, o_j) = true$. More precisely, the input for the next recursive iteration is the result of a SQL-Query to data that is a more detailed view of the data.

### 3.5 *Step 5: Presentation*

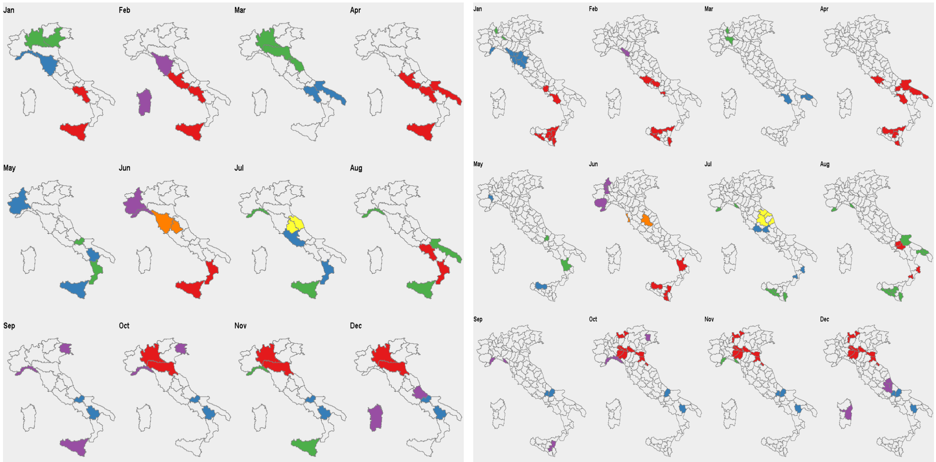
The final step is to present the result to the data analyst. Our interactive approach allows a data analyst to adjust visualizations interactively to satisfy data exploration objectives (see section 2).

## 4 Support Interactive Decision Support

Our interactive approach supports a combination of automatic data analysis and visualization that addresses the following three levels in the process of interactive decision making.

### 4.1 *Tracking Space-Time Pattern*

The goal is to visualize objects, resources and their activities within a combined temporal and geo-spatial display. Figure 4(a) and 4(b) show the interconnectedness of the sales data example. In practice, the data analyst displays at first the overview highlighting and then identifies some potential interesting pattern for investigating further details. The aim is to identify impact factors of the different patterns.



(a) Highlighting based on regions (Overview)

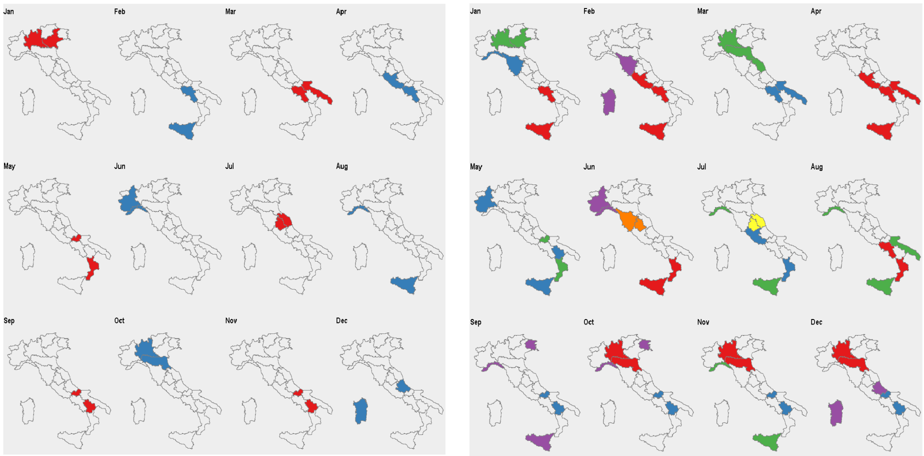
(b) Top-Down Highlighting into provinces (Detail)

Figure 4. *Tracking Space-Time Pattern* – Highlighting Space-Time Pattern at different geographic scales. The sales pattern of the first product type starts (*red cluster*) in January and ends in April and is located in the south of Italy (Sicilia, Campania, Lazio). It shifts to the north regions Lombardi and Emilia-Romagna in the last three months.

**4.1.1** ***Customer Sales Analysis.*** Figure 4(a) shows the global spread of two major customer sales patterns. The sales pattern of the first product type (*red cluster*) starts in January and ends in April and is located in the south of Italy (Sicilia, Campania, Lazio). It shifts to the north regions Lombardi and Emilia-Romagna (see figure 4(b)) in the last three months. We can identify two important sales periods of this product type and it may allow a data analyst to adjust sales policies. We can see that the *green cluster* sales pattern (second product type) has a vice versa behavior.

### 4.2 *Comprehension of Space-Time Pattern*

The aim is to improve visual awareness that enable insight into the patterns. The basic idea is to show at first the clusters on the top level. Here, the data analyst gets an overview about the clusters and their distribution over the different regions. An important task in interactive decision making is (a) to explore the factors that create clusters and (b) to explore issues explaining their interestingness. Our interactive approach (see figure 5) allows the data analyst to explore interestingness of pattern by navigating through the levels of the hierarchical clustering.




(a) Lowest Hierarchy Level

(b) Highest Hierarchy Level

Figure 5. *Comprehension of Space-Time Pattern* – explore the interestingness of patterns by navigating through the levels of the hierarchical clustering. The figure shows the spread of major customer sales patterns at lowest level and the highest level. The data analyst can navigating through the clustering process to understand the impact factors of different sales pattern.

**4.2.1** ***Customer Sales Analysis.*** Figure 5(a) and 5(b) show the spread of major customer sales patterns at the lowest and the highest level. We can see that sales pattern of the first product type (*red cluster*) expands into more sub clusters in April and May and it shifts more to the north regions Lombardi and Emilia-Romagna in the last three months (see figure 5(b)).

### 4.3 *Projection of their Future Status*

For a successful decision making it is crucial to understand all dimensions in space and time. Figure 6 shows the sales periods of 5 important product types. We can find three important sales periods in Italy. Each sales period consists of at least four months; in each a totally different set of articles is sold.

**4.3.1** ***Customer Sales Analysis.*** We extracted all products from our database that have a continuous sales period of at least 3 months. More precisely, we extracted all clusters that are inter-connected in at least 3 different tree nodes (see section 3.2 for more details) with more than 50 percent of constant cluster members. This allows us to project these sales periods to their near future because it is very likely that these periods can be identified every year.

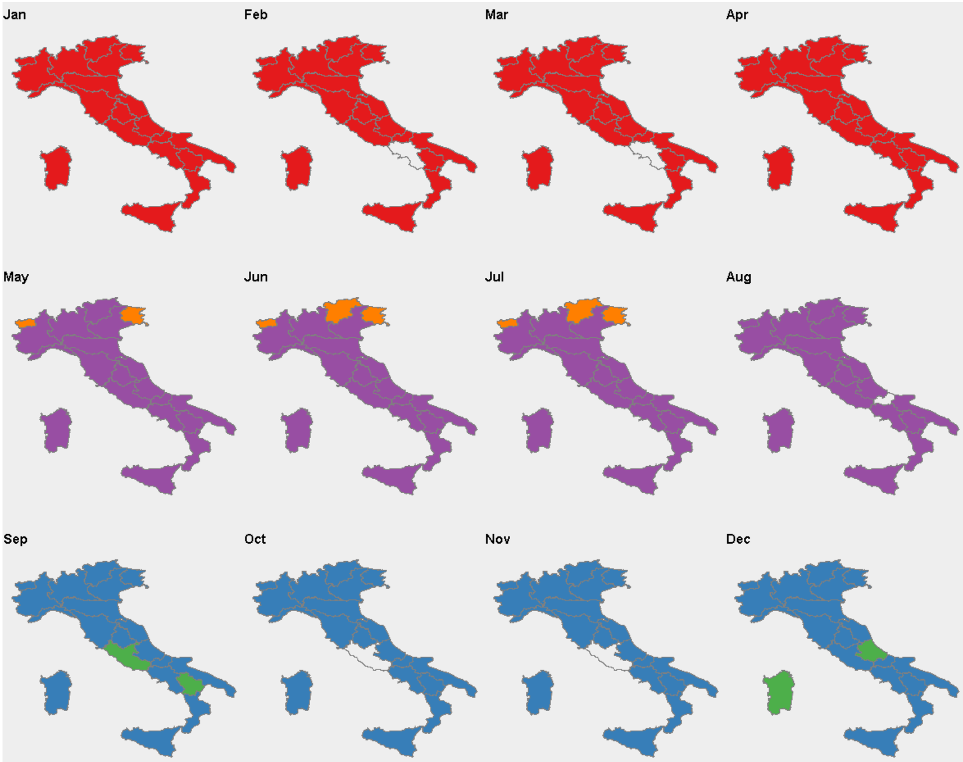



Figure 6. *Projection to their future status* – Show the spread of sales slots of 5 important product types over a year that are inter-connected in at least 3 different tree nodes (see section 3.2 for more details) with more than 50 percent of constant cluster members. We can detect three different sales periods (*red cluster*), (*blue*) and (*violet*) product types and it is very likely that these periods can be identified every year.

We can observe that the *red cluster* has a strong sales period in the first three months of the year in entire Italy. After the first three months there is no market anymore for this product type. The same can be observed for the (*blue*) and (*violet*) product types. Additionally, we can observe two small interesting sales patterns. The (*orange*) cluster is only active in the north of Italy from May until July. We know that these months are the major tourist season in the north. This product type could be related to the tourists. The green one is defined from September until December. We can notice that the big blue cluster overlaps the green at some time steps.

## 5 Health and Disease Analysis

Government agencies publish a wealth of statistical information that data analysts can apply to key problems in public health and safety. Service providers,

such as managed care organizations, hospitals, nonprofit corporations, schools, faith organizations, and businesses are reliant on the rapid identification of epidemics or diseases. Essential public health issues are to monitor health status, to identify local health problems, and to diagnose and investigate local health hazards in the community.
Figure 4(a) and 4(b) could also be used to show the growth of the outbreak of a dangerous epidemic.

## 6 Related Work

Exploring and analyzing large spatio-temporal data sets is an challenging task because of data complexity and scalability issues. The work in Livnat et al. (2005) presents a novel paradigm for situational awareness based on a generalization of the what, when, where attributes.
An interesting approach is proposed by Chen et al. (2005) Chen et al. (2006) Guo et al. (in press). The authors propose a novel inquiry system for exploring space-time pattern. They construct overviews about the data by using both computational (self-organizing maps) and visual methods (reorderable matrix and map matrix).
Successful analysis of space-time-attribute data requires the tight integration of the user into the exploration process. Seo and Shneiderman (2005), Seo and Shneiderman (2006) propose a framework to enable interactively mining for multi-variate pattern. Some efforts have been made in visually mining spatio-temporal patterns with focus on spatial distribution of temporal behavior Andrienko and Andrienko (2005).

## 7 Conclusions

This research studies how we can adequately analyze large and heterogeneous data sets from multiple sources such as sales data to enable interactive decision making. Interesting questions about space-time pattern are:
How to adequately analyze changing events in time over a geo-spatial context? A great challenge is to develop methods that significantly improve the perception of activities, events and links as they change in time over a geo-spatial context (important for intelligent emergency management).
How to enable intelligent situational awareness and visual emergency management? Many difficulties in intelligent visual emergency management occurs because of the complex dependencies among influencing factors in large heterogeneous data such as population, poverty, environmental safety, clean air, safe water etc. A great challenge in geo-visualization is to successfully develop

methods for visualization events in time that clearly shows the broad picture. Our new interactive approach supports a combination of automatic data analysis and visualization that addresses the following three levels: tracking and comprehension of space-time pattern and their projection to their future status.

**Acknowledgement**

The authors thank the anonymous referees for their comments toward the improvement of this report. The authors thank Jacob Haddick for implementing the framework. Thanks to Stefano Rizzi for providing the data.

## REFERENCES

Andrienko, G. L. and Andrienko, N. V.: 2005, Visual exploration of the spatial distribution of temporal behaviors., *9th International Conference on Information Visualisation, IV 2005, 6-8 July 2005, London, UK*, pp. 799–806.

Chen, J., Guo, D. and MacEachren, A. M.: 2005, Space-time-attribute analysis and visualization of us company data (infovis05 contest first place entry). `http://www.cs.umd.edu/hcil/InfovisRepository/contest-2005/files/`, 2005.

Chen, J., MacEachren, A. M. and Guo, D.: 2006, Visual inquiry toolkit - an integrated approach for exploring and interpreting space-time, multivariate patterns, *AutoCarto 2006, Vancouver, WA, June 26-28, 2006*.

Furnas, G.: 1988, Generalised fisheye views, *ACM SIGCHI '86 Conference on Human Factors in Computing Systems*, ACM.

Guo, D., Chen, J., MacEachren, A. M. and Liao, K.: in press, Visual inquiry system for space-time and multivariate patterns (vis-stamp), *IEEE Transactions on Visualization and Computer Graphics*.

Han, J. and Kamber, M.: 2001, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.

Jain, A. K. and Dubes, R. C.: 1988, *Algorithms for Clustering Data*, Prentice Hall.

Keim, D. A.: 2005, Scaling visual analytics to very large data sets, *Workshop on Visual Analytics, Darmstadt, Germany, 2005*.

Keim, D. A. and Schneidewind, J.: 2005, Scalable visual data exploration of large data sets via multiresolution, *JUCS Special Issue on Visual Data Mining* **11**(11), 1766–1779.

Livnat, Y., Agutter, J., Moon, S. and Foresti, S.: 2005, Visual correlation for situational awareness, *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05)*, IEEE Computer Society, p. 13.

Seo, J. and Shneiderman, B.: 2005, A rank-by-feature framework for interactive exploration of multidimensional data., *Information Visualization* **4**(2), 96–113.

Seo, J. and Shneiderman, B.: 2006, Knowledge discovery in high-dimensional data: Case studies and a user survey for the rank-by-feature framework, *IEEE Transactions on Visualization and Computer Graphics* **12**(3), 311–322.

Sips, M., Schneidewind, J., Keim, D. A. and Schumann, H.: 2006, Scalable pixel-based visual interfaces: Challenges and solutions, *Information Visualization (IV 2006), July 5-7, London, United Kingdom*, IEEE Press.

Stolte, C., Tang, D. and Hanrahan, P.: 2002, Polaris: A system for query, analysis, and visualization of multidimensional relational databases, *IEEE Transactions on Visualization and Computer Graphics* **8**(1), 52–65.

Thomas, J. and K.A.Cook: 2005, Illuminating the path: Research and development agenda for visual analyics, *IEEE*, pp. 79–86.