Temporal-Spatial Association Analysis of Ocean Salinity and Temperature Variations

Yo-Ping Huang and Jung-Shian Jau
Department of Electrical Engineering
National Taipei University of Technology
Taipei 10608, Taiwan
Email: yphuang@ntut.edu.tw

Frode Eika Sandnes
Faculty of Engineering
Oslo University College
Oslo, Norway
Email: frodes@hio.no

ABSTRACT
Ocean circulation plays an important role in global climate change. In an effort to monitor ocean circulation an infrastructure of more than 3,000 buoys have been deployed in the open water to measure ocean salinity and temperature variations. Some of these data are made freely available by Argo. The focus of this study is extracting previously unknown patterns of abnormal ocean salinity and temperature variations from Argo data that can be further applied to predict ocean current variations. First, Argo data are converted to market-basket type data that are used to find temporal-spatial association rules. The discovered rules reveal the associations of abnormal ocean salinity and temperature variations. Next, the discovered temporal and spatial variation patterns are used to predict future ocean salinity and temperature variations surrounding Taiwan. A 3-D visualization model is developed to present a) the interactions between events at different dates, concentric circles and ocean depths, and b) relationships between ocean temperature and salinity variations. The proposed 3-D visualization model help researchers determine whether the ocean temperature and salinity variations occurred in the same water mass in the relative importance of each attribute. Having established the informative relationships among different attributes an early warning system for climate changes can be established such that their impact on property and loss of life is reduced. The discovered association rules are compared with traditional association rules to illustrate their strength in analyzing global climate change.

Keywords
Data mining, temporal-spatial association rules, climate changes, ocean temperature and salinity variations.

1. INTRODUCTION
Global warming is a threat that has received much attention recently. Rising sea levels are leading to lowland being submerged in seawater. The National Snow and Ice Data Center (NSIDC) and NASA found that the Arctic Ocean melts more rapidly than before. Fig. 1 shows how the Arctic ice cap has shrunk. The sea level is currently rising at a rate of 3 mm per year.

Argo is an international research program where autonomous floats have been used to collect temperature and salinity data since 2000. The infrastructure comprises more than 3,000 floats deployed in the oceans to collect data for oceanography research. Each float can collect data at various depths as they ascend from depths of up to 2000 meters to sea surface. The temperature and salinity variation data are collected and transmitted to satellite repeatedly every two weeks [1].

Fig. 1. Environmental changes in the Arctic Ocean from 1979 to 2003. (Source: http://svs.gsfc.nasa.gov/vis/a000000/a003400/a003464/index.html)

The observed ocean temperature and salinity variations data can be used to predict the ocean current variations. In this study, an efficient data mining model is used for discovering previously unknown patterns of abnormal ocean salinity and temperature variations relationships from Argo data. Moreover, a 3-D visualization model is developed to display the interactions among occurrence of events at different dates, in different concentric circles and at different ocean depths. Moreover, the relationships between ocean temperature and salinity variations are visualized. The proposed information theoretic 3-D model can be used to analyze whether the ocean temperature and salinity variations occurred in the same water mass. Next, the 3-D model can be used to determine the relative importance of each attribute in the overall system.

The rest of paper is structured as follows: Section 2 describes the problems at hand and introduces related work. Section 3 describes the proposed algorithm, and section 4 provides experimental evidence. The conclusion and future work are discussed in section 5.

2. PROBLEM DEFINITION
The goal of this study is to discover temporal and spatial variation patterns that subsequently can be used to predict ocean temperature and salinity variations patterns in the waters surrounding Taiwan. Mining salinity and temperature patterns is a difficult task due to the temporal-spatial nature of the data [2], even though there is a vast literature on how to extract spatial and temporal patterns from scientific data. This section discusses the
challenges involved in preprocessing and analyzing the data. Necessary definitions are also provided.

2.1 Transactions in Argo data

The raw Argo data are first transformed into a market-basket data representation to facilitate association rules mining. Association rule mining algorithms assume that a finite set of disjoint transactions are given as inputs to the algorithms [3-4]. However, there may not be an explicit finite set of transactions in a spatial data set. This study adopts the strategy proposed in [3], where the spatial transaction is defined around the instances of a special reference feature.

This rule expresses that a school close to a sport center has 80% chance of being close to a park, and 0.5% of the data belong to this category. The definition of distance is “close” or “far” is subjective and may influence the mining results. For complex instance, the association rule in the above example is as follows:

\[
\text{is\_a}(X, \text{"forest"}) \land \text{is}(X, \text{"high\_temperature"}) \\
\land \text{is}(X, \text{"very\_dry"}) \\
\Rightarrow \text{result}(X, \text{"wildfire"})[0.2\%, 75%]
\]

This rule shows that a forest with a high temperature and dry environment has 75% chance of causing fire, and 0.2% of the data belong to this category. Spatial association rules are important in exploratory data analysis, where the main effort involves analyzing data along several dimensions according to progressive refinement approaches [7-8]. These methods first perform an initial fast and rough pass over the large datasets, then using a more careful and time-consuming algorithm on selected parts of the data. Huang, et al. proposed a strategy for spatial co-location of patterns [9]. This transaction-free approach mines by the means of proximity neighborhoods. Spatial co-location patterns define relationships between events and locations.

2.3 Inter-transaction association rules

Initial research into association rules mining focused on intra-transaction association rules [10], then inter-transaction association rules were considered [11]. Here, salinity and temperature variations associated with different locations and transaction days are revealed, that is, the rule associates itemsets across different transactions. In this section, fundamental definitions are introduced.

Inter-transaction is the associations among items from different transaction records. It is different to an intra-transaction.

Example 2.1. Intra-transaction association rules: if the ocean salinity rises, then the temperature rises.

Example 2.2. Inter-transaction association rules: if area A salinity anomaly rises little, then area B temperature will also anomaly rise little in the next month.

Data mining algorithms cannot directly derive rules such as the one in Example 2.2. The inter-transaction concept is used to find association rules among itemsets in different records. If the recorded transaction data occurred at different times in sequence, they can be used to predict future events. In this study the goal is to determine association rules such as the one in Example 2.2. Some fundamental definitions are given to describe inter-transaction association rules formally.

Definition 2.1. Let \( I = \{ i_1, i_2, i_3, \ldots, i_n \} \) be a set of items. Let \( D \) be a dimensional attribute and \( Dom(D) \) be the domain of \( D \). A transaction database is a database containing records in the form \((d, I_j)\), where \( d \in Dom(D) \) and \( I_j \subseteq I \). We call this type is \( D \)-dimensional database.

The dimensional attribute usually describes properties associated with the item time or place. It is assumed that the domain of the dimensional attribute is ordinal and can be divided into equal length intervals. For example, time can be divided into day, week, month, etc., and distance into meter, mile, etc. These intervals can be represented by integers 0, 1, 2, etc., without loss of generality [2].
An inter-transaction association rule that spans \( p \) intervals is found if an association exists between items that are \( p \) intervals apart. While an inter-transaction association rule may cover many intervals, discovering all such rules is time-consuming. In order to minimize the effort involved in mining uninteresting rules, a mining parameter called sliding window denoted by \( w \) is introduced. When mining inter-transaction association rules, only the rules spanning shorter than or equal to \( w \) intervals are considered. The sliding window is thus used to avoid mining rules that span many consecutive intervals.

Each sliding window forms a mega-transaction. A mega-transaction \( M \) that is contained within \( W \) can be described as follows:

**Definition 2.2.** Let \( W \) be a sliding window with \( w \) intervals and \( u \) is number of items in \( I = \{i_1, i_2, i_3, \ldots, i_u\} \).

\[
M = \{i_j | i_j \in W[j], 1 \leq k \leq u, 0 \leq j \leq W - 1\}.
\]

To distinguish the items in a mega-transaction from traditional transaction items, the mega-transaction items are called extended items. The set of all possible extended items are denoted \( I' \).

Given \( I' \) and \( w \), then:

\[
I' = \{i_1(0), \ldots, i_1(w-1), i_2(0), \ldots, i_2(w-1), \ldots, i_u(0), \ldots, i_u(w-1)\}.
\]

The following is the definition of an inter-transaction association rule.

**Definition 2.3.** An inter-transaction itemset is a set of extended items \( B \subseteq I' \) such that \( \exists i_k(0) \in B, 1 \leq k \leq u \).

**Definition 2.4.** An inter-transaction association rule has the form \( X \Rightarrow Y \), where

1. \( X \subseteq I' \) and \( Y \subseteq I' \)
2. \( \exists i_k(0) \in X, 1 \leq k \leq u \)
3. \( \exists i_j(0) \in Y, 1 \leq k \leq u, j \neq 0 \)
4. \( X \cap Y = \{} \)

**Definition 2.5.** Let \( T_{xy} \) be the set of mega-transactions that contains a set of extended items \( X \cup Y \) and \( T_{xy} \) be the set of mega-transactions that contains \( X \). Let \( S \) be the number of transactions in the transaction database. Then, the support and confidence of an inter-transaction association rule \( X \Rightarrow Y \) can be defined as:

\[
support = \frac{|T_{xy}|}{S}, \quad confidence = \frac{|T_{xy}|}{|T_x|}.
\]

As with intra-association rules mining algorithms, a minimum support, minsup, and a minimum confidence, minconf, are given and the task is to discover the inter-transaction association rules from the transaction database with support and confidence greater than or equal to the minimum requirements.

### 3. MINING METHOD AND PROCESS

The mining of inter-transaction association rules from a multi-dimensional database can be divided into four steps: Data preprocessing, quantitative attribute transformation, the discovery of frequent inter-transaction itemsets using the Reduced Prefix-Projected Itemsets method (RPPI) [12], and association rule generation.

#### 3.1 Data preprocessing

This study first gathered Argo data from January 2006 to April 2009. Second, a reference model with concentric circles surrounding Taipei, Taiwan (see Fig. 2) was built. The concentric circles are also used to annotate the location of abnormal events. The sub-areas defined by the concentric circles represent different distances and directions to the reference site and transform market-basket type data.

#### 3.2 Quantitative attribute transformation

The first step is to map each quantitative attribute into intervals. Let \( I = \{i_1, i_2, i_3, \ldots, i_k\} \) be the set of all items that belongs to the original database. For the sake of simplicity, the database is assumed to only have one quantitative attribute, \( i_j, 1 \leq j \leq k \), if \( i_j \) is mapped to \( l \) intervals, then the new set of all items \( J_f \) becomes \( \{i_1, i_2, \ldots, i_{l,1}, i_{l,2}, \ldots, i_{l,j}, i_{l,i_1}, \ldots, i_{l,i_k}\} \), where \( i_r \) is a binary attribute and \( i_{l,j}, i_{l,1}, i_{l,2}, \ldots, i_{l,i_k} \) are attributes that are transformed from quantitative attribute \( i_j \).

#### 3.3 RPPI algorithm and association rule generation

Since the quantitative attribute transformations and inter-transactions results in more data, the algorithm employs a reduced prefix-projected itemsets method based on the PrefixSpan algorithm to expedite the efficient search for the frequent itemsets, instead of using the widely used Apriori algorithm which is more resource demanding.

The processing cost of finding frequent 1- and 2-item inter-transaction itemsets (i.e., obtaining L1 and L2 from C1 and C2, respectively) dominates the total mining cost; therefore, a straightforward implementation is to reduce the size of C1 and C2.

The strategy employs a 3-dimensional sliding window to find association rules (see fig. 3). For example, assume \( B \) happens on the third day after \( A \), then \( A(0) \rightarrow B(2) \).

The proposed method is termed the Reduced Prefix-Projected Itemsets method (RPPI). Following these steps, the RPPI recursively generates the reduced projected database for each frequent \( k \)-itemset to find the frequent \( (k+1) \)-itemsets.

Inter-transaction rules are extracted from important data established using the RPPI algorithm on the multi-dimensional inter-transaction database to expedite the efficient search for the frequent itemsets, instead of using the widely used Apriori algorithm which is more resource demanding.

Our purpose is to record event occurring times and relative positions in inter-transactions mining. We define 3-dimensional sliding windows in Fig. 3 to find the inter-transaction association rules between itemsets. Transaction data with given temporal and spatial factors can be associated for prediction purpose. For example, assume \( B \) happens on the third day after \( A \) happens, then the fundamental rule is \( A(0) \rightarrow B(2) \). One example is given to illustrate how RPPI works.
Example 3.1. Table 1 is the inter-transaction database from Fig. 3. Assuming that the minimum support count is 2, RPPI first scans the database to find the frequent 1-inter-itemset with form $i_0(0,0,0)$, i.e., $<a(0,0,0)>:2$, $<b(0,0,0)>:2$ and $<c(0,0,0)>:2$. Then, the projected database is generated for each frequent 1-itemset. Table 2 shows the final result. This is how the PrefixSpan algorithm recursively generates the projected database for each frequent $k$-itemset to find the frequent $(k+1)$-itemsets.

Fig. 3. A sliding window applied to a 3-dimensional transaction database.

4. EXPERIMENTAL RESULTS
The experimental salinity and temperature data used in this study were taken from the Argo delayed-mode database that can be downloaded from Oceanographic Products in France. The monthly salinity and temperature data were collected from January 2006 to April 2009. We put concentric circles on each map and treated all the abnormal events that occurred inside the concentric circles on each map as a transaction. The radii for the inner and outer circles were 380 km and 760 km, respectively. Concentric circles allow the location context information such as direction and distance to the reference site of Taipei to be maintained. Table 3 shows the mapping intervals for salinity and temperature variations.

In case no special events occurred in a sub-area then it is denoted a T/SNO event. After the quantitative attributes in Table 4 are mapped to the intervals shown in Table 3, the PrefixSpan algorithm is employed to find large inter-itemsets, where the maxspan window size is set to the first and second ten-day period of a month and the minimum support is set to 7% (the minimum item numbers is 47), the minimum confidence is set to 70%. Fig. 4 illustrates how the RPPI algorithm calculates association rules with Prefix(2) and Prefix(3). Table 5 shows some of the inter-transaction association rules. In Fig. 5(a) the association rule $H_2.TDL(0,0,0)\rightarrow A_2.TDM(0,1,1)$ (the three elements represent year, month and depth) says that if the salinity rose little (0 – 0.15 psu) in area $H_2$, then the temperature will drop much (-1.3 – -2°C) in $A_2$ during the next ten-days at substratum depth. In Fig. 5(b) the association rule $A_1.TDL(0,0,1)\rightarrow H_2.SRL(0,0,0)\& H_1.SRL(0,1,0)$ says that if temperature dropped little (0 – -1.2°C) in $A_1$, then the salinity will rise little (0 – 0.15 psu) in $H_1$ during the next ten-days at substratum depth, whereas the salinity will rise little for $H_2$, too.

Fig. 5(a). Illustration of unusual event of concentric circles (2-item association rules).

Fig. 5(b). Unusual events at concentric circles (3-item association rules).

5. CONCLUSION AND FUTURE WORK
In this study, data mining techniques are used to facilitate the analysis of interesting association rules in Argo data. Usually, statistical methods are employed to construct a prediction model of salinity and temperature variations. However, it is difficult to apply statistical analysis to spatial and temporal relationships prediction. For example, if one of the discovered rules is “if area A temperature variation rises little, then area B salinity variation rises little during the next ten-days”, then the associated salinity and temperature variations that cover different locations and transaction days are revealed. Therefore, a reduced prefix-projected itemset method is employed with inter-transaction association rules mining algorithm to reduce time complexity.
The experimental results verify that the proposed model is effective in predicting the occurrence of salinity and temperature variations. For example, one of the inferences of our experimental results is "if area H₂ salinity rose little (0 – 0.15 psu), area A₂ temperature rise much (-1.3 – -2°C) in the next ten-day at substratum depth". Such results demonstrate that satisfactory results can be achieved without the domain knowledge of experts. Future work includes incorporating ocean patterns. It is hoped that this research can allow people to better understand ocean climate changes.

6. ACKNOWLEDGMENTS

This work is supported by National Science Council, Taiwan, R.O.C. under Grants NSC 97-2221-E-027-034-MY3. The research data were collected and made freely available by the International Argo Project and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). Argo is a pilot program of the Global Ocean Observing System.

7. REFERENCES


---

![Fig. 4. The RPPI algorithm of Prefix itemsets processes (minimum support is set to 0.9%).](image)
### Table 1. An inter-transaction database transformed from Fig. 3.

<table>
<thead>
<tr>
<th>Transaction ID</th>
<th>Transaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>{a(0,0,0), b(0,0,0), c(0,0,0), d(0,0,0), f(1,0,0), a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
</tr>
<tr>
<td>M2</td>
<td>{a(0,0,0), b(0,0,0), c(0,0,0), c(0,1,1), a(1,1,1)}</td>
</tr>
</tbody>
</table>

### Table 2. The reduced projected database from Table 1.

<table>
<thead>
<tr>
<th>Prefix(1)</th>
<th>Reduced projected (postfix) database(1)</th>
<th>Prefix(2)</th>
<th>Reduced projected (postfix) database(2)</th>
<th>Prefix(3)</th>
<th>Reduced projected (postfix) database(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\langle a(0,0,0)\rangle</td>
<td>{b(0,0,0), c(0,0,0), d(0,1,1), e(0,0,0), a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
<td>\langle a(0,0,0)\rangle</td>
<td>{c(0,0,0), a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
<td>\langle a(0,0,0)\rangle</td>
<td>{a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
</tr>
<tr>
<td>\langle b(0,0,0)\rangle</td>
<td>{c(0,0,0), a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
<td>\langle b(0,0,0)\rangle</td>
<td>{d(0,0,0), a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
<td>\langle b(0,0,0)\rangle</td>
<td>{a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
</tr>
<tr>
<td>\langle c(0,0,0)\rangle</td>
<td>{a(0,1,1), b(0,1,1), c(0,1,1), b(0,1,0)}</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>\langle a(0,0,0)\rangle</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3. The mapping intervals for salinity and temperature quantitative values.

<table>
<thead>
<tr>
<th>Event</th>
<th>Meaning</th>
<th>Event</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRL</td>
<td>Temperature rose little (0 – 1.2°C)</td>
<td>SRL</td>
<td>Salinity rose little (0 – 0.15 psu)</td>
</tr>
<tr>
<td>TDL</td>
<td>Temperature dropped little (0 – -1.2°C)</td>
<td>SDL</td>
<td>Salinity dropped little (0 – -0.15 psu)</td>
</tr>
<tr>
<td>TRM</td>
<td>Temperature rose much (1.3 – 2°C)</td>
<td>SRM</td>
<td>Salinity rose much (0.16 – 0.25 psu)</td>
</tr>
<tr>
<td>TDM</td>
<td>Temperature dropped much (-1.3 – -2°C)</td>
<td>SDM</td>
<td>Salinity dropped much (-0.16 – -0.25 psu)</td>
</tr>
<tr>
<td>TNO</td>
<td>No event happen</td>
<td>SNO</td>
<td>No event happen</td>
</tr>
</tbody>
</table>

### Table 4. Mapping results of salinity and temperature variations from Table 3.

<table>
<thead>
<tr>
<th>Date</th>
<th>Depth 200m (Loc. &amp; Tem.)</th>
<th>Depth 200m (Loc. &amp; Sal.)</th>
<th>Depth 400m (Loc. &amp; Tem.)</th>
<th>Depth 400m (Loc. &amp; Sal.)</th>
<th>Depth 600m (Loc. &amp; Tem.)</th>
<th>Depth 600m (Loc. &amp; Sal.)</th>
<th>Depth 800m (Loc. &amp; Tem.)</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Jan., 2006</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SNO</td>
<td>H2.TNO</td>
<td>…</td>
</tr>
<tr>
<td>Mid-Jan., 2006</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SNO</td>
<td>H2.TNO</td>
<td>…</td>
</tr>
<tr>
<td>Late Jan., 2006</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SNO</td>
<td>H2.TNO</td>
<td>…</td>
</tr>
<tr>
<td>Early Feb., 2006</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SNO</td>
<td>H2.TNO</td>
<td>…</td>
</tr>
<tr>
<td>Mid-Feb., 2006</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SRL</td>
<td>H2.TNO</td>
<td>H2.SNO</td>
<td>H2.TNO</td>
<td>…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

### Table 5. Partially discovered inter-transaction association rules.

<table>
<thead>
<tr>
<th>2-item rules</th>
<th>3-item rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2.TNO(0,1,1) \rightarrow A2.TDM(0,1,1)</td>
<td>A1.TDL(0,1,1) \rightarrow H2.SRL(0,0,0) &amp; H1.SRL(0,1,0)</td>
</tr>
<tr>
<td>H2.SNO(0,0,0) \rightarrow B1.SNO(0,0,0)</td>
<td>A2.TNO(0,0,0) \rightarrow B2.TNO(0,1,0) &amp; A2.TNO(0,1,0)</td>
</tr>
<tr>
<td>H2.SNO(0,0,0) \rightarrow B1.SNO(0,0,0)</td>
<td>A2.TNO(0,0,0) \rightarrow B2.TNO(0,1,0) &amp; A2.TNO(0,1,0)</td>
</tr>
<tr>
<td>H2.TNO(0,0,0) \rightarrow G2.TNO(1,0,0)</td>
<td>A2.TDM(0,1,1) \rightarrow H2.SRL(0,0,0) &amp; H1.SRL(0,1,0)</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>