Visual Analytics of Causal Relationship Between Fish Catches Data in Adjacent Sea Areas Using Convergent Cross Mapping

Chi Zhang\textsuperscript{1,2}, Masaaki Iiyama\textsuperscript{3}, Koji Koyamada\textsuperscript{3}

\textsuperscript{1}Graduate School of Engineering, Kyoto University, Japan
\textsuperscript{2}School of Electrical Engineering and Automation, Wuhan University, China
\textsuperscript{3}Academic Center for Computing and Media Studies, Kyoto University, Japan

\textsuperscript{*}zhangchi6789@outlook.com

Received: December 19, 2019; Accepted: July 14, 2020; Published: August 5, 2020

Abstract. In this paper, a visual analytics system that are used to analyze the causal relationship between fishing catches data in adjacent sea areas is proposed. By using the proposed system, the distribution of sea surface temperature is visualized so that the sea areas can be divided accordingly. Next, the daily fishing efficiency in each sea area are calculated and the convergent cross mapping is used to analyze the causality between the data in adjacent sea areas. The experimental results show that the possible causation between the data in two of the divided sub sea areas are detected. In addition, the causal variable can be further identified by referring to the prediction skills.

Keywords: visual analytics, fish catches, convergent cross mapping, empirical dynamic modeling

1. Introduction

As the development of science and technology, and the improvement of productivity, the impact on global climate is deeper and deeper due to certain human activities [1]. Global climate change usually manifests itself as global warming [2], which is influencing our production and living in many aspects, such as meteorology, environment, ecological systems and so on [3, 4, 5, 6]. The influence of climate change on global fish production has also been studied during these years [7, 8, 9, 10]. Furthermore, the relationship between the fishing grounds distribution and sea surface temperatures (SST) has been revealed [11, 12]. Therefore, SST can be considered to be a significant factor that influences fish catches. Our data record the fishing locations and hours of fishing vessels all over the world. If the fishing hours are considered to be the reflection of fish catches, the data can be deemed as global fish catches data. Meanwhile, some public SST data are also accessible. The difference is that
in our research, we focus on analyzing the relationship, especially the causation between fishing catches in adjacent sea areas, which are divided based on the distribution of SST. We believe that the identification of the relationship can help to contribute to marine fishery resources management and cooperation among maritime nations under the circumstance of global climate change.

If the fishing catches in different sea areas are considered to belong to a dynamic system, a method that is called empirical dynamic modeling (EDM) [14] can be used to analyze the relationship between two time series. Although EDM has been applied to the analysis of ecosystems, it has not been used for analyzing the fishing catches between two sea areas. EDM includes various utilities for investigating dynamical systems [15]. Among these utilities, convergent cross mapping (CCM) [16, 19] is usually used to analyze the causation between two time series. Because our source data are daily fish catches data, the daily fishing hours in a specific sea areas can be summed. Thus, it is capable of generating time series with respect to fish catches in a sea areas. In our system, an interface based on Folium [17] provides users with intuitive experience on dividing a large sea area into relatively small areas by referring to the distribution of SST. Then, to confirm the causation between the fishing catches in adjacent sea areas, CCM needs to be applied.

The rest of this paper is organized as follows. In Section 2, the related work with respect to empirical dynamic modeling, ocean data visualization and the library of Folium is summarized. Section 3 describes the proposed system in detail, including the system composition and the algorithms that are used to analyzing the causal relationship between fish catches in adjacent sea areas. The experimental results and the related discussions are presented in Section 4 and 5, respectively. Finally, Section 6 presents the conclusion and future work.

2. Related Work

EDM [14] can be considered to be an approach for detecting causality in complex deterministic systems. It is based on Takens’s theorem [18] and nonlinear state space reconstruction [19]. To investigating the causation between time series, some algorithms are provided, such as simplex projection [20, 21], sequential locally weighted global linear map (S-map) [22] and CCM [16, 19]. The EDM framework has been applied to the analytics of ecosystems. For example, Deyle et al. used EDM to measure and forecast changing interactions in real systems, and identify the underlying mechanisms [23]. Grziwotz et al. used EDM to analyze the relationship between mosquito populations and environmental variables [24].

The analytics of fishing catches data plays an important role in the study on ecosystems. Some researches that are related to the influences on fishing catches due to global climate change have been mentioned in the previous section [7, 8, 9, 10]. These researches mainly focus on the change of fishing catches, fishing grounds locations and fish distributions. Specifically, as the conclusion of the paper [7], recent changes in the distribution and productivity of a number of fish species could be ascribed with high confidence to regional climate variability, such as the El Niño–Southern Oscillation. There were strong interactions
between the effects of fishing and the effects of climate because fishing reduces the age, size, and geographic diversity of populations and the biodiversity of marine ecosystems, making both more sensitive to additional stresses such as climate change. In the paper [8], as the prediction result of the changes in global catch potential for 1066 species of exploited marine fish and invertebrates from 2005 to 2055 under climate change scenarios, climate change might lead to large-scale redistribution of global catch potential, with an average of 30% to 70% increase in high-latitude regions and a drop of up to 40% in the tropics, which indicated the need to develop adaptation policy that could minimize climate change impacts through fisheries. In the paper [9], according to the analysis of cod data, comparing to the circumstance in the 20th century, there was a significant change in the distribution of the cod in North Sea, which manifested as the northward shift and the eastward shift. These changes were attributed to warming and fishing pressure. In the paper [10], as the simulation result, climate change decreased the modelled global fish community biomass by as much as 30% by 2100.

Moreover, systems have also been developed to visualize fishing catches data and ocean data. For instance, Ocean Data View (ODV) was developed for the visual analytics of ocean data and other types of geoscience data [25, 26]. Williams et al. extracted eddies by using the Okubo-Weiss parameter and visualized the extracted eddies in three-dimensional space [27]. Similarly, as our previous work, a visualization system was developed for visualizing ocean data and extracting vortices [28]. Another method that was used for visualizing the spatial distribution of ocean temperatures was proposed by Feng et al. [29]. In this method, the pseudo-octree model was applied to the organization of uniformly sampled temperature data and the spatial distribution of temperatures was rendered by ray casting. Recently, browser-based visualization tools become more and more popular. Comparing to some traditional visualization tools, the advantages of cross-platform and relatively less environmental dependency improve the usability of browser-based visualization tools. As a popular library under Python, the Folium library is based on the leaflet.js [30] library. It has been used to visualize various types of data, such as social media data [31], earthquake data [32] and lightning data [33]. By using the Folium library, multi-dimensional data can be visualized on different layers by different methods. Because of the convenience to add a map layer by the library, it is useful to visualize geographic data. It also provides some interactive features, which will be explained in detail in the following sections.

3. System Details

The proposed system consists of two components, a visualization module and a causation identification module. The visualization module visualizes the distribution of SST as an overlay on the world map. Hence, the target sea areas that are used for analysis can be selected by referring to the SST distribution. Then, the daily fishing efficiency in each sea area are calculated to generate the time series data. By using the causation identification module, the causation between the fish catches in adjacent sea areas is verified.
3.1. Visualization Module

The visualization module in the proposed system provides an interactive interface based on the Folium library to explore the world map and the distribution of SST, which help to divide the target sea area into a number of sub sea areas. By default, a world satellite map is displayed with the SST distribution shown as a semitransparent overlay. By referring to the world map and the SST distribution, the sub sea areas can be segmented. Specifically, when clicking on the world map, the latitude and the longitude of the corresponding location are recorded. In addition, this location can be selected as a vertex of a sub sea area. After all of the vertices are selected, the segmentation of the sub sea areas are completed. It is also possible to display the polygon range of the sub sea areas as a semitransparent overlay. Moreover, the display content of the map layer and the overlays is switchable. For example, the map layer can be switched between different types of maps, and the overlays can be switched between the distributions of different variables. In this case, a hover menu that lists the names of different layer types is added to the interface. The display content of a layer can be changed by clicking the corresponding layer name in the menu, as shown in Fig. 1.

After all of the sea areas are selected, the daily fishing efficiency in each sea area are calculated and the time series data are generated. Here, the daily fishing efficiency of a sea area is calculated by using the number of fishing boats to divide the fishing catches in the sea area. Because the fishing catches in a sea area can be easily increased if more fishing boats work in the area, fishing efficiency is selected to measure the concentration of fish.
3.2. Causation Identification

As mentioned above, CCM [16, 19] is used to identify the causation between two time series of fish catches in our system. CCM can be considered to be an approach to verify the causation between variables from time series data that belong to the same dynamic system. It uses an embedding of one variable to predict another, then identifies the causation and the causal variable according to the prediction skill. In this sub section, to explain the algorithm of CCM, the algorithm of simplex projection [20, 21] and S-map [22] will be explained in advance.

3.2.1. Simplex Projection

Simplex projection [20, 21] is an approach that is capable to make short-term predictions about the trajectories of chaotic dynamic systems. The main idea of the algorithm is to use the similar patterns in previous time steps to predict the value in the present time step. Specifically, assume that a time series data is expressed as \(X = \{x_n, x_{n-1}, \ldots, x_1\}\), and the value of a certain time step \(x_{t+1}\) in the time series \(X\) needs to be predicted. The algorithm can be summarized as follows.

1. From the time series \(X\), a number \(E\) of values \(X_t = \{x_t, x_{t-\tau}, \ldots, x_{t-(E-1)\tau}\}\) are selected as a new data point (time series).

2. Calculate the distance between the selected data point \(X_t\) and other data points that are selected from the time series \(X\) with the form of \(X_k = \{x_k, x_{k-\tau}, \ldots, x_{k-(E-1)\tau}\}\), where \(k \neq t\). Arrange the data points in short-distance order and select the first \(m\) data points. Here, assume that \(m = 3\) and the three points \(X_a, X_b\) and \(X_c\) are selected. By default, Euclidean distance is used as the metric for distance. Depending on the situation, Manhattan distance and Chebyshev distance can also be used. Euclidean distance \(D\) between the two points \(X_t(x_{t,1}, x_{t,2}, \ldots, x_{t,n})\) and \(X_k(x_{k,1}, x_{k,2}, \ldots, x_{k,n})\) is defined as \(D(X_t, X_k) = \sqrt{\sum_{i=1}^{n}(x_{t,i} - x_{k,i})^2}\).

3. According to the source data, the value of the next step of a point \(X_k = \{x_k, x_{k-\tau}, \ldots, x_{k-(E-1)\tau}\}\) can be obtained as \(\{x_{k+1}, x_{k+\tau}, x_{k+2\tau}, \ldots, x_{k+(E-2)\tau}\}\). Hence, for the three points \(X_a, X_b\) and \(X_c\), the values of the next steps can be obtained. In other words, the values of \(x_{a+1}, x_{b+1}\) and \(x_{c+1}\) are known.

4. Calculate the mean of \(x_{a+1}, x_{b+1}\) and \(x_{c+1}\). The result can be considered to be the prediction value of \(x_{t+1}\). Note that weighted means are usually used. The values from the points that are relatively closer to \(X_t\) are weighted relatively more heavily.

In conclusion, simplex projection finds some nearest neighbours of the data point at the previous time step of the target time step, and then uses the data points at the next time step of these nearest neighbours to predict the value at the target time step. The distance between two data points is used to evaluate the similarity of the patterns of the two data points. If some similar patterns in the previous time steps can be obtained, the following changes of the patterns are used for prediction. Based on the above description of the simplex projection
algorithm, there are some related parameters. The number of elements $E$ in a data point (the length of the vector) is defined as the embedding dimension. The number of time steps $\tau$ between two adjacent selected values in a data point is defined as the lag (time delay). The data points that need to be found and used for prediction is defined as the nearest (closest) neighbours. Theoretically, the minimal number of nearest neighbours equals to the number of embedding dimension plus 1. Because at least $E + 1$ nearest neighbours are needed to form a geometric structure for surrounding a point with the embedding dimension $E$. As our understanding, for example, for a two-dimensional point, at least three two-dimensional points are needed to form a triangle to surround it. For a three-dimensional point, a tetrahedron with four vertices is needed, etc.

In our experiments, the optimal values for the embedding dimension and time delay are determined by comparing the prediction skill of the model. There are also some other methods that are proposed to determine the optimal values for the embedding dimension and time delay. For example, the time delay can be found from the first minimum in the mutual information between the time series and a shifted version of itself. The embedding dimension can be calculated by using a false near neighbours test [34]. After the parameters are defined, the above calculation can be expressed as equations. Assuming that the next time step of the vector $X_t$ is $X_{t+1}$, $\hat{X}_{t+1}$ is the prediction value of $X_{t+1}$, $X_{t,i}$ is the $i$th nearest neighbour of $X_t$ and $w_{t,i}$ is the $i$th weight, $\hat{X}_{t+1}$ can be calculated as

$$\hat{X}_{t+1} = \sum_{i=1}^{E+1} \frac{w_{t,i} X_{t+1,i}}{\sum_{i=1}^{E+1} w_{t,i}}. \quad (1)$$

Moreover, the weight $w_{t,i}$ can be calculated as

$$w_{t,i} = \exp -\frac{D(X_t, X_{t,i})}{D(X_t, \hat{X}_{t+1})}, \quad (2)$$

where $D(X_t, X_{t,i})$ indicates the distance between $X_t$ and $X_{t,i}$.

In the simplex projection, the source data are embedded by using the embedding dimension $E$ and time delay $\tau$. The group of embedded data points that are used for prediction is named as library. Library length (library size) indicates the number of sample points in the library. A specified library size splits the data into a training set and a testing set, which is the case of CCM. In the simplex projection, as an extreme circumstances, the data points except for the predicted one are grouped in the library. Another related parameter is the sampling rate. Usually, it is set to 1, meaning that it is possible for each data point to be selected as a nearest neighbour. A smaller sampling rate makes fewer data points be used for prediction, which also reduces the library size.

### 3.2.2. Sequential Locally Weighted Global Linear Map (S-map)

Other than the simplex projection that uses a fixed number of near neighbours to predict the target data point, the S-map [22] uses all data in the library for prediction. By using S-map, the prediction $\hat{X}_{t+1}$ of a data point $X_{t+1}$ can be calculated as

$$\hat{X}_{t+1} = \sum_{i=1}^{n} w_{t,i} X_{t+1,i}, \quad (3)$$
where \( n \) is the library size and \( w_i \) is the weight. The calculation of \( w_i \) can be expressed as

\[
 w_i = \exp\left(-\theta \frac{D(X_{t,i}, X_t)}{D_n}\right)
\]  
where \( D(X_{t,i}, X_t) \) is the distance between the \( i \)th data point in the library and the point \( X_t \), \( D_n \) is the mean distance of all paired library points, namely

\[
 D_n = \frac{1}{n} \sum_{i=1}^{n} D(X_{t,i}, X_t).
\]

The parameter \( \theta \) is used to evaluate the non-linearity of the system with the given embedding dimension. A larger \( \theta \) corresponds to larger weight values for the data points that are closer to the target point. If \( \theta = 0 \), all data points in the library are weighted equally, and the model becomes the linear auto regression. Furthermore, if the prediction accuracy increases with the increasing of \( \theta \), it can be deduced that the system is non-linear.

### 3.2.3. Convergent Cross Mapping (CCM)

CCM [16, 19] can be considered to be an approach to verify the causation between variables from time series data that belong to the same dynamic system. This algorithm extends the simplex projection to the multivariate case. CCM is based on Takens’s theorem [18]. It uses an embedding of one variable to predict another, then identifies the causation and the causal variable according to the prediction skill.

Assume that there are two time series \( X \) and \( Y \):

\[
 X = \{x_n, x_{n-1}, \ldots, x_t, \ldots, x_1\},
\]

\[
 Y = \{y_n, y_{n-1}, \ldots, y_t, \ldots, y_1\}.
\]

In CCM, if the value of \( y_t \) needs to be predicted, the calculation can be summarized as follows.

1. Determine the optimal values for the embedding dimension \( E \) and time delay \( \tau \).
2. The shadow manifolds of \( X \) and \( Y \) are constructed with the embedding dimension and time delay.
3. In the shadow manifold of \( X \), select \( E + 1 \) nearest neighbours of \( X_t = \{x_t, x_{t-\tau}, \ldots, x_{t-(E-1)\tau}\} \) (e.g. \( X_j, X_k \) and \( X_l \)), calculate the weights based on the distances between these data points and \( X_t \).
4. Predict \( Y_t = \{y_t, y_{t-\tau}, \ldots, y_{t-(E-1)\tau}\} \) by using \( Y_j, Y_k \) and \( Y_l \). The weights of \( Y_j, Y_k \) and \( Y_l \) are equal to the weights of \( X_j, X_k \) and \( X_l \), respectively.

Assuming that \( \hat{Y}_t \) is the prediction value of \( Y_t \), \( X_{t,i} \) is the \( i \)th nearest neighbour of \( X_t \), \( w_{t,i} \) is the weight of \( X_{t,i} \), and \( Y_{t,i} \) is the vector that corresponds to \( X_{t,i} \), the calculation of \( \hat{Y}_t \) can be expressed as

\[
 \hat{Y}_t = \frac{\sum_{i=1}^{E+1} w_{t,i} Y_{t,i}}{\sum_{i=1}^{E+1} w_{t,i}}.
\]
The weights $w_{t,i}$ can be calculated as

$$w_{t,i} = \exp\left(-\frac{D(X_t, X_{t,i})}{D(X_t, X_{t,1})}\right),$$

(8)

where $D(X_t, X_{t,i})$ indicates the distance between $X_t$ and $X_{t,i}$.

Theoretically, if the prediction skill (forecasting ability) increases as the library size increases, and converges to a significantly positive value, it can be inferred that there is a causal relationship between $X$ and $Y$. Here, prediction skill can be considered to be a reflection of the prediction accuracy. It is defined as the Pearson correlation coefficient [35, 36] between the prediction values and the observed values of $Y$. The Pearson correlation coefficient $\rho$ of a pair of random variables $(A, B)$ can be calculated as

$$\rho_{A,B} = \frac{\text{cov}(A, B)}{\sigma_A \sigma_B},$$

(9)

where $\text{cov}(A, B)$ is the covariance between $A$ and $B$, $\sigma_A$ and $\sigma_B$ are the standard deviations of $A$ and $B$, respectively. Thus, the range of the prediction skill is from $-1$ to $1$. Moreover, the causal variable can be determined by comparing the prediction skills when using the two time series to predict each other. If time series $A$ can be properly predicted by using time series $B$, and comparing to using $A$ to predict $B$, $B$ has higher prediction skill to predict $A$, $A$ can be considered to be a reason of $B$.

4. Experimental Results

The data that are used in our experiments includes global SST data from National Oceanic and Atmospheric Administration (NOAA) [37] and global fishing catches data. The dimensions in the SST data are date, longitude, latitude and SST, while the dimensions in the fishing catches data are date, longitude, latitude, fishing vessel number and fishing hours. In the experiments, the data from July 1st, 2016 to September 30th, 2016 are used. The sea area to the west of Japan and the sea area to the east of Japan are selected and analyzed. As the first step, the distribution of the average SST during the specified period is visualized. As shown in Fig. 2, the region in dark purple represents the sea area with relatively high SST, whereas the region in light blue represents the sea area with relatively low SST. According to the distribution of SST, the target sea area to the east of Japan is divided into four smaller sea areas. The regions visualized as semi-transparent blue overlays indicate the sub sea areas. For convenience, the sea areas are indexed from Area 1 to Area 4, as shown in the figure. At present stage, only SST is used as the dividing evidence. Because the SST is mainly distributed in the north-south direction, vertical lines are used as the boundaries in the east-west direction. Therefore, the sub sea areas are not divided in detail. Actually, the system supports more numbers of vertices. For each time, after all vertices of a sub sea area are selected, the sea area with corresponding shape can be generated. The system also supports the segmentation of nonadjacent sea areas. In order to obtain more detailed sub sea areas, it is necessary to apply more dividing evidences. However, only appropriate parameters can be used, which will be discussed in the following section. After the approximate
regions of sub sea areas are obtained, the causation between adjacent sea areas are identified by using CCM. The steps of sub sea area segmentation and causation identification can be repeated until causally related sea areas are found.

As preprocessing of the fish catches data, the fish catches are converted to fishing efficiency. Because the fishing hours in the source data are considered to be a reflection of fishing catches, the fishing hours for each fishing boat in a sub sea area are calculated and regarded as the fishing efficiency. After the daily fishing efficiencies of the four sub sea areas are obtained, the non-linearity of the four time series is verified by using S-map [22]. Before applying the S-map algorithm, it is necessary to set the optimal values of the embedding dimension and the time delay. In the experiment, because the ranges of the embedding dimensions and the time delay are not very wide, the prediction skills of CCM [16, 19] are compared by scanning all of the possible parameters, and the parameters that correspond to the highest prediction skill are selected as the optimal parameters. Otherwise, the other methods mentioned in Section 3.2.1. can be used. The results of the non-linearity check are shown in Fig. 3. In the figure, the horizontal axes indicate the parameter $\theta$, and the vertical axes indicate the prediction skill $\rho$. According to the figure, the non-linearity of the time series in Area 2 and 3 can be verified, while the cases for Area 1 and Area 4 is relatively complicated. The results will be further discussed in the following section.

Despite the results of non-linearity check, the causation between each two adjacent sea areas is identified by using CCM. The optimal parameters that are previously obtained are used. The calculation results are shown in Fig. 4. In the figure, the horizontal axes indicate the library size, while the vertical axes indicate the prediction skill. For each group of two time series, each time series is used to predict the other in turn. Thus, the prediction needs to be made twice. The results are represented as red lines and blue lines in the figure. According to the figure, as the library size increases, an increasing trend can be found between Area 2 and Area 3, and the prediction skill for using the data in Area 2 to predict the data in Area 3 reaches approximate 0.4. Therefore, there is a relatively high possibility that causal relationship exists between the fish catches in Area 2 and Area 3. Although increasing trends can also be found in other sub figures of Fig. 4, the prediction skills are relatively low. In addition, the non-linearity of time series in Area 1 and Area 4 is not verified, which does not satisfy the precondition. Comparing to using the data in Area 3 to predict the data in Area 2, the prediction skill for using the data in Area 2 to predict the data in Area 3 is higher, meaning that the data in Area 3 can be considered to be the causal variable.

5. Discussion

In the experiment, fishing efficiency, rather than total fish catches is used to express the fish catches in a sea area. Because total fish catches can be influenced by the number of fishing boats in a sea area, fishing efficiency is calculated and used for the causation identification. When dividing the target sea area into sub sea areas, the reference data is the distribution of SST, which is an important factor that affect the fish catches. The limitation of SST distribution is that the sea areas with different SST ranges are usually distributed in the north-south direction. Thus, if no other data are used, it is difficult to determine the boundary
Figure 2: Visualization of SST distribution and sea areas division.

Figure 3: Non-linearity verification of the four time series by using S-map. (a) Fishing efficiency in Area 1. (b) Fishing efficiency in Area 2. (c) Fishing efficiency in Area 3. (d) Fishing efficiency in Area 4.
Figure 4: Causation identification results of the sea areas to the east of Japan. (a) CCM between Area 1 and Area 2. (b) CCM between Area 2 and Area 3.
Figure 4: Causation identification results of the sea areas to the east of Japan (cont.). (c) CCM between Area 3 and Area 4. (d) CCM between Area 1 and Area 4.
between the sub sea areas in the east-west direction. It is possible that many attempts have to be made in order to obtain a appropriate boundary. In the experiment, as a simple case, vertical boundaries are applied. There is also an option that introduces other variables, such as salinity and ocean currents, as the reference data for the division of sub sea areas. However, the distribution of sea surface salinity (SSS) dose not change significantly, except for the SSS near estuaries. As future work, the division of sub sea areas by referring to both SST and ocean currents data will be tested, if appropriate ocean currents data can be obtained. In non-linearity check, the S-map [22] curves for the data in Area 2 and Area 3 better meet the conditions of non-linearity, while for the Area 1 and Area 4, though the S-map curves show rising trends in later parts, in the foreparts, there are obvious downswing trends. Since the criterion for the non-linearity is that the prediction skill increases with the increasing of $\theta$, the non-linearity of the two time series cannot be deduced. However, for comparison, the data in Area 1 and Area 4 are included in the causation check. As for the causal relationship investigation by using CCM [16, 19], the following three aspects are focused on: the shape of the CCM curves, the values of the prediction skills and the positional relation of the two CCM curves in a sub figure. On the one hand, as the library size increases, a rising trend of the prediction skill should be obtained. On the other hand, the prediction skill should converge to a significantly positive value as the library length increases. However, at present, there is not an unambiguous definition for the significantly positive value in CCM. In some cases, 0.3 is defined as the threshold value, but it is not widely acknowledged. According to the experimental results, the result in Fig. 4b satisfies the criteria in the greatest degree. Hence, there is highest possibility that the data in Area 2 and Area 3 are causally related. To further verify the causality, other approaches can be employed to identify causation from other aspects.

6. Conclusion

The proposed system that used to analyze the causal relationship of fish catches data between two adjacent sea areas consists of two modules. The visualization module shows the distribution of SST as an overlay on two-dimensional world map, which helps to divide a target sea area into a number of sub sea areas. The map layer and the overlay can also be switched to show different types of maps or variables. The analysis module identifies the causal relationship by using CCM [16, 19]. The possible causality can be recognized according to the prediction skill by using one embedded time series fish catches data to predict another. In the experiment, by using the proposed system, the possible causal relationship is found between the fish catches in two adjacent sub sea areas. In addition, the causal variable is also identified. In the future, other ocean data will be tried adding to the referring data to check if the division of sub sea areas can be more reasonable and the prediction skill can become higher. Moreover, since the theory of EDM [14] is developing, other methods can be added to the system to verify the causation more comprehensively from different angles.
Acknowledgement

This work was supported by JST CREST Grant Number JPMJCR19F1, Japan. The author (Chi Zhang) appreciates the financial support of China Scholarship Council during his study at Kyoto University. NOAA High Resolution SST data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/.

References

[1] National Research Council: *Advancing the Science of Climate Change*, National Academies Press, Washington, D.C., United States, 2010.

[2] D. L. Hartmann, et al.: Observations: atmosphere and surface, *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom, 2013.

[3] G. C. Hegerl, et al.: Understanding and attributing climate change, *Climate Change 2007: The Physical Science Basis. Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) Volume Chapter 9*, Geneva, Switzerland, 2007.

[4] E. M. Fischer, R. Knutti: Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes, *Nature Climate Change*, 5:6 (2015), 560–564.

[5] C. S. Watson, et al.: Unabated global mean sea-level rise over the satellite altimeter era, *Nature Climate Change*, 5:6 (2015), 565–568.

[6] C. Deutsch, et al.: Climate-forced variability of ocean hypoxia, *Science*, 333:6040 (2011), 336–339.

[7] K. M. Brander: Global fish production and climate change, *Proceedings of the National Academy of Sciences*, 104:50 (2007), 19709–19714.

[8] W. W. Cheung, et al.: Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change, *Global Change Biology*, 16:1 (2010), 24–35.

[9] G. H. Engelhard, et al.: Climate change and fishing: a century of shifting distribution in North Sea cod, *Global Change Biology*, 20:8 (2014), 2473–2483.

[10] D. A. Carozza, et al.: Metabolic impacts of climate change on marine ecosystems: Implications for fish communities and fisheries, *Global Ecology and Biogeography*, 28:2 (2019) 158–169.

[11] H. U. Solanki, et al.: Fishery forecast using OCM chlorophyll concentration and AVHRR SST: validation results off Gujarat coast, India, *International Journal of Remote Sensing*, 24:18 (2003) 3691–3699.
[12] M. Zainuddin, et al.: Detection of potential fishing ground for albacore tuna using synoptic measurements of ocean color and thermal remote sensing in the northwestern North Pacific, *Geophysical Research Letters*, 31:20 (2004).

[13] G. Sugihara, et al.: Detecting causality in complex ecosystems, *Science*, 338:6106 (2012), 496–500.

[14] H. Ye, et al.: Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling, *Proceedings of the National Academy of Sciences*, 112:13 (2015) E1569–E1576.

[15] C. W. Chang, et al.: Empirical dynamic modeling for beginners, *Ecological Research*, 32:6 (2017) 785–796.

[16] H. Ye, et al.: Distinguishing time-delayed causal interactions using convergent cross mapping, *Scientific Reports*, 5:14750 (2015).

[17] Folium: Python data, leaflet.js maps, available at: https://github.com/python-visualization/folium, June, 2019.

[18] F. Takens: “Detecting strange attractors in turbulence”, *Dynamical Systems and Turbulence, Warwick 1980*, Springer, Berlin, Heidelberg, 1981, 366–385.

[19] G. Sugihara, et al.: Detecting causality in complex ecosystems, *Science*, 338:6106 (2012) 496–500.

[20] G. Sugihara, R. M. May: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature*, 344:6268 (1990) 734.

[21] C. H. Hsieh, et al.: Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean, *Nature*, 435:7040 (2005) 336.

[22] G. Sugihara: Nonlinear forecasting for the classification of natural time series, *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 348:1688 (1994) 477–495.

[23] E. R. Deyle, et al.: Tracking and forecasting ecosystem interactions in real time, *Proceedings of the Royal Society B: Biological Sciences*, 283:1822 (2016) 20152258.

[24] F. Grziwotz, et al.: Empirical Dynamic Modelling Identifies different Responses of Aedes Polynesiensis Subpopulations to Natural Environmental Variables, *Scientific Reports*, 8:1 (2018) 16768.

[25] R. Schlitzer: Interactive analysis and visualization of geoscience data with Ocean Data View, *Computers & Geosciences*, 28:10 (2002) 1211–1218.

[26] R. Schlitzer: Data analysis and visualization with Ocean Data View, *CMOS Bulletin SCMO*, 43:1 (2015) 9–13.
[27] S. Williams, et al.: Visualization and analysis of eddies in a global ocean simulation, in Computer Graphics Forum, Oxford, UK: Blackwell Publishing Ltd, 2011, 991–1000.

[28] C. Zhang, et al.: Extraction of vortices and exploration of the ocean data by visualization system, in International Computer Science Conference, Springer, Berlin, Heidelberg, 2012, 114–123.

[29] Y. Feng, B. Han.: Ocean Temperature Field 3D Visualization Key Technology Research Based on Pseudo-octree Model, Journal of Physics: Conference Series, 1064:1 (2018) 012064.

[30] Leaflet: JavaScript library for mobile-friendly interactive maps, available at: https://leafletjs.com, June, 2019.

[31] M. Bonzanini: Users, Followers, and Communities on Twitter, Mastering social media mining with Python, Packt Publishing Ltd., United Kingdom, 2016, 118–126.

[32] A. Chourasia, et al.: Visual Exploration and Analysis of Time Series Earthquake Data, in Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact, New Orleans, 2017, 48.

[33] P. Sarajcev, H. Matijasevic: Python in Lightning Detection Network Data Analysis, in 25th International Lightning Detection Conference & 7th International Lightning Meteorology Conference, Florida, 2018.

[34] M. B. Kennel, et al.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A, 45:6 (1992) 3403.

[35] J. L. Rodgers, W. A. Nicewander: Thirteen ways to look at the correlation coefficient, The American Statistician, 42:1 (1988) 59–66.

[36] S. M. Stigler: Francis Galton’s account of the invention of correlation, Statistical Science, 4:2 (1989) 73–79.

[37] R. W. Reynolds, et al.: Daily high-resolution-blended analyses for sea surface temperature, Journal of Climate, 20:22 (2007) 5473–5496.