

Studies of Large-Scale Data Visualization: EXTRAWING and Visual Data Mining

Project Representative

Fumiaki Araki

Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

Authors

Fumiaki Araki^{*1}, Shintaro Kawahara^{*1}, Daisuke Matsuoka^{*1}, Takeshi Sugimura^{*1}, Yuya Baba^{*1} and Keiko Takahashi^{*1}

^{*1} Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology

A research and development project, EXTRAWING, is introduced for attractive representations and transmissions of results of geophysical and environmental fluid simulations. EXTRAWING is based on the use of Google Earth. In this project, a novel technique to make results of geophysical fluid dynamics simulations possible to be represented on Google Earth is proposed. A web application program using Google Earth API is developed in order to transmit simulation results to the general public. A software tool to facilitate making of a KML content of the simulation result using the proposed technique is also developed.

Visual data mining techniques in terms of composition of transfer functions for feature emphasis and semi-automatic feature extraction of oceanic currents have been also studied, aimed at analysis of oceanic simulations. A way to make a multivariate color map to emphasize all of oceanic currents equally, taking into account flow speed distribution, is proposed. As a further approach, semi-automatic extraction of oceanic currents using one of cluster analyses based on non-hierarchical algorithms is examined.

Keywords: EXTRAWING, Google Earth, web application, tool development, feature extraction

1. EXTRAWING

EXTRAWING is a research and development project, started in February 2010, for novel and attractive representations of results of geophysical and environmental fluid simulations and effective transmission of those results to the general public. The name "EXTRAWING" is a coined acronym consisting of the initial letters of the following words: EXploring and TRAvelling the World INSIDE Geoscientific data. In the following three-subsections, we denote a novel technique of representation and software development of both a web application and a content-making tool.

1.1 Three-dimensional representation on Google Earth

Most of graphical contents for geoscientific information placed on Google Earth [1] are roughly classified into following geometric shapes: point objects (zero-dimensional form) as earthquake centers, line object (one-dimensional form) as tracking results of tagged pelagic fishes and surface object (two-dimensional form) with textures as images by satellite observations. All of these contents are written in KML [2] format which can be opened by Google Earth. On the other hand, results of geophysical fluid simulations are, in many cases, obtained as spatially three-dimensional volume datasets of scalar and vector fields. Although some kinds of three-dimensional objects such as terrain, buildings and statistical charts can be

found, it is not easy to apply such ways to represent complex features extracted from the volume data.

A straightforward way to do it is to reconstruct the features using geometric elements like lines and polygons and write down in COLLADA [3], which is a kind of modeling format and used for townscape elements including buildings. This approach is applicable to several kinds of data visualization techniques such as isosurfaces for scalar fields and streamlines for vector fields. However, the graphic object obtained by this way is not good in appearance on Google Earth, because edges of polygons of the object stand out, when the resolution of the volume data is not very high. So it is inevitable to segment the object into smaller polygons in order to smooth the edges. In case of isosurface reconstruction of three-dimensional scalar fields, particularly, increasing of smaller polygons might seriously affect performance of the Google Earth.

We propose another approach to representing features of the scalar fields. It is achieved by laminating surfaces where color contour images visualized on each slice of a volume dataset are mapped, as shown in Fig. 1. The important point here is that suitable value of opacity is assigned to each pixel of the images. A PNG format is useful to make such the images because it has an alpha channel for opacity. The advantage of this approach is that anyone can easily see the features, when viewing it from the direction approximately same as the perpendicular direction

to the surfaces. This representation requires fewer polygons than isosurface's in many cases.

For convenience, the normal directions to the laminated surfaces should be conformed to any one of the axes of simulation grids, usually corresponding to latitude, longitude and vertical directions, respectively. The surfaces perpendicular to the vertical direction, shown in Fig. 2 (a), can be represented on Google Earth, only using the 'GroundOverlay' method of KML. When one views scene from high in the sky, the aim can be accomplished only by this representation. On the other hand, when viewing it from horizontal direction, the laminated surfaces perpendicular to the latitude or longitude direction shown in Fig. 2 (b) should be needed. For these surfaces, COLLADA previously described is used partly together with KML, because the KML has no method defining these surfaces. Notice that COLLADA's coordinates are defined on the Cartesian coordinate system, whereas KML's are on the spherical coordinate system.

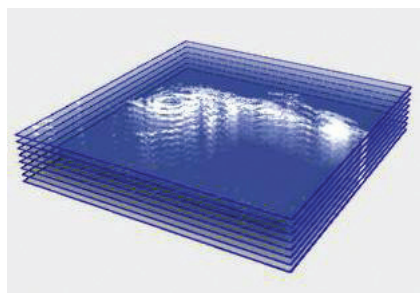


Fig. 1 A schematic figure of the proposed approach.

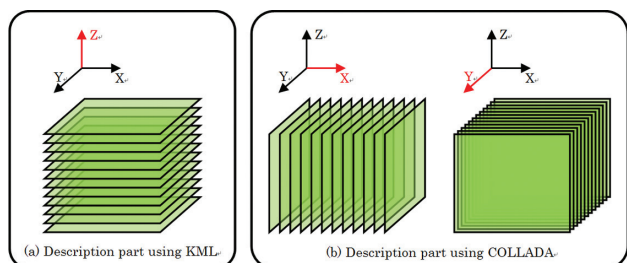


Fig. 2 Laminating direction and available data format.

1.2 Development of a web application program

The simplest way to transmit KML contents for Google Earth is to install a web server and prepare download site of those contents on the Internet. However, this way imposes burdensome tasks on the user of those contents, such as installing Google Earth, downloading the content data and reference material and operating Google Earth's functions. Such tasks are not easy for the people who are unfamiliar with Google Earth. Therefore, it is important to construct a system that everyone can easily access those datasets and comprehend the meaning of our simulation results without such difficulties. Under this consideration, we developed a web application program based on an Ajax (Asynchronous JavaScript and XML) framework with the Google Earth API [4]. The layout

of this web application is shown in the Fig. 3. In designing the program, we concerned about computer environment (monitor size, browser type and version) and prepared several text documents (Google Earth, operating instructions, FAQ, etc.).



Fig. 3 Appearance of the web application program.

As the first release on the web page, we prepared two types of contents by simulation results by MSSG model [5]; (1) temperature distribution in central Tokyo (Fig. 4 (a)) and (2) Typhoon No.4 of 2007 (Fig. 4 (b)). In the case (1), the growth and collapse streaky structure in the temperature distribution are found in the graphical area of the figure. In the case (2), three-dimensional structures of the Baiu front lying on the Japanese archipelago and the typhoon located at southwest of Kyushu is visible.

This web application was opened to the public in December 2010. The URL is as following:

<http://www.jamstec.go.jp/esc/extrawing/>

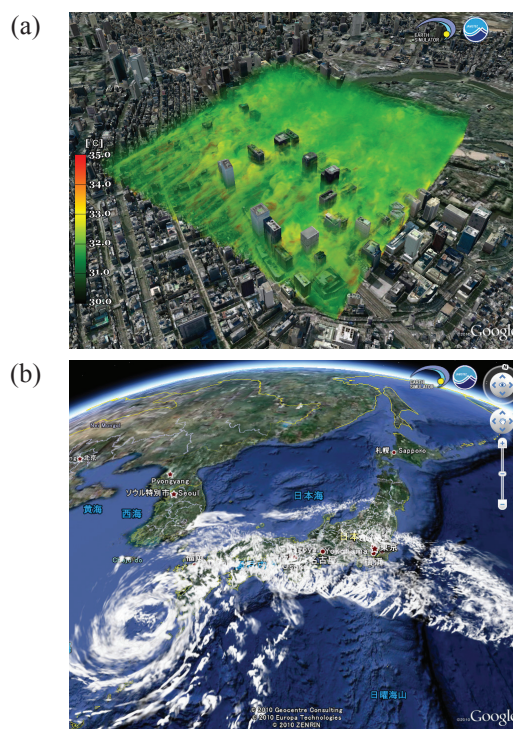


Fig. 4 Contents of EXTRAWING web page.

1.3 Development of a content making software tool

We are developing GUI-based software tool to facilitate making of volume visualization contents for Google Earth. This enables us to determine visualization parameters such as color transfer functions, visualized range, scale along each coordinate axis, etc., and generate a resulting KML file, in accordance with the way described in the subsection 1.1. This tool is built using Qt SDK [6], which provides a software development environment of cross-platform applications on the basis of a unique framework for user-interface creation. Fig. 5 shows a snapshot of the main window of this program. The right hand side (RHS) of this figure is a parameter-setting area, and the left hand side (LHS) is a viewer area of the parameter-tuning result of the content. In the RHS, four broad color-bands is the region enabling to deform four transfer functions corresponding to RGBA, respectively, using a simple mouse action. Every white curve drawn in each band region, which denotes the transfer function, is formed as a set of some Ferguson curves continuously connected. User can deform the white curve by dragging each of the control points, the endpoints of those curve segments, with the mouse. Furthermore, he/she can also add or remove control points on the curve clicking and dragging the mouse. Several color map samples are prepared as preset patterns and located on the lower side of the color-bands. These are useful to determine quickly the forms of transfer functions. Other effective functions, such as sliders for setting the number of slice planes and time sequence, are also implemented in this application. The viewer area in the LHS of Fig. 5 displays a visualization result of applying the transfer functions tuned at the right hand side. It is updated in real time, according to the latest transfer functions modified by the user. A color bar, where the range and the color map are given by the above parameter setting, is placed in the left side of this area.

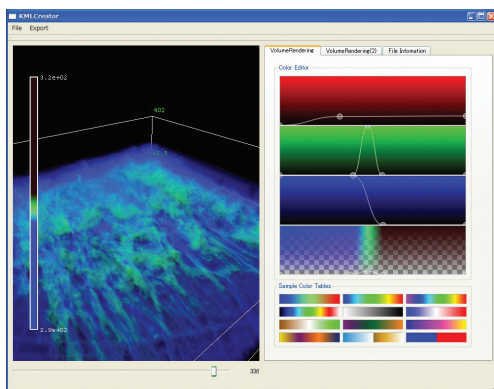


Fig. 5 A snapshot of the developed software.

2. Case studies of visual data mining for oceanic simulation

In this section, we introduce our developments of the transfer function for feature emphasis and the semi-automatic feature extraction from oceanic simulation data.

2.1. Feature emphasis technique to visualize the oceanic simulation

We propose multivariate color map in order to emphasis the features such as ocean currents and vortices. In general, a transfer function for color mapping is used as a vector-valued function which assigns one set of colors such as RGB, HSV, etc. (and also opacity in three-dimensional visualization) to an input value. The color map is, therefore, usually represented one-dimensionally like a straight bar. As examples, let the distributions of sea surface temperature (SST) and flow speed (FS) of the ocean be shown in Fig. 6 (a) and (b), respectively. Both of them are obtained by OFES (Ocean general circulation model For the Earth Simulator) [7] with horizontal resolution of 1/10 degree. For simplicity, let SST and FS be denoted with hue (H of HSV) and brightness values (V of HSV), respectively.

It is also possible to divide the transfer function into component functions and also use them in combination. Figure 6 (c) is such an example, the simple mixture of both (a) and (b). The SST at the region where the FS is faster is emphasized.

The approach has, however, disadvantage that slower currents are relatively suppressed whereas faster currents are emphasized. In order to equally emphasize them, it is important to control brightness values to fit the flow distributions as shown in Fig. 7 (a). In this case, brightness values, denoted by a dashed white line in this figure, has been set taking into account the balance of flow speed distributions between Kuroshio (faster current) and Oyashio (slower current). The result using this color map is shown in Fig. 7 (b) (in this case, the dataset of OFES with horizontal resolution of 1/30 degree is used).

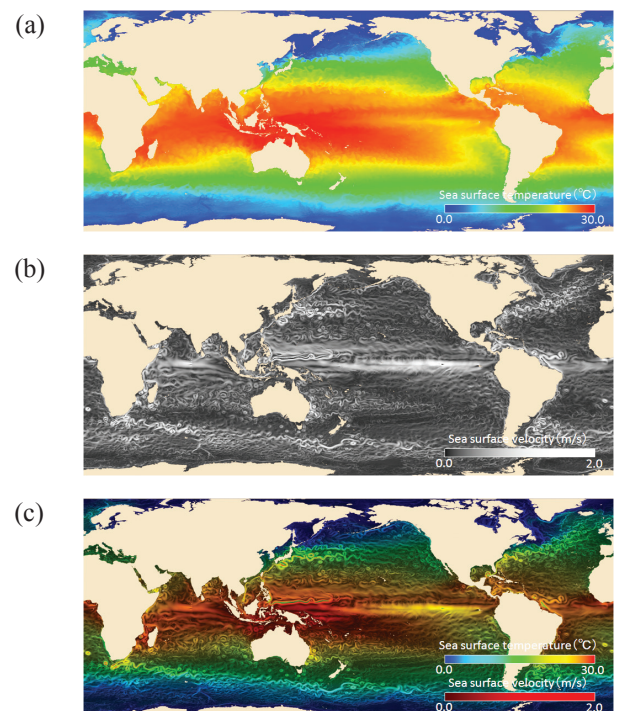


Fig. 6 (a) sea surface temperature, (b) flow speed and (c) the mixed image of them.

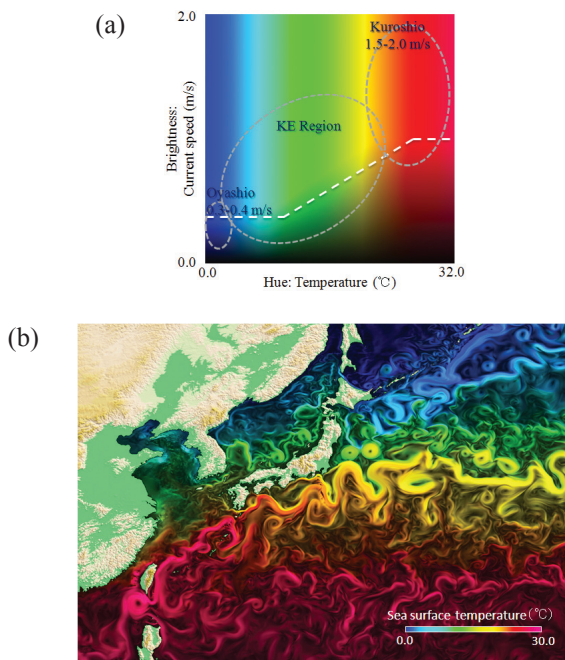


Fig. 7 (a) multivariate color map to emphasize the ocean currents and (b) the resulting image.

2.2. Semi-automatic extraction of the ocean currents

As a further approach in order to emphasize any currents equally, we try to extract oceanic currents individually from the simulation data, using one of cluster analyses based on non-hierarchical algorithms. This approach clusters elements, each of which has an attribute, a set of values coming from several variables at a corresponding simulation grid point, into several groups. An initial cluster centroid of each group should be selected manually from those elements. A cluster consists of elements being regarded as mutually similar, beginning from the centroid. Therefore, one of the important steps is how to define similarity of the elements. We use a distance defined in the attribute space of the elements as a degree of the similarity.

In the trial, we selected four variables, a temperature, a current speed and position coordinates (latitude and longitude) as the attribute variables, and sixteen centroids on the latitude-longitude subspace as shown in Fig. 8. The result of this trial is shown by color-coding in the same figure. Not only typical currents such as the Kuroshio ((1) and (2)), the Kuroshio

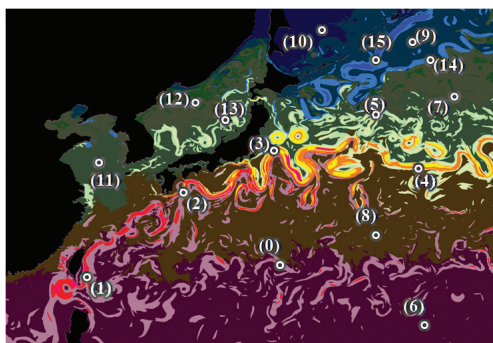


Fig. 8 Sixteen centroids and the result of non-hierarchical cluster analysis.

extension region ((3) and (4)) and the Oyashio ((14) and (15)), but also the mixed layer in the Kuroshio/Oyashio extension region in the Sea of Japan are clustered almost correctly. In continuing this work, it is important to select the centroids and evaluate the clustering results.

Summary

The EXTRAWING project was introduced. In this project, one novel representation technique was proposed and two types of developed programs, the web application and the KML content-making tool were introduced. The web application has been public to general since December 2010. The content-making program is equipped a viewer and GUI for deforming transfer functions and effectively makes KML contents based on the proposed technique which laminates color contour images with opacity.

A multivariate color map to emphasize all of oceanic currents equally, taking into account flow speed distribution, was proposed. As the further approach, semi-automatic extraction of oceanic currents using one of cluster analyses based on non-hierarchical algorithms was also examined. Typical currents around Japan islands were successfully clustered almost correctly.

References

- [1] Google Earth, <http://earth.google.com>
- [2] KML, <http://www.khronos.org/standards/kml/>
- [3] COLLADA, <http://www.khronos.org/collada/>
- [4] Google Earth API, <http://code.google.com/intl/ja/apis/earth/>
- [5] Keiko Takahashi, Xindong Peng, Kenji Komine, Mitsuru Ohdaira, Koji Goto, Masayuki Yamada, Hiromitsu Fuchigami, and Takeshi Sugimura, "Non-hydrostatic atmospheric GCM development and its computational performance", Use of High Performance computing in meteorology, Walter Zwiefelhofer and George Mozdzynski Eds., World Scientific, pp. 50-62 (2005).
- [6] Qt SDK, <http://qt.nokia.com/>
- [7] Y. Masumoto, H. Sasaki, T. Kagimoto, N. Komori, A. Ishida, Y. Sasai, T. Miyama, T. Motoi, H. Mitsudera, K. Takahashi, H. Sakuma, and T. Yamagata, "A fifty-year eddy-resolving simulation of the world ocean: Preliminary outcomes of OFES (OGCM for the Earth Simulator)", J. Earth Simulator, Vol.1, pp. 35-56 (2004).

大規模データ可視化研究 : EXTRAWING と ビジュアルデータマイニング

プロジェクト責任者

荒木 文明 海洋研究開発機構 地球シミュレータセンター

著者

荒木 文明^{*1}, 川原慎太郎^{*1}, 松岡 大祐^{*1}, 杉村 剛^{*1}, 馬場 雄也^{*1}, 高橋 桂子^{*1}

^{*1} 海洋研究開発機構 地球シミュレータセンター

地球シミュレータを用いた地球・環境流体シミュレーションによる結果を魅力的に表現し一般社会へ発信するプロジェクト、EXTRAWING について紹介する。EXTRAWING は Google Earth をベースにしている。Google Earth 上でシミュレーションデータの 3 次元可視化表現と情報発信を可能にする、カラーコンター画像を層状に積み上げた流体場の分布の表現、社会へ向けた情報発信のための Web アプリケーション開発、複雑なコンテンツを GUI 上で容易に作成可能なツールの開発、について述べる。

また、海洋シミュレーションのデータを用いて、特に海流の抽出をターゲットとしたビジュアルデータマイニングに関する 2 つの試みについて紹介する。ここでは、流速や水温の異なる複数の海流を多変数伝達関数によっていずれも同じ程度に強調する方法、およびクラスタ分析を用いた海流の半自動的な抽出法について述べる。

キーワード: EXTRAWING, Google Earth, Web アプリケーション, ツール開発, 特徴抽出