CrossMark

REGULAR PAPER

Yi Cao · Zeyao Mo · Zhiwei Ai · Huawei Wang ·
Li Xiao

An efficient and visually accurate multi-field visualization framework for high-resolution climate data

Abstract Recent advances in high-performance computing technologies, applied in climate science, are allowing increases in model complexity, model resolution, and the number of simulations run. The interactive exploration and analysis of these complex, multi-field climate data sets have been identified as one of the major current challenges in scientific visualization. For example, without direct 3D multi-field visualization, it is difficult to recognize the important correlative effects between vertical wind velocities and transport of the volumetric atmosphere. As such data have complicated 3D structures and are highly time-dependent, a visualization approach must handle these dynamic data in a highly interactive way. In this paper, an efficient multi-field visualization framework is proposed for Earth climate simulation data. A novel visualization pipeline is presented for on-demand data processing, enabling scalable handling of large-scale climate data sets. The hardware-accelerated multi-field visualization method used in the framework allows interactive and accurate visualization of multiple intersecting climate phenomena. An information-theoretic-based wind-field analysis method is also implemented within the visualization framework to help scientists gain a deeper understanding of the underlying multi-field climate data.

1 Introduction

Climate science is a discipline in which scientific progress is critically dependent on the availability of a reliable infrastructure for managing and accessing large and heterogeneous data sets. It is an inherently collaborative and multi-disciplinary effort that requires sophisticated modeling (Kindler et al. 1996) of the physical processes and exchange mechanisms between multiple Earth domains (e.g., atmosphere, land, ocean, and sea ice), and comparison and validation of these simulations (Washington and Parkinson 1986) with observational data from various sources, sometimes collected over long periods.

Climate scientists learn from such simulations by comparing modeled and observed data. A variety of grid schemes and temporal and spatial resolutions makes this task challenging, even for small data sets. As the need to understand and project climate change becomes increasingly critical, climate model complexity, model resolution, and the number of simulations required all increase. To keep pace with the growing size and complexity of climate simulations, new software or methods are required to enhance and accelerate the manipulation and analysis of both climate data and model output.

Multiphysics (Lethbridge 2005) refers to simulations involving multiple physical models such as magneto hydrodynamics, fluid structure interaction, or fluid flow combined with chemical reactions. A

Y. Cao (✉) · Z. Mo · Z. Ai · H. Wang · L. Xiao
High Performance Computing Center, Institute of Applied Physics and Computational Mathematics, Beijing, China
E-mail: cao_yi@iapcm.ac.cn

coupled climate model is a typical example of a multiphysics model, comprising interdependent models that simulate the Earth's atmosphere, ocean, cryosphere, and biosphere. Multiphysics simulations are a source of multi-field data sets containing an even larger number of components than are obtained from single-field simulations.

Nowadays, climate output is usually presented as three-dimensional, time-dependent, and multi-field data sets (Keherer and Hauser 2013). Although standard 2D visualization techniques (e.g., time charts, 2D maps, and scatter plots) are still frequently used in current climate data analyses (Nocke et al. 2008), domain scientists have found they require additional capability for the visualization of complex simulations. The visualization community has produced numerous techniques for examining single scalar, vector, and tensor fields. However, rather than considering single fields in isolation, climate scientists need to explore the interaction of several fields concurrently to gain a deeper understanding of the underlying processes and relationships, e.g., the important correlative effects between vertical wind velocities and transport of atmospheric species between layers.

As climate models become increasingly complex and the resolutions of climate models become even higher, it becomes increasingly difficult for researchers to understand the raw data. Thus, the creation of tools suitable for the exploration of these large-scale, multi-field, and time-dependent data sets is one of the most significant challenges in the field of interactive visualization and analysis. The visual effects, performance, and scalability will be vital to the applicability of the multi-field visualization systems. Hence, a multi-field visualization framework for climate simulations is presented in this paper, which is intended to accomplish the following:

1. The creation of a demand-driven visualization pipeline for the visual analysis of large data sets, to provide a scalable strategy by reducing the cost of massive data loading, filtering, and intermediate data generation within the pipeline.
2. GPU hardware-accelerated multi-field visualization for coupling the atmospheric volume with the complex three-dimensional structure of climate models to implement efficient and accurate visualization.
3. Information-theoretic-based vector field analysis of climate data to enhance the visualization quality of arrow representation by extracting the interesting characteristics from the vector field.

1.1 Related work

The process of loading and filtering data is one of the most important steps in the process of climate data visualization. Although there are predefined structures and strategies for managing model output (e.g., parallel loading and visualizing partial time-domain data for each processor from a separate file), usually these rules are inefficient for the visual exploration of high-resolution data over a relatively small area of domain. Consequently, the visualization pipeline for multi-field climate data needs to be re-designed. The visualization pipeline (Lucas et al. 1992) provides the crucial internal mechanism for managing the ingestion, transformation, display, and recording of data in many visualization applications, e.g., VisIt (VisIt 2005), ParaView (Squillacote 2007), and VisTrails (Bavoil et al. 2005). A visualization pipeline embodies a dataflow network (Ken 2013) in which the computation is described as a collection of executable modules, which are connected in a directed graph that represents how the data move between the individual modules. However, such visualization pipelines only provide general execution schemes that lack the optimization capability necessary for the application-based complex data structures of climate models.

Climate visualization has been an area of active research for many years (Kniss et al. 2002). Climate models integrated with several equations based on differential laws such as physics, fluid motion, and chemistry will output complex data sets. Many of these complexities can be subsumed under the concept of multi-fields and therefore a new visualization method that can be used to describe climate data better needs to be developed. Volume rendering (Xin et al. 2014) has been broadly applied in applications such as cloud simulation and geoscience data visualization. However, because of the level of complexity in rendering clouds (e.g., realistic shading and light scattering), considerable previous work (e.g., Hibbard et al. 1994) has been undertaken regarding non-interactive techniques for cloud rendering.

Graphic processing units (GPUs) have largely been used to accelerate volume rendering for interactive applications (Engel et al. 2001). However, conventional GPU volume-rendering algorithms (Kruger and Westermann 2003) do not provide a specific solution for multi-field rendering and therefore a simply coupled rendering method cannot reflect accurately the deep relationship between the physical phenomena

or geometric structures for multi-field data. Most visualization applications can only provide a pure volumetric-rendering environment that lacks the ability to integrate various data sources (Berberich et al. 2009; Michael et al. 2007). A few attempts have been made to implement hybrid volume rendering that can merge volume data with various multi-source spatial data in climate applications (Liang et al. 2014).

The Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT) (Williams et al. 2013) represent an advanced toolset well suited for the analysis of large-scale climate data. UV-CDAT offer multi-field rendering integrated with the VisIt software. However, VisIt is mainly implemented with the CPU ray-casting algorithm and it cannot realistically be used for interactive rendering, even when run in parallel. Liang et al. (2014) presented a GPU-based volume-ray-casting method to visualize atmospheric volume data directly with other integrated data. Unfortunately, their scheme was only suited to small sets of climate data.

Multi-field climate data consist of considerable amounts of information and thus multi-field methods often suffer from cluttering if the number of fields becomes too large. For example, the vector field is an important component of multi-field climate data. The wind-field distribution plays a critical role in characterizing regional weather or the effects of atmospheric circulation on the Earth's surface. The wind field is also a type of vector field, and the arrows method is commonly used as a form of representations in the direct visualization of the structure of a vector field. However, the arrows method will often lead to icon clustering, which causes confusion in the image.

An efficient means of reducing arrow cluttering is to concentrate on relevant structures or regions, i.e., the so-called features. Topological techniques (Pobitzer et al. 2011) can be quite effective in capturing the global features of a flow field. Helman and Hesselink (1991) proposed to analyze vector field topologies based on critical point classification. Critical points are those points that have zero vector magnitude. However, some practical issues can affect the effectiveness of this approach (Crawfis et al. 2000). A new approach for the detection of important features within a vector field is based on information-theoretic concepts (Cover and Thomas 1991). The benefit of using the entropy field is that it can highlight regions near the critical points but also regions that contain other flow features (Xu et al. 2010). However, few studies have reported on the application of entropy to the arrow method, which remains the most frequently used approach in climate data analyses.

2 Background

2.1 Climate model

A climate model is an expression of mathematical models of ancient, modern, and future climates. It uses quantitative methods to simulate the interaction between the atmosphere, oceans, land surface, and ice. A global model is a description of the climate of the Earth as a whole, for which all the regional differences are averaged, and the climate at a given location on Earth is the regional model. The climate models used in this study were the Grid-point Atmospheric Model of IAP LASG (GAMIL) (Li et al. 2013) and the Advanced Regional Eta-coordinate Model (AREM) (Wu et al. 2013). These are two climate models at global and regional scales, respectively, and both were developed by the National Key Laboratory for Numerical Modeling of Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences (CAS). GAMIL is an atmospheric general circulation model based on a finite difference dynamical core. It has been used widely for studies of the 20th-century climate change and seasonal predictions. AREM is designed for simulating moisture advection based on regional numerical prediction models. It can handle the actual topographic variations incorporated within the regional models. The AREM has been used widely for forecasting heavy rain during the flood season by China's scientific and business communities, such as in the fields of meteorology, hydrology, and the environment.

2.2 Climate simulation

To improve the numerical fidelity of climate predictions, climate science needs to increase the resolution of both modeling and simulation. Fine resolution in climate simulation usually requires high-performance computing to handle the complex theoretical models. The high resolution of GAMIL and AREM is achieved using J Adaptive Structured Meshes applications Infrastructure (JASMIN) (Mo et al. 2010), which is a parallel software infrastructure for scientific computing developed by the Institute of Applied Physics and

Table 1 Summary of on-going development of climate models

Model	Country	Horizontal resolution (km)
Global	World average (by IPCC AR5)	110
	China	220
	GAMIL, China	20
Regional	NCEP, USA	3.33/2.5
	Navy/FNMOC/NRL, USA	45/15/5
	JMA, Japan	5/2
	CMA, China	10/5
	AREM, China	8

Table 2 The climate dataset used in the experiments

Climate model	Resolution	Data size per time step
GAMIL	$1440 \times 616 \times 26$	2.7 GB
AREM	$801 \times 1001 \times 42$	900 MB

Computational Mathematics (IAPCM). The main objective of JASMIN is to accelerate the development of parallel programs for large-scale simulations of complex applications on parallel computers. Table 1 shows that the resolutions of GAMIL and AREM have exceeded or nearly reached international levels. Based on the multiphysics parallel computing middleware provided by JASMIN, the applications of these two models can solve high-resolution models with one to tens of thousands of processor cores.

3 Data description

The climate data sets used in this work were generated from the applications of GAMIL and AREM (Table 2). GAMIL is modeled based on a zonally uniform latitude–longitude grid and its vertical grid represents the height above the Earth's surface. AREM is built with an E-grid (Yu 1995), which is used to guarantee high-precision calculations of the divergence and vortices. The data sets of GAMIL and AREM are both generated in HDF5 format. To help researchers navigate the global climate data intuitively, a coordinate transformation from spherical to Cartesian coordinates was executed as part of this work.

4 Multi-field visualization framework for high-resolution climate data

4.1 Visualization pipeline for visual analysis of large data sets

To improve the pipeline performance, such as the data loading and processing, when researchers perform complicated spatial explorations (e.g., using clip, slice, or iso-surface filters) in a multi-field visualization system, a novel demand-driven visualization pipeline is employed. Similar to VisIt (2005), this proposed pipeline consists of a file reader (source), several operators (data filters), and an image renderer (sink), as shown in Fig. 1a. Execution is initiated at the bottom of the pipeline. The sink's upstream modules satisfy this request by first requesting data from their own upstream modules and so on up to the sources. Once execution reaches a source, it produces data and returns execution back to its downstream modules. The execution eventually unrolls back to the originating sink. The advantage of this pipeline is that it can launch an execution in response to requests for data.

A patch-based data structure (Mo et al. 2010) is the crucial data structure of the GAMIL and AREM applications. The patch is a local region on which the simulated system and related physical variables are well defined. Figure 1b shows a two-dimensional structured mesh consisting of 20×20 cells. It is decomposed into seven patches and each patch is defined by a logical index box. In the climate simulation output, one or several patches can be incorporated into a block for fitting the buffer size of the storage (e.g., all the patches in Fig. 1b can be called a single block). When adopting parallel visualization of these climate data in the proposed system, each processor will read and deal with the block data.

Based on this mesh structure, a dual-granularity optimized strategy is presented for the visualization pipeline. This can be very efficient for the visual exploration of high-resolution spatial data in a relatively
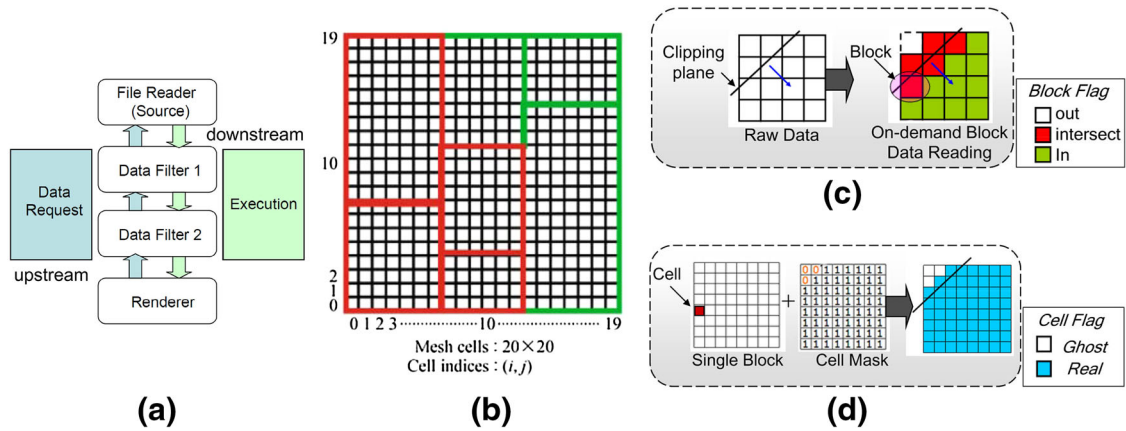
Fig. 1 The application of GAMIL and AREM. **a** Pipeline overview, **b** application mesh structure, **c** block-based pre-filtering strategy, **d** cell-based fine-filtering strategy

small area. First, a block-based pre-filtering strategy is used in the upstream modules. This is a type of coarse-granularity data filtering for the optimization of data loading. When a user performs a spatial exploration procedure, such as using a clip filter (as shown in Fig. 1c), the blocks outside the clipping plane do not need to be loaded into the system memory. Hence, all the blocks can be marked according to the attributes of the clip filter and only the required block data loaded into the pipeline during the downstream execution. When applying the block-based strategy, each block will be marked with a flag that corresponds to its internal, intersecting, or external attributes.

After applying the coarse-level data filtering, a cell-based fine-filtering strategy is subsequently introduced in the upstream modules. This strategy is performed inside each block, and it is a type of fine-granularity data filtering for the optimization of data processing in the visualization pipeline. The purpose of adopting this strategy is that when a user uses some non-orthogonal spatial exploration procedure on the entire structured mesh (e.g., an inclined clipping), usually those meshes inside the clipping plane will be converted to the unstructured mesh in the pipeline. This leads to considerable overheads regarding the memory size and execution times of the visualization algorithms. As shown in Fig. 1d, each cell in a block data set will be marked according to the attributes of the clip filter. This is helpful in ensuring that the original mesh structure remains unchanged in the downstream execution. Hence, large amounts of intermediate data can be reduced in the pipeline, which saves memory and data processing time.

4.2 Hardware-accelerated multi-field visualization

In this paper, the multiple data fusion is implemented with our multi-field visualization framework, as shown in Fig. 2a. In the framework, each variable field is mapped into a plot, which is responsible for the corresponding rendering of a physical phenomenon. During the multi-pass rendering stage, each plot will be fed into the pipeline in sequence. In the first rendering stage, the framework makes it a priority to send geometry-rendering plots into the pipeline. The multi-field data fusions between the geometries are performed automatically by the graphics pipeline based on the depth information. Ultimately, intermediate data such as the RGBA image of geometry and the corresponding depth image can be obtained. In the second stage, the volume plot and the intermediate data are both sent into the pipeline. The intermediate geometries are then blended with the volume data during ray accumulation.

The visualization accuracy of multi-field data is obtained by deferred shading (Hargreaves and Harris 2004) strategy. In our framework, the deferred shading is used to control the execution of the GPU shader pipeline, such that the image-based multi-field data fusion is delayed until the ray-accumulation stage of the GPU volume rendering. In contrast to the traditional image level intermixing, deferred shading provides a mechanism for the data intermixing in accumulation level (Cai and Sakas 1999). This ensures that the correct depth ordering during occlusion is achieved in the final image. The high-precision depth comparison is the key factor to maintain accuracy when applied deferred shading. But the default depth value retrieved from OpenGL buffer is a non-linear scale, which give more precision close to the eye and less precision far
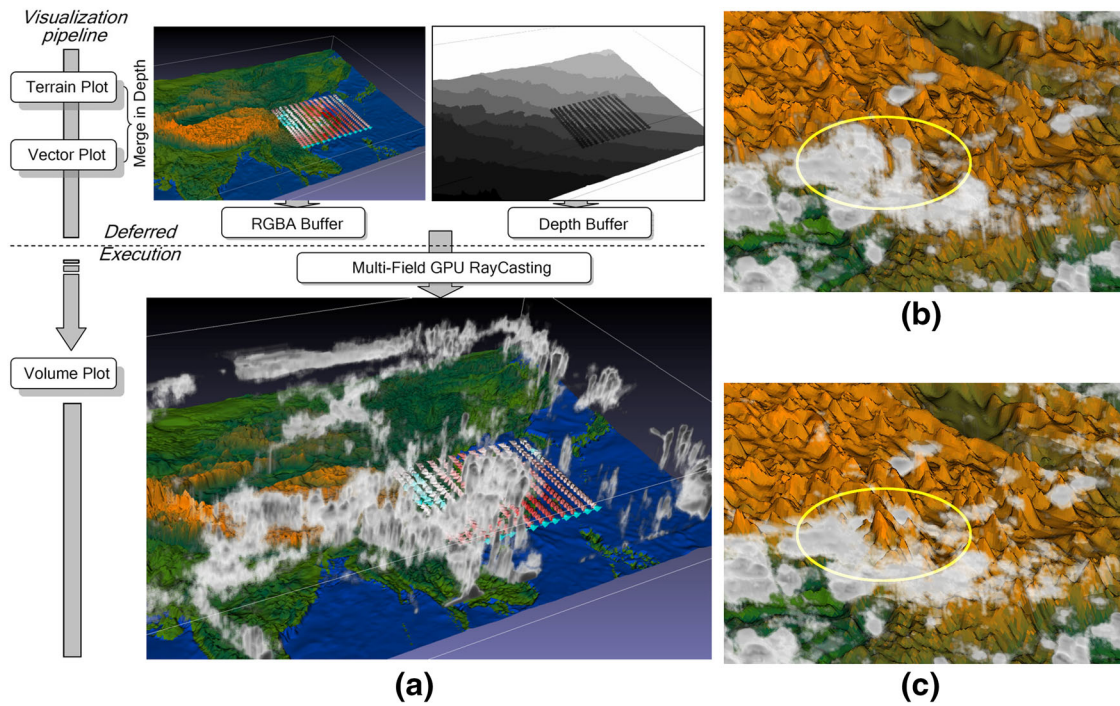
Fig. 2 The accurate multi-field visualization schematic. **a** Visualization framework overview, **b** incorrect result with image level intermixing, **c** the visual accurate result with our method

from the eye. Non-linear depth introduces the occlusion problems because it is impossible for the ray to judge the correct distance in the scene. Therefore, according to the near and far clipping plane in perspective view, we implemented the effective linear depth mapping to provide accurate depth comparing. A contrast effect between image level and our multi-field intermixing can be seen in Fig. 2b, c. As can be seen from Fig. 2c, the meaningful relations of mutual occlusion between the mountain terrain and the translucent cloud volume are reflected.

For scalable visualization of high-resolution climate data, a multi-node hardware-accelerated parallel rendering scheme is also implemented. This scheme incorporates three parts: (1) a hardware-accelerated multi-field data renderer in the intra-node layer, (2) a multi-node parallel volume-rendering scheme in the inter-node layer, and (3) a scalable parallel-image compositer in the inter-node layer. The multi-field data renderer is handled by the CPU and GPU in each parallel node. On top of the intra-node layer, the multi-node parallel volume-rendering algorithm performs data partitioning and schedules the workload for each renderer. The parallel-image compositer deals with the massive communication between the parallel nodes, and it provides the final rendered image to the viewer.

The partition data typically have to overlap with each other to prevent seams at the block boundaries. The amount of overlap depends on the filtering scheme used. For example, if tri-linear interpolation were requested, the blocks would have to overlap by a layer of voxels one voxel thick in each direction; if cubic interpolation were requested, three layers would be needed in each direction. Another factor that determines the width of the block overlap is the gradient computation. When the gradient is computed, the x, y, and z dimensions for each voxel are graded over $x + 1$ to $x - 1$, $y + 1$ to $y - 1$, and $z + 1$ to $z - 1$. If any of these values fall outside the brick, the overlap might have to be greater. A gradient-oriented overlap strategy is implemented in the proposed parallel-visualization system.

The multi-field GPU ray-casting pseudocode in GLSL style is illustrated below:

```
float linearDepth ( float farPlane, float nearPlane, float d )
  {   return (2*nearPlane)/(farPlane+nearPlane-d*(farPlane-nearPlane)); }
ivec3 texSize        = textureSize3D(volume_tex, 0)
float texThickness = 1/float(texSize)
float rayDepth       = linearDepth(viewfar, viewnear, nolinearDepth)
For each pixel() {
        // step 1: compare current ray depth with geometry depth
        if(rayDepth >= depth of geometries){
            // blending geometry image with the ray-casting result
            // termination }
        else
            // accumulate the rayDepth in ray direction    }
        //step 2:   GPU ray-casting
        sample                = texture3D(volume_tex, vec).xy;
        color_sample = texture2D(ColorMap, sample);
        // step 3:   skip the overlapping thickness of the texture data
        if( (ray_entry>=texThickness) &&(ray_entry<=1-1.0*texThickness))
        {// ray accumulation}
        ray_entry      = ray_entry + delta_dir;
        length_acc = length_acc + delta_dir_len;
        // alpha control
        //   ray length control    }
```

4.3 Vector field analysis in climate data

The representation by arrows is a direct visualization method that is commonly used to reflect the structure of a vector field. However, the arrow method has certain limitations, e.g., the arrows can lead to icon clustering, causing confusion in the image. Therefore, if the interesting characteristics or distribution can be extracted from the vector field, the contradiction between accurate descriptions of a vector field and confused images caused by overlapping icons can be overcome.

The typical topological structure of a vector field is formed by critical points, as shown in Fig. 3, however, some practical issues can affect the effectiveness of the topology approach. To obtain a distribution of characteristics with greater uniformity, similar to the work introduced by Xu et al. (2010), we apply Shannon's entropy $H(x)$ to measure these characteristics of vector fields. If velocity direction information is used as a random variable X, we can examine the corresponding histogram of the distribution of these vectors. The vectors around the critical points always spread across a relatively large number of
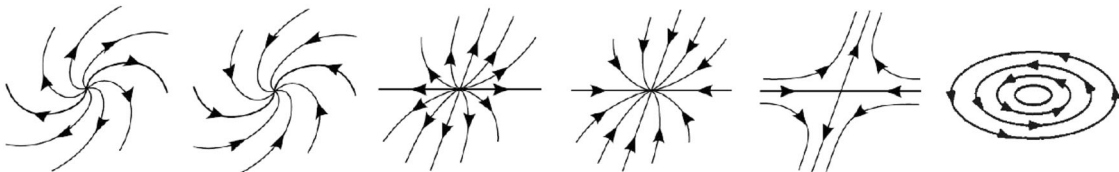

Fig. 3 The typical topological structures of vector fields

histogram bins; hence, the entropy will have a relatively high value. To construct the probability distribution function of a vector field, the angle of the vector at each mesh node is used as a random variable. If $C(x_i)$ is taken as the number of vectors in bin x_i, then the probability of vectors in bin x_i can be computed as following equation:

$$P(x_i) = \frac{C(x_i)}{\sum_{i=1}^{n} C(x_i)} \tag{1}$$

An eddy is a current of water or air, moving contrary to the direction of the main current, particularly characterized as a circular motion. Eddies that are between about 10 and 500 km in diameter are known as mesoscale eddies, and they play a major role in atmospheric circulations. It would be meaningful if the areas of such mesoscale eddies could be captured during wind field visualization. However, not only mesoscale eddies would be captured but also relatively small-scale eddies, because the velocity direction information is only used for the statistical characterization. In fact, both the magnitude and the direction of surface winds play important roles in determining eddy formation. Usually, higher velocity magnitudes are considered indicative of the presence of a mesoscale eddy and therefore, they are retained in the analysis. It is possible to evaluate the representativeness of mesoscale eddies for the wind visualization by combining the statistical analysis of the direction field and the magnitude field. To consider the distribution of the two fields, Shannon's joint entropy can be used. Assuming the velocity direction is represented as a random variable X, and the velocity magnitude is denoted as a random variable Y, the joint entropy of (X, Y) is defined as:

$$H(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y). \tag{2}$$

To calculate the joint entropy between the two fields of velocity, a two-dimensional histogram is constructed, where each field is represented by an axis. The histogram cell (x_i, y_i) is incremented by one if the velocity direction x of vector and the velocity magnitude y of vector from the same location of the domain fall into the histogram bin (x_i, y_i). The value $H(x, y)$ is a measure of the average uncertainty of the random variables x and y. It indicates how much information in the two fields of velocity remains unknown. The joint entropy will tend toward a large value if areas of mesoscale eddies have been detected. Figure 4 shows the vector analysis workflow for the proposed arrow visualization method. Figure 5 shows the
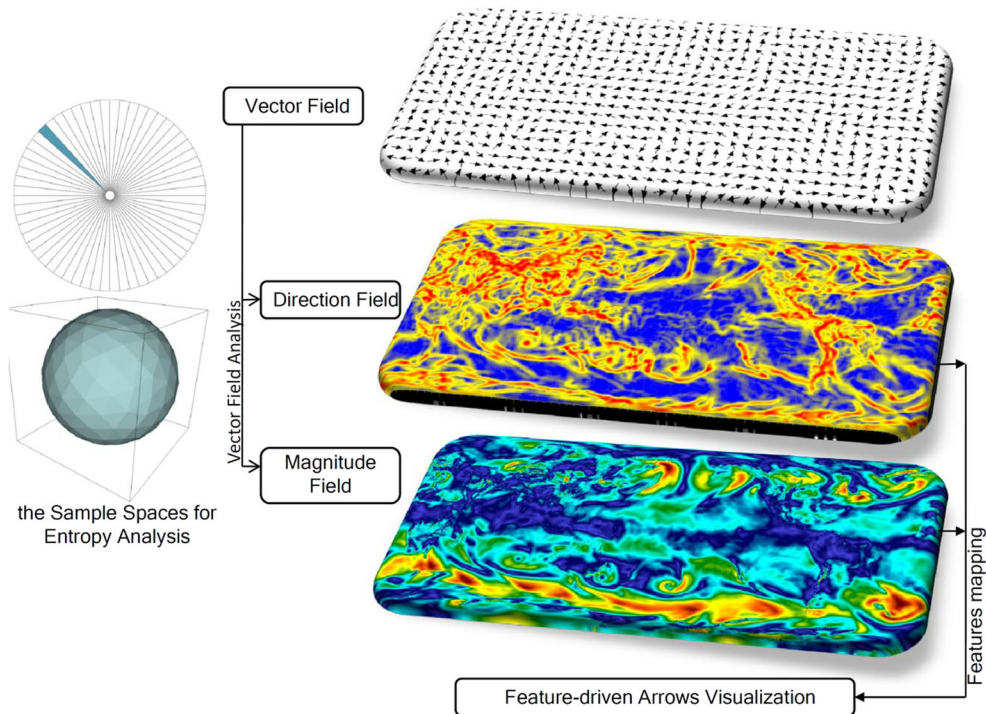

Fig. 4 Workflow of feature-driven arrow-visualization method. The GAMIL wind-field distribution is shown
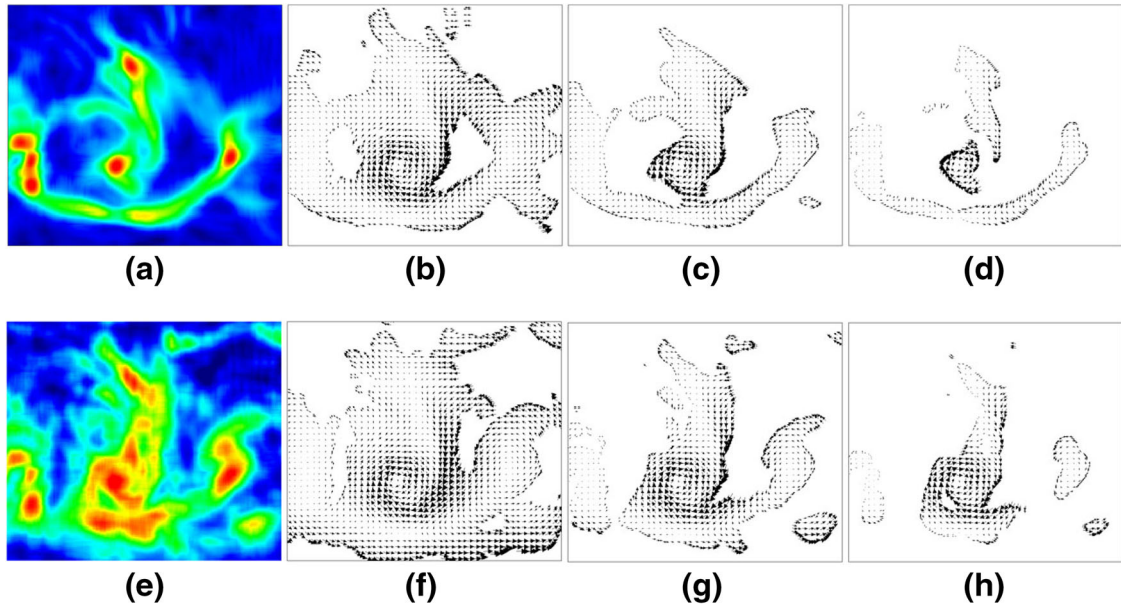
Fig. 5 Comparisons between the entropy method (*top*) computed from velocity direction and the joint entropy method (*bottom*) computed from velocity direction and magnitude. **a** Entropy field (0–5.4), **b** $t > 2$, **c** $t > 2.8$, **d** $t > 3.5$, **e** joint entropy field (0–7.3), **f** $t > 4$, **g** $t > 5$, **h** $t > 5.8$
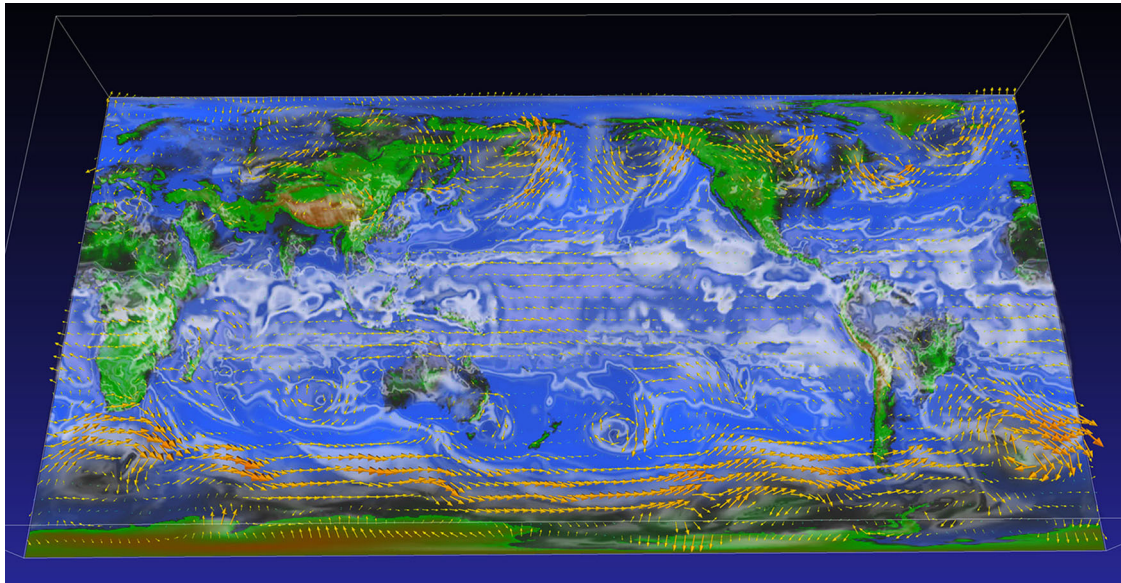

Fig. 6 The multi-field visualization result of the global climate model simulation

Table 3 Interactive Performance (second/frame)

	1 GPU node	8 GPU nodes	1 CPU core	128 CPU cores	256 CPU cores
Multi-field rendering	0.85	0.46	137.74	8.72	4.09

characteristics of the distributions of wind fields with different threshold values, where t is the threshold value of the statistical field. By comparison with Fig. 5f–h, it can be seen that the areas of mesoscale eddies are captured more accurately using the joint entropy method, and the results are less affected by relatively small eddies.
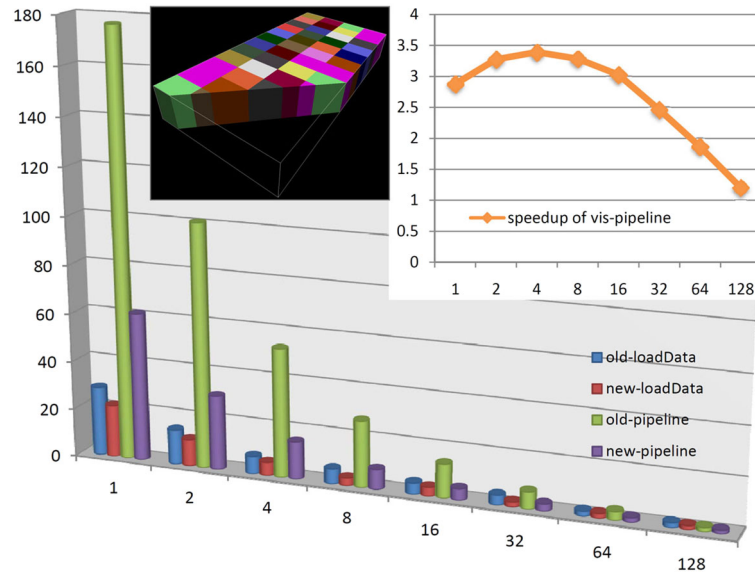
Fig. 7 The performance of the proposed visualization pipeline
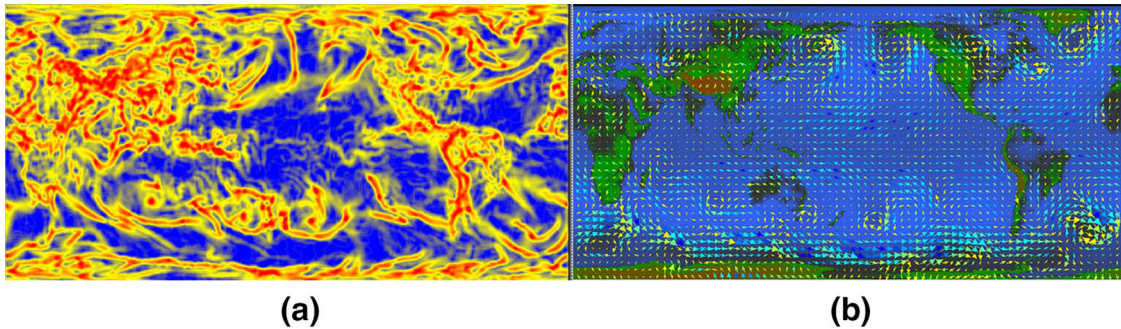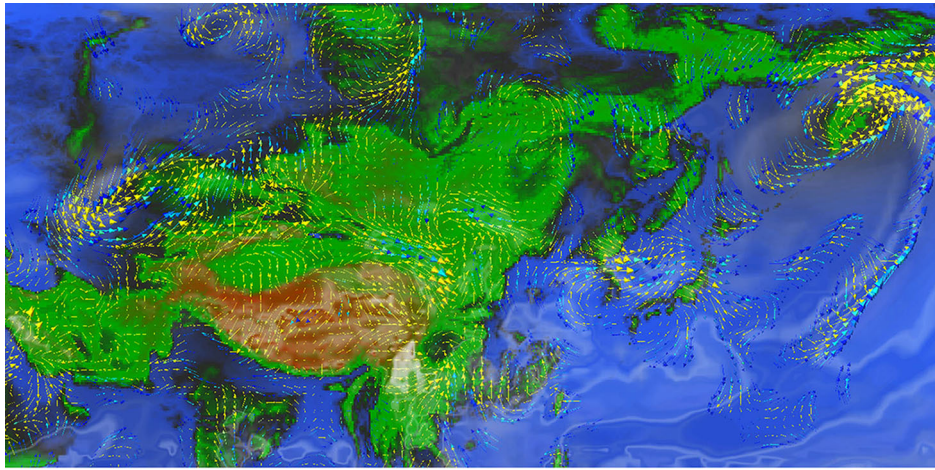

Fig. 8 The entropy distribution of wind field characteristics in GAMIL. **a** The characteristics of vectors is mapped into the *arrow colors*, **b** the *arrow* based visualization result

5 Results and discussion

The multi-field climate data visualization was run on Inspur's high-performance server system (Inspur 2015), which is suited to large-scale high-performance computing in the science and engineering computing fields. The graphics card in the server node was an NVIDIA Quadro K6000 graphics card with 12 GB RAM. The core volume ray-casting was written in C++ and GLSL shaders. The proposed visualization pipeline supports the remote visualization mode, i.e., the massive computing required for the visualization was performed on a remote server node, and users were able to explore and analyze the results interactively on their local desktop computers.

Figure 6 shows the global climate result visualized using the multi-field visualization framework presented in this paper. The figure contains three data fields: geographic elevation, distribution of atmospheric water vapor in January, and the 850-hPa wind field. The volume-rendering method was used to describe the translucent water vapor data. Because of the visual means of multi-field coupling, the correlative effects between the transport of the volumetric atmosphere and wind velocities at spatial resolutions of 20 km are easily discernible. Cloud simulation is always an important and difficult component within a climate model. Using the proposed visualization framework, the high-quality results clearly exhibit the high-resolution cloud simulation capability of GAMIL. Figure 6 shows that the proposed visualization method can reasonably represent important weather processes, such as mesoscale eddy activities at high latitudes in the Southern Hemisphere, strong westerlies and storm activities over the Southern Ocean, and equatorial easterlies and equatorial clouds.
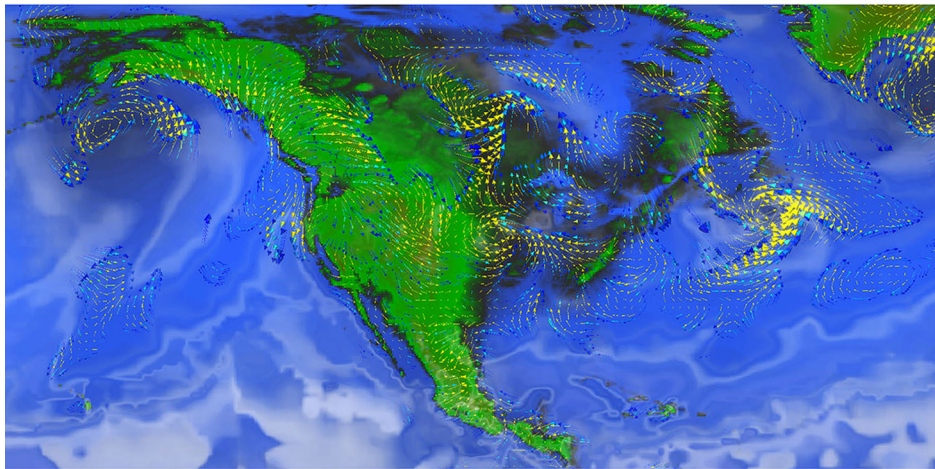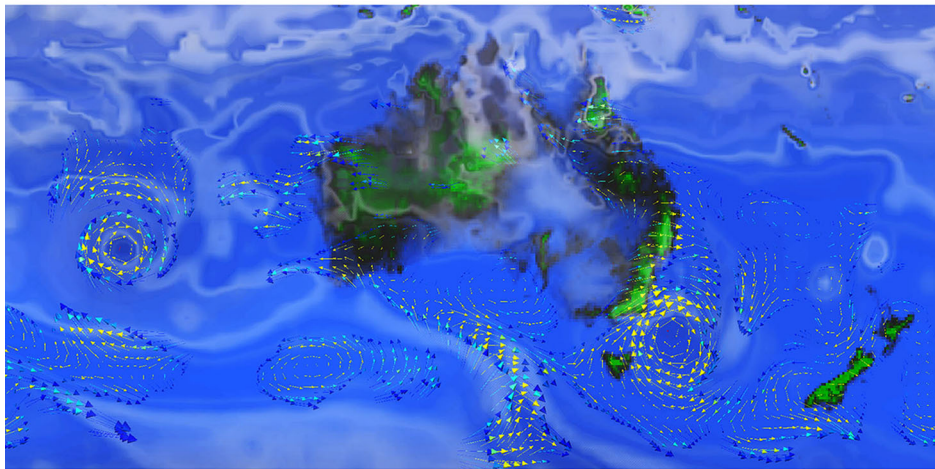
Fig. 9 The joint-entropy distribution of wind-field characteristics in GAMIL. **a** The distribution characteristics of wind-field over Asia. **b** The distribution characteristics of wind-field over North America. **c** The distribution characteristics of wind-field over the Southern Ocean
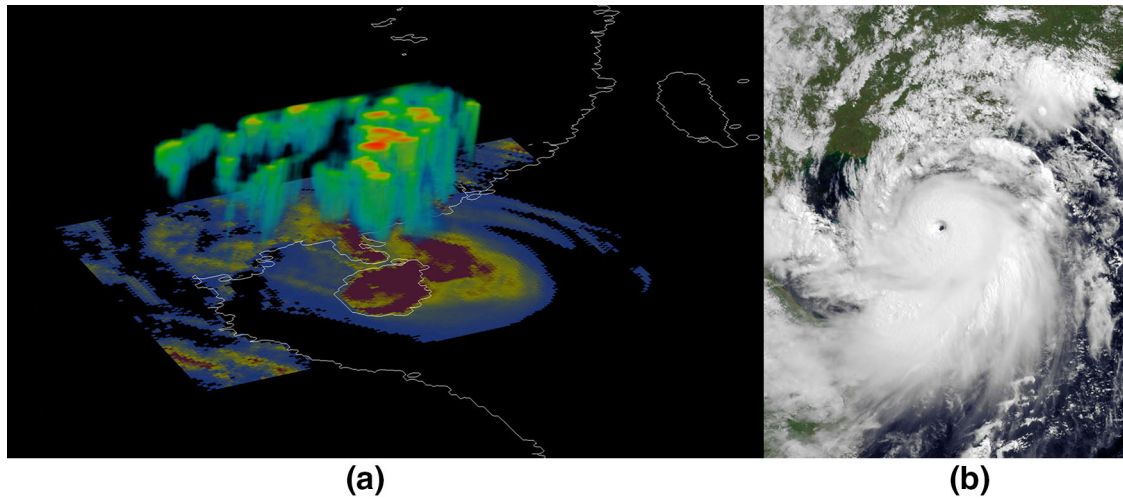
Fig. 10 **a** The distribution of *vertical thickness* of cloud system and precipitation, **b** the satellite cloud image
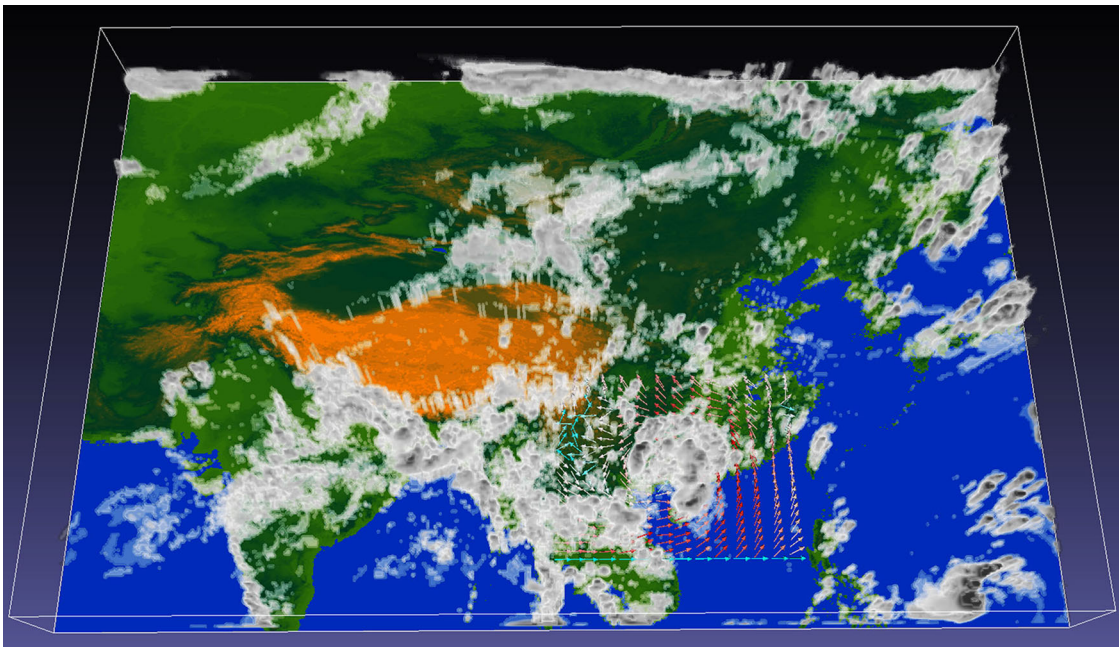

Fig. 11 The multi-field weather data visualization of Super Typhoon Rammasun (2014)

The GAMIL climate data were generated from tens of thousands of processor cores in the Tianhe-1 supercomputer at the National Supercomputer Center in Tianjin. However, Table 3 shows that the rendering performance of the cloud visualization in Tianhe-1 cannot provide suitable interactive visual analysis capability, even when using hundreds of CPU cores. The proposed method can provide a satisfactory interactive experience for climate data analysis using the GPU node of the Inspur high-performance computer. The test results can be seen in Table 3. The GAMIL data sets were used for the performance test and the image resolution reaches 1920×1080 pixels. The results show that the performance of the proposed method was at least one or two orders of magnitude faster than the traditional method that uses a large number of CPU cores on the Tianhe-1 supercomputer at the National Supercomputer Center in Tianjin. However, because of the remote parallel rendering is used in our distributed environments, nearly 80 % of the time is dominated by the network delay. Hence, the parallel efficiency of the GPU test is about 23 %.

Figure 7 shows the performance of the proposed visualization pipeline. The "old-loadData" and "old-pipeline" represent the data loading and data processing times, respectively, needed in the previous non-optimized pipeline. The "new-loadData" and "new-pipeline" correspond to the performance in the optimized pipeline. When performing a clipping spatial exploration, as shown in Fig. 7, the results indicate that the proposed visualization pipeline can provide a speedup of at least three times. Although the curve of speedup declines when additional CPU cores are used, it is meaningful because the parallel scale of the GPU-based visualization is usually smaller than in the CPU approach.

Figure 8a shows the entropy distribution of the wind-field direction extracted from the GAMIL data through the information-theoretic based data analysis method in the proposed visualization framework. Observation shows that the entropy will be larger (red areas in Fig. 8a) when the wind field reflects specific characteristics, i.e., vortex phenomena. In contrast, when the entropy value of the wind field is small (shown by the blue areas in Fig. 8a), the corresponding vortex phenomena will not be obvious. This corresponding relation between entropy and the characteristics of the wind field can be seen by comparing Fig. 8a, b. Using the entropy-based method for visualization with large sets of simulation data, the massive and confusing array of vector symbols will be less influential, and a more objective understanding of the global wind distribution patterns will be achievable. Figure 9 illustrates the distribution of the wind-field characteristics using joint-entropy-based vector field analysis, which clearly shows the important mesoscale eddies in the Northern and Southern hemispheres.

Figure 10 shows the three-dimensional regional weather pattern of Super Typhoon Rammasun (2014), simulated by AREM and visualized using the multi-field visualization framework presented in this paper. Typhoon Rammasun made landfall near Wenchang on the island province of Hainan on 17 July 2014. Rammasun is considered the severest typhoon to hit the city in 41 years. Overall, 51,000 homes were destroyed, at least 62 people were killed throughout the country, and economic losses amounted to US $6.25 billion. Using the proposed multi-field visualization method, the distribution and movement of the typhoon cloud systems are clearly represented. The volume-rendering method was used to describe the red translucent typhoon cloud systems. Figure 10a shows the distribution of the vertical thickness of the cloud systems in the near-Earth space and the corresponding precipitation intensity distribution on the ground. Figure 10b is the satellite cloud image captured by NASA. Based on the visualization pipeline for on-demand data processing presented in this paper, the proposed visualization framework could implement interactive visual analysis for both global and regional climate data sets. Figure 11 shows that the position of the vortex center of the typhoon could be predicted using the 48-h weather forecast generated by AREM.

6 Conclusions

Climate scientists learn from simulations by comparing modeled and observed data. A variety of grid schemes and temporal and spatial resolutions makes the task of climate visualization challenging, even for small data sets. The multi-field climate data visualization framework presented in this paper has been tested using two applications: GAMIL and AREM. The user case study with these two applications proved that the multi-field data fusion method presented here could be useful and effective with large data sets. GPU-accelerated volume rendering is the critical method involved in the multi-field visualization. Further study is required regarding scalable strategies for GPU-based multi-field visualization methods, such as feature-driven data reduction, multi-resolution rendering, and load-balancing strategies for parallel rendering.

References

Bavoil L, Callahan SP, Crossno PJ, Freire J, Scheidegger CE, Silva CT, Vo HT (2005) VisTrails: enabling interactive multiple-view visualizations. In: Proceedings of IEEE visualization, pp 135–142

Berberich M, Amburn P, Moorhead R, Dyer J, Brill M (2009) Geospatial visualization using hardware accelerated real time volume rendering. In: OCEANS 2009, MTS/IEEE Biloxi—Marine Technology for our future: global and local challenges, pp 1–5, 26–29

Cai W, Sakas G (1999) Data intermixing and multivolume rendering. Comput Graph Forum 18(3):359–368

Cover TM, Thomas JA (1991) Elements of information theory, 99th edn. Wiley, New York

Crawfis R, Shen H W, Max N (2000) Flow visualization techniques for CFD using volume rendering. In: 9th international symposium on flow visualization, Edinburgh, Scotland

Engel K, Kraus M, Ertl T (2001) High-quality preintegrated volume rendering using hardware-accelerated pixel shading. In: Proceedings of the ACM Siggraph/Eurographics workshop on graphics hardware 2001, pp 9–16

Hargreaves S, Harris M (2004) Deferred rendering. NVIDIA Corporation, Santa Clara

Helman JL, Hesselink L (1991) Visualizing vector field topology in fluid flows. IEEE Comput Graph Appl 11(3):36–46

Hibbard W, Paul B, Santek D, Dyer C, Battaiola A, Voidrot-Martinez MF (1994) Interactive visualization of Earth and space science computations. Computer 27:65–72

Inspur. http://www.inspur.com/. Accessed 13 Aug 2015

Keherer J, Hauser H (2013) Visualization and viusal analysis of multi-faceted scientific data: a survey. IEEE Trans Vis Comput Graph 19(3):495–513

Ken M (2013) A survey of visualization pipelines. IEEE Trans Vis Comput Graph 19(3):367–378

Kindler T, Schwan K, Silva D, Trauner M, Alyea F (1996) A parallel spectral model for atmospheric transport processes. Concurr Pract Exp 8:639–666

Kniss J, Hansen C, Grenier M, Robinson T (2002) Volume rendering multivariate data to visualize meteorological simulations: a case study. In: IEEE visualization symposium, pp 189–194

Kruger J, Westermann R (2003) Acceleration techniques for GPU-based volume rendering. In: Proceedings of the 14th IEEE visualization 2003(VIS'03), Washington, DC, USA, pp 287–292

Lethbridge P (2005) Multiphysics analysis. Ind Phys 1:26–29

Li LJ et al (2013) Evaluation of grid-point atmospheric model of IAP LASG version 2 (GAMIL 2). Adv Atmos Sci 30:855–867. doi:10.1007/s00376-013-2157-5

Liang JM, Gong JH, Li WH, Ibrahim AN (2014) Visualizing 3D atmospheric data with spherical volume texture on virtual globes. Comput Geosci 68:81–91

Lucas B, Abram GD, Collins NS, Epstein DA, Gresh DL, McAuliffe KP (1992) An architecture for a scientific visualization system. In: Proceedings of IEEE visualization, pp 107–114

Michael D, Harris Jr FC, Sherman WR, McDonald PA (2007) Volumetric visualization methods for atmospheric model date in an immersive virtual environment. In: Proceedings of high performance computing systems (HPCS'07), Prague, Czech

Mo ZY, Zhang AQ, Cao XL, Liu QK, Xu XW, An HB, Pei WB, Zhu SP (2010) JASMIN: a parallel software infrastructure for scientific computing. Front Comput Sci China 4(4):480–488

Nocke T, Sterzel T, Bottinger M, Wrobel M (2008) Visualization of climate and climate change data: an overview. In: Proceedings of Digit Earth Summit on Geoinformatics (2008), pp 226–232

Pobitzer A, Peikert R, Fuchs R, Schindler B, Kuhn A, Theisel H (2011) The state of the art in topology-based visualization of unsteady flow. Comput Graph Forum 30(6):1789–1811

Squillacote AH (2007) The ParaView guide: a parallel visualization application. Kitware Inc. http://www.paraview.org. Accessed 13 Aug 2015

VisIt User's Manual, Lawrence Livermore National Laboratory, October 2005, technical report UCRL-SM-220449

Washington WM, Parkinson CL (1986) An introduction to three-dimensional climate modeling. Oxford University Press, Oxford

Williams D et al (2013) The Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT): data analysis and visualization for geoscience data. IEEE Computer 99. http://doi.ieeecomputersociety.org/10.1109/MC.2013.119.3

Wu SQ, Xu YP, Hu BH et al (2013) The application and experimentation of a new hydrostatic extraction of reference atmosphere in AREM. Torrential Rain Disasters 32(2):132–141

Xin L et al (2014) Efficient quadratic reconstruction and visualization of tetrahedral volume datasets. J Vis 17(3):167–179

Xu LJ, Lee TY, Shen HW (2010) An information-theoretic framework for flow visualization. IEEE Trans Vis Comput Graph 16(6):1216–1224. doi:10.1109/TVCG.2010.131

Yu R (1995) Application of a shape-prreserving advection scheme to the moisture equat ion in an E-grid regional Forecast model. Adv Atmos Sci 12(1):13–19