Visual Analysis of Relevant Fields in Geoscientific Multifield Data

Anatoliy Antonov and Lars Linsen

Abstract Exploration of multifield geoscientific data sets is a complex task involving the investigation of individual fields and correlations between fields. We present an approach to analyze the importance of fields and their correlations in multifield datasets by treating given or derived fields as multidimensional objects and projecting these objects to a 2D space, and visually investigating the fields using the projected layout. We demonstrate how our approach supports the analysis of atmospheric simulation data in two different settings.

Keywords Visualization · Interactive visual data analysis · Multifield · Dimension projection

1 Introduction

Present technical capabilities enable scientists to produce much more data than can be carefully analyzed. In addition, the visualization techniques that are commonly employed in the geoscientific research do not allow to effectively extract features from more than a few variables: the independent variables are usually visualized side-by-side, and changes in values of a single variable over time are usually depicted with animations. Both approaches do not scale well and impose an excessive cognitive load on the scientist.

We present an approach to analyze relationships between scalar fields, with the goal of finding and choosing fields, both simulated and derived, that add most information to each other. When applied to climate data such an approach allows, among others, for the detection of seasonal changes. For example, temperature

A. Antonov (✉) · L. Linsen
Jacobs University Bremen, Bremen, Germany
e-mail: an.antonov@jacobs-university.de

L. Linsen
e-mail: l.linsen@jacobs-university.de

values in January and February are highly correlated, while temperatures in January and July are more distinct.

The main contribution of this paper is the design and application of field-based projection methods to geoscientific multifield datasets.

Our approach is based on two core ideas. First, we derive a number of fields that could potentially be of interest for the exploration of the data. Second, we define a global distance measure between pairs of fields, and generate a difference-based overview of all fields. Here, we interpret fields as high-dimensional data points and project them to a 2D space, where they are visualized in the form of a scatterplot.

We demonstrate how our approach supports the analysis process of a scientist in an interactive visual set-up.

2 Related Work

When exploring a multifield dataset, attention is often given to measuring per-point similarity between fields (Edelsbrunner et al. 2004; Nagaraj and Natarajan 2011; Sauber et al. 2006). Many of these approaches are based on gradients. Gosink et al. (2007) used the normalized dot product between gradients and visualized it over statistically important isosurfaces of a third field. Sauber et al. (2006) introduced the gradient similarity measure (GSIM), which is combined from directional similarity and magnitude similarity. Nagaraj et al. (2011) developed a measure as the norm of the matrix that comprises the gradient vectors, and showed that it is robust to noise in input fields.

Projection methods are commonly used to describe similarities between spatial samples of a multifield, by placing points with similar multivariate attributes close to each other in respective visualizations. There exist various linear (Kandogan 2001) and non-linear (Jänicke et al. 2008; Sammon 1969) projection algorithms.

Our approach, however, depicts global similarities between data fields. Thus, it is closer to Turkay et al. (2011, 2012), who introduced a dual-space analysis of multivariate data using linked visualizations of the item space (where objects are entities represented by their values in different attributes) and dimension space (where objects are attributes represented by their values for the different entities). In their approach, the item space represented results of multivariate analyses. The dimension space was visualized with scatterplots, showing either multidimensional scaling (MDS) projection results with a correlation-based distance measure, or 2D scatterplots of two selected dimension statistics (e.g. mean value vs. standard deviation).

Yuan et al. (2013) introduced a dimension projection matrix, which builds on the concept of scatterplot matrices by assigning a group of dimensions to each row or column and using projections instead of simple 2D plots. It leverages symmetric property of the matrix to create a dual space visualization: the cells in the upper triangle of the matrix contain projections of items in combined set of respective dimensions, and the lower triangle contains projections of dimensions themselves.

Neither Turkay et al. (2011, 2012) nor Yuan et al. (2013) were investigating spatial data stemming from scientific simulations. Thus, our approach is the first to

apply dimension visualization for the analysis of multifields. Similar to the dual-space approaches, we apply MDS to dimensions. However, we consider spatial data visualizations to be more appropriate for this scenario than the data item projections of other dual-space approaches.

3 Field Similarity Plot

Our approach focuses on the uniform qualitative comparison of the fields present in a dataset, which requires all fields to be normalized. In the following, we assume that all fields are normalized to the unit interval.

First, a number of commonly studied and newly derived fields are included in the analysis. Specifically, we add gradient magnitudes, Hessian determinants, and fields based on the gradient similarity measure (GSIM) introduced by Sauber et al. (2006). For two gradients g_i and g_j, GSIM is defined as

$$s(g_i, g_j) = (s_d(g_i, g_j) \cdot s_m(g_i, g_j))^r, \tag{1}$$

$$s_d(g_i, g_j) = \left(\frac{g_i^T g_j}{\|g_i\| \cdot \|g_j\|}\right)^2, \tag{2}$$

$$s_m(g_i, g_j) = 4 \frac{\|g_i\| \cdot \|g_j\|}{(\|g_i\| + \|g_j\|)^2}, \tag{3}$$

where s_d is the direction similarity, s_m is the magnitude similarity, and r regulates sensitivity of the measure (set to 1.3 as recommended (Sauber et al. 2006)).

The following GSIM fields are computed: first-order similarity between gradients, second-order similarity between eigenvectors of the principal eigenvalue of the Hessians, and mixed similarity between the first- and the second-order derivatives estimate (i.e. gradients and Hessian eigenvectors, respectively). In the set of derived fields, the gradients and their similarities carry first-order relationships, while other fields might indicate more complex relationships. All derived fields are normalized as well.

Next, for each data field, we interpret the vector of its values at all spatial locations as a multidimensional data point and compute integrated differences between pairs of fields. In the following, we use the global Euclidean distance measure defined on the *-dimensional space, where $* = n_x \times n_y \times n_z$, and n_x, n_y, n_z are respective sizes of the spatial grid in x, y, z dimensions. This distance is computed as

$$d_{ij}^* = \sqrt{\sum_{xyz} [v_{j_{xyz}} - v_{i_{xyz}}]^2}, \tag{4}$$

where $v_{j_{xyz}}$ denotes the value of the jth field at the spatial position xyz. The distance value is 0 for two identical fields, and increases proportional to the differences between the compared fields.

The *-dimensional points are then projected to points in a 2D space while trying to maintain the computed pairwise distances (Eq. 4) as much as possible. We have selected Sammon's mapping (Sammon 1969) as a projection technique. It starts with random point coordinates in the 2D space and then iteratively moves the points to minimize the error given by the equation:

$$E = \frac{1}{\sum_{i<j}\left[d_{ij}^*\right]} \sum_{i<j}^{N} \frac{[d_{ij}^* - d_{ij}^2]^2}{d_{ij}^*},$$ (5)

where d_{ij}^2 are the Euclidean distances of projected 2D points in the 2-dimensional space, and d_{ij}^* are the distances between original points in the multidimensional space.

Finally, the results of the projection are visualized as a scatterplot. By definition, this scatterplot has no inherent axes (i.e. it is only unique up to rotation), while the distances between the 2D points indicate how similar the respective fields are (Figs. 1, 4 and 5). Colors are used to indicate the types of the fields (original, derivative or similarity fields).

4 Interactive Visual Analysis

The field similarity plot described above helps the scientist to identify informative fields. The plot serves as an interaction widget, where individual fields can be clicked at and investigated using linked views (selected fields are highlighted by increased point size).

The linked views are (1) slice-based volume visualizations for the spatial investigation of field value distributions (Figs. 2a, 3a and 6a) and (2) 1D histograms of the normalized field values for understanding the field value distribution within its range (Figs. 2b, 3b and 6b). In the individual views, we decided to avoid using the typically default rainbow colormap because of its misleading perceptual properties (e.g. it introduces artificial sharp contrasts at the color transitions) (Borland and Taylor 2007; Rogowitz and Treinish 1998; Silva et al. 2007). The selected black body radiance colormap represents data without such issues.

The slice-based visualization renders axis-aligned slices through the volumetric dataset. The position and the orientation of the cutting plane, as well as the orientation and scaling of the 3D view, can be changed interactively. Multiple slice-based visualizations for different data fields are coordinated: all of them use the same view on the volume data as well as the same cutting plane.
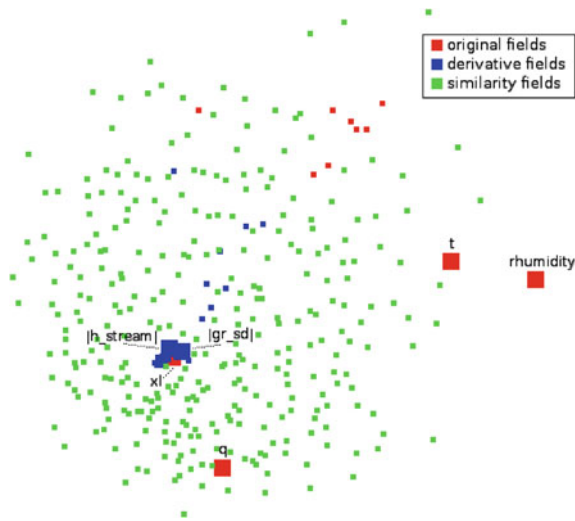
Fig. 1 The field similarity plot for the multifield scenario. There is a tendency towards separation of different types of fields, but large differences exist within the groups of fields of each type. Big points with labels correspond to the fields which have individual views in Figs. 2 and 3. Derivative fields refer to gradient magnitudes and Hessian determinants, similarity fields refer to all the fields computed with GSIM

5 Use Cases

In this section, we demonstrate two possible geoscientific applications, focusing on the atmospheric part of a climate simulation output. The data are a sample run of a climate model setup with a pre-industrial configuration, similar to the pre-industrial control setup described by Zhang et al. (2013), but with different settings of orbital parameters. Details of the employed Community Earth System Models COSMOS (consisting of ECHAM5 for the atmosphere, MPIOM for the ocean, and JSBACH for the vegetation) are outlined, for example, by Stepanek and Lohmann (2012).[1]

The analyzed dataset contains monthly means of ECHAM5 output for 13 volumetric climate variables at the spatial grid resolution of $96 \times 48 \times 19$. The z-axis is given in terms of hybrid sigma-pressure levels.

Our first scenario addresses the multifield constellation with 13 variables at a single time step, while our second scenario is concerned with a single field at all time steps of one year.

[1] Data courtesy of Christian Stepanek and Gerrit Lohmann from Alfred-Wegener Institute Helmholtz Centre for Polar and Marine Research in Bremerhaven, Paleoclimate Dynamics research group.
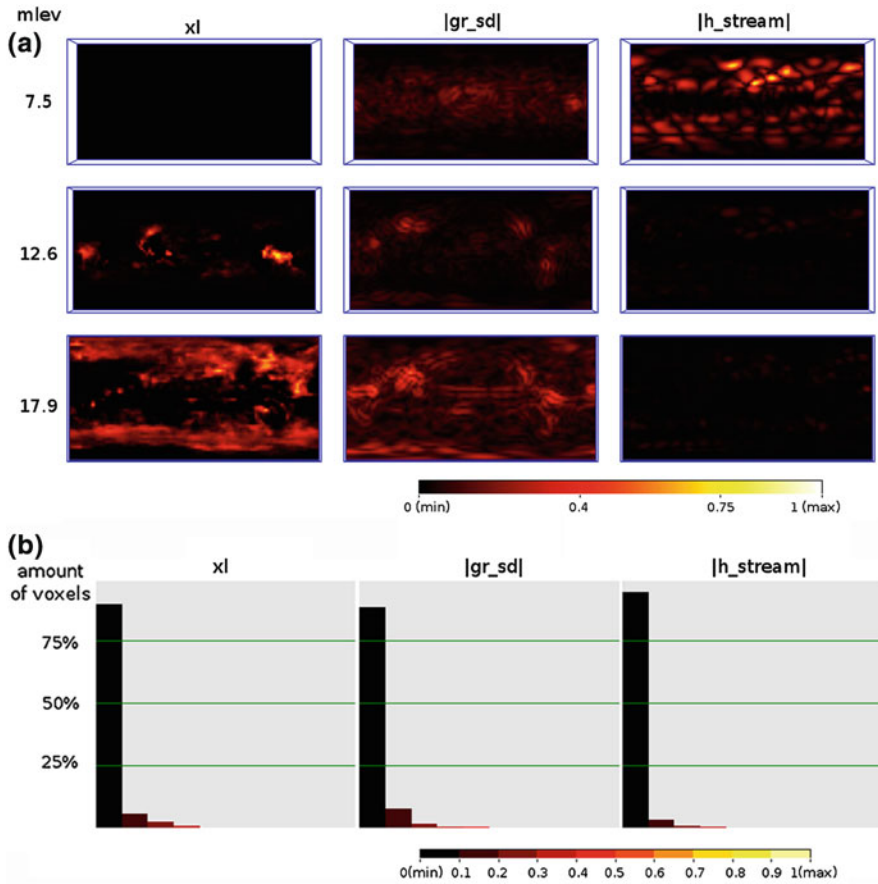
Fig. 2 Individual views for closely located points in Fig. 1: cloud water field *xl*, gradient magnitude of divergence field $|gr\_sd|$ and Hessian determinant of streamfunction field $|h\_stream|$. **a** Slice views, *mlev* is the interpolated layer index of the slice (1—top layer, 19—surface layer). **b** 1D histograms

5.1 Multifield

The 13 fields of ECHAM5 describe temperature (*t*), wind velocity components (*u*, *v*, *omega*), specific and relative humidity (*q* and *rhumidity*), cloud water (*xl*) and cloud ice (*xi*), vorticity (*svo*) and divergence (*sd*), streamfunction (*stream*), velocity potential (*velopot*) and geopotential height (*geopoth*). We consider one time step, namely the monthly means of April of the first year.

In the field similarity plot after adding all the discussed derived fields (Fig. 1) one can observe a wide spread of the original fields. However, we can also see a very dense group of points including all Hessian determinant fields. A few gradient magnitude field points as well as the two original fields of cloud water *xl* and cloud ice *xi* (occluded by *xl* in Fig. 1) are placed very closely to this group. The plot indicates high similarity between the aforementioned fields with respect to Euclidean distance. By looking at the linked views of slice-based volume visualizations and 1D histograms for a few fields from this area (Fig. 2), we see that, while the fields exhibit different patterns, the distribution of the data values are very much in the lower range. Using the Euclidean distance for the projection, the fields are closer to each other than to other fields with values in the upper range.

When observing outliers among the original fields, we can see that each of them has strong unique features (Fig. 3). The overall distribution of the values of the temperature field *t* and the relative humidity field *rhumidity* are more similar than of the values of the specific humidity field *q*. The latter field is more similar to the group described above.

5.2 Time-Varying Field

In the second scenario, we investigate the change of the temperature field throughout the first year by treating each month as a separate, independent field.

The projection of only the original fields arranges the points in a loop (Fig. 4) that corresponds to the annual cycle, which documents, again, the feasibility of the overall approach.

The loop has a pendular behavior with the winter months on one side and the summer months on the other. The spring and fall months are close together (there is even a crossing) and form transitional phases. Moreover, we can conclude that changes over months are gradual.

The full projection of original and derived fields (Fig. 5) clearly separates the five types of derived fields as well as the original dataset fields. Thus, we can conclude that each of the chosen types of derived fields conveys distinctly different information from the original data and from each other, i.e. differences within a group are much smaller than between the groups. It can easily be confirmed by looking at the individual views (Fig. 6).

6 Discussion

The presented approach describes a conceptual workflow, where a design choice among many alternatives is made at each step. In this paper, we followed some of the common decisions. In the following, we discuss other possibilities. A thorough
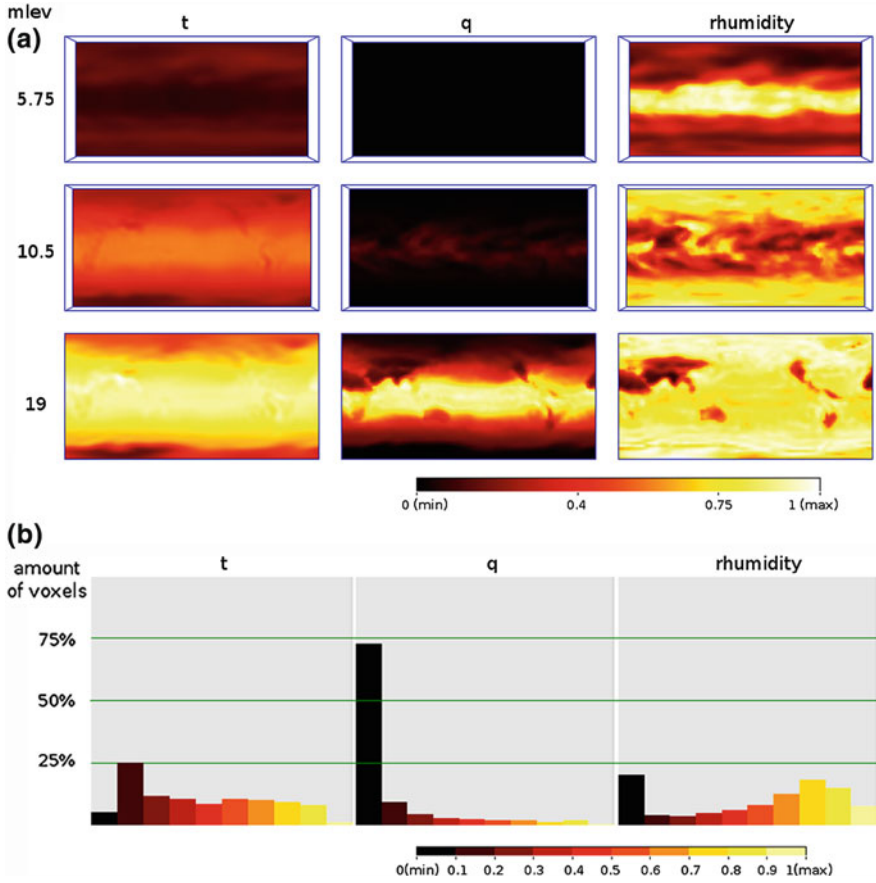
Fig. 3 Individual views for outliers among original fields in Fig. 1: temperature *t*, specific humidity *q*, and relative humidity *rhumidity*. **a** Slice views, *mlev* is the interpolated layer index of the slice (1—top layer, 19—surface layer). **b** 1D histograms

Fig. 4 The field similarity plot for twelve months of the temperature field. Yearly cycle is clearly visible, separating months in three groups: winter, summer, and transitional seasons
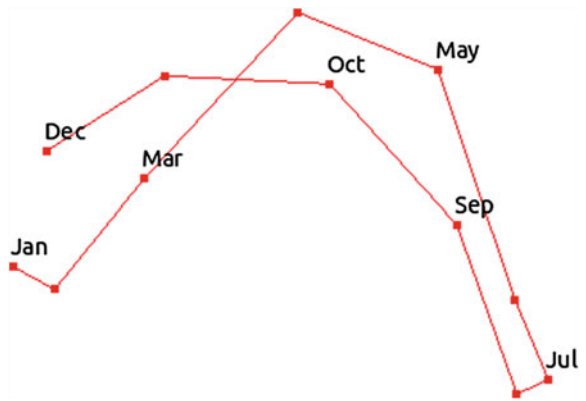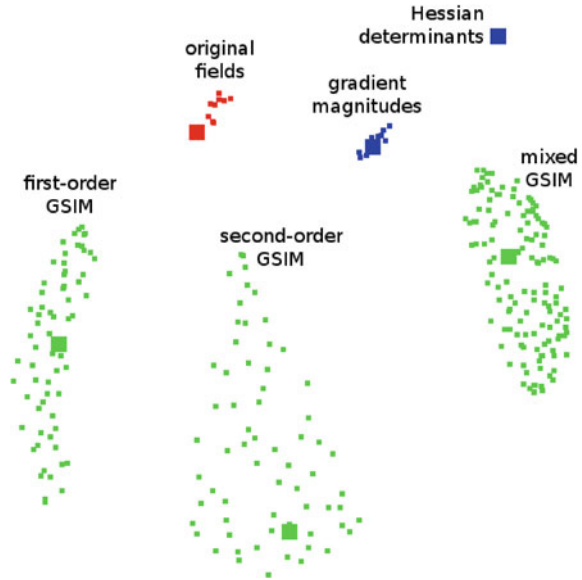
Fig. 5 The field similarity plot in the case of a single time-varying field. Different types of fields are clearly separated. Individual views for the highlighted points are shown in Fig. 6

analysis on which alternatives work best for which application is beyond the scope of this conceptual paper and left for future work.

In this paper, we normalized the range of each field to the [0, 1] interval. In case of noisy data, a distribution-based normalization method can be employed. It puts the mean value of the distribution to 0 and standard deviations to ±1. The latter method is more robust against outliers.

Sammon's mapping is a generally applicable projection method that produces suitable results for many applications. However, our approach is not tied to this particular technique, and other distance-based projection methods (Minghim et al. 2006; Paulovich and Minghim 2006; Paulovich et al. 2008) can be used.

The field similarity plots in Sect. 5 were computed with the Euclidean distance measure. In our experiments, we also employed a correlation-based distance measure. In the multifield case, the correlation measure produced quite different results. Here, the denser groups of original and derivative fields shown in Fig. 1 were widely spread, while the fields chosen for Fig. 3 were placed closer to each other. The reason is that the correlation measure is less sensitive to the scaling of fields. Thus, fields with similar histograms but different spatial patterns are judged to be more different. For the case of a single time-varying field, the results produced by the correlation-based measure were similar to those with the Euclidean distance, only that the groups were overlapping and not so clearly separated.

From the purely information-based point of view, the common recommendation to analyze the general behavior of the data is to take fields that are spread in the projected space and cover it well. However, it is clear that specific targeted questions require inclusion of certain fields that capture the respective information.
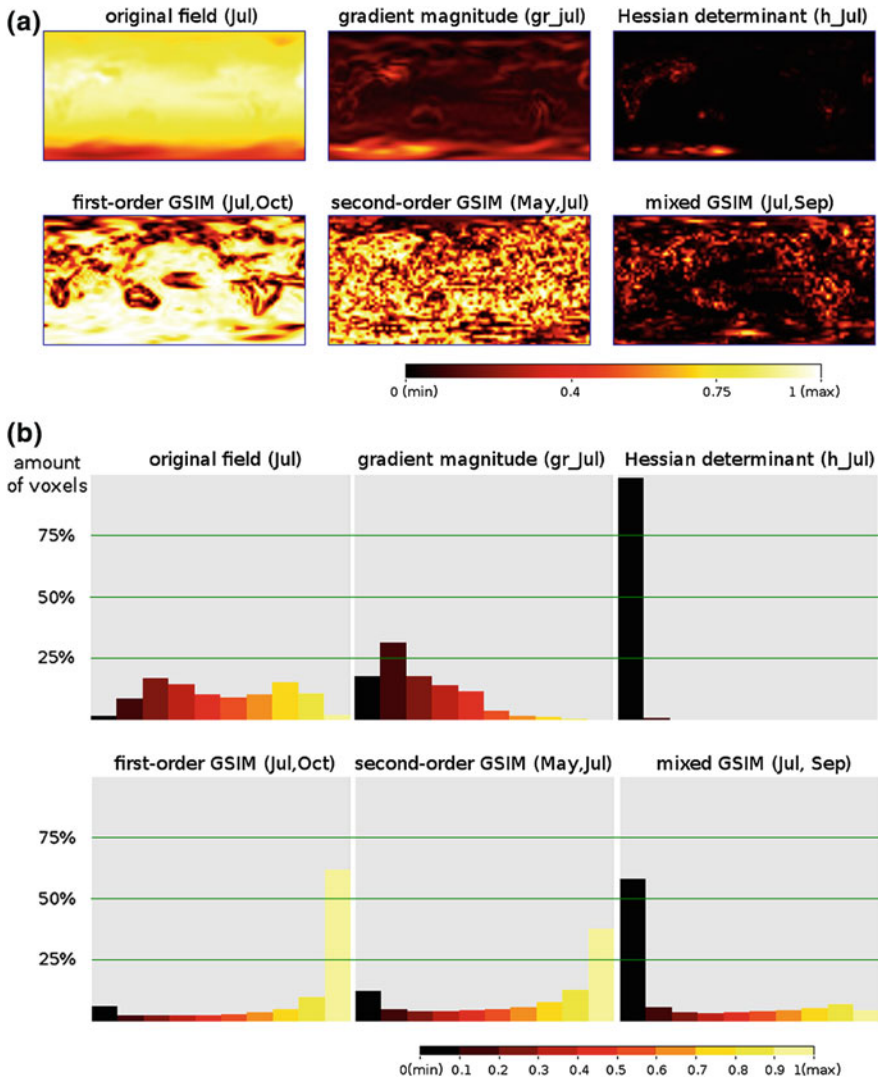
Fig. 6 Individual views for different types of fields, corresponding to the highlighted points in Fig. 5. **a** Slice views at the surface layer. **b** 1D histograms. It is clearly visible that different types of derived fields convey distinctly different information

Depending on the needs of an application, instead of slicing views and 1D histograms, other visualizations can be used for the exploration of the overview, e.g. direct visualization as volume rendering of individual selected fields (Abellán and Tost 2008; Drebin et al. 1988), multimodal volume rendering (Woodring and Shen 2006), plotting correlations or other statistics in scatterplots or parallel

coordinates (Inselberg 1985), side-to-side volumetric visualizations, as well as commonly used scatterplot matrices.

7 Conclusion

In this paper, we described a two-step approach to support the analysis process by providing overview of the correlation between scalar fields in the data. At the first step, a number of predefined fields is derived from a given dataset. At the second step, a similarity-based overview of all fields is presented to the user, by employing a multidimensional projection technique.

The described approach allows for the interactive analysis of relationships between multiple fields, including simulated and derived fields. It provides an overview over all fields in the projection view, allows the user to investigate individual fields with coordinated views using slice-based volume visualization and 1D histograms, and provides means to choose fields for further investigation.

References

Abellán P, Tost D (2008) Multimodal volume rendering with 3d textures. Comput Graph 32(4):412–419

Borland D, Taylor RM (2007) Rainbow color map (still) considered harmful. IEEE Comput Graph Appl 27(2):14–17

Drebin RA, Carpenter L, Hanrahan P (1988) Volume rendering. In: Proceedings of ACM SIGGRAPH '88, the 15th annual conference on computer graphics and interactive techniques, pp 65–74

Edelsbrunner H, Harer J, Natarajan V, Pascucci V (2004) Local and global comparison of continuous functions. IEEE Vis 2004:275–280

Gosink LJ, Anderson JC, Bethel EW, Joy KI (2007) Variable interactions in query-driven visualization. IEEE Trans Vis Comput Graph 13(6):1400–1407

Inselberg A (1985) The plane with parallel coordinates. Visual Comput 1(2):69–91

Jänicke H, Böttinger M, Scheuermann G (2008) Brushing of attribute clouds for the visualization of multivariate data. IEEE Trans Vis Comput Graph 14(6):1459–1466

Kandogan E (2001) Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In: Proceedings of KDD '01, the seventh ACM SIGKDD international conference on knowledge discovery and data mining, pp 107–116

Minghim R, Paulovich FV, de Andrade Lopes A (2006) Content-based text mapping using multi-dimensional projections for exploration of document collections. In: Proceedings SPIE, 6060:60600S–60600S–12

Nagaraj S, Natarajan V (2011) Relation-aware isosurface extraction in multifield data. IEEE Trans Vis Comput Graph 17(2):182–191

Nagaraj S, Natarajan V, Nanjundiah RS (2011) A gradient-based comparison measure for visual analysis of multifield data. Comput Graph Forum 30(3):1101–1110

Paulovich FV, Minghim R (2006) Text map explorer: a tool to create and explore document maps. In: IV 2006, Tenth international conference on information visualization, pp 245–251

Paulovich FV, Nonato LG, Minghim R, Levkowitz H (2008) Least square projection: a fast high-precision multidimensional projection technique and its application to document mapping. IEEE Trans Vis Comput Graph 14(3):564–575

Rogowitz B, Treinish LA (1998) Data visualization: the end of the rainbow. IEEE Spectr 35(12):52–59

Sammon JW Jr (1969) A nonlinear mapping for data structure analysis. IEEE Trans Comput C-18 (5):401–409

Sauber N, Theisel H, Seidel H-P (2006) Multifield-graphs: an approach to visualizing correlations in multifield scalar data. IEEE Trans Vis Comput Graph 12(5):917–924

Silva S, Madeira J, Santos BS (2007) There is more to color scales than meets the eye: a review on the use of color in visualization. In: IV '07, 11th international conference on information visualization, pp 943–950

Stepanek C, Lohmann G (2012) Modelling mid-pliocene climate with COSMOS. Geosci Model Dev 5:1221–1243. doi:10.5194/gmd-5-1221-2012

Turkay C, Filzmoser P, Hauser H (2011) Brushing dimensions—a dual visual analysis model for high-dimensional data. IEEE Trans Vis Comput Graph 17(12):2591–2599

Turkay C, Lundervold A, Lundervold AJ, Hauser H (2012) Representative factor generation for the interactive visual analysis of high-dimensional data. IEEE Trans Vis Comput Graph 18(12):2621–2630

Woodring J, Shen H-W (2006) Multi-variate, time varying, and comparative visualization with contextual cues. IEEE Trans Vis Comput Graph 12(5):909–916

Yuan X, Ren D, Wang Z, Guo C (2013) Dimension projection matrix/tree: interactive subspace visual exploration and analysis of high dimensional data. IEEE Trans Vis Comput Graph 19(12):2625–2633

Zhang X, Lohmann G, Knorr G, Xu X (2013) Different ocean states and transient characteristics in last glacial Maximum simulations and implications for deglaciation. Clim Past 9:2319–2333. doi:10.5194/cp-9-2319-2013