Workshop on Visualisation in Environmental Sciences (EnvirVis) (2026)
K. Feige and B. Nsonga and K. Rink (Editors)

Comparison of Categorical Data from Meteorological Models and Observations using a Pattern-Based Approach

S. Pandey[1,2], M. Sondag[2,3], C. Krauter[2,4], S. Zellmann[2] and T. von Landesberger[2]

[1]University of Münster, Germany
[2]University of Cologne, Germany
[3]Maastricht University, Netherlands
[4]University of Stuttgart, Germany

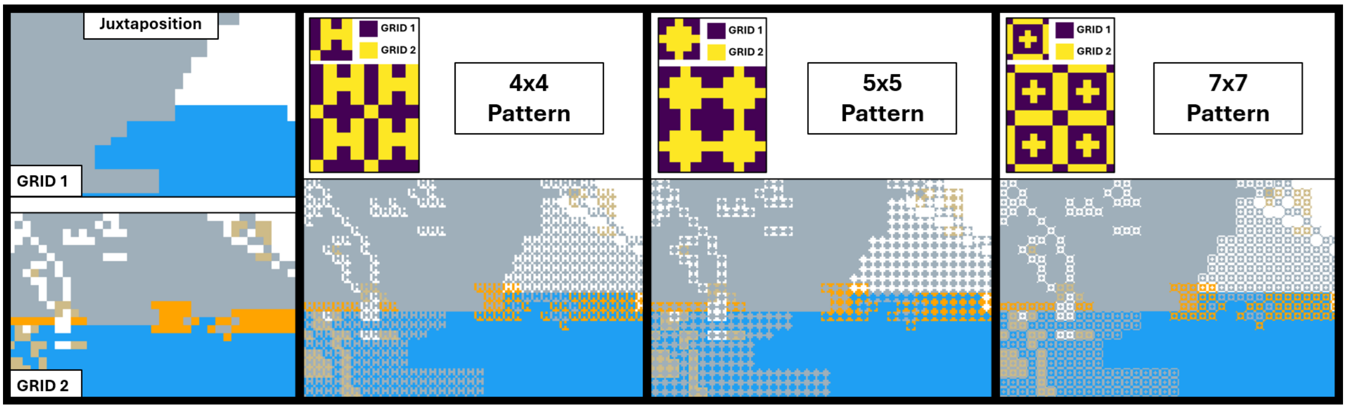


Figure 1: Comparative visualization of meteorological model and observation data given as time-height plots. Left: Juxtaposition as a reference (Grid 1: model, Grid 2: observation), followed by 4x4, 5x5, and 7x7 design patterns. Legends in the top-left corners show the pattern at the top and a multi-view representation of the pattern in the middle. Below are examples with real data.

Abstract

This paper introduces a pattern-based approach for comparing large categorical gridded data. Our goal is to compare meteorological models and observations given as time-height data. The data come with constraints, including heavy noise and model and observation uncertainties. We define requirements with domain experts and propose design criteria tailored to the application domain. We explore the design space of feasible patterns and propose three patterns designed to resolve the drawbacks of juxtaposition with estimating (dis)agreement and superposition with disambiguation of the comparison source. We also conduct a crowdsourced experiment with a small group of non-expert users, providing initial insights into how the proposed approach might advance the field of comparative visualizations for categorical data.

1. Introduction

Comparative visualization of regularly gridded data with categorical values plays an important role in meteorology to enable visual comparison of model predictions and observations. Besides its application in meteorology, similar problems also arise with two-dimensional spatial data [AK15], in matrix data [ZLD*15], and in time-varying data, for example in time-height plots [SE20].

Visual comparison of the data provides a more intuitive way to detect variation than numerical comparison [Gle18]. In this paper, we focus on comparing categorical, regularly gridded data from large (> 1 M cells) grids, where each cell belongs to one of (up to ten) categories. Our motivating use case is a meteorological dataset with observations (reference) and model results (compared). These are shown in so-called time-height plots, which are gridded scatterplots (horizontal axis: time, vertical axis: height) with color-coded categories that correspond to meteorological phenomena, e. g., rain

DOI: 10.2312/envir.20261003

and ice. Climate prediction models such as ICON [Deu25] contain millions of grid cells even when comparing a single day at a single point in space to the observational data of this point . This makes it difficult to identify patterns where the model performs well and where it should be improved.

In meteorology, visual comparison of gridded categorical data of observation against prediction has several goals. The comparison should identify commonalities (agreements) and differences (disagreements) between the datasets. (Dis) agreement is defined as the same (different) color between two cells at the same grid location. In the comparison, it is also important to know which value is the reference and which is the prediction. After consulting with domain experts in meteorology, this leads to the following requirements for visualizing the two grids:

R1 *Agreement*: how much and which categories agree. Corresponds to finding commonalities, i. e., a match of model and observation in the whole dataset or a selected region, in all categories or a selected category. Example question: "How do model and observation match for ice in the morning hours at low altitudes?".

R2 *Disagreement*: how much and which categories disagree. Corresponds to finding differences, i. e., a mismatch of model and observation in the whole dataset or a selected region, either in all categories or in selected categories. Example question: "How much do model and observation disagree for rain droplets?"

R3 *Source of disagreement*: in case of disagreement, identify which category corresponds to which grid, i. e., what is the value in the model and what is the category in the observation. Corresponds to identifying the source of disagreement: what is the value of the model, and what is the value of observation in causing the mismatch, in the whole dataset or a selected region? Example question: "What does the model predict for the observation values of ice?"

R4 *Suitability for larger grids* (> 1 M cells, ca. ten categories). In meteorological grids, typical time resolution is in seconds, and altitude in meters. The data can have up to 10 categories, e. g., ice, rain, rain droplets, or dust. Thus, a visualization needs to be able to show the comparison of up to 10 categories for grids with millions of cells.

The standard strategy for comparisons like this is juxtaposition, i. e. showing two separate visualizations next to each other, one for each dataset [LJS21]. Although useful in smaller grids with few elements [Gle18], comparing many cells across plots becomes increasingly difficult. Juxtaposition requires more space than superposition, and incurs high cognitive load. Increased size and complexity of the plot leads to subtle yet potentially important differences that become difficult to discern. Comparison by eye movement between grids further increases the load. Moreover, difficulty in exactly matching grid positions can lead to imprecision, and in sum renders this infeasible for us.

An alternative strategy is superposition [NM17], where data is layered on top of each other with different techniques, such as color blending. Due to limitations in human color perception, it is challenging to use it to recognize categories across two different grids. An additional problem is that, with color blending, it is not possible to distinguish whether a cell has a red color in reference grid 1 and a blue color in the compared grid 2, or vice versa, because the blended colors are identical.

We present a novel domain-independent visualization approach based on patterns enhancing the concepts of weaving [LRS10] to superimpose large categorical gridded data and ensure that the design requirements (**R1–R4**) are met. For this purpose, we propose a set of design criteria for the patterns to meet. Based on these, we explore the design space of patterns and create suitable examples.

The main contributions of this paper are:

1. Domain-independent design criteria for creating patterns for visual comparison of categorical gridded data.
2. Exploration and construction of novel patterns to compare large categorical gridded data.
3. Application of novel patterns to meteorological data showing their usefulness.
4. An initial, crowdsourced user study evaluating three patterns of different sizes against juxtaposition.

We use the time-height data from the ICON model on its corresponding observational counterpart [Fin25]. We visualize both the model and the observation grids in a single plot, enabling identification of categories. (see Fig. 1). This approach creates solid colors on canvas where the observation and the model are in agreement, and patterns appear upon disagreement.

2. Related Work

Our work relates to three areas: comparative visualization, grid-based visualization, and scatterplots with multi-categorical data. We focus on how the comparison can be visualized. Research on fast rendering and topics related to drawing large scatterplots and grids, e. g. [THS*20, HM23] are out of the scope. For an overview of multi-scale visualization techniques, we refer to [CJS*21].

Gleicher et al. [Gle18] identify four key considerations for comparisons: defining comparison elements, addressing scaling challenges, selecting strategies, and mapping visual designs. Deng et al. [DCM*22] build upon this by identifying general design patterns and evaluating how they have been used so far. In our case, we will use a variant of a superimposed design via an overlay with a coordinate design pattern, using the gridded data as a coordinate anchor for the pattern.

For data with fixed coordinates, e. g., for spatial data, overlaying data is common. Gunther et al. [GSFT16] visualized the impact of volcanic eruptions on climate by mapping trajectories to a grid on the Earth and overlaying the point measurements for visual comparison. Asadie et al. [AK15] compare trends in extreme precipitation using juxtaposed maps with the gridded data overlaid on top of them. Röber et al. [RB20] use stippling on top of the data to directly show the agreement of multiple models for quantitative data. Pomme et al. [PBGA22] evaluate juxtaposed, superimposed, and explicitly encoded designs of confusion matrices.

In grid visualizations, the grid is fixed, and every cell has data. Miller et al. [Mil07] use glyphs to visualize multiple attributes of a grid cell. Ware et al. [WP13] employ glyphs to visualize gridded weather data. Liu et al. [LS15] investigated how to juxtapose multiple grids. Janicke et al. [JBS08] use stippling combined with color to show which category a cell belongs to. Fernandez et al. [FGDD15] use colored, oriented shapes on a grid to indicate categories. This is closely related to glyph-like design on maps, such as that from Nguyen et al. [NLHS20]. Wickman et al. [WHWC12] use more complicated glyphs to visualize multiple dimensions. The patterns in this paper can be seen as simplified glyphs, using only color and complementary texture to denote category memberships of multiple grids for large-scale data.

The main difficulty in scatterplot visualizations is preventing overplotting. Chen et al. [CCM*14] introduce a sampling technique to overcome overplotting in scatterplots, aiming to preserve relative density while overcoming the creation of additional colors through transparent overlap. Li et al [LSL*22] introduce the overlap-free scatterplot method using a dual-spaced coupling model to prevent overdraw. For large datasets, Chen et al. [CZF*22] created a hierarchical sampling method that progressively refines the visualization. Hu et al. [HSK*20] design a sampling method for multicategorical data using Z-order curves for scatterplots. Their method preserves relative data and category densities while highlighting major outliers. While many of these approaches could apply to grids, overplotting in grids is slightly different from that in scatterplots, as there are exactly two datapoints overplotted at each cell. We focus on visualizing these, though sampling could help reduce the grid size.

Variants of scatterplots can ensure that points in multiple categories are visible. Sarikaya et al. [SG18] and Heimerl et al. [HCSG18] explored the design space of scatterplots for various data characteristics and analysis tasks. Weaving as presented by Luboschik et al. [LRS10] and Jo et al. [JVDF19] ensures all overlapping objects remain visible. We used this approach for our patterns, where, on a small scale, all objects are visible, while on a large scale, they resemble blending due to the limited display space.

3. Approach

We introduce a novel pattern-based approach to view two scatterplot grids (short: grids) within a single, combined visualization. Our design preserves distinguishability between categories within each grid, allowing for the clear and intuitive interpretation of data from both sources simultaneously. We aim to produce static, non-interactive visualizations that meet these requirements. This allows embedding in static media, such as scientific papers, and improves performance and replicability for large-scale datasets.

Fig. 2 shows our approach. For each cell in the two grids, we generate a distinct pattern P colored by Grid 1, and its inverted counterpart P^{-1} colored by Grid 2. Superimposing the patterns for each cell gives us the Resultant Grid. The black cross pattern denotes transparency. Solid colors indicate agreement (**R1**), patterns becoming visible means disagreement (**R2**). The colors allow identifying the categories from each grid, i. e., the source of disagreement (**R3**). Fig. 3 shows this approach applied to a large grid (**R4**).

Pattern Design Criteria to create patterns based on perceptual considerations and to meet **R1-R4** from Sec. 1.

- **Large scale (L)** Visualize (dis)agreement in multiple cells of the resultant pattern grid. All non-uniform patterns fulfill this criterion. This is necessary to identify (dis)agreement.

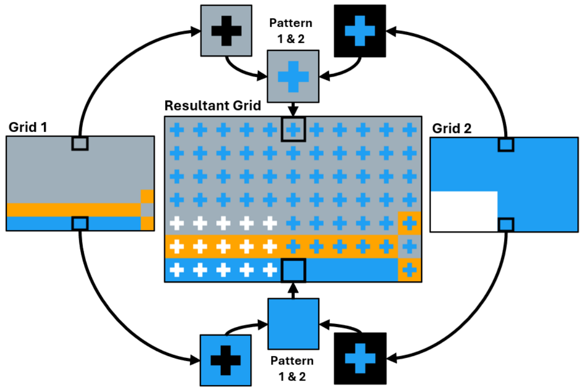


Figure 2: Using patterns to show (dis)agreement (**R1**, **R2**), and disagreement source (**R3**) for large grids with multiple categories (**R4**). Colors denote categories. Pattern shape shows the source of disagreement. Pattern composition shows (dis)agreement.

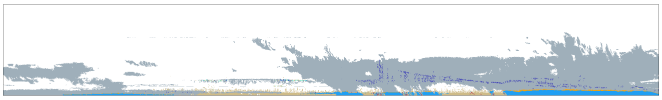
- **Balanced (B)** The pattern balances the number of pixels for each source grid. If the pattern does not balance the color data, then one grid would be overrepresented (see Fig. 6c). This is important for being able to quantify (dis)agreement accurately.
- **Identifiable (I)** To identify which grid the data corresponds to. A counter-example is a vertical 2x2 pattern: an individual pattern is easy to identify. But when tiled, a symmetric "super-pattern" arises that makes source identification impossible (see Fig. 5b).
- **Aspect ratio (A)** Pattern grids must maintain a 1:1 aspect ratio to preserve visual consistency. Uneven expansion would lead to spatial misrepresentation through squeezing or stretching.
- **Shape (S)** Using shapes like symbols or letters makes the patterns easy to recognize, increasing readability.
- **Pattern (P)** Pattern must be contained within a single grid, to make it easy to discern in the larger grid. A counterexample is shown in Fig. 4, where adjacent grid cells form a super-pattern.
- **Optical Illusion (O)** Optical illusions such as the Hermann [Spi94] effects should be avoided, as users may perceive objects that are not actually present (see Fig. 7c).
- **Resolution (R)** To construct a pattern of a desired resolution, each grid cell must be expanded to the same resolution. Hence, smaller patterns with lower resolution are preferred.

3.1. Design Space and Pattern Selection

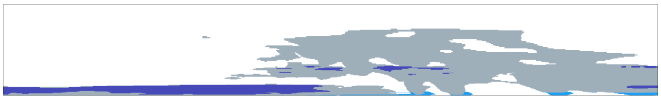
Our objective is to generate patterns that adhere to the design criteria while maintaining a minimal grid size. We test resolution patterns from 2×2 to 7×7. Moreover, most of the perceptually distinguishable patterns such as a cross, a plus-sign, or square [SG18] can be drawn in this range. We refrain from using larger patterns due to data size, display resolution, and similar concerns.

The theoretical number of all squared patterns with constraints **R**, **A** ($a \times a$, e. g. 4×4) is very large ($2^{a \times a}$-1, e. g. 2^{16}-1 for 4×4 grid). To reduce the number of patterns that we need to check for suitability, we can first filter them based on our design criteria. The **B** criterion can be automatically applied. The exclusion of grids leading to the Moire effect (**O** criterion) reduces this number fur-

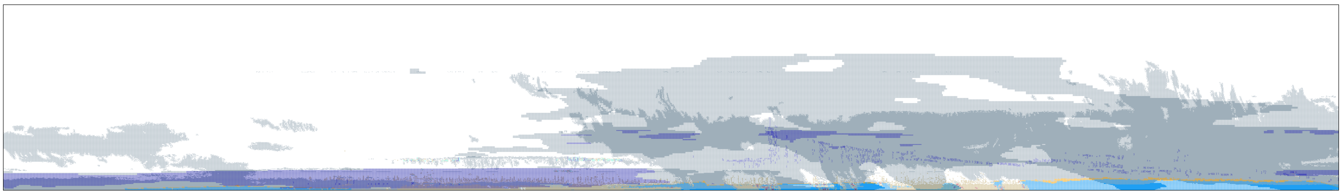
Proceedings published by Eurographics - The European Association for Computer Graphics.



(a) Grid plot of observation data

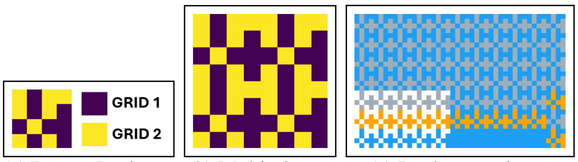


(b) Grid plot of model data



(c) Grid plot of observation and model grid plots in pattern design

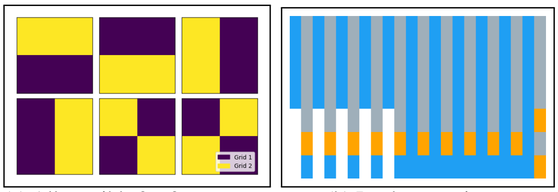
Figure 3: A 24-hour grid plot depicting cloud particle categories in three variations. On a large scale, the approach resembles a superimposition-based approach. Agreement in cells is shown by solid colors. Disagreeing cells show blended colors. Cells with symmetric disagreeing categories ((a, b) vs. (b, a) as values for grid 1 and 2) have slight color differences through criteria "L" to disambiguate them.



(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	✓	✓	✓	✓	×	×	✓	✓

Figure 4: A 4×4 Design where adjacent cells form super-patterns violating **P**, making it hard to match the cell location to the pattern.



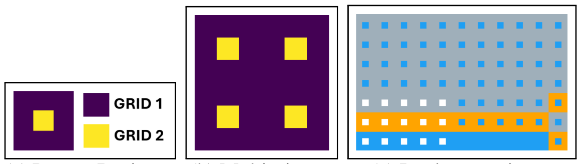
(a) All possible 2×2 patterns (b) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	✓	✓	×	✓	×	✓	✓	✓

Figure 5: Example of a 2×2 pattern failing the **I** criterion. Tiled patterns are indistinguishable as they are symmetrical.

ther – all patterns with colored corners are excluded. In addition, we dismiss rotation-equivalent patterns that add no information, as well as symmetric patterns as they do not allow us to distinguish the source grid and violate criterion **I** (see Fig. 5b). When designing patterns, we focus on those with easily recognizable shapes such as +, H, and O (criterion **S**).

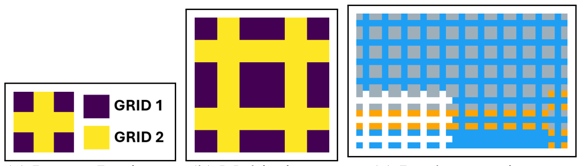
2×2 **Patterns** can accommodate colors from both grids equally. They have equal numbers of pixels for both colors (criterion **B**) and, as the smallest pattern, satisfy the **R** criterion optimally. The designs are, however, limited to sole diagonal, horizontal, and vertical shapes (see Fig 5a) that all fail the identification criterion **I**. When the patterns are tiled, the source grid cannot be identified anymore (see Fig 5b). Thus, all 2×2 patterns need to be excluded.



(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	×	✓	×	✓	×	✓	✓	✓

Figure 6: An imbalanced 3×3 pattern violating criterion **B**.



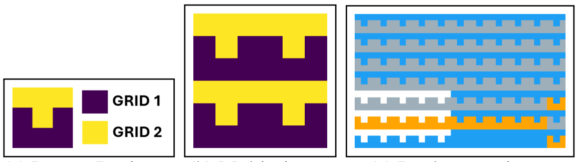
(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	✓	✓	✓	✓	✓	✓	×	✓

Figure 7: A pattern causing an optical illusion [Spi94], failing **O**.

3×3 **Patterns** have 126 possibilities that satisfy criteria **A** and **R**, putting five cells of Grid 1 and 4 cells of Grid 2 in a 3×3 patterned grid. Most of these patterns do not have two disjoint and identifiable shapes (S,I). Exceptions are the two patterns (and their rotated and translated versions) shown in Fig. 8, which have a "U"-shape and a "T"-shape, and in Fig. 7, which has a "plus" shape and a "square" shape. For the "T" and "U" shape, the pattern is not clearly visible when multiple patterns are adjacent due to a lack of contrast around the shapes (**I**). For the "plus" and "square" shape, the mesh-like structure creates an optical illusion **O** (see Fig. 7).

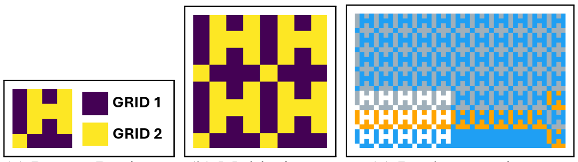
4×4 **Patterns** Beyond a 3×3 resolution, more varied designs can be generated. From the approximately 13K possible designs, many variations do not fulfill our design criteria. Testing every individual pattern design becomes impractical. Therefore, we manually construct these with the goal of obtaining easily identifiable shapes. The 4×4 grid allows us to create a pattern that fulfills all the criteria



(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	×	✓	✓	✓	✓	✓	✓	✓

Figure 8: Example of a 3×3 Pattern Grid



(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	✓	✓	✓	✓	✓	✓	✓	✓

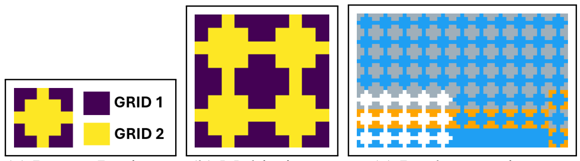
Figure 9: Example of a 4×4 Pattern Grid fulfilling all criteria.

(see Fig. 9). It allows for identification (**I**) and enables the creation of asymmetric patterns (**P**) that do not form super-patterns, without optical illusions (**O**), and with perceivable shapes (**S**). We evaluate this design in our evaluation. We note that while 4×4 patterns are necessarily non-centered, and do not fulfill the optional centering criterion **C**. However, these patterns are smaller in resolution (**R**) than the 5×5 patterns.

5×5 **Patterns and Larger Resolutions** For grid resolutions of 5×5 and beyond, creating simple, yet effective designs that remain unique becomes increasingly challenging. Design patterns such as a plus, a cross, or a letter, as in previous cases, do not offer additional benefits yet occupy more space. Patterns as in Fig. 10b meet the design criteria; they also generate secondary shapes when tiled, which may draw attention to the secondary symbol. As visual perception varies among individuals [JM11], the presence of two distinct shapes could allow users to choose which shape to focus on when observing the resultant plot, which may facilitate grid identification. In addition, the patterns are centered, making them easier to read. We include this pattern in our evaluation.

We expand the canvas to a maximum of 7×7 cells to see whether there is any added benefit to using a more complex design while maintaining equal proportions of color data between the two grids. One advantage is that we can use two very distinct shapes with sufficient contrast between them. However, this comes at the cost of additional space (**R**), and it is unclear whether the trade-off is favorable. Hence, we also evaluate this design.

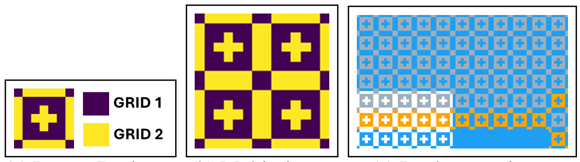
Both 5×5 (Fig. 10a) and 7×7 (Fig. 11a) designs satisfy our design criteria. Additionally, both feature a symmetric pattern symbol that closely resembles a plus sign, which is easy to identify. The data is irregular due to its variable height, so we resample it to a regular grid by splatting. The resulting grid has horizontal units of time (in sec.) and vertical units of height (in meters). To reduce the size, compression is applied to both datasets for consistency and to preserve the fine details.



(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	✓	✓	✓	✓	✓	✓	✓	✓

Figure 10: Example of a 5×5 Pattern Grid



(a) Pattern Design (b) Multi-view (c) Implementation

Criteria	L	B	I	A	S	P	O	R
Fulfilled	✓	✓	✓	✓	✓	✓	✓	✓

Figure 11: Example of a 7×7 Pattern Grid

4. Evaluation

To gain initial insights about our pattern-based approach, we conducted a crowdsourced between-subjects study with 84 non-experts. We compare 3 pattern designs against juxtaposition as a baseline across six tasks described in detail below. Here, *grid* describes a visualization of a whole day, such as in Fig. 3, and *subgrid* describes a zoomed-in view of a subset of the data, like in Fig. 1.

T1 *"Estimate category agreement between grids"*: Estimate the percentage of agreement of a given category in the grid.
T2 *"Estimate category agreement between subgrids"*: Estimate the percentage of agreement of a given category in a subgrid.
T3 *"Estimate color agreement between grids"*: Estimate the percentage of agreement of a given color in the grid.
T4 *"Estimate color agreement between subgrids"*: Estimate the percentage of agreement of a given color in a subgrid.
T5 *"Estimate percentage of a color in a subgrid"*: Find the source of a specific color and estimate its percentage in this subgrid.
T6 *"Estimate the amount of disagreeing colors between subgrids"*: Identify the source of disagreement and estimate the number of colors that disagree with the source.

We derive the following four hypotheses from these tasks:

H1 The pattern-based technique is faster and more accurate than juxtaposition for grids and subgrids (**T1, T2, T3, T4**). For each task, the hypothesis is tested separately.
H2 The pattern-based technique is not significantly less accurate than juxtaposition for estimating the percentage in a subset of either the model or observation dataset. (**T5**).
H3 The pattern-based technique is not significantly slower than juxtaposition for estimating the values in a subset of either the model or observation dataset (**T5**).
H4 The pattern-based technique allows for quicker and more accurate detection of disagreements against a specific category when comparing a subset of the model and observation datasets (**T6**).

Proceedings published by Eurographics - The European Association for Computer Graphics.

We use observational and model-based data representing cloud properties, categorizing different cloud types based on ground-based remote sensing measurements. We use cloud particle classification data for the Ny-Ålesund region, Norway, and ICON model data. A single dataset (12km in height, 24h of data) has $12,000 * 86,400 = 1,036,800,000$ cells. We use minimum frequency pooling [ZYW13] with a window of 30×30, resulting in a $400 * 2880 = 1,152,000$ grid cells.

The study was preregistered on OSF [PSvLK25]. Participants were recruited via Prolific and were paid £10.80 per hour. They were at least 18 years old, had no color vision deficiencies, and spoke English fluently. We randomly assigned 21 participants to each design.

Each task is performed three times by every participant, and each time, a different set of grids is used to increase the consistency. Grid 1 represents observation and Grid 2 prediction. Dates range from 14.06.2017 to 16.06.2017 (3 days). As the tasks are sufficiently different, we reuse datasets across tasks. Tasks **T1** and **T3** use the same grids; Tasks **T2**, **T4**, **T5**, and **T6** all use small subgrids of the **T1** grids, with Tasks **T2** and **T5** using the same subgrids. Questions within the tasks are randomized.

For each question, we measure the response error (the difference between the participant's answer and the correct answer) and the response time. For each task-design combination, we average the error across all stimuli (average response error) and the response times (average response time).

To increase reliability, we exclude participants from the analysis who fail both attention checks or the task designed to assess understanding of the technique. Data points where participants take >3x the average time to answer a question, or that are extraordinarily fast (<2 sec. on average) are excluded. Further excluded are responses for tasks in which participants stated they did not understand the example question. Responses for tasks in which participants stated they did not include the color white during estimation are also excluded.

To test our hypotheses, we use a Shapiro–Wilk test for each task to check for normality of the data. Upon success, we do a two-sided t-test (error and time). If the test fails, we perform a two-sided Wilcoxon rank-sum test (error and time). We use Bonferroni correction to compensate for performing multiple tests. We consider alpha values below 0.05 as significant. Our target sample size is 84 participants (21 per group) after filtering for exclusion criteria, determined via a power analysis.

4.1. Results

We present the results of our analysis per task. We deviated from the original analysis plan by not excluding "participants who fail the task designed to assess understanding of the technique", of which an example is shown in Fig. 12. Responses to this task were too varied to exclude it (see Fig. 13).

T1: Estimate category agreement between grids. Fig. 14a (error) and Fig. 14c (time). Error and response time are not normally distributed. No significant differences were detected.

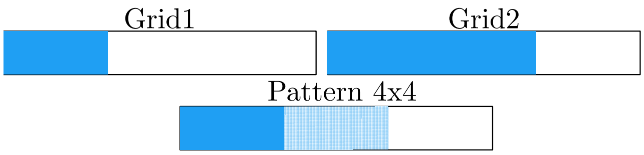


Figure 12: Example stimuli for task understanding. The two grids (top) and the resulting juxtaposition with a 4x4 pattern (bottom). Grid size: 2880x400, i. e. patterns are visible as superposition.

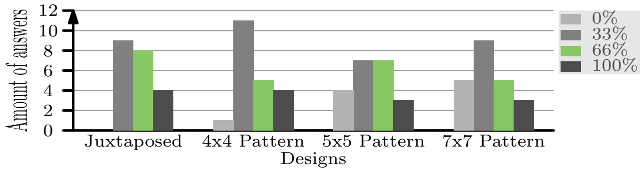


Figure 13: Results from task understanding.

T2: Estimate category agreement between subgrids. Fig. 14b (error) and Fig. 14d (time). The error is normally distributed for Patterns 4×4 and 7×7. No significant differences were detected for the error. A significant difference in average response time was detected between the 4×4 ($\mu = 27s$) and the 5×5 ($\mu = 16s$) patterns.

T3: Estimate color agreement between grids. Fig. 14e (error) and Fig. 14g (time). Estimation for the 5×5 Pattern is normally distributed. No significant differences were detected.

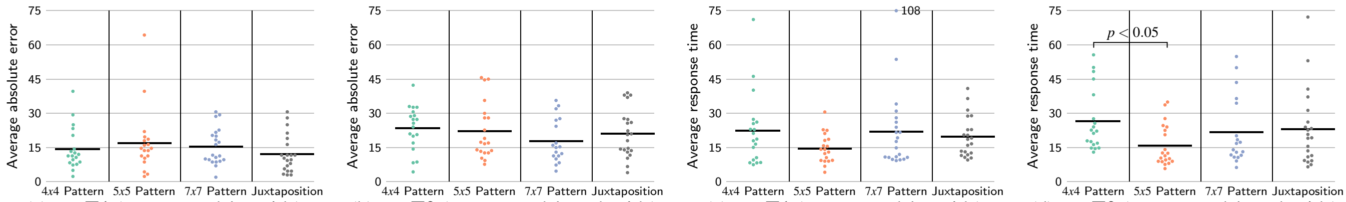
T4: Estimate color agreement between subgrids. Fig. 14f (error) and Fig. 14h (time). Estimation for Pattern 5×5 is normally distributed. No significant differences were detected.

T5: Estimate percentage of a color in a subgrid. Fig. 14i (error) and Fig. 14k (time). Error and response time are not normally distributed. No significant differences were detected in the response time. A significant difference in average absolute error was detected between Juxtaposition ($\mu = 7.6$) and all patterns (4×4 ($\mu = 18$), 5×5 ($\mu = 18.6$), 7×7 ($\mu = 18.1$)).

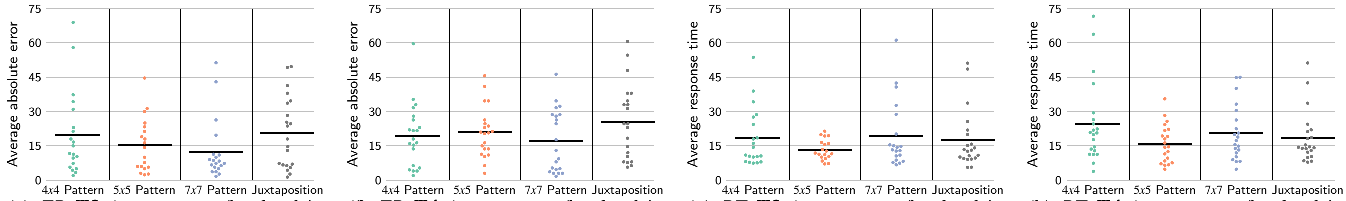
T6: Estimate the amount of disagreeing colors between subgrids. Fig. 14j (error) and Fig. 14l (time). Error and time are not normally distributed. No significant differences were detected.

4.2. Discussion

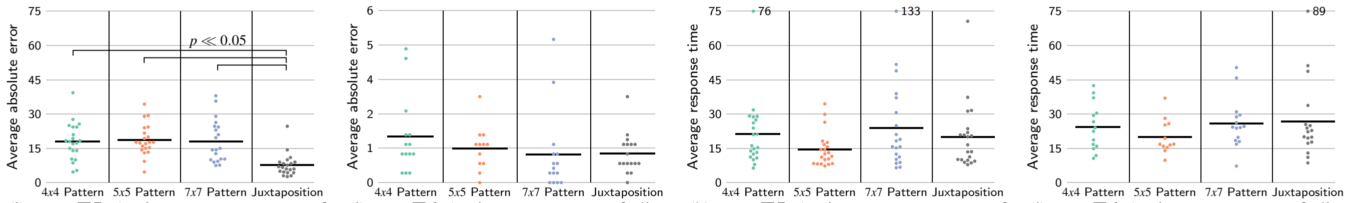
Based on the results, we cannot confirm our hypotheses, as only a few significant effects were found. This is a surprising result given the literature and design. Thus, we analyzed the error distribution in more detail. The results show high average absolute errors and high absolute-error variance across all tasks and designs, except for T5 (Juxtaposition). For T1-T5, the average absolute error is 15-20%. The only exception is the juxtaposed design in T5, where participants needed to estimate the percentage of a specific color in *one* subgrid. For small but complete grids, similar large variances in the errors have been observed in both juxtaposed and superimposed



(a) *ER* **T1** (agreement b/w grids). (b) *ER* **T2** (agreement b/w subgrids). (c) *RT* **T1** (agreement b/w grids). (d) *RT* **T2** (agreement b/w subgrids).



(e) *ER* **T3** (agreement of color b/w grids). (f) *ER* **T4** (agreement of color b/w subgrids). (g) *RT* **T3** (agreement of color b/w grids). (h) *RT* **T4** (agreement of color b/w subgrids).



(i) *ER* **T5** (estimate percentage of color in one subgrid). (j) *ER* **T6** (estimate amount of disagreeing colors b/w subgrids). (k) *RT* **T5** (estimate percentage of color in one subgrid). (l) *RT* **T6** (estimate amount of disagreeing colors b/w subgrids).

Figure 14: Results of the crowdsourced user study with *ER* for average absolute error and *RT* for average response time. Each circle represents the average error of one participant per task per design. The black horizontal bar represents the mean value. For all figures, lower values are better. All significant differences between means are shown.

designs [PBGA22] across different tasks. For sparse grids in the context of network visualizations, however, significant differences have been found [MGK11], with superimposed designs performing better, although error rates remain relatively high.

5. Conclusion and Future Work

We presented a carefully crafted set of visual patterns to support visual comparison tasks for large, gridded data with categorical values. We support assessing commonalities and differences, and identifying the source of differences. The patterns are based on criteria for faithful and perceptually oriented data visualization.

We identified and evaluated three patterns in comparison to juxtaposition. The initial results remain somewhat inconclusive, but one very interesting insight is the low level of comprehension of basic visualizations of comparison in the crowdsourced study, in contrast to the pre-study. We intend to complement the evaluation with an expert user study to further confirm our hypotheses.

Our work focuses on gridded data with millions of cells and up to ten categories. We used a static design motivated by the domain and basic design goals. In the future, we wish to increase scalability by exploiting and enhancing approaches on multi-scale and interaction. This would include hierarchical multi-scale approaches, high-level aggregated patterns, and interactive exploratory views or progressive visual analytics (VA) approaches. It would also be interesting to investigate which color combinations are best suited for pattern design.

Acknowledgments

This work was funded by the German Federal Ministry of Research, Technology and Space under the funding code 01LK2204A. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Christian Krauter. The responsibility for the content of this publication lies with the authors.

References

[AK15] Asadieh B., Krakauer N. Y.: Global trends in extreme precipitation: climate models versus observations. *Hydrology and Earth System Sciences 19*, 2 (Feb. 2015), 877–891. doi:10.5194/hess-19-877-2015. 1, 2

[CCM*14] Chen H., Chen W., Mei H., Liu Z., Zhou K., Chen W., Gu W., Ma K.-L.: Visual Abstraction and Exploration of Multi-class Scatterplots. *IEEE Trans. Visualization Computer Graphics 20*, 12 (12 2014), 1683–1692. doi:10.1109/TVCG.2014.2346594. 3

[CJS*21] Cakmak E., Jäckle D., Schreck T., Keim D. A., Fuchs J.: Multiscale visualization: A structured literature analysis. *IEEE Trans. Visualization Computer Graphics 28*, 12 (2021), 4918–4929. 2

[CZF*22] Chen X., Zhang J., Fu C.-W., Fekete J.-D., Wang Y.: Pyramid-based scatterplots sampling for progressive and streaming data visualization. *IEEE Trans. Visualization Computer Graphics 28*, 1 (Jan. 2022), 593–603. doi:10.1109/tvcg.2021.3114880. 3

[DCM*22] Deng D., Cui W., Meng X., Xu M., Liao Y., Zhang H., Wu Y.: Revisiting the design patterns of composite visualizations. *IEEE Trans. Visualization Computer Graphics 29*, 12 (2022), 5406–5421. doi:10.1109/TVCG.2022.3213565. 2

[Deu25] Deutscher Wetterdienst (DWD): ICON - description of the numerical weather prediction model, 2025. URL: https://www.dwd.de/EN/research/weatherforecasting/num_modelling/01_num_weather_prediction_modells/icon_description.html. 2

[FGDD15] FERNÁNDEZ Á., GONZÁLEZ A. M., DÍAZ J., DORRONSORO J. R.: Diffusion maps for dimensionality reduction and visualization of meteorological data. *Neurocomputing 163* (Sept. 2015), 25–37. doi:10.1016/j.neucom.2014.08.090. 3

[Fin25] FINNISH METEOROLOGICAL INSTITUTE (FMI): CloudNet documentation, 2025. URL: https://docs.cloudnet.fmi.fi/. 2

[Gle18] GLEICHER M.: Considerations for visualizing comparison. *IEEE Trans. Visualization Computer Graphics 24*, 1 (Jan. 2018), 413–423. doi:10.1109/tvcg.2017.2744199. 1, 2

[GSFT16] GUNTHER T., SCHULZE M., FRIEDERICI A., THEISEL H.: Visualizing volcanic clouds in the atmosphere and their impact on air traffic. *IEEE Computer Graphics Applications 36*, 3 (May-Jun 2016), 36–47. doi:10.1109/MCG.2015.121. 2

[HCSG18] HEIMERL F., CHANG C.-C., SARIKAYA A., GLEICHER M.: Visual designs for binned aggregation of multi-class scatterplots. *arXiv preprint arXiv:1810.02445* (Oct. 2018). doi:10.48550/ARXIV.1810.02445. 3

[HM23] HEER J., MORITZ D.: Mosaic: An architecture for scalable & interoperable data views. *IEEE Trans. Visualization Computer Graphics 30*, 1 (2023), 436–446. doi:10.1109/TVCG.2023.3327189. 2

[HSK*20] HU R., SHA T., KAICK O. V., DEUSSEN O., HUANG H.: Data sampling in multi-view and multi-class scatterplots via set cover optimization. *IEEE Trans. Visualization Computer Graphics 26*, 1 (Jan. 2020), 739–748. doi:10.1109/tvcg.2019.2934799. 3

[JBS08] JANICKE H., BOTTINGER M., SCHEUERMANN G.: Brushing of attribute clouds for the visualization of multivariate data. *IEEE Trans. Visualization Computer Graphics 14*, 6 (Nov. 2008), 1459–1466. doi:10.1109/tvcg.2008.116. 2

[JM11] JENSEN M. S., MATHEWSON K. E.: Simultaneous perception of both interpretations of ambiguous figures. *Perception 40*, 8 (2011), 1009–1011. doi:10.1068/p6880. 5

[JVDF19] JO J., VERNIER F., DRAGICEVIC P., FEKETE J.-D.: A declarative rendering model for multiclass density maps. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (Jan. 2019), 470–480. doi:10.1109/TVCG.2018.2865141. 3

[LJS21] LYI S., JO J., SEO J.: Comparative layouts revisited: Design space, guidelines, and future directions. *IEEE Trans. Visualization Computer Graphics 27*, 2 (2021), 1525–1535. doi:10.1109/TVCG.2020.3030419. 2

[LRS10] LUBOSCHIK M., RADLOFF A., SCHUMANN H.: A new weaving technique for handling overlapping regions. In *Int. Conf. Advanced Visual Interfaces* (May 2010), ACM. doi:10.1145/1842993.1842999. 2, 3

[LS15] LIU X., SHEN H.-W.: The effects of representation and juxtaposition on graphical perception of matrix visualization. In *Conf. Human Factors in Computing Systems* (New York, NY, USA, 2015), ACM, p. 269–278. doi:10.1145/2702123.2702217. 2

[LSL*22] LI Z., SHI R., LIU Y., LONG S., GUO Z., JIA S., ZHANG J.: Dual space coupling model guided overlap-free scatterplot. *IEEE Trans. Visualization Computer Graphics* (2022), 1–11. doi:10.1109/tvcg.2022.3209459. 3

[MGK11] MELVILLE A. G., GRAHAM M., KENNEDY J. B.: Combined vs. separate views in matrix-based graph analysis and comparison. In *Int. Conf. Information Visualisation* (2011), IEEE, pp. 53–58. doi:10.1109/IV.2011.49. 7

[Mil07] MILLER J. R.: Attribute blocks: Visualizing multiple continuously defined attributes. *IEEE Computer Graphics Applications 27*, 3 (May 2007), 57–69. doi:10.1109/mcg.2007.54. 2

[NLHS20] NGUYEN Q. V., LIN HUANG M., SIMOFF S.: Enhancing Scatter-plots with Start-plots for Visualising Multi-dimensional Data. In *Int. Conf. Information Visualisation* (Sept. 2020), IEEE, pp. 80–85. doi:10.1109/IV51561.2020.00023. 3

[NM17] NARAGINO Y., MISUE K.: Visualization techniques for the comparative analysis of weighted free trees. In *Int. Conf. Information Visualisation* (2017), pp. 45–51. doi:10.1109/iV.2017.30. 2

[PBGA22] POMMÉ L.-E., BOURQUI R., GIOT R., AUBER D.: Relative confusion matrix: Efficient comparison of decision models. In *Int. Conf. Information Visualisation* (2022), IEEE, pp. 98–103. doi:10.1109/IV56949.2022.00025. 2, 7

[PSvLK25] PANDEY S., SONDAG M., VON LANDESBERGER T., KRAUTER C.: Pre-registration for comparison of categorical data on grids using a pattern-based approach, 2025. OSF Pre-registration. URL: https://osf.io/p9qzg. 6

[RB20] RÖBER N., BÖTTINGER M.: Visuelle analyse großer daten in der klimaforschung. *Informatik Spektrum 42*, 6 (Jan. 2020), 410–418. doi:10.1007/s00287-019-01222-w. 2

[SE20] SCHEMANN V., EBELL K.: Simulation of mixed-phase clouds with the ICON large-eddy model in the complex Arctic environment around Ny-Ålesund. *Atmospheric Chemistry and Physics 20*, 1 (jan 2020), 475–485. URL: https://acp.copernicus.org/articles/20/475/2020/, doi:10.5194/acp-20-475-2020. 1

[SG18] SARIKAYA A., GLEICHER M.: Scatterplots: Tasks, data, and designs. *IEEE Trans. Visualization Computer Graphics 24*, 1 (Jan. 2018), 402–412. doi:10.1109/tvcg.2017.2744184. 3

[Spi94] SPILLMANN L.: The Hermann grid illusion: A tool for studying human perception. *Perception 23*, 6 (1994), 691–708. doi:10.1068/p230691. 3, 4

[THS*20] TAO W., HOU X., SAH A., BATTLE L., CHANG R., STONEBRAKER M.: Kyrix-s: Authoring scalable scatterplot visualizations of big data. *IEEE Trans. Visualization Computer Graphics 27*, 2 (2020), 401–411. doi:10.1109/TVCG.2020.3030372. 2

[WHWC12] WICKHAM H., HOFMANN H., WICKHAM C., COOK D.: Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics 23*, 5 (2012), 382–393. doi:10.1002/env.2152. 3

[WP13] WARE C., PLUMLEE M. D.: Designing a better weather display. *SAGE Information Visualization 12*, 3-4 (2013), 221–239. doi:10.1177/1473871612465214. 2

[ZLD*15] ZHAO J., LIU Z., DONTCHEVA M., HERTZMANN A., WILSON A.: MatrixWave: Visual Comparison of Event Sequence Data. In *Conf. Human Factors in Computing Systems* (04 2015), ACM, pp. 259–268. doi:10.1145/2702123.2702419. 1

[ZYW13] ZUBAIR S., YAN F., WANG W.: Dictionary learning based sparse coefficients for audio classification with max and average pooling. *Digital Signal Processing 23*, 3 (2013), 960–970. doi:10.1016/j.dsp.2013.01.004. 6