Semantically Aggregated Ensemble Visualization In Meteorology

Author: Kaspar Tvedt Moberg

Supervisors: Helwig Hauser

UNIVERSITETET I BERGEN
Fakultet for naturvitenskap og teknologi

April, 2025

Abstract

Lorem ipsum dolor sit amet, his veri singulis necessitatibus ad. Nec insolens periculis ex. Te pro purto eros error, nec alia graeci placerat cu. Hinc volutpat similique no qui, ad labitur mentitum democritum sea. Sale inimicus te eum.

No eros nemore impedit his, per at salutandi eloquentiam, ea semper euismod meliore sea. Mutat scaevola cotidieque cu mel. Eum an convenire tractatos, ei duo nulla molestie, quis hendrerit et vix. In aliquam intellegam philosophia sea. At quo bonorum adipisci. Eros labitur deleniti ius in, sonet congue ius at, pro suas meis habeo no.

Acknowledgements

Est suavitate gubergren referrentur an, ex mea dolor eloquentiam, novum ludus suscipit in nec. Ea mea essent prompta constituam, has ut novum prodesset vulputate. Ad noster electram pri, nec sint accusamus dissentias at. Est ad laoreet fierent invidunt, ut per assueverit conclusionemque. An electram efficiendi mea.

<div align="right">

Your name

Monday 7th April, 2025

</div>

Contents

List of Figures

List of Tables

Listings

Chapter 1

Introduction

Most people regard weather forecasts as a small part of the end of the daily news on TV—a person waving their arms with pictures of clouds on a map behind them. Many do not think about the amount of work and the extent of the data involved in making these forecasts. Each day, multiple times per day, meteorologists try to predict how the weather going forward will develop. They do this by running complex simulations based on the current variables in the atmosphere and see how they affect each other and change over time. These variables, however, are highly volatile, and minor changes to the starting parameters may provide drastic differences as the simulation time increases. These simulations are imperfect, and tiny errors in the starting parameters may result in a wrong prediction. To combat this, meteorologists may run as much as 50 of these simulations simultaneously, but with minor changes in the starting parameters, creating an ensemble of simulations. By doing this, they have a higher chance of capturing reality and seeing how each parameter affects the result. However, now they have 50 predictions, and one prediction has to be chosen and compared to all others. Let us say that one ensemble member describes 15 features in the form of lines on a map; this makes for a total of $15 \times 50 = 750$ lines across the entire ensemble. A common way to visualize this is using a Spaghetti Plot. A problem with Spaghetti Plots is that they get cluttered quickly, making it hard to find features of interest or identify similar features across ensemble members. The high dimensionality of the data also makes standard methods for visualizing the descriptive statistics of ensembles non-viable. We require a better way to visualize data of this magnitude to aid domain experts in making decisions and better their understanding of the data.

In this thesis, we will start by outlining the tasks we want to accomplish by visualizing this data and the requirements for doing so. We will then review the necessary background

to understand this thesis's contents, mainly regarding meteorology. Next, we will propose our solution to the problem, explain how we solved the tasks we set out to solve, and describe the results of our efforts. We will explain important implementation details before we discuss possible shortcomings and possibilities for future work with our method. Lastly, we will summarize and briefly discuss the conclusions of this work.

1.1 Tasks

Munzner[1] recommends anyone designing a visualization to define and abstract the tasks they want to accomplish clearly. In this thesis, we focus on meteorology, but by abstracting these tasks into more general goals, we can apply the same methods to other fields. In this thesis, we will divide our tasks into three categories:

1.1.1 Exploration

When a user is unfamiliar with the data, they will begin exploring it. By doing this, they want to get an overview of the data and locate areas of interest. We want to support the following exploration tasks:

- Easy detection of outliers. Outliers include lines in irregular places and lines of irregular shapes.
- Identification of interesting patterns.
- Examine how lines change over time.

1.1.2 Analysis

When the user has a hypothesis about some part of the data, they often want to analyze it. By delving deeper into the data, they can validate their hypothesis or develop new ones. We want to support the following analysis tasks:

- Easy detection of outliers within a cluster. Outliers include lines in irregular places and lines of irregular shapes.
- Show a measure of centrality and spread in a cluster of lines providing a cleaner visualization of the cluster as opposed to other methods like the spaghetti plot.
- Compare or identify relationships between clusters.

1.1.3 Presentation

Once the user makes a conclusion about the data, they may want to present it to others. Here, it is important to be clear about their audience and what they want to show. We want to support the following presentation tasks for a general audience:

- Show a statistically aggregated view of each cluster of lines in an intuitive way.
- Incorporate interactive elements to make viewing and exploring the entire data easier.
- Provide descriptions of the visual design to aid in understanding the features and patterns.

1.2 Requirements

1.3 Background

Chapter 2

Related Work

2.1 Statistics

2.1.1 Contours

2.1.2 Bands

2.1.3 Depth

2.2 Functional Data Analysis (FDA)

2.3 Other

Chapter 3

Visual Design

3.1 The Data

3.1.1 Lines

3.1.2 Time

3.2 Representation

3.2.1 Functional Data Analysis

3.3 Clustering

3.3.1 Clustering Across Time

3.3.2 Clustering Across Ensembles

3.4 Parameterization

3.5 Descriptive Statistics

3.5.1 Statistical Depth

3.5.2 Centrality

3.5.3 Spread

Chapter 4

Results and Evaluation

4.1 Results

4.1.1 Splines

4.1.2 Clustering

4.1.3 Descriptive Statistics

4.2 Evaluation

4.2.1 Tasks

Tasks Accomplished

Future Work

4.2.2 Field Expert Feedback

Chapter 5

Implementation

5.1   Background

Chapter 6

Discussion

6.1   Future Work

Chapter 7

Summary and Conclusions

Glossary

Spaghetti Plot A Spaghetti Plot is a way to visualize many lines by overlaying them on top of each other, often in different colors..

Bibliography

[1] Tamara Munzner. *Visualization Analysis and Design*, chapter 3. CRC Press, 2014.

Appendix A

Generated code from Protocol buffers

Listing A.1: Source code of something

```
1 System.out.println("Hello Mars");
```