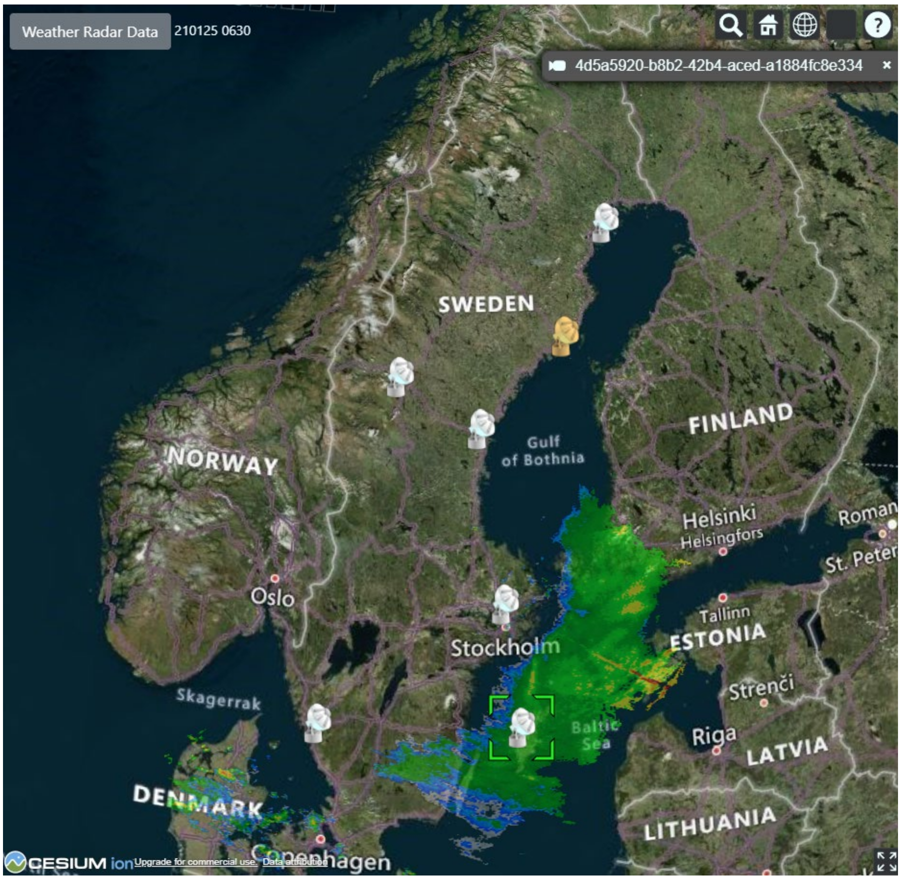
# Interactive visual data analysis of system monitoring data in Air Traffic Management (VDM)


Main authors:

Katerina Vrotsou

Additional project participants and contributors:

Aida Nordman, Carlo Navarra, Anders Dahlgren, Anders Andersson, Geraud Lemmet, Billy Josefsson

# Contents

# 1 Introduction

The project "Interactive visual data analysis of system monitoring data in Air Traffic Management" (VDM) is concerned with exploring the potential of using visual data mining approaches for visually analyzing system data from ATM surveillance systems aiming towards better planning and predictive maintenance.

## 1.1 Svensk projektsammanfattning

VDM- förstudien har syftat till att undersöka potentialen i att använda metoder för visuell datautvinning (Visual Data Mining) för att hantera larm och avvikelser från sensorövervakningssystem, specifikt från radarsystemdata. Syftet har varit att utifrån systemhändelser och metadata, utveckla visuella gränssnitt för att identifiera och analysera mönster i data. Sådana metoder har potential att leda till bättre planering av underhåll genom att: (1) förutsäga fel eller problem och schemalägga tidigare underhåll eller felsökning, (2) bidra med underlag för att senarelägga underhållet till längre intervall.

Projektets första fas fokuserade på kartläggning av behov och datatillgänglighet från de olika övervakningssystem som LFV har i sina anläggningar och potentiella användningsfall togs fram utifrån dessa. Man beslutade att fokusera på radarmeddelanden och undersöka potentialen för att kunna förutsäga avvikelser utifrån sådana data, av särskilt intresse är avvikelser kopplade till väderförhållanden. Metoder för förbearbetning och förberedande datautforskning tillämpades i successiva omgångar för att skapa en grund för vårt tillvägagångssätt. Förberedande dataanalysmetoder tillämpades och utvärderades tillsammans med experter från LFV.

En hypotes formulerades att avvikelser i radarsystemets meddelanden kunde observeras under tider av låga temperaturer och ökad nederbörd. Resonemanget är att isbildning eller förekomsten av klibbig snö kan påverka radarns rotationstid som kan potentiellt observeras innan ett stopplarm inträffar. Liknande beteende kan potentiellt förväntas under perioder med hård vind.

Under VDM-projektet utvecklades två visuella gränssnitt för visuell analys av relevanta radarsystemdata och väderdata under två succesiva faser. Först implementerades ett gränssnitt för att visualisera och jämföra tidsserier av olika relevanta datavariabler för enstaka radarstationer. Därefter utökades gränssnittet med ett geografiskt element som kartlade flera radarstationers lägen och möjliggjorde spatio-temporal utforskning av data från användarvalda radarstationer i detta sammanhang.

VDM-projektet resulterade i visuella gränssnitt som tillåter en användare att kors jämföra flera faktorer som påverkar radarsystem. Inom dessa kan hypoteser om samband mellan olika variabler utforskas. Inga konsekventa korrelationer kunde dock observeras mellan de utforskade variablerna.

## 1.2 English summary

The VDM pre-study has aimed to investigate the potential of using methods for visual data mining to handle alarms and deviations from sensor monitoring systems, specifically from radar system data. The aim has been to, based on system events and metadata, develop visual interfaces to identify and analyze patterns in the data. Such methods have the potential to lead to better

planning of maintenance by: (1) predicting errors or problems and scheduling earlier maintenance or troubleshooting, (2) providing evidence for postponing maintenance to longer intervals.

The first phase of the project focused on the mapping of needs and data availability from the various monitoring systems that LFV has in their facilities and potential use cases were outlined based on these. It was decided to focus on radar messages and investigate the potential of being able to predict deviations based on such data, of particular interest being deviations linked to weather conditions. Pre-processing and preparatory data exploration methods were applied in successive rounds to create a basis for our approach. Preparatory data analysis methods were applied and evaluated together with experts from LFV.

A hypothesis was formulated that deviations in the radar system messages could be observed during times of low temperatures and increased precipitation. The reasoning being that ice-building, or the presence of sticky snow could have an effect on the rotation time of the radar which could be observed before a stopping alarm occurs. Similar behavior could potentially be expected for periods of strong winds.

During the VDM project two visual interfaces were developed for visual analysis of relevant radar system data and weather data during two consecutive phases. First, an interface for visualizing and comparing time series of various relevant data variables for single radar stations was implemented. Thereafter, the interface was extended with a geographical element mapping several radar stations' locations and allowing for spatio-temporal exploration of data from user-selected radar stations in this context.

The VDM project resulted in visual interfaces that allow a user to cross-compare several factors influencing radar systems. Within these, hypotheses regarding relations between different variables can be explored. However, no consistent correlations could be observed between the explored variables.

# 2 Background and Motivation

In the Air Traffic Management (ATM) target area, research and innovation work within the following two themes has been prioritized: (1) Increased automation for increased safety and efficiency in air traffic management, (2) Develop information services and standards for sharing information between ATM actors, which provides a better picture of situational circumstances and reduces system costs.

Sensor monitoring systems today provide large amounts of time series data that can be analyzed to find patterns and relationships in the data. By analyzing floating train data it is possible, for example, to predict rare pitch and draft errors. Although these techniques have proven valuable in achieving longer availability and better reliability of rail transport systems, less work has focused on investigating new predictive maintenance methods based on the analysis of data obtained from ATM system monitoring (e.g. error measurements in radar system). This has been the motivation behind and focus of the VDM project.

LFV has facilities spread across the country where data collection takes place. It includes radar facilities and corresponding sensors for positioning aircraft, radio facilities for air traffic controllerpilot voice communication and navigation facilities, and more recently with remote air traffic control, data is also collected from the camera tower. Today, the facilities are visited regularly for planned preventive maintenance where various service measures are taken at fixed intervals. In addition, the facilities are visited when necessary, when errors occur.

With today's data explosion, the use of machine learning (ML) and data mining (DM) techniques has increased dramatically across disciplines. Such techniques have proven effective in studying underlying processes in systems and identifying patterns within them. At the same time, many ML and DM algorithms act as a "black box" approach that hides the process from the user. By integrating visualization into the DM process, both in the algorithm calculation and to represent the resulting patterns, it becomes possible to effectively include the human in this process (human-in-the-loop) and in a way improve the analysis significantly by producing results that are easier to understand and trust.

# 3 Aims and Delimitations

Based on the motivation above, the overall aim of the VDM project has been to research the use of interactive visual analysis techniques applied to various system and meta- data in order to map and explore the underlying relations and processes within and between monitoring systems as well as identify patterns of events leading to interesting or unwanted outcomes.

To achieve this, the following overall objectives were outlined:

1. A review of the current system monitoring status and the data available
2. Identification of relevant questions and use case scenarios to focus on.
3. Collection of relevant system monitoring data and pre-processing of these for input into pattern mining algorithms for initial analysis, in particular in the visual sequence mining tool, ELOQUENCE, previously developed by the project participants (Vrotsou and Norman, 2019).
4. Iterative adjustment, refinement and testing of the visual sequence mining prototype for enabling sequential pattern identification in the data at hand.
5. Outline of potential future research and development of classification approaches suitable for predicting errors or failures based on the extracted patterns.

The delimitations the project was subject to were mainly concerned with availability of system data for analysis as well as the availability of experts for discussion and validation of results. In particular:

- Being a data-driven project, the availability and accessibility of data has been an essential prerequisite for the success of the project.
- The project aimed to explore the development of sequential pattern mining techniques for the identification of patterns in system data. Consequently, the performance and outcome of any data mining / machine learning depends heavily on the quality and content of the input data.
- Finally, as the LiU researchers conducting the work are themselves not air traffic management experts, the project relies on the collaborative efforts between LiU researchers and LFV experts.

# 4 The VDM project

The VDM project evolved in 3 main phases. An initial *preparatory phase* which included: (1) a survey of the current monitoring systems of LFV and review of the data that they produce and the information these data carry, and (2) the definition of a use case scenario relevant to the aim of the project. A *data collection & acquaintance* phase followed. This phase involved: (1) retrieval, preparation and initial exploration of the data for better understanding it, and (2) initial analysis in search of examples of deviations for subsequent model building. After, an iterative *prototype development* phase took place which involved: (1) implementation & refinement of visual analysis prototypes, (2) exploration of data in these in collaboration with experts.

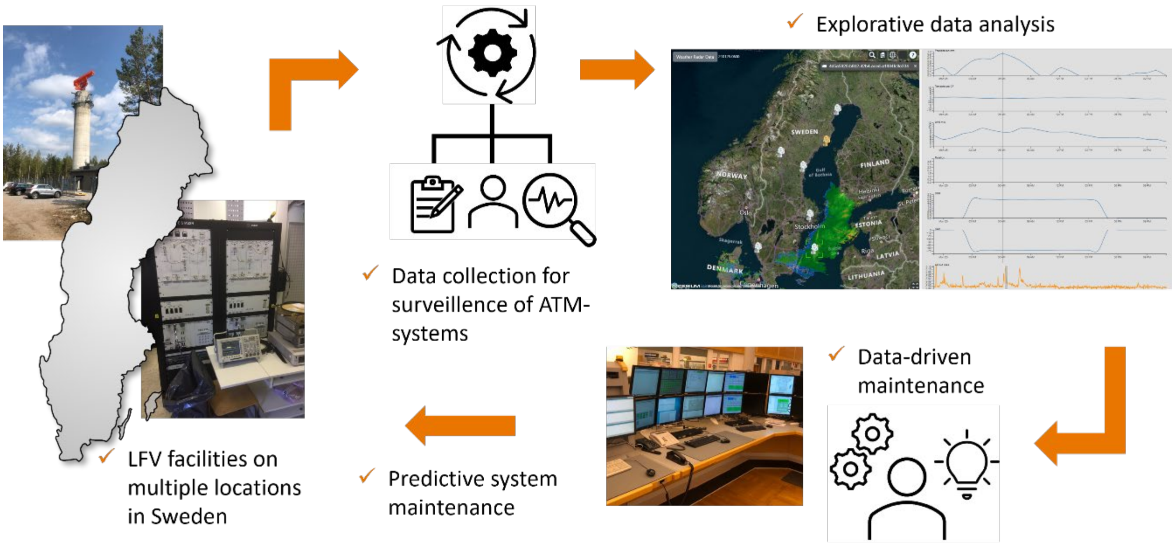



*Figure 1: Envisioned VDM project process outline*

## 4.1 Preparatory phase

The VDM project started with a preparation phase.

### 4.1.1 System monitoring survey

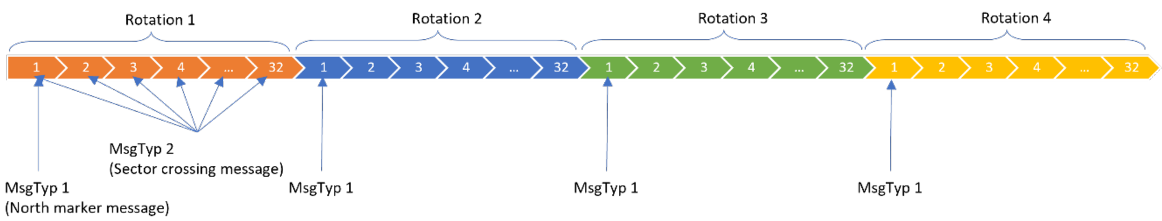
A workshop was held for introducing the project participants to ATM concepts and systems related to surveillance. The types of systems present in air-traffic towers, air-traffic centers and various LFV locations were presented together with their system monitoring capabilities. Their appropriateness and potential relevant to the focus of the project was assessed and the prioritization and interest for LFV was discussed. Finally, surveillance was prioritized as an issue of high importance that was quite under-researched.

Vast amounts of sensor data are produced by surveillance radars (Primary and Secondary Surveillance Radars (PSR, SSR)) and collected by LFV, but little analysis is being performed on them. LFV computes and monitors some values such as biases in azimuth and range, as indicators of diverging radar behavior. A decision was made to make the analysis of radar messages the focus of the project in order to explore whether useful information can be mined from this large dataset. Of particular interest being patterns that could predict a potential stop and/or malfunction of the radar system.

### 4.1.2 Surveillance use case

During a second workshop, a surveillance use case scenario was defined that laid out the main hypotheses to be investigated.

Radar systems transmit service messages with fixed intervals during each rotation. The messages carry information specified in the ASTERIX standard using categories 34[1] and 48[2] depending on the type of radar. *North Marker crossing* messages (type 1 message) and *Sector crossing* messages (type 2 messages) in category 34 were identified relevant for our analysis. Type 1 messages inform of when the antenna crosses the north marker while type 2 messages include information about the: radar station identifier, sector number, time of day, and antenna rotation period among others. Per each radar rotation 32 type 2 messages are transmitted and one type 1 message. If messages are not transmitted, this can be an indicator of malfunction and if several consecutive messages are dropped the radar gets disconnected until the message transmission recovers.




In discussions with surveillance experts from LFV the question of if and how effects of weather conditions (precipitation, wind) can be observed in the radar system message transmission became prominent. Of high interest was the potential prediction of radar disturbance due to heavy ice building that affects its rotation, since this becomes an issue in the winter, particularly in the north of Sweden, and it is an expensive and time-consuming task to reset the radar in function.

Consequently, the overall hypothesis to be tested in the project was whether deviating patterns in a radar's transmitted messages and rotation speed can be observed as precursors for a malfunction and/or failure, especially during severe weather conditions such as high wind speed, low temperatures, increased precipitation amount.

## 4.2 Data collection & acquaintance phase

Following the selection of topic area (*surveillance*), the definition of the use case scenario (*analysis of radar system messages*), and the formulated project hypothesis (*exploration of deviating patterns in message transmission sequences in relation to weather*), the main research phase of the project started. To investigate the project hypothesis, one needs to (1) define what a deviation is; i.e. understand how the sought deviations present themselves in the radar system

[1] https://www.eurocontrol.int/publication/cat034-eurocontrol-specification-surveillance-data-exchange-part-2b
[2] https://www.eurocontrol.int/publication/cat048-eurocontrol-specification-surveillance-data-exchange-asterix-part4

data, and (2) apply analytic approaches to look for similarities/patterns in the message transmission sequence occurring prior to this outcome. Both these steps involve preparation, preprocessing and analysis of data:

(1) *Deviations definition*. Two approaches were formulated for understanding how/when deviations appear in the transmission messages: (1) search for deviations in the message transmission data and then connect these deviations to the weather conditions at the time, and (2) identify periods of severe weather conditions and look for variations in the radar's message transmission during these periods.

(2) *Preparation for sequence analysis*. The initial intention of the project was to apply visual data mining, in particular the visual sequence mining approach previously developed by the project participants (Vrotsou and Nordman, 2019) to discover patterns in the transmission of messages that lead to the previously identified deviations and/or malfunctions. For achieving this, however, first the message transmission data needed to be converted to a discrete format appropriate for sequence mining. Following this, the existing analysis prototypes could be tested for pattern identification and successively adjusted accordingly.

These two data processing steps were initiated in parallel in the project and are described in more detail in the subsections below.

### 4.2.1 Deviations definition

A random sample of data from 2 radar stations was initially retrieved, processed, and visualized to observe the message transmission behavior. The number of dropped messages (consecutive and total) over certain time intervals was traced and the antenna rotation period was plotted. Overall, it was observed that the radar data was very stable and consistent. Rarely were more than a couple of consecutive messages dropped and the average antenna rotation period was keeping stable over time at around 3.99 sec ± 0.1. We found no indication of deviation in these initial datasets.

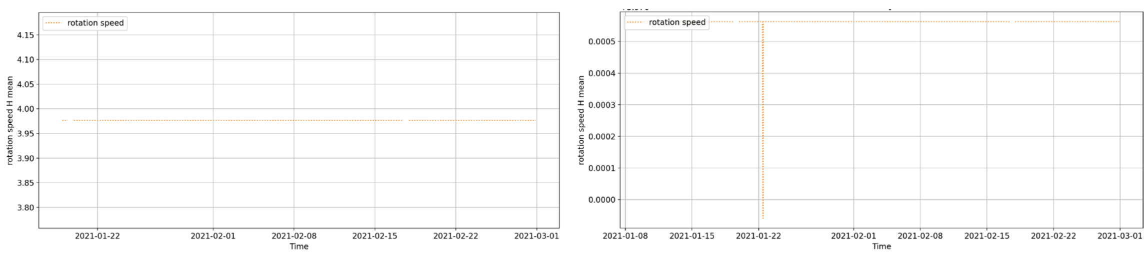



*Figure 2: Example of rotation period timelines.*

Next, we identified time periods when it was known that there has been high wind speed. We retrieved historic weather parameter data from the SMHI archives[3] where hourly data on air temperature, precipitation amount and wind speed are available among others. The weather data were retrieved from the closest weather station to each radar station. We retrieved message

[3] https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer#param=airtemperatureInstant,stations=core

transmission data from these time periods for 3 radar stations. However, no deviations were visible in these time periods either.

A third data collection round was initiated when data was retrieved from radar stations during times of known radar disturbances. These radar disturbances were visible through the monitoring of the radar azimuth bias which was computed by LFV. Interestingly, no significant deviations in the rotation speed of the antenna were observed in these datasets either. Processing of the data indicated that deviations cannot be predicted based on message information regarding dropped messages and rotation speed alone.

We investigated the possibility of processing additional relevant data to understand the course of events in the event of radar deviations. We included the hourly average of azimuth bias that indicates deviations in radar and range bias per hour, and considered variables related to data transmission and power load that can indicate mechanical failures. After exploring data from 3 radar stations during winter, we saw indications that deviations in azimuth bias values correlate with low temperature and increased precipitation (which can then probably indicate "sticky" snow and ice formation) and in some cases also with high values of wind speed (storm). To confirm/reject this hypothesis, our next step was to investigate this with additional data from several radar stations.

Overall, since we did not observe significant variations in the message information that we considered, it was not possible to define explicitly and consistently what constitutes a deviation in the data at hand. Based on the indications we saw that deviation in azimuth bias values may correlate with certain weather conditions, we decided to explore closer the relationship between the message transmission data and several weather parameters, and we initiated the development of a prototype for doing so which is described below in section 4.3.

### 4.2.2 Preparation for sequence analysis

Simultaneously with the data exploration process described above, we initiated the process of preparing the data for application of sequence mining approaches. Since the transmission message data are collected with very high resolution (32 messages per ~4 second rotation) they result in long time series data which had to be discretized before further processing.

We formulated a pipeline to aggregate the very high-resolution radar message transmission time series into discrete event sequences for further analysis using our previously developed sequence mining approach, ELOQUENCE (Vrotsou and Nordman, 2019). This pipeline allowed us to map radar messages in a way that has not been done before, which has provided a better understanding of this data.

The idea has been to combine relevant transmission message information with weather information into compound events that will form a single event sequence representing the radar system function over time. The following figure depicts this idea.

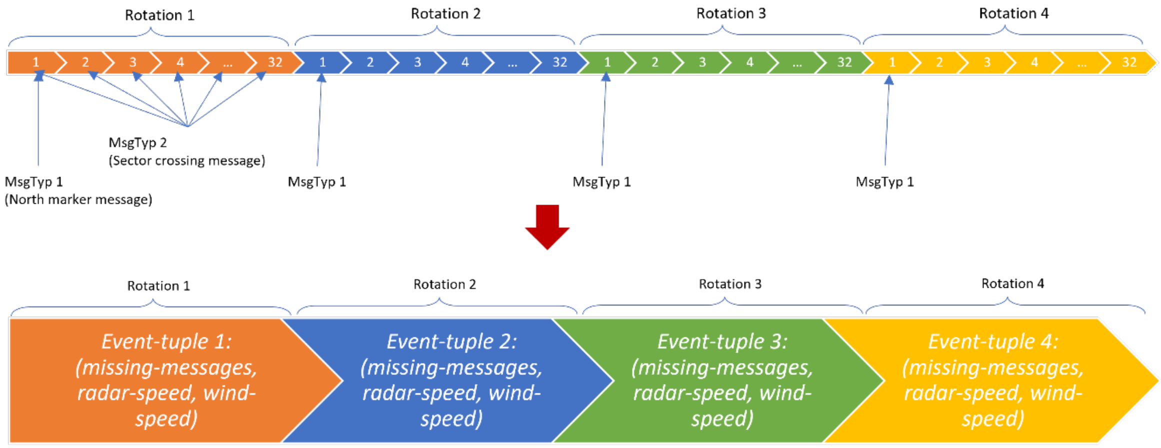



*Figure 3: Aggregation of transmission message related information to discrete sequence of compound events.*

Each event in this event sequence is built up of the values of the separate variables it is composed of, in this case missing message number, rotation period, and wind speed. However, for events to be comparable and for patterns to be able to be identified more efficiently the level of detail that the values are measured with need also to be aggregated. Therefore, instead of using the actual values of the different variables we need to classify these into classes of value.

For dropped messages, the classes were defined based on the number of consecutive messages dropped per rotation cycle into: normal (0 dropped), moderate (1-10 dropped) and severe (over 15 dropped).

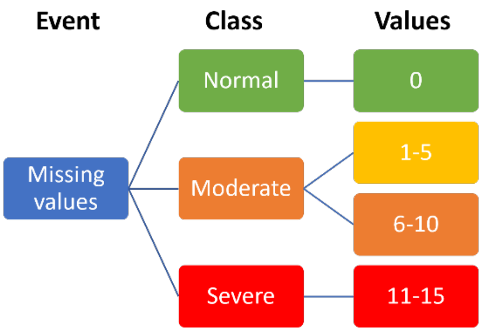



*Figure 4: Severity classification of dropped transmission messages.*

Antenna rotation period was classified based on the computed change of rotation speed into: no change, increasing and decreasing. And the same classification approach was applied to wind speed.

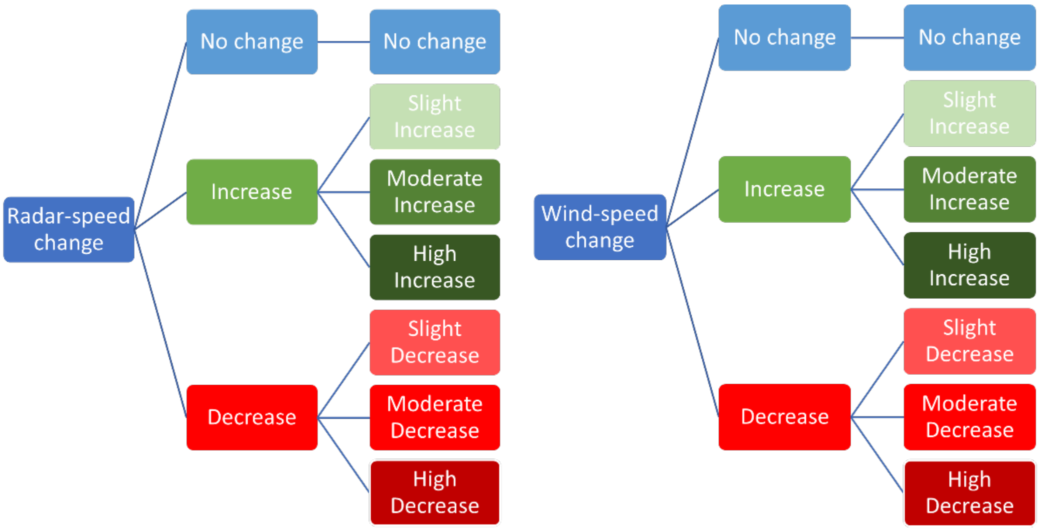



*Figure 5: Classification of rotation speed and wind speed variables.*

Following this process would enable single event sequences to be created per radar station and allow for patterns to be identified that would lead to events indicating deviations in the data.

The initial exploration that was performed on the transmission data, however, did not display enough variation for the planned aggregation to event sequences and the subsequent planned sequence analysis using our existing prototype. Therefore, alternative analysis approaches were instead explored through the creation of new, customized visual interface prototypes. These are described in the following subsection (sec. 4.3).

## 4.3 Prototype development phase

The project diverged from the initial plans of applying sequence mining and redirected towards alternative approaches for visually exploring and analyzing the radar and weather data at hand. The idea of creating a prototype interface for cross-comparing data related to radar systems from various sources in a single view was discussed. The idea was regarded as a relevant approach to flexibly explore any potential relations that were indicated in the initial data processing phase.

A visual interface was created in python for representing data from multiple sources reflecting information for a single radar station. These include information collected directly from the transmitted messages, weather related information, and computed biases in the azimuth and range value. The interface, hence, incorporates line plots for the following variables averaged per hour: number of messages, amplitude of messages, antenna rotation speed, wind speed, air temperature, precipitation amount, radar azimuth bias mean value, radar range bias mean value. A screenshot of the interface is seen in Figure 6.

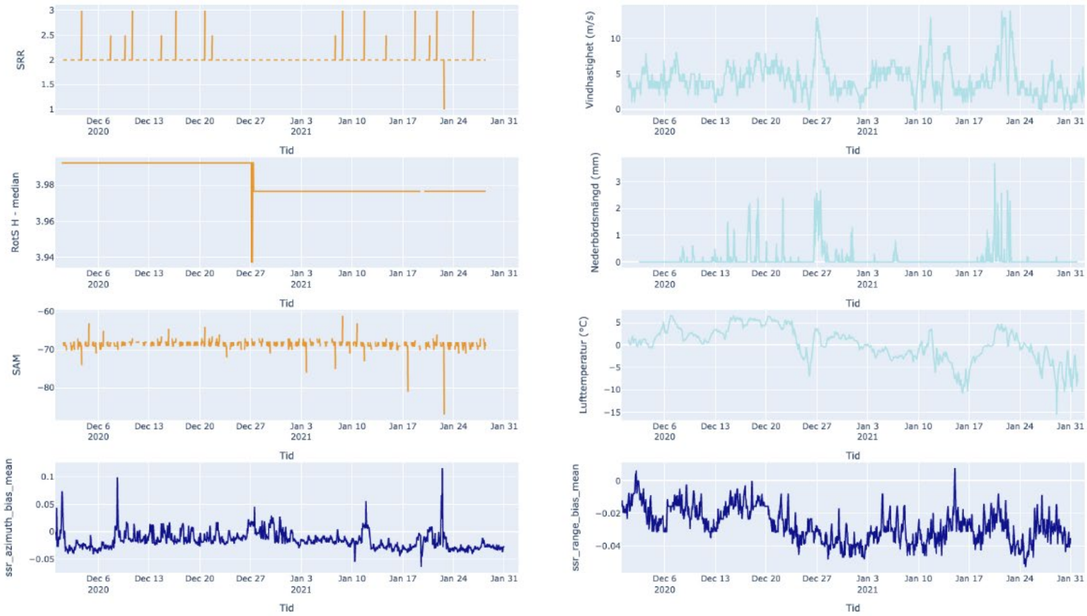



*Figure 6: First visual exploration prototype developed.*

All plots are interactive, so hovering over a lines returns the value in a popup. Also, all plots are linked meaning that if a time period is put in focus by zooming in in one plot, all other plots get zoomed to the same time period so that cross-comparison between plots becomes easier.

This interface provided flexibility to simultaneously view and interact with all these variables which are relevant to understanding the behavior of the radar system.

In a second implementation step of the project a second interface was developed which included a visualization of the spatial distribution of the radar stations making it possible for a user to flexibly change between radar stations to explore their data in an integrated manner. The second interface is composed of a map with markers on the radar system locations. Clicking on a radar, the corresponding data for that radar is represented. Similar to the previous interface all plots of the variable are linked. A screenshot of the interface is seen in Figure 7.

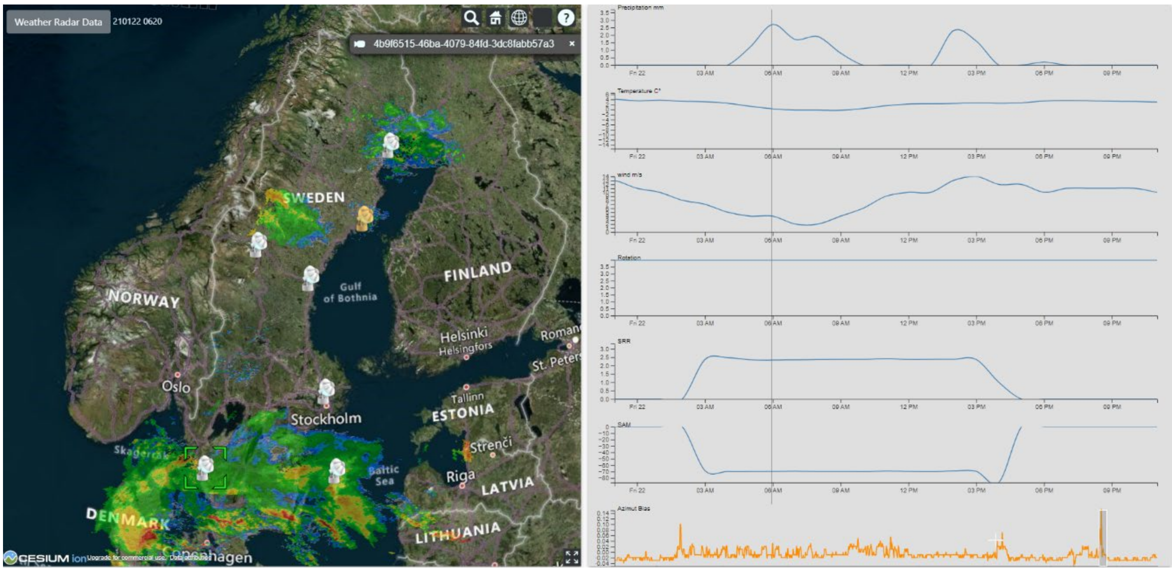



*Figure 7: Second visual exploration prototype developed.*

In addition to the radar system related data, in this second prototype, historic weather radar data, collected from SMHI, has also been incorporated into the visualization. A user can interactively select a time window of interest in the graphs and the radar data over that period are visualized animated over the map. This provided additional input to the corresponding weather conditions relative to the radar system data.

# 5 Deviations from initial project plan

The goal of the project has been to search for patterns in a radar's transmitted messages that lead to deviations in the radar's function.

The initial plan for achieving this was to first define what a deviation is and how it appears in the radar system data (transmitted messages). Secondly, to apply data mining approaches, in particular test, adapt and extend a visual sequence-mining prototype previously developed by the project. Finally, we identify patterns in the message transmission sequence that leads to such deviations.

However, the message transmission data that was collected and explored from various radar stations and under different periods, did not show any concrete disturbances in the number of transmitted messages and rotation speed which was our starting premise. This meant that we could not successfully apply our initially designed sequence mining based analytic approach. Therefore, working towards the same goal, the project redirected and adopted an alternative approach of interactive visual analysis through the two developed visual interfaces.

# 6 Conclusions and future work

The VDM project has successfully worked towards its initially set goal despite the adjustments in the adopted analytic approach.

The project has proposed two successive prototypes for visually exploring and analyzing data variables relevant for monitoring of radar systems. The interactive prototypes make it possible for several influencing factors to be inspected simultaneously and cross-compared.

During the project, we processed and explored data from different radar stations and over different periods of time. While we found indications of correlations between deviating azimuth bias values and low temperature plus increased precipitation, these were not consistent between all the radar/weather stations surveyed. Furthermore, no significant variations could be observed in the number of transmitted messages and the rotations speed of the radar antenna, even during periods of known malfunction. Our initial hypothesis could therefore not be confirmed so far. However, a positive aspect to note is that the lack of observed deviations in the data that was surveyed in this project can be seen as contributing evidence for the robustness of the currently used surveillance systems.

Further exploration of radar system data over additional time periods of known malfunction would be interesting to perform to gather additional evidence. Moreover, a broader user-based evaluation of the developed prototypes with experts in surveillance would be highly relevant to perform as future work. This could provide suggestions for further extensions of the interfaces.

# 7 References

L. Meyer, S. Boonsong, B. Josefsson, A. Nordman, K. Vrotsou, C. Westin, K. J. Klang, J. Lundberg, “Mapping the Decision-Making Process of Conflict Detection and Resolution in En-Route Control: An Eye-tracking based approach”, in SESAR Innovation Days, 2022.

P. K. Muthumanickam, K. Vrotsou, A. Nordman, J. Johansson, and M. Cooper, “Identification of temporally varying areas of interest in long-duration eye-tracking data sets”. In IEEE transactions on visualization and computer graphics, vol. 25, no. 1, pp. 87-97, 2018.

A. Nordman, K. Vrotsou, K.J. Klang, J. Lundberg, L. Meyer, “Extraction of CD&R work phases from eye-tracking and simulator logs: a topic modelling approach”, submitted to 15th USA/Europe Air Traffic Management Research and Development Seminar (ATM2023), 2023.

K. Vrotsou and A. Nordman, “Exploratory Visual Sequence Mining Based on Pattern-Growth,” in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 8, pp. 2597-2610, 1 Aug. 2019, doi: 10.1109/TVCG.2018.2848247.

C. Westin, K. Vrotsou, A. Nordman, J. Lundberg, and L. Meyer, “Visual Scan Patterns in Tower Control: Foundations for an Instructor Support Tool,” in SESAR Innovation Days, 2019.