Identifying false alarms and bird tracks in a full scale radar tracks database using clustering algorithms and SQL Server 2008 Analysis Services

Roald Vang, Roel May & Frank Hanssen
Norwegian Institute for Nature Research (NINA), NO-7047 Trondheim, Norway

Background

All data from the MERLIN Avian Radar System from April 2008 until March 2011 has been processed and automatically stored in a SQL Server 2008 database. This database, however, contains both bird tracks and false alarms. By studying each track drawn out on a map and studying its signature, it is possible to make an educated guess whether a track is a bird or false alarm. However, due to the huge number of tracks (April 2008 – March 2011; horizontal database: 130 million track points ~ 45 GB), it is impossible to do this job manually, and a too time-consuming task to work with the entire database in tools like Excel, SPSS, R, etc. We therefore wanted to take advantage of the powerful quad-core processor server which the database resides on, and develop an automated method for filtering the radar-data.

Since the start of the project, about 2.000 bird track segments have been ground-truthed manually (visually confirmed) within the wind-power plant. These tracks have been used to verify the accuracy of the models created during our work.

Data mining tools used

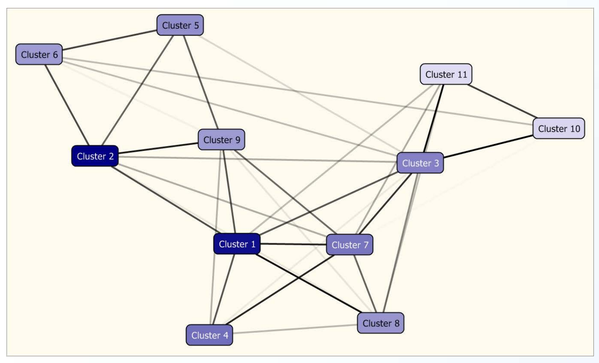
- Microsoft SQL Server 2008
- SQL Server 2008 Analysis Services
- Microsoft Clustering Algorithm
- Statistical programme R version 2.10.1

Data mining steps

1. Create flattened dataset of the entire database to use as basis. A subset of 484,088 tracks was used for testing purposes. For each track a spatial column (Latitude/Longitude) was generated and proportions near roads (<20m) and within radial clutter areas were calculated.

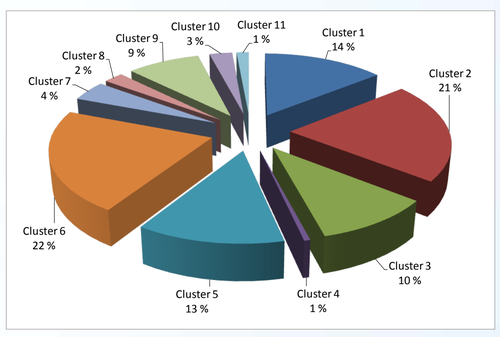
2. For each track average, variance and delta values of parameters deemed biological or radar-technological relevant were calculated (speed, heading, turning angle, track length, reflectivity, target area and shape parameters, etc.). Inclusion of wind speed and direction, and precipitation.

3. The tracks in the test-dataset were clustered using the Microsoft Clustering Algorithm. The Microsoft Clustering algorithm is a segmentation algorithm that uses iterative techniques to group cases in a dataset into clusters that contain similar characteristics. The algorithm first identifies relationships in a dataset and generates a series of clusters based on those relationships.



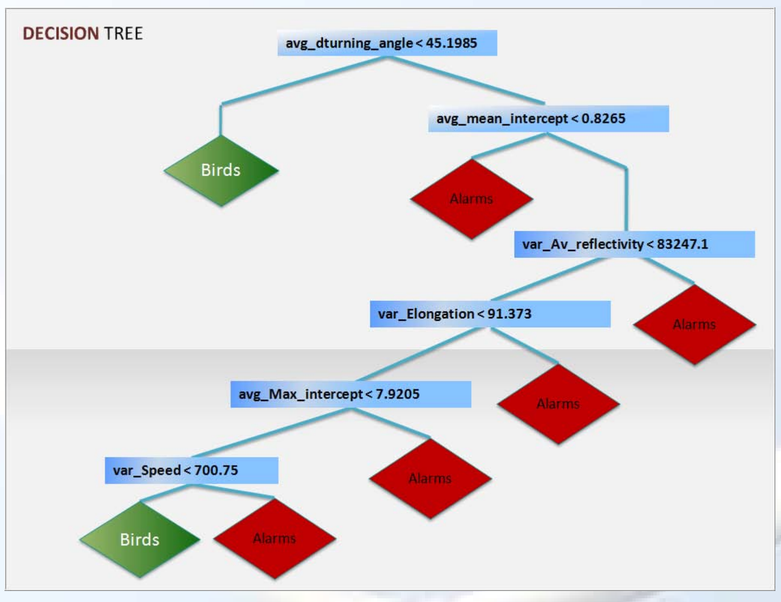
The various clusters in the figure are natural groupings of the tracks based on the chosen parameters.

4. Each cluster was classified as signifying either birds or false alarms (i.e. precipitation, interference, outliers) in a semi-quantitative manner by comparing each cluster's signature (e.g. long straight tracks are likely birds; highly varying track clusters may signify false alarms). Our ground-truthed tracks helped us classify the tracks in the different data mining clusters.



Ground-truthed data divided in clusters

4. We defined the splits between these reclassified clusters based on their signature using a decision tree model. This resulted in a misclassification error rate of 4 % (14,959 of 385,116) for birds, and 2 % (19 of 888) (excluding vehicles and other non-bird tracks; track length > 4) for the ground-truthed tracks.



The figure shows the most significant parameters and their threshold values.

Results

The chosen model was implemented on the entire tracks database and each track was classified as bird, vehicle (i.e. targets following a road) or false alarm.

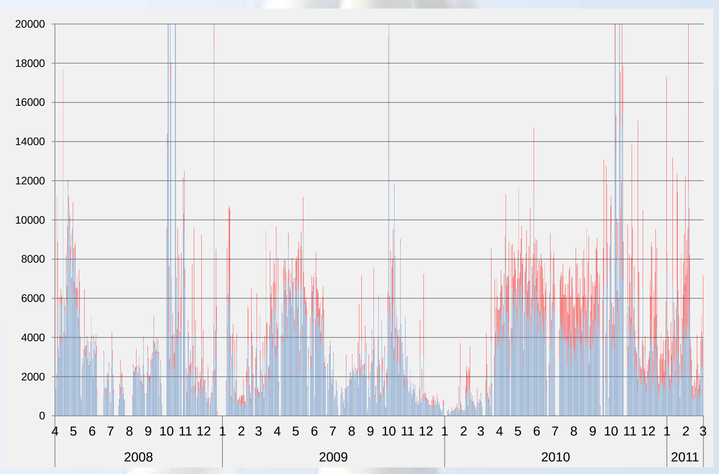
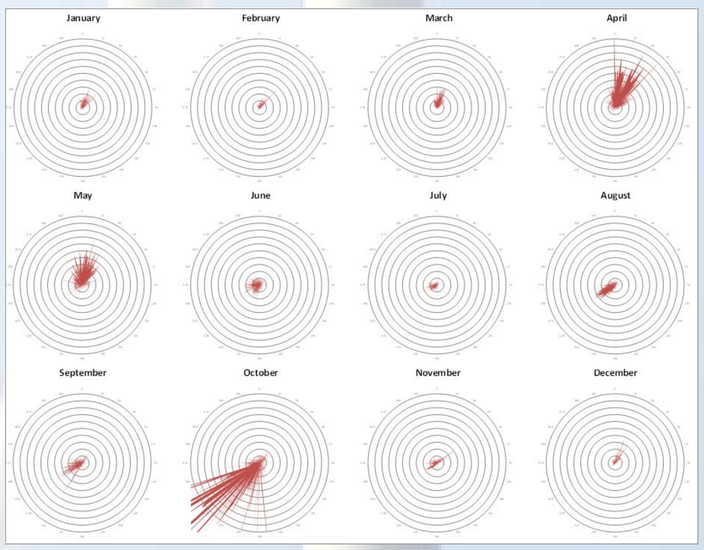


Figure showing all bird tracks (blue) and false alarms (red) per month from April 2008 until March 2011.



Monthly changes in the number of bird tracks by heading.

NINA Norwegian Institute for Nature Research
CEDREN Centre for Environmental Design of Renewable Energy
www.nina.no www.cedren.no