

STATISTICAL LEARNING FOR TRAIN DELAYS AND INFLUENCE OF WINTER CLIMATE AND ATMOSPHERIC ICING

The Department of Mathematics and Mathematical Statistics at Umeå University and The Atmospheric Science Group at Luleå University of Technology jointly worked on a study that investigated the influence of winter climate and atmospheric icing on arrival delays of high-speed passenger trains in northern Sweden.

The results showed that temperature, snow depth, ice/snow precipitation, and train operational direction significantly impact the arrival delay, and the prediction error is less than 10%.

INTRODUCTION

This study is an in-depth investigation of one previous research work that was presented in the *NoICE Highlights No. 2*. It mainly investigates the influence of consecutive winter climate and atmospheric icing from December 2016 to February 2018 on arrival delays of high-speed passenger trains between Umeå and Stockholm. Novel statistical learning approaches, including inhomogeneous Markov chain model and stratified Cox model, are adopted to account for the time-varying risks of train delays. The inhomogeneous Markov chain modelling for the arrival delays uses several covariates, including weather variables, such as temperature, humidity, snow depth and ice/snow precipitation, train operational direction, and findings from the primary delay analysis through stratified Cox model.

The train line under the investigation is shown in *Figure 1*. The total length of the train line is about 711 km, including 116 measuring spots. The weather data is simulated from the Weather Research and Forecasting (WRF) model along the train line with a spatial resolution of 3 km and a temporal resolution of 1 hour. The performance of the fitted inhomogeneous Markov chain model is evaluated by a four-time expanding window walk-forward validation method illustrated in *Figure 2*.



Figure 1. Train line in the region



*Figure 2:
Illustration of the
validation method*

RESULTS

A survival plot from the fitted stratified Cox model shows how survival probabilities of a train running from Umeå to Stockholm vary for the first and second occurrence of primary delays in Figure 3.

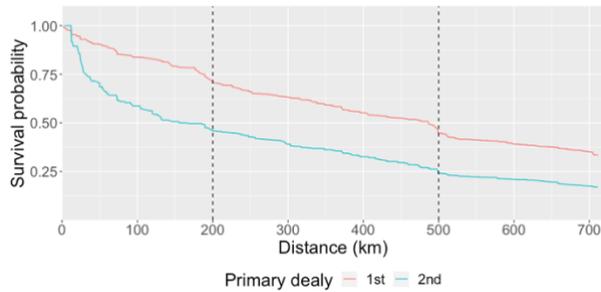


Figure 3: Survival plot for 1st & 2nd primary delays

There are two noticeable probability reductions at 200 km and 500 km on both curves in the figure. It provides evidence that transition intensity in the inhomogeneous Markov chain model varies at these two places.

Table 1: HR from punctuality to arrival delay

Variable	Hazard ratio
Direction	0.586
Temperature	0.958
Snow depth	1.026
Ice/snow precipitation	1.142

Table 2: HR from arrival delay to punctuality

Variable	Hazard ratio
Direction	0.751
Temperature	1.017
Snow depth	0.984
Ice/snow precipitation	0.765

The hazard ratios (HR) of significant variables in the model are presented in Table 1 and 2, for transition of states from punctual to delayed, and from delayed to punctual, respectively.

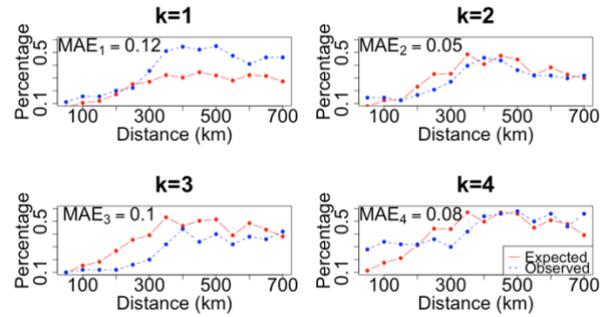


Figure 4: Comparisons between observed and predicted rates of arrival delays over the trip for the four validation periods

Finally, mean absolute error (MAE) is used to measure the performance of the fitted inhomogeneous Markov chain model. Comparisons between observed and predicted rates of arrival delays are plotted in Figure 4. The averaged MAE over the four periods is 0.088, which implies approximately 9% of trains may be misclassified as having arrival delays by the fitted model at a measuring point on the train line.

AUTHOR:

Jianfeng Wang¹, Roberto Mantas Nakhai² and Jun Yu¹

¹ Department of Mathematics and Mathematical Statistics, Umeå University

² Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology



EUROPEAN UNION



European Regional Development Fund



Regional Council of Ostrobothnia

