Additional Material for the Paper:

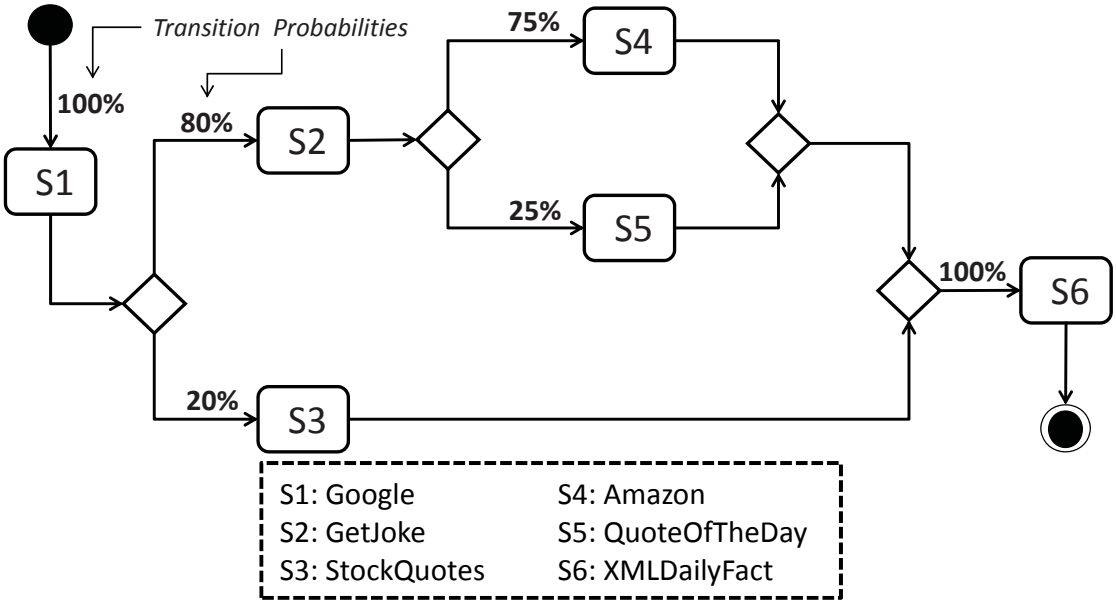
Accurate Service Failure Prediction through Online Testing

This document includes additional information for the experimental setup and results that were not presented in the paper due to space limitations.

1. Experimental Setup:

In this section, we present how we computed the usage rates we presented in TABLE 1 in the paper for the different services.

Our experiments are based on the following workflow:



Workflow and Web services used in our experiments

We simulate the execution of this workflow with the third-party Web services provided by Cavallo et al[1]. The workflow is given in a Markov-like representation, i.e., in addition to the service invocations and control constructs we have added transition probabilities. Those have been used such that we can get the wide range of usage rates by simulating 100, 200 and 400 executions of this workflow over the 2000 data points from the data set. Based on those transition probabilities, the simulation determines the actual execution path of the workflow. At each branch in the workflow, a random variable is used to choose one of the possible paths. As an

[1] B. Cavallo, M. Di Penta, and G. Canfora, "An empirical comparison of methods to support qos-aware service selection," in Proc. 2nd Int'l Workshop on Principles of Engineering Service-Oriented Systems (PESOS'10), 2010, pp. 64–70.

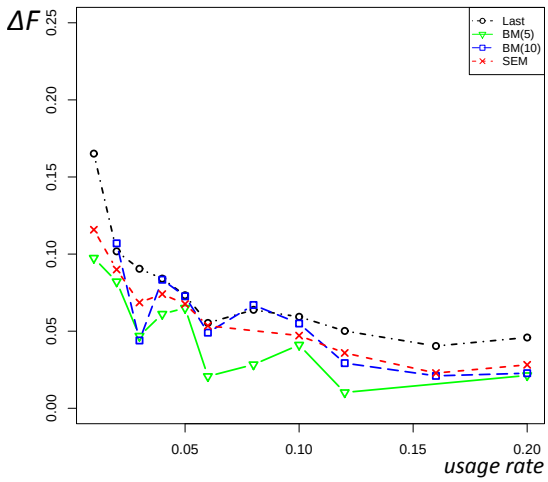
example, the branch from Google to StockQuotes is chosen with a probability of 20%. Using 100 executions of the workflow we obtain the following usage rates: 0.01 for StockQuotes and QuoteOfTheDay, 0.03 for Amazon, 0. 04 for GetJoke, and 0.05 for Google and XMLDailyFact. The remaining usage rates covered in our experiments are computed similarly using 200 and 400 executions of the workflow. As summarized in TABLE 1 of the paper, altogether, we cover the following range of usage rates: 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.10, 0.12, 0.16, and 0.20.

2. Execution and Results:

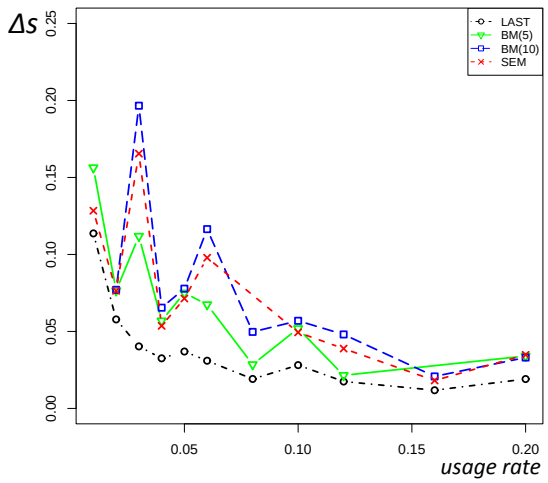
In this section we present the comprehensive experimental results we have computed for answering the research questions RQ 2-RQ 4 in our paper. In particular, due to space limitations, we have only shown a subset of the combinations of the influential factors and we have not been able to present the box plots and the results using specificity (*s*) for the research questions RQ 2-RQ 4.

RQ 2 (*influence of prediction models*)

The below figures show the accuracy gains in terms of the F-measure (*F*) and specificity (*s*) for the different prediction models we have used in our experiments.

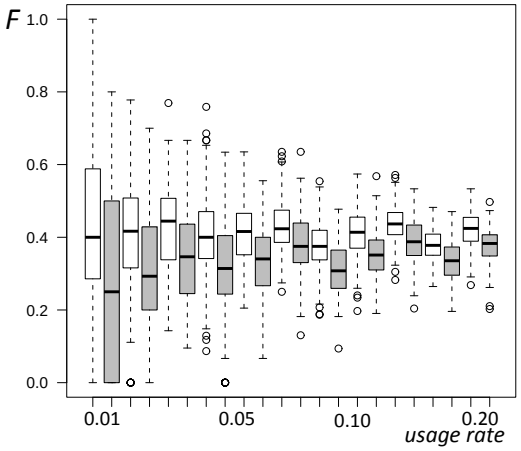


Accuracy gains in terms of *F* using different prediction models, 0.25 failure rate and 0.30 test rate

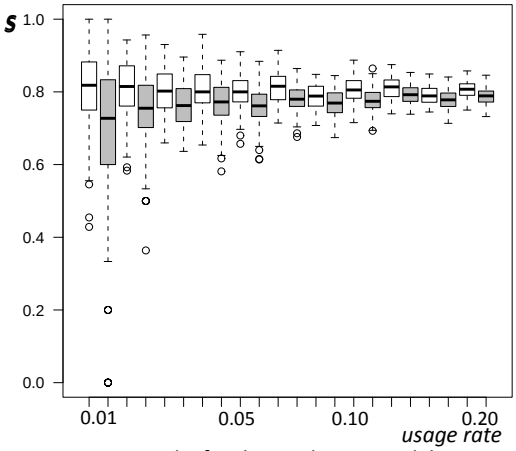


Accuracy gains in terms of *s* using different prediction models, 0.25 failure rate and 0.30 test rate

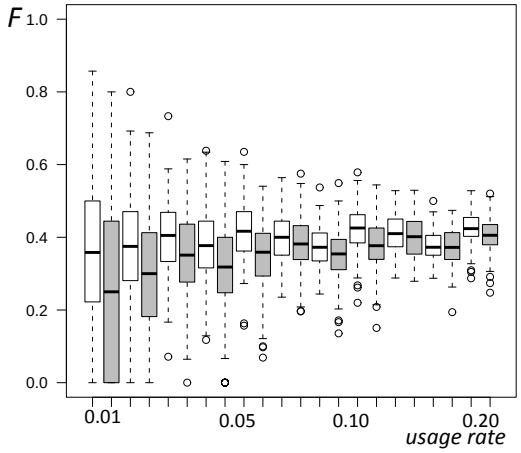
The following figures show the box plots for both *F* and *s* results presented above. White boxes represent results of predication using monitoring and online testing (*M&OT*) and the grey boxes represent results of predication using monitoring only (*M*).



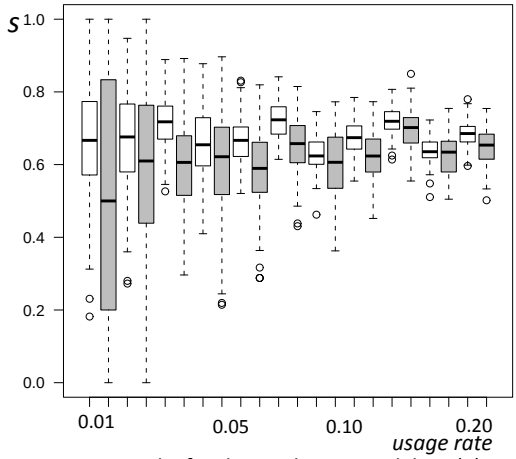
F results for the prediction model Last



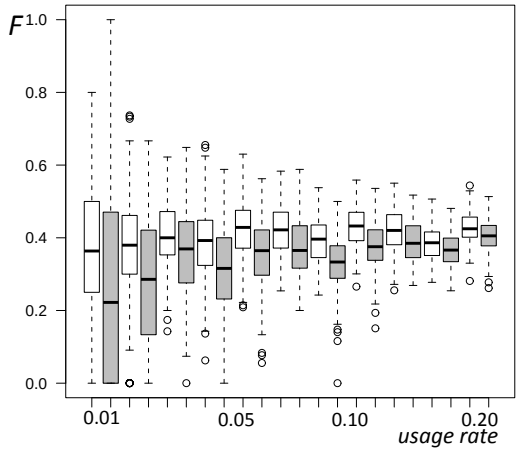
s results for the prediction model Last



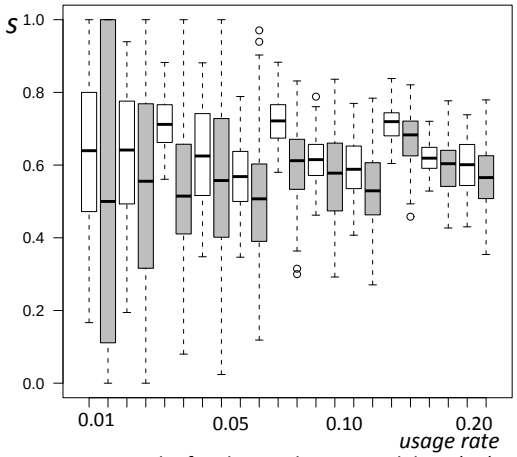
F results for the prediction model BM(5)



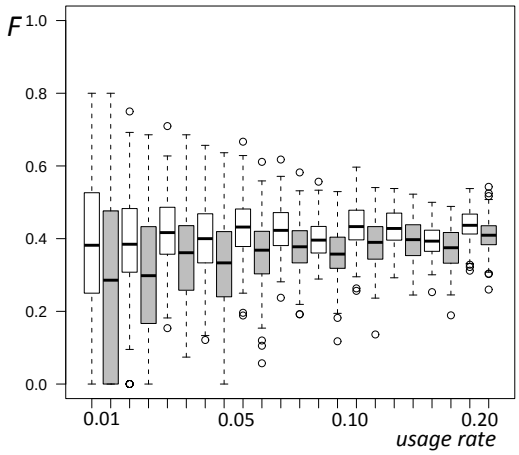
s results for the prediction model BM(5)



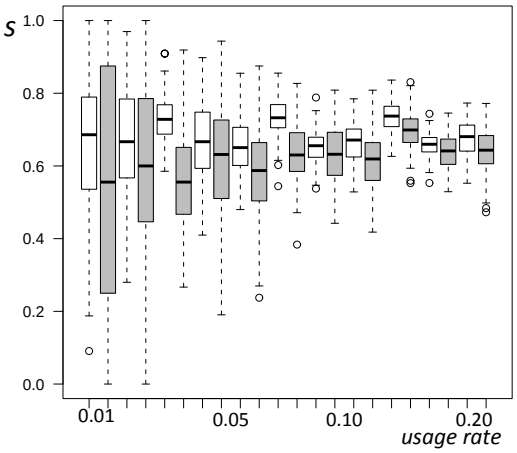
F results for the prediction model BM(10)



s results for the prediction model BM(10)



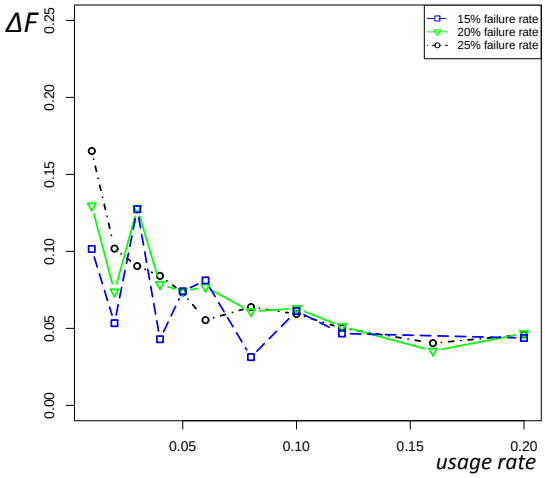
F results for the prediction model SEM



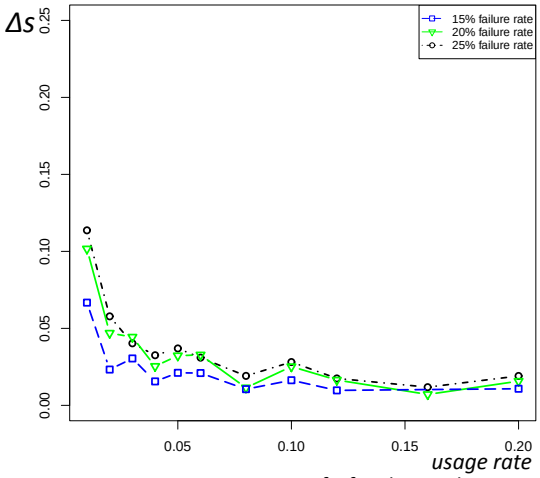
s results for the prediction model SEM

RQ 3 (*influence of failure rate*)

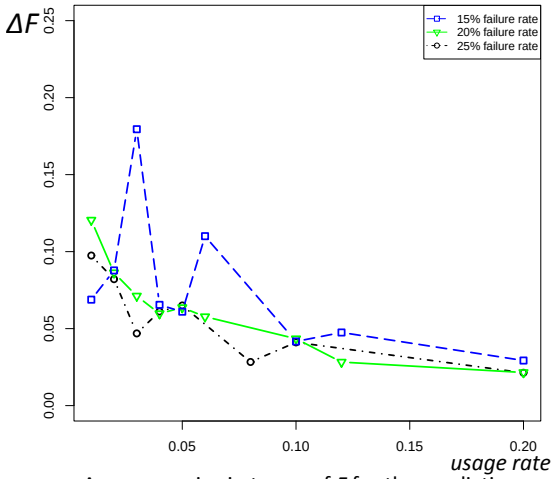
The below figures show the accuracy gains in terms of the F-measure (*F*) and specificity (*s*) for the different prediction models using different failure rates (25%, 20% and 15%) we have used in our experiments.



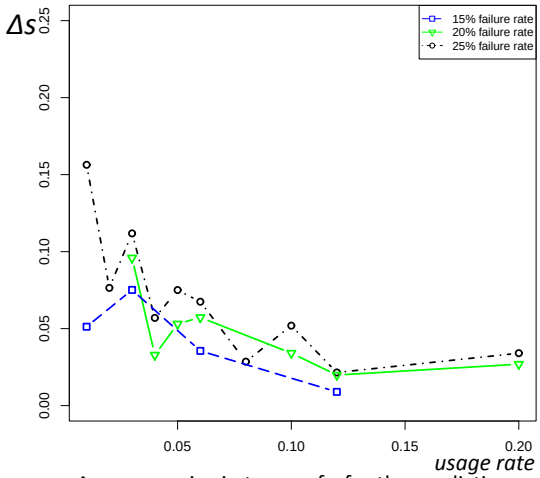
Accuracy gains in terms of *F* for the prediction model Last using 0.30 test rate



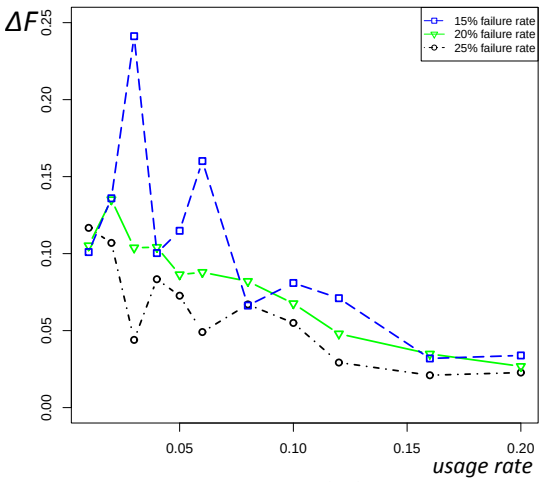
Accuracy gains in terms of *s* for the prediction model Last using 0.30 test rate



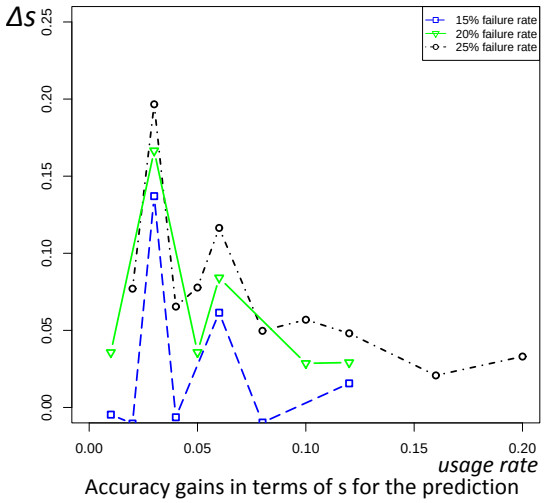
Accuracy gains in terms of *F* for the prediction model BM(5) using 0.30 test rate



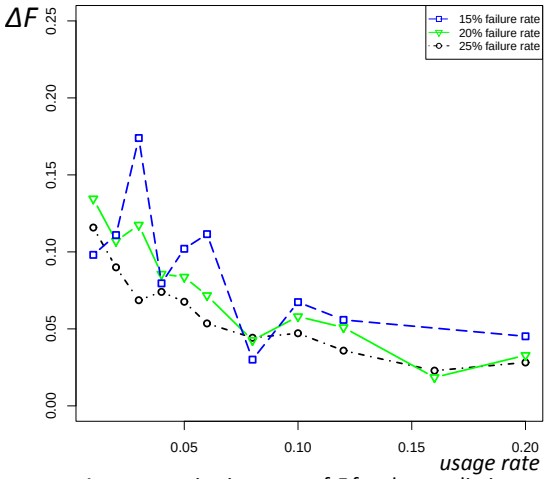
Accuracy gains in terms of *s* for the prediction model BM(5) using 0.30 test rate



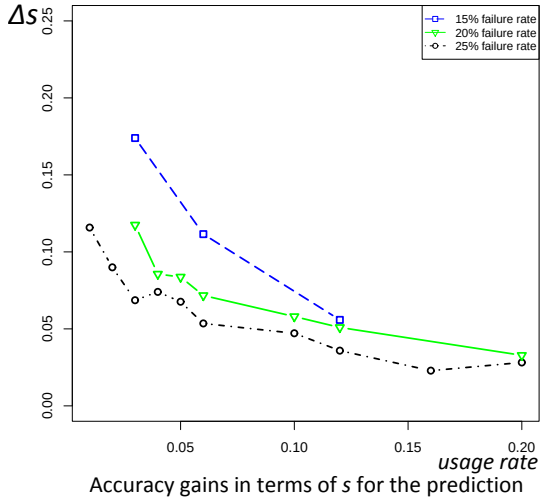
Accuracy gains in terms of *F* for the prediction model BM(10) using 0.30 test rate



Accuracy gains in terms of s for the prediction model BM(10) using 0.30 test rate

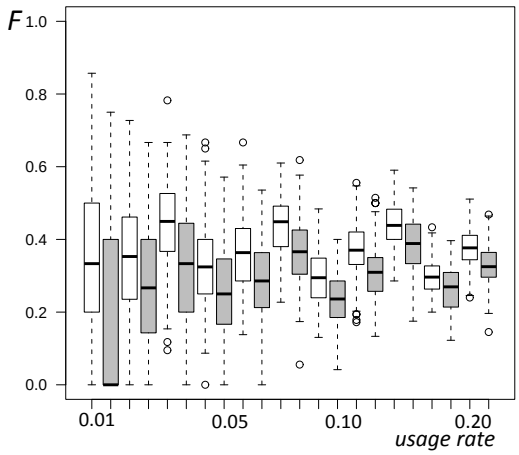


Accuracy gains in terms of *F* for the prediction model SEM using 0.30 test rate

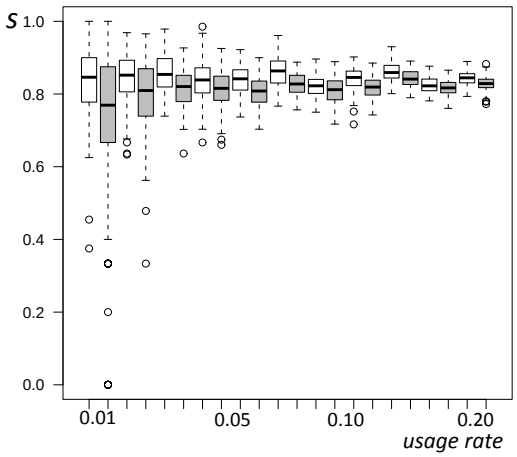


Accuracy gains in terms of *s* for the prediction model SEM using 0.30 test rate

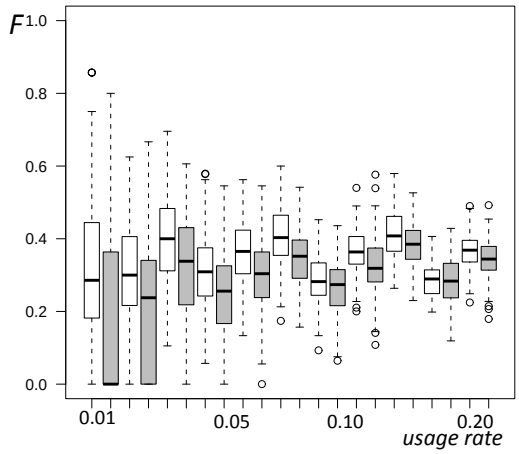
The following figures show the box plots for both *F* and *s* results shown above. White boxes represent results of predication using monitoring and online testing (*M&OT*) and the grey boxes represent results of predication using monitoring only (*M*). Please note that we have already presented the boxplots for the results using 25% failure rate in RQ2 above.



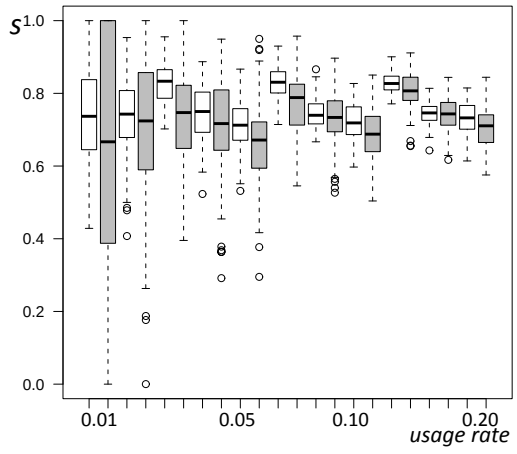
F results for the prediction model Last using 20% failure rate and 0.30 test rate



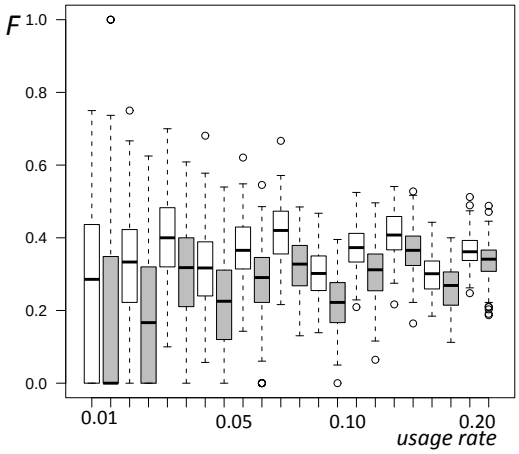
s results for the prediction model Last using 20% failure rate and 0.30 test rate



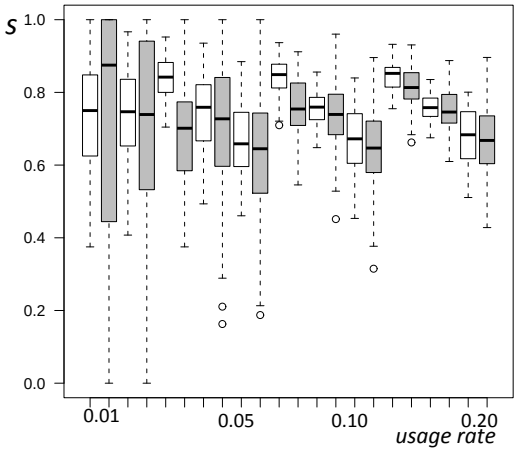
F results for the prediction model BM(5) using 20% failure rate and 0.30 test rate



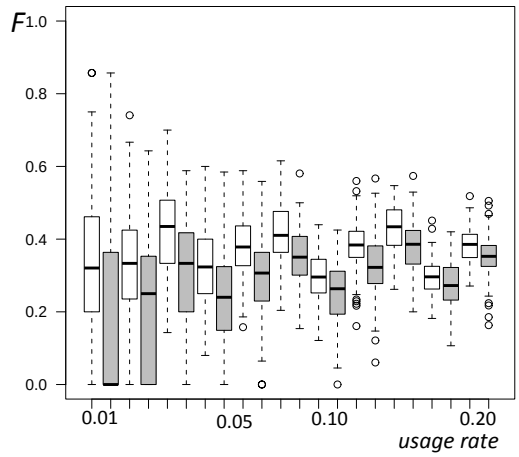
F results for the prediction model BM(5) using 20% failure rate and 0.30 test rate



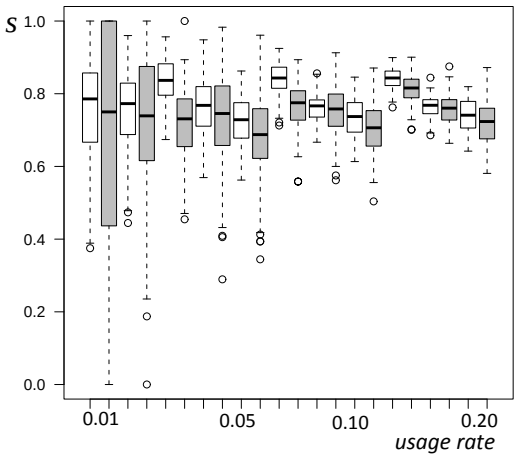
F results for the prediction model BM(10) using 20% failure rate and 0.30 test rate



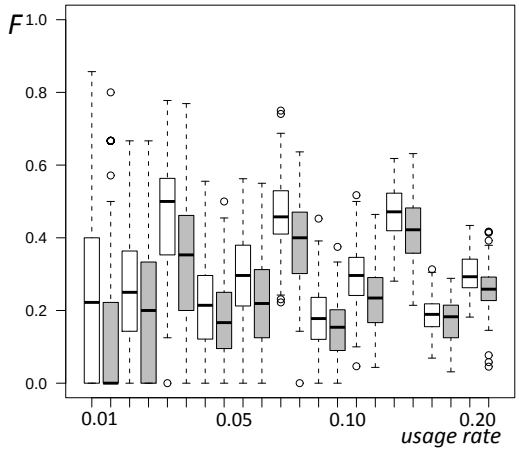
s results for the BM(10) prediction model using 20% failure rate and 0.30 test rate



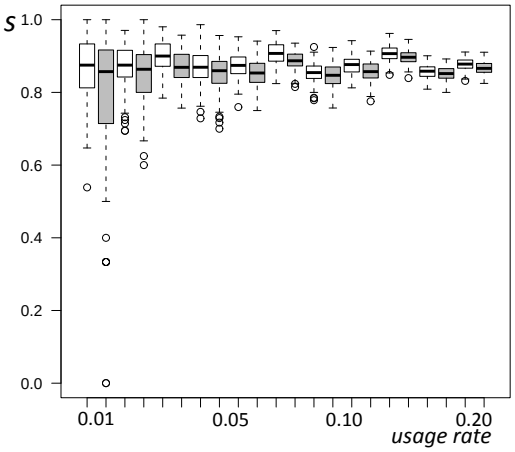
F results for the prediction model SEM using 20% failure rate and 0.30 test rate



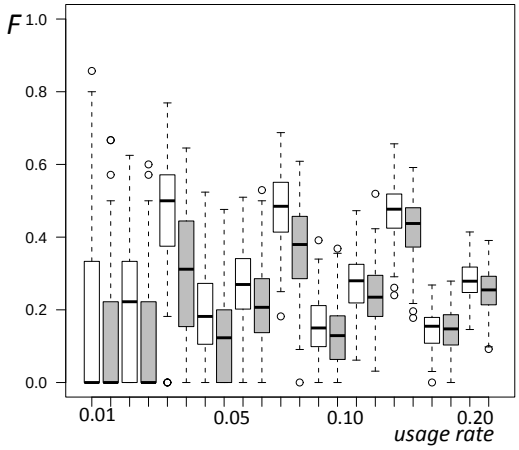
s results for the prediction model SEM using 20% failure rate and 0.30 test rate



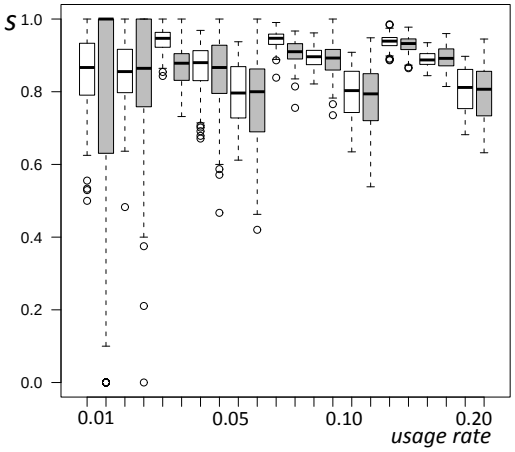
F results for the prediction model Last using 15% failure rate and 0.30 test rate



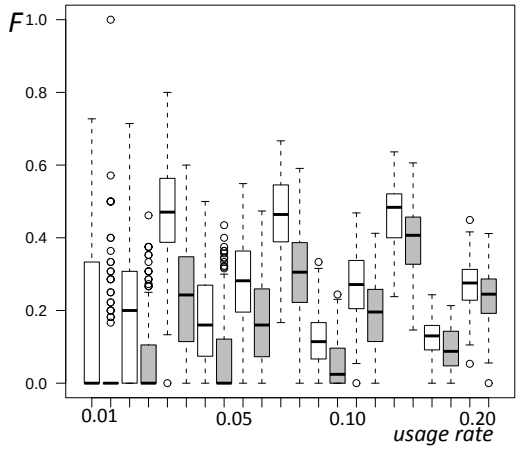
s results for the prediction model Last using 15% failure rate and 0.30 test rate



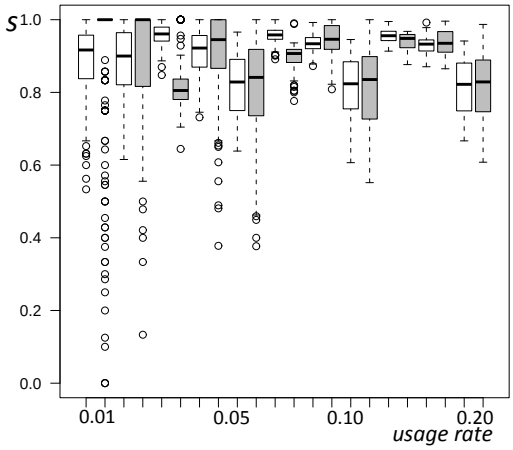
F results for the prediction model BM(5) using 15% failure rate and 0.30 test rate



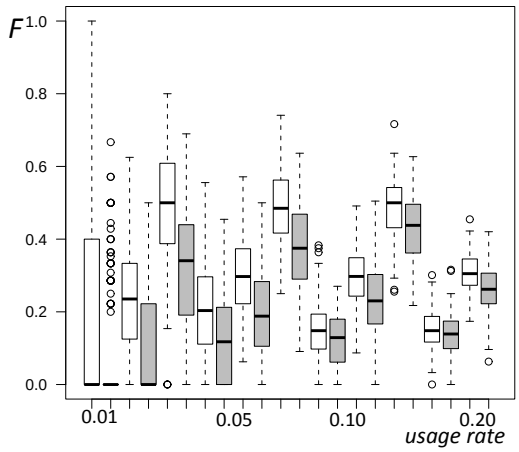
s results for the prediction model BM(5) using 15% failure rate and 0.30 test rate



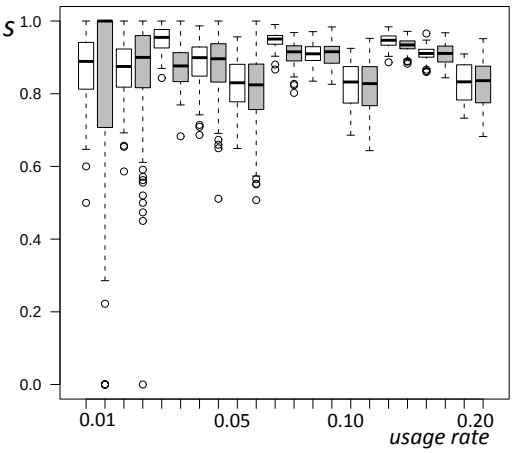
F results for the prediction model BM(10) using 15% failure rate and 0.30 test rate



s results for the prediction model BM(10) using 15% failure rate and 0.30 test rate



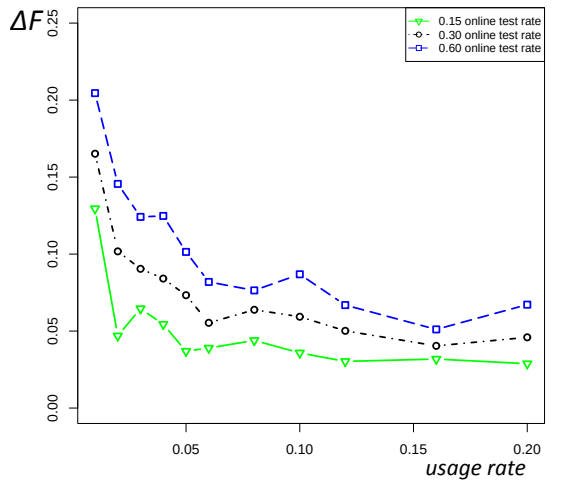
F results for the prediction model SEM using 15% failure rate and 0.30 test rate



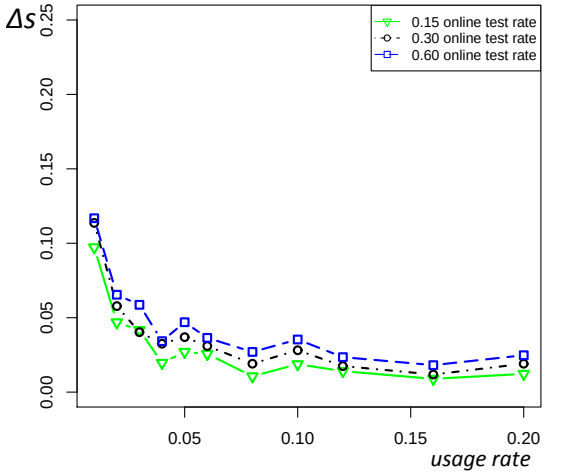
s results for the prediction model SEM using 15% failure rate and 0.30 test rate

RQ 4 (*influence of test rate*)

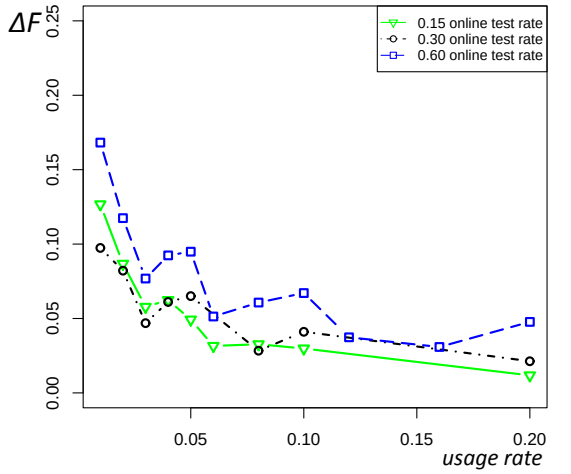
The below figures show the accuracy gains in terms of the F-measure (*F*) and specificity (*s*) for the different prediction models using different online test rates (0.15, 0.30 and 0.60) we have used in our experiments.



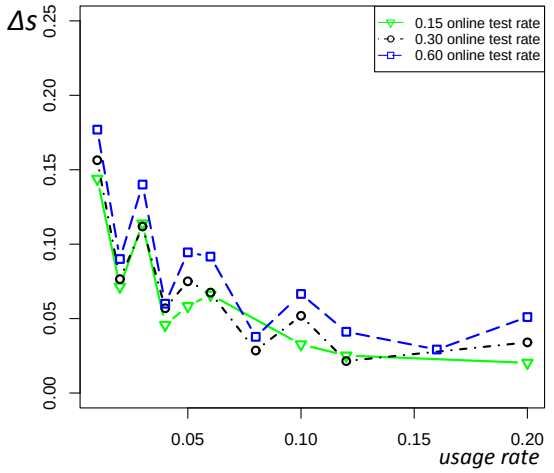
Accuracy gains in terms of *F* for the prediction model Last for different test rates and 25% failure rate



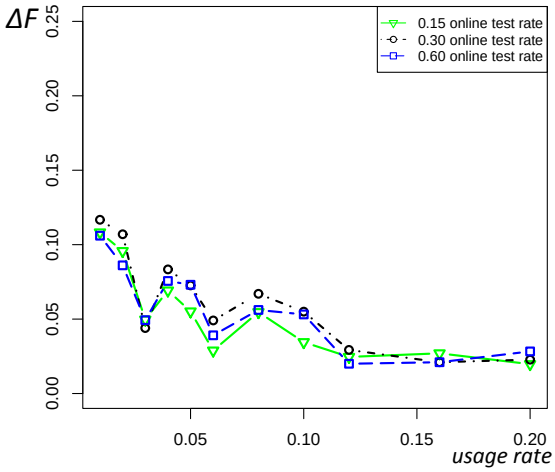
Accuracy gains in terms of *s* for the prediction model Last for different test rates and 25% failure rate



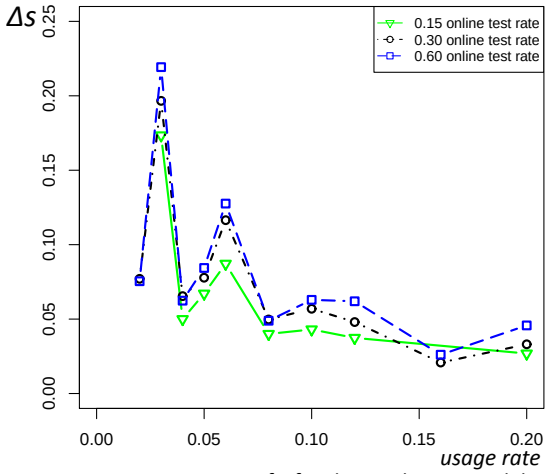
Accuracy gains in terms of *F* for the prediction model BM(5) for different test rates using 25% failure rate



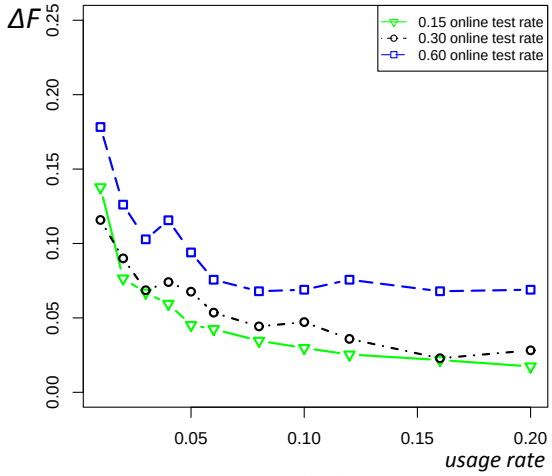
Accuracy gains in terms of *s* for the prediction model BM(5) for different test rates using 25% failure rate



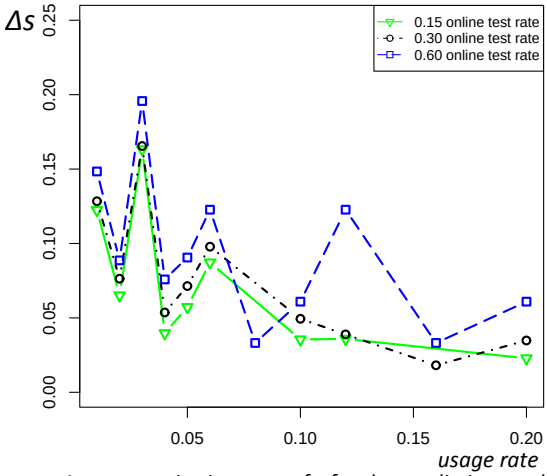
Accuracy gains in terms of *F* for the prediction model BM(10) for different test rates and 25% failure rate



Accuracy in terms of *s* for the prediction model BM(10) for different test rates and 25% failure rate

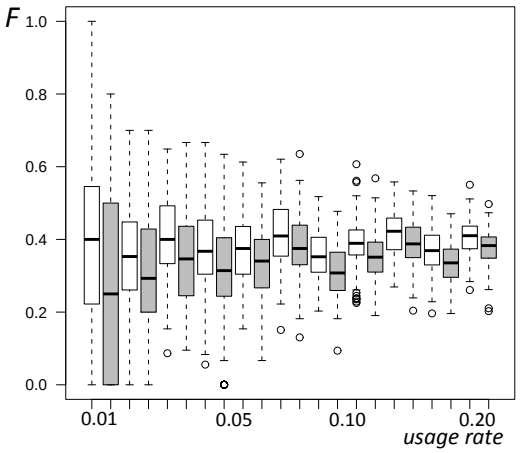


Accuracy gains in terms of *F* for the prediction model SEM for different test rates and 25% failure rate

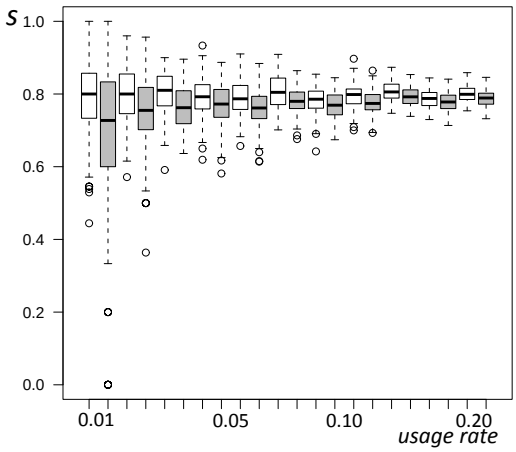


Accuracy gains in terms of *s* for the prediction model SEM for different test rates and 25% failure rate

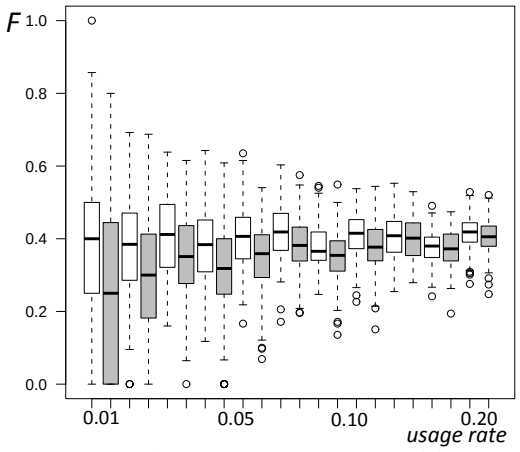
The following figures show the box plots for both *F* and *s* results shown above. White boxes represent results of predication using monitoring and online testing (*M&OT*) and the grey boxes represent results of predication using monitoring only (*M*). Please note that we have already presented the boxplots for the 0.30 test rate in RQ2 above.



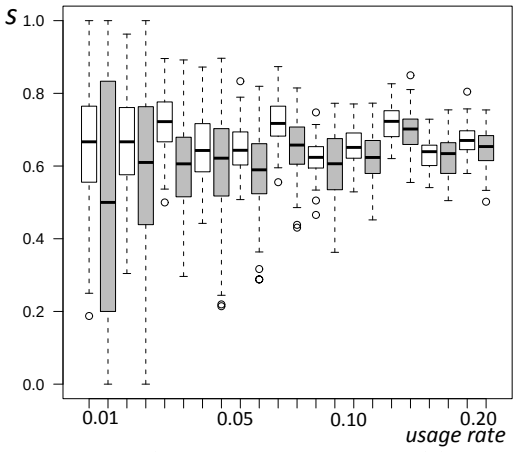
F results for the prediction model Last using 0.15 test rate



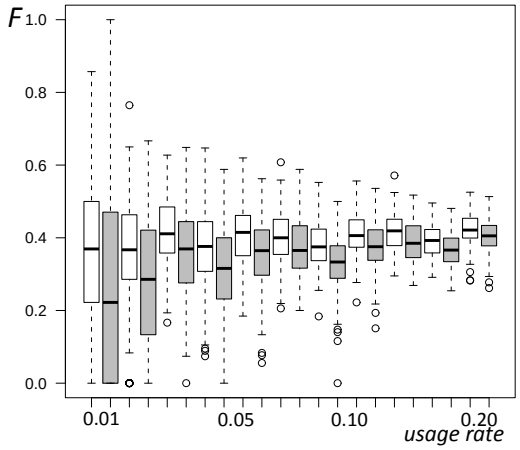
s results for the prediction model Last using 0.15 test rate



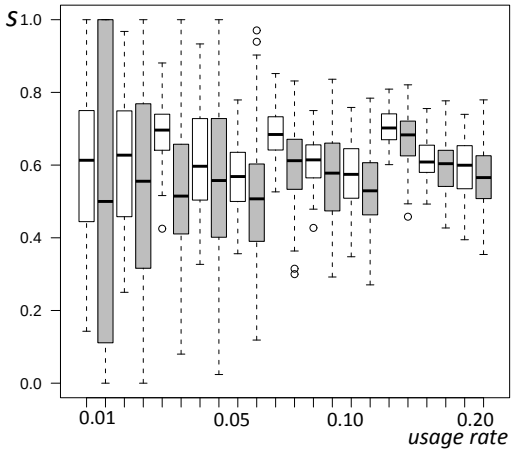
F results for the prediction model BM(5) using 0.15 test rate



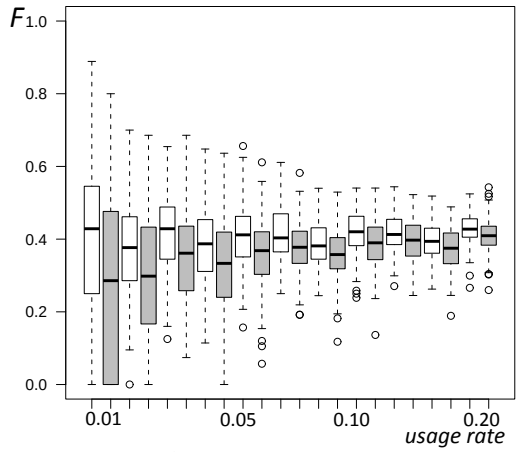
s results for the prediction model BM(5) using 0.15 test rate



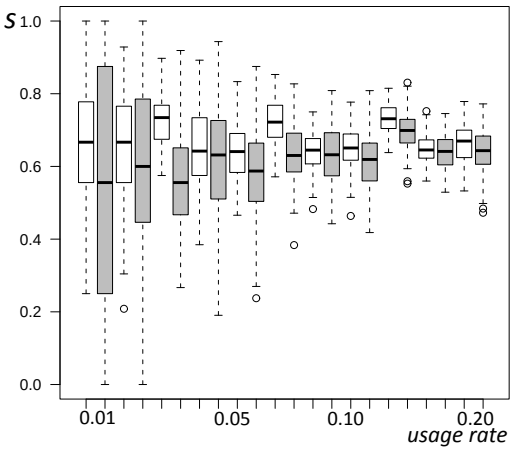
F results for the prediction model BM(10) using 0.15 test rate



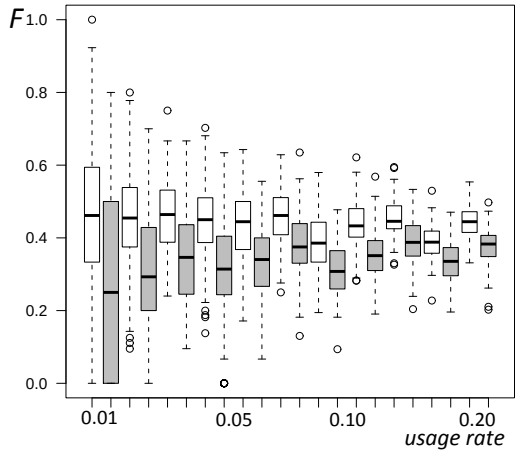
s results for the prediction model BM(10) using 0.15 test rate



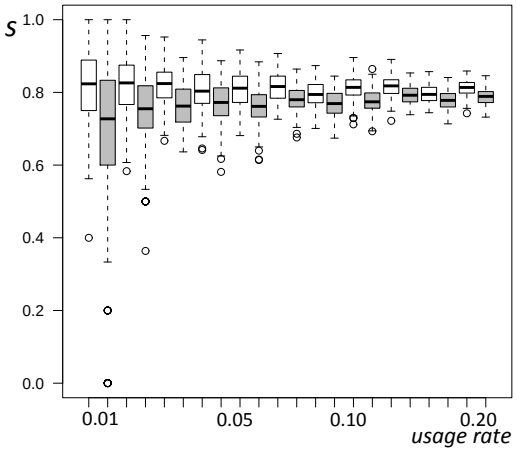
F results for the prediction model SEM using 0.15 test rate



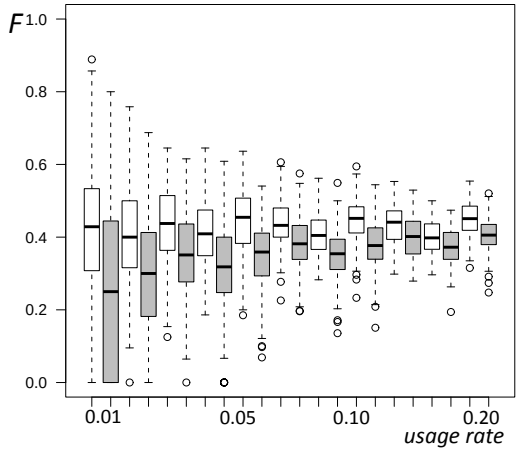
s results for the prediction model SEM using 0.15 test rate



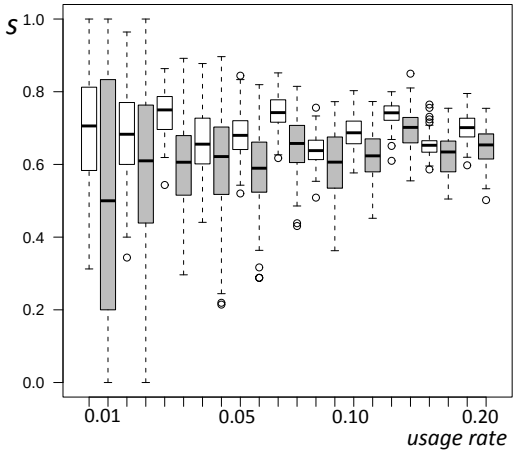
F results for the prediction model SEM using 0.60 test rate



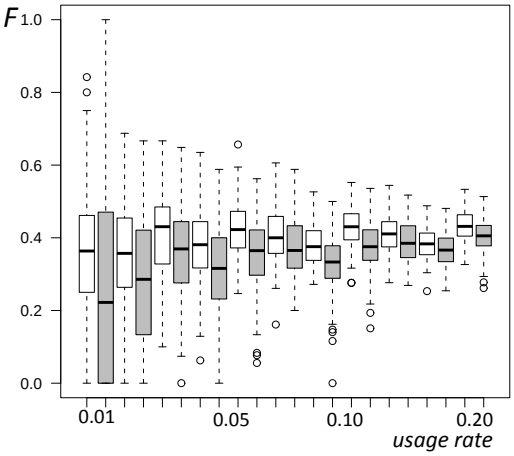
s results for the prediction model SEM using 0.60 test rate



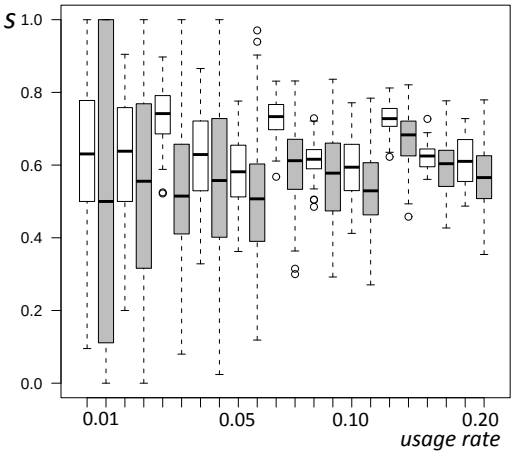
F results for prediction model BM(5) using 0.60 test rate



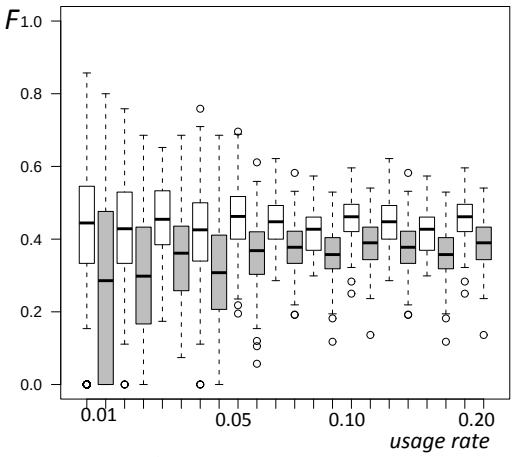
s results for prediction model BM(5) using 0.60 test rate



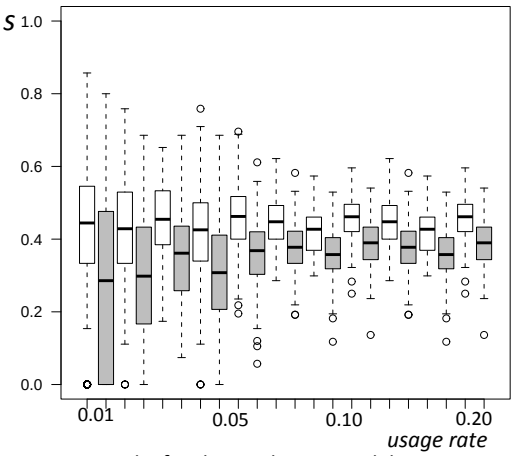
F results for the prediction model BM(10) using 0.60 test rate



s results for prediction model BM(10) using 0.60 test rate



F results for the prediction model SEM using 0.60 test rate



s results for the prediction model SEM using 0.60 test rate