Supplementary materials to the paper:

Prediction approaches for partly missing multi-omics covariate data: An empirical comparison study

Roman Hornung*, Frederik Ludwigs, Jonas Hagenberg, Anne-Laure Boulesteix

*To whom correspondence should be addressed: hornung@ibe.med.uni-muenchen.de

Contents

A Detailed reasons for unsuccessful predictions in the cases of some prediction methods

In Section 3.3 of the main paper we described that `ComplcRF` was not applicable for tebmp 4 in the cases of trbmps 2, 3, and 4 because in these cases there were no complete cases in the training data. The same problem of no complete cases occurred for this method in the case of combination trbmp 4 and tebmp 3. For combinations of tebmp 3 with trmps 2 and 4 it also frequently happened that there were no complete cases in the training data. Whether or not this happened depended on which block was missing in tebmp 3. There were also two errors for `ComplcRF` in the case of trbmp 5.

As described in Section 3.3 of the main paper, the R package `multisForest` (version 0.1.0) implementing `MultisRF` does not allow training sets without missing values, which is why we obtained no results for trbmp 1 in the case of this method. In addition, the method also delivered no predictions in many cases for tebmp 1. This is due to the fact that, for tebmp 1, only the clinical block is available in the test data. To understand why `MultisRF` does not deliver predictions in many cases if only the clinical block is available we have to consider the following: Starting with the first splits, `MultisRF` prunes all trees, removing all splits for which variables are used that do not occur in the test set. Because the clinical covariates are so few in comparison to the omics covariates it is very unlikely that the first splits use clinical covariates. As a consequence, if the test data feature only clinical covariates, the likelihood is high that all trees are removed from all forests. And if all trees are removed no predictions can be performed. This explains why `MultisRF` did not deliver predictions in many cases for tebmp 1. There were also some cases where there are no results for tebmp 2, that is, the tebmp for which only one omics block was available in the test data in addition to the clinical block. Here, the same mechanism as described above was at work. For some datasets, there were memory issues for trbmp 5 in the case of `MultisRF`, which were likely due to the peculiarity that this trbmp featured eight subsets.

`PrLasso` also did not deliver predictions for 10 repetitions, where all of these but one occurred for the dataset ESCA.

B Global performance

B.1 Rank values

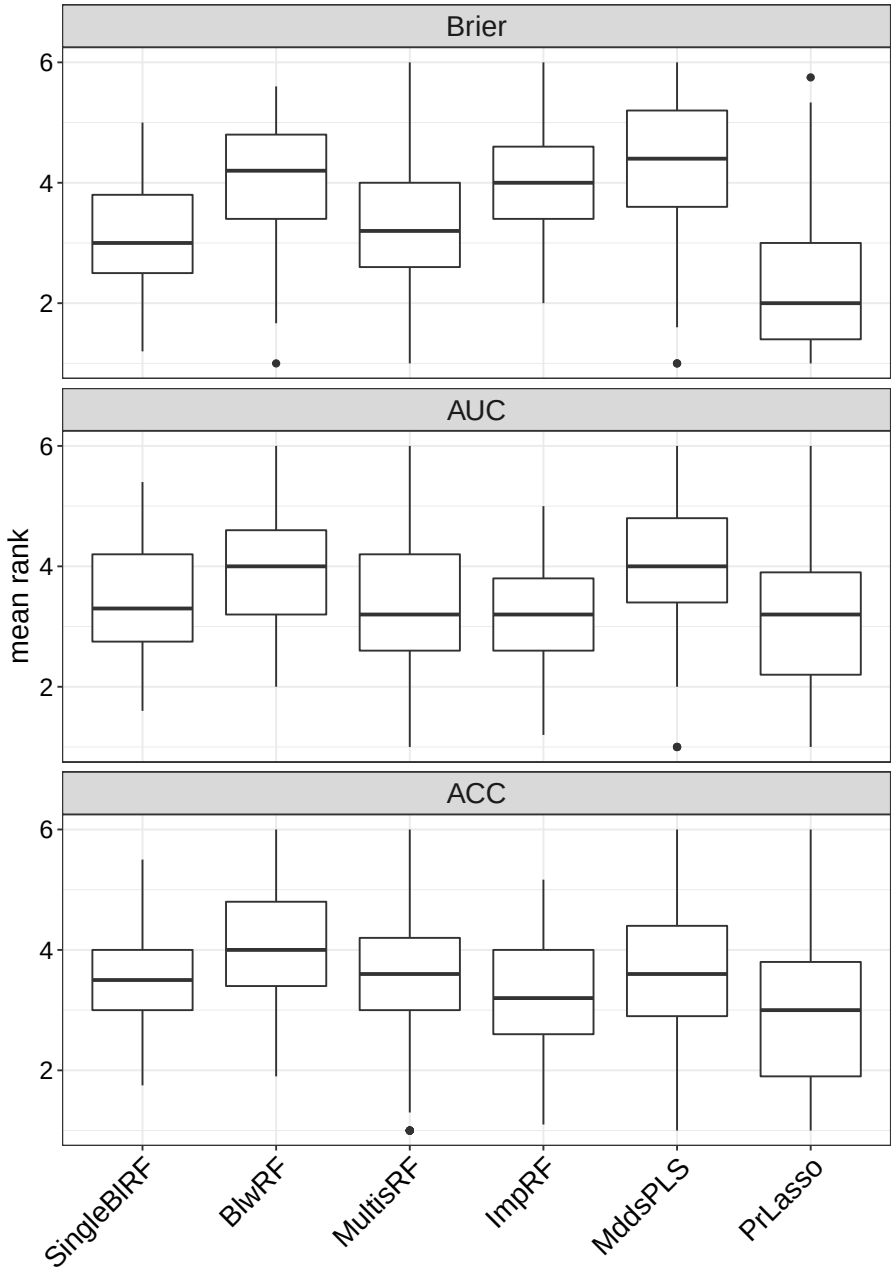

Fig. S1: Ranks each method achieved among the other methods in terms of the three considered performance metrics – global performance without `ComplcRF`. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

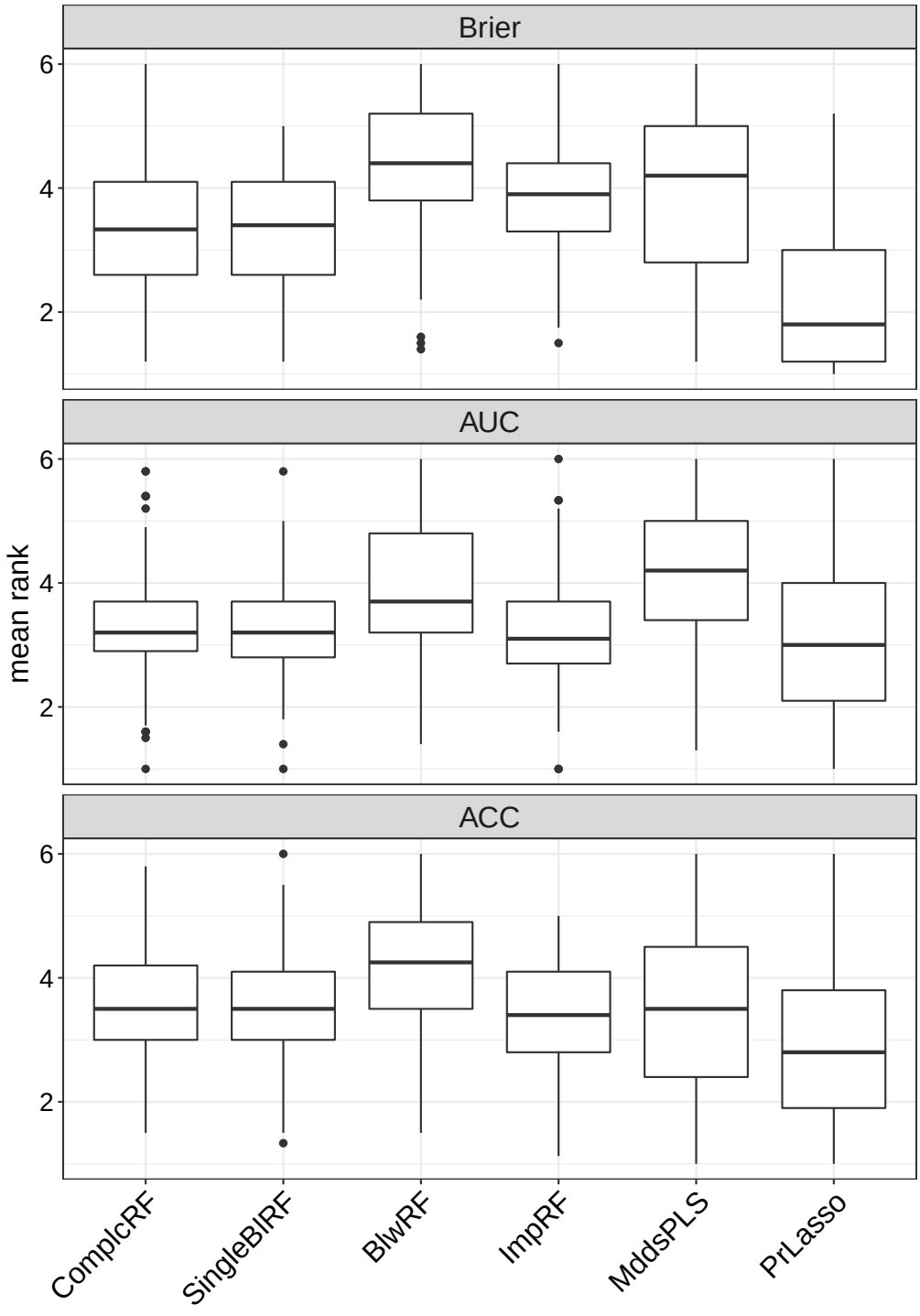

Fig. S2: Ranks each method achieved among the other methods in terms of the three considered performance metrics – global performance without `MultisRF`. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

B.2 Performance metric values

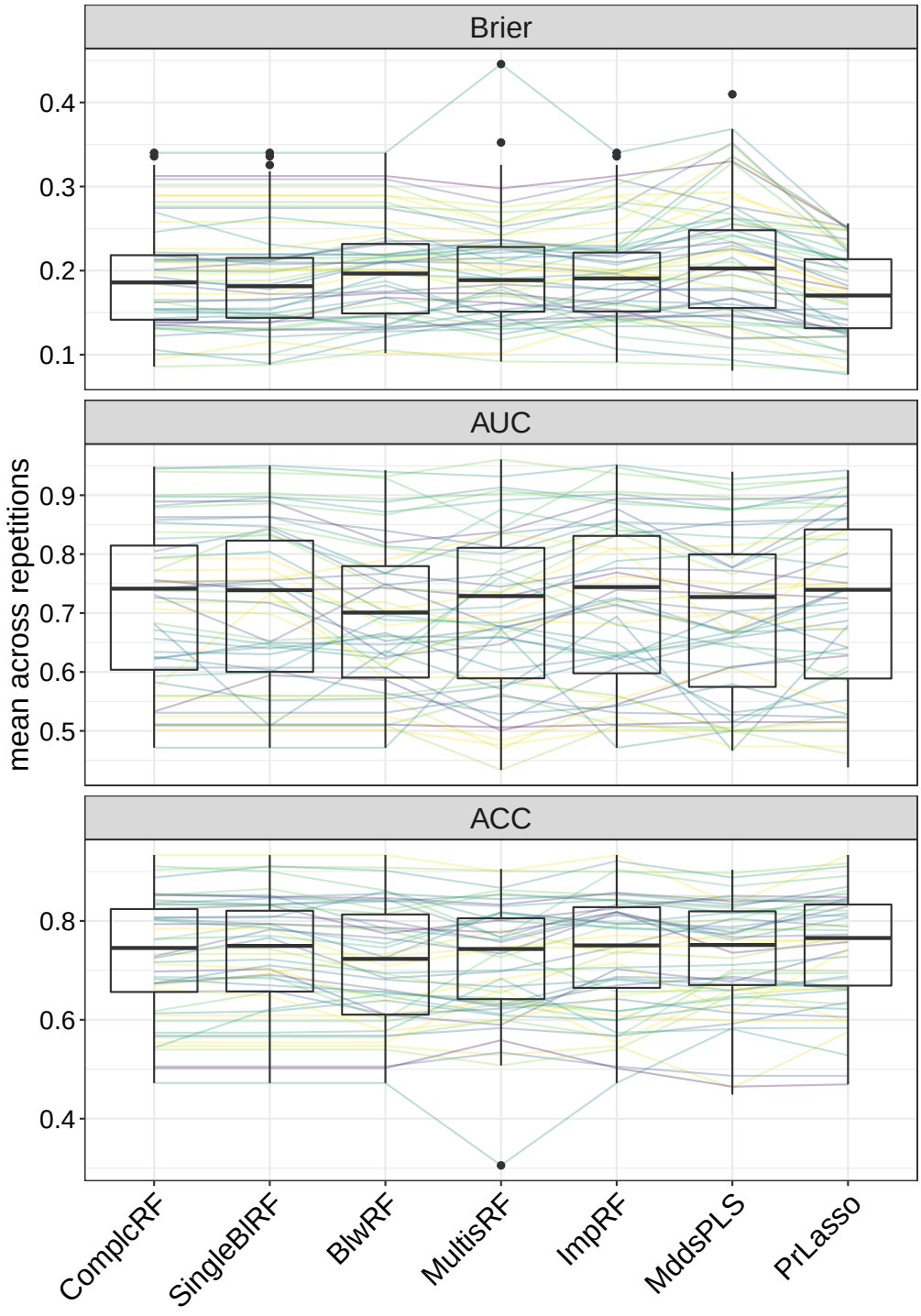

Fig. S3: Performance metric values – global performance. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions. For reasons of clarity, the lines only show a subset of 50 repetitions.

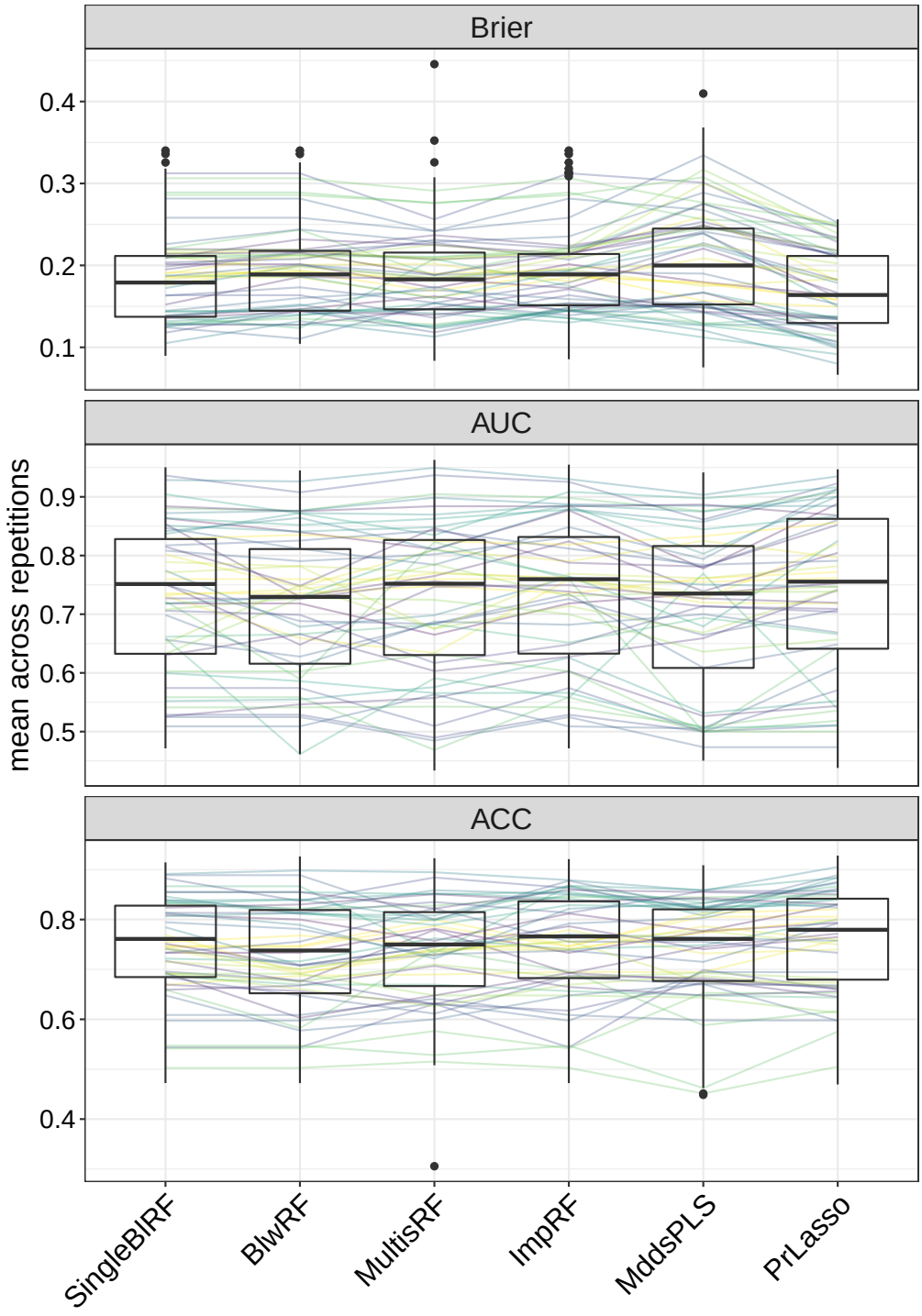

Fig. S4: Performance metric values – global performance without `ComplcRF`. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions. For reasons of clarity, the lines only show a subset of 50 repetitions.

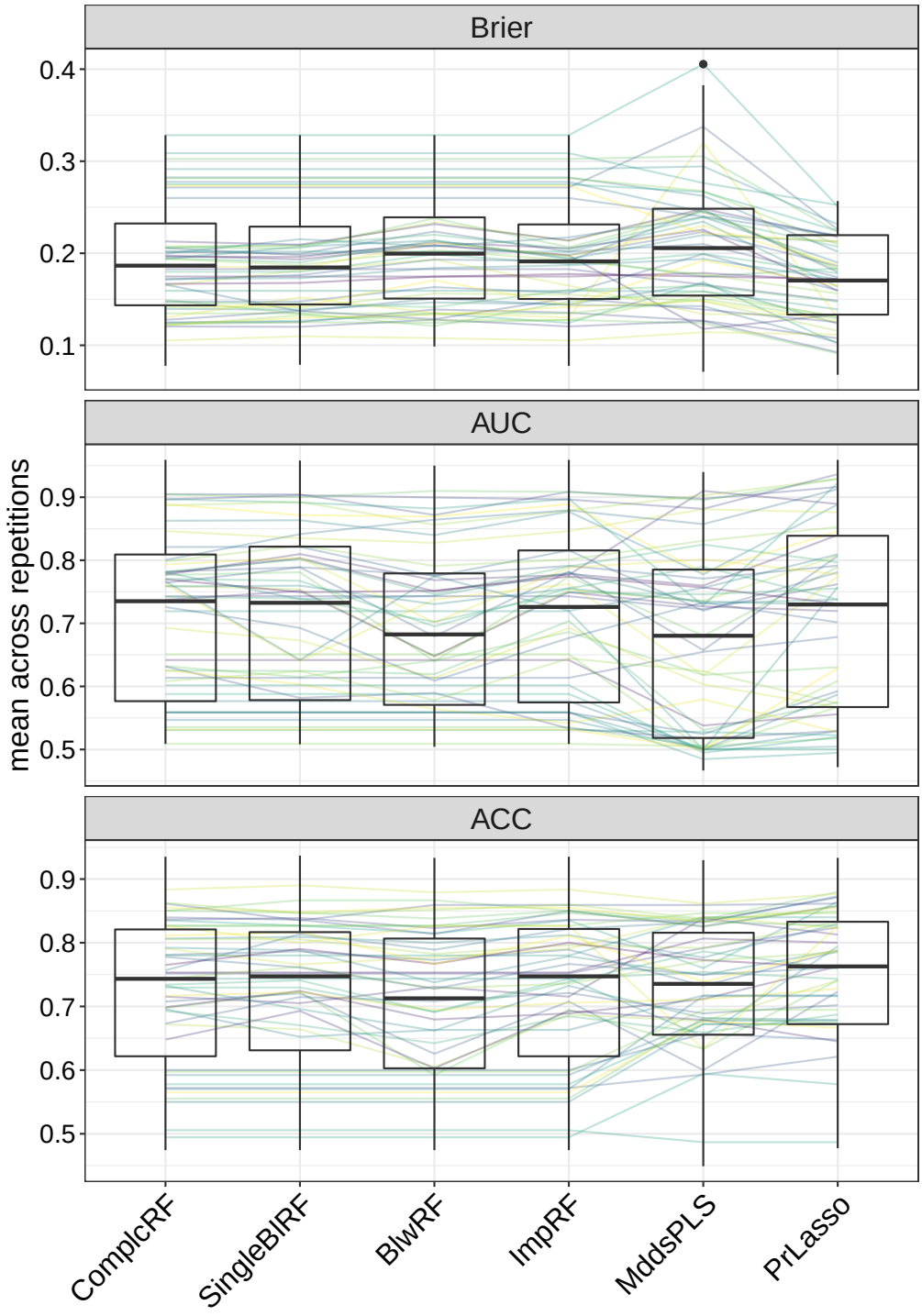

Fig. S5: Performance metric values – global performance without `MultisRF`. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions. For reasons of clarity, the lines only show a subset of 50 repetitions.

C Performance separately by trbmp

C.1 Rank values

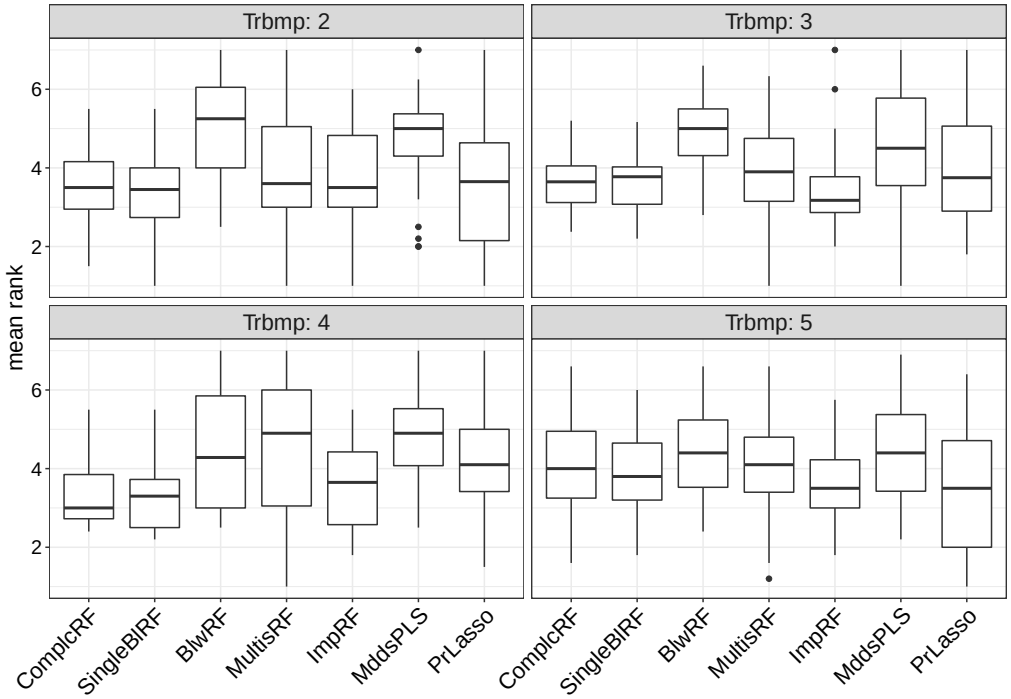

Fig. S6: Ranks each method achieved among the other methods in terms of the AUC – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all seven methods. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

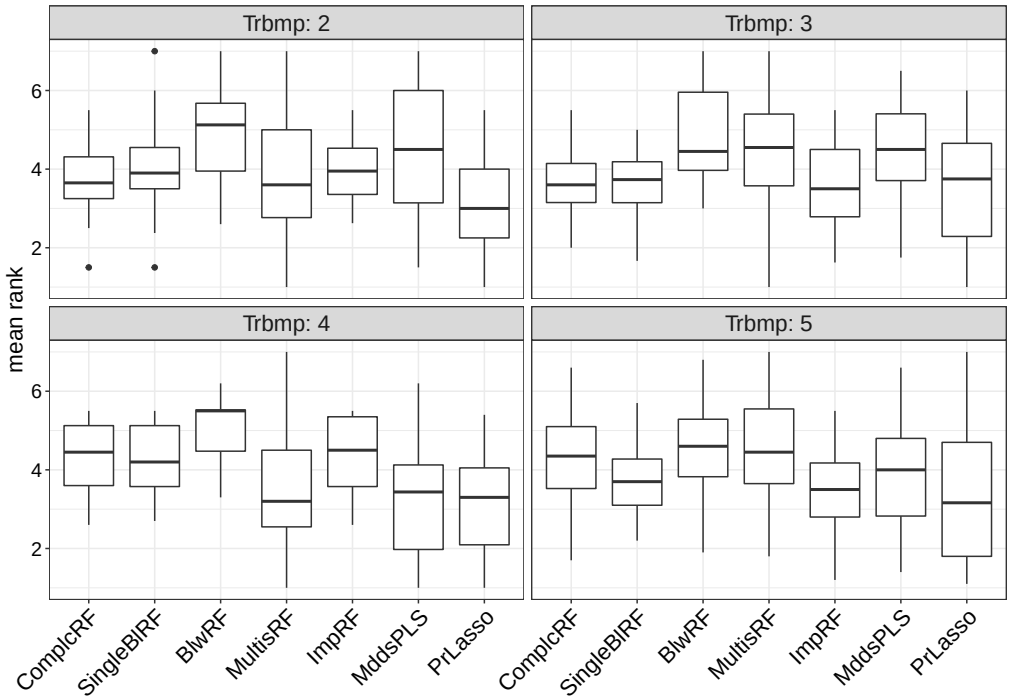

Fig. S7: Ranks each method achieved among the other methods in terms of the accuracy – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all seven methods. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

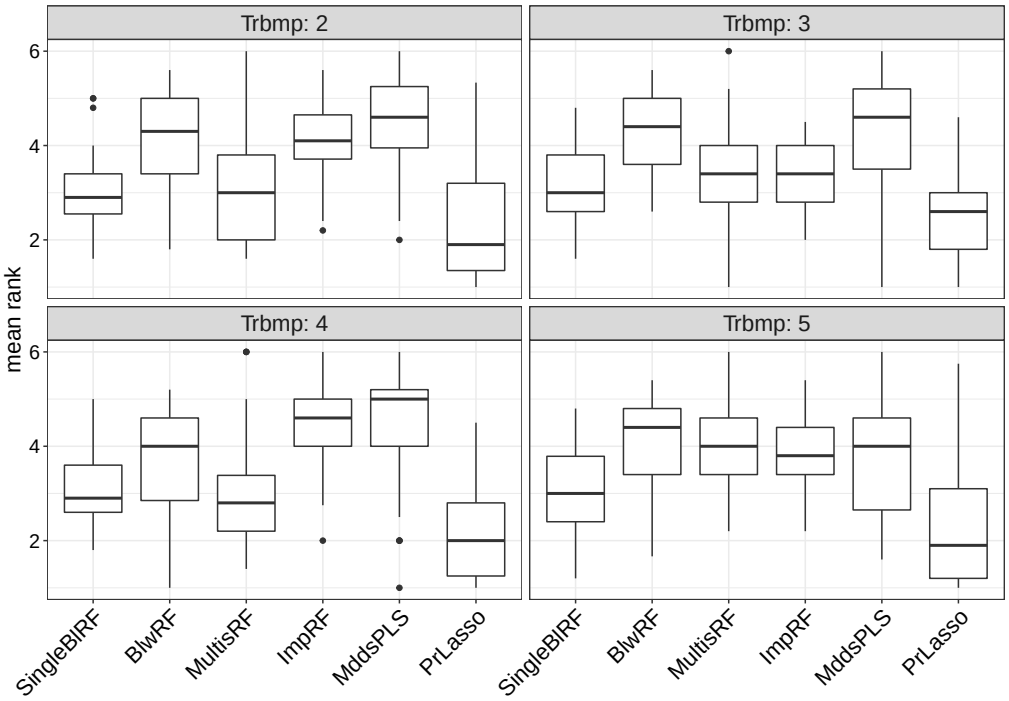

Fig. S8: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the Brier score – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

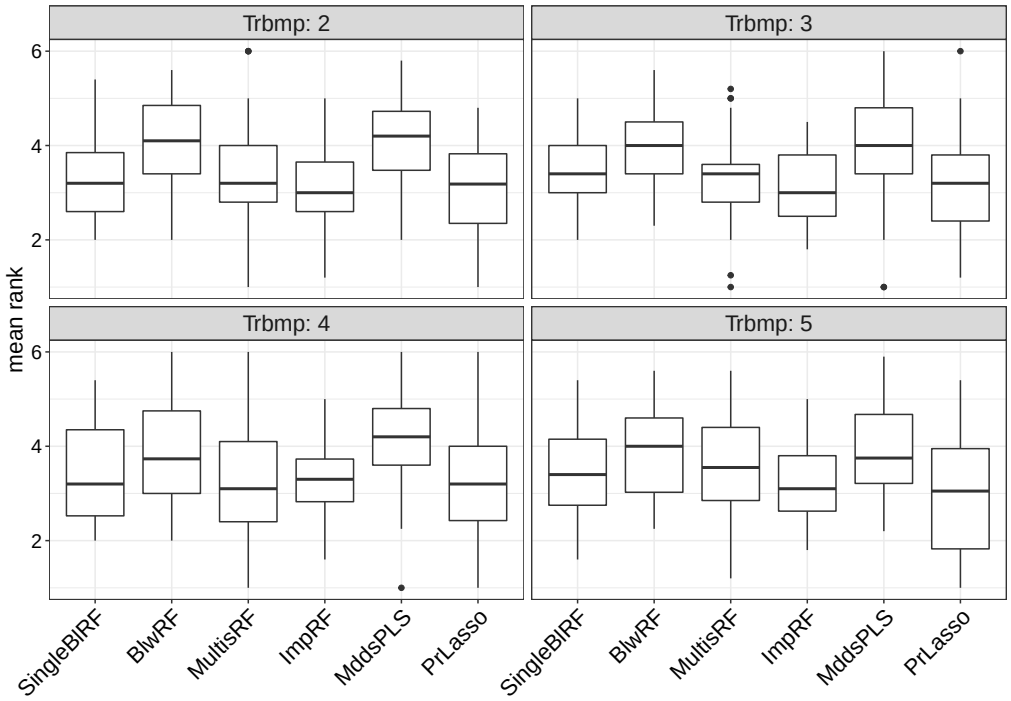

Fig. S9: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the AUC – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

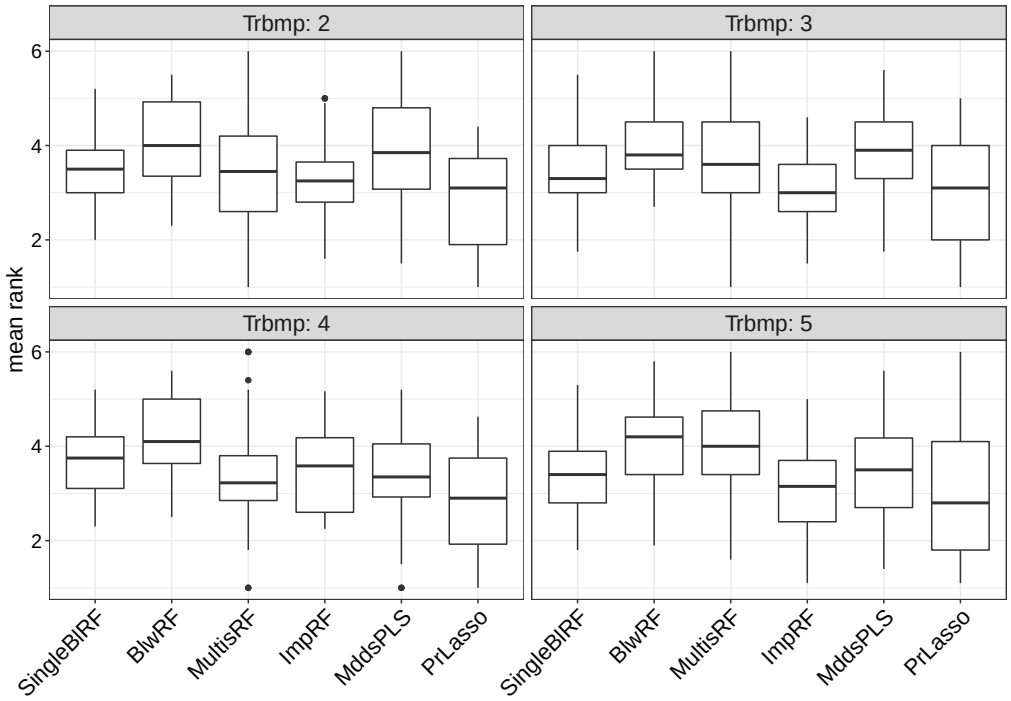

Fig. S10: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the accuracy – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

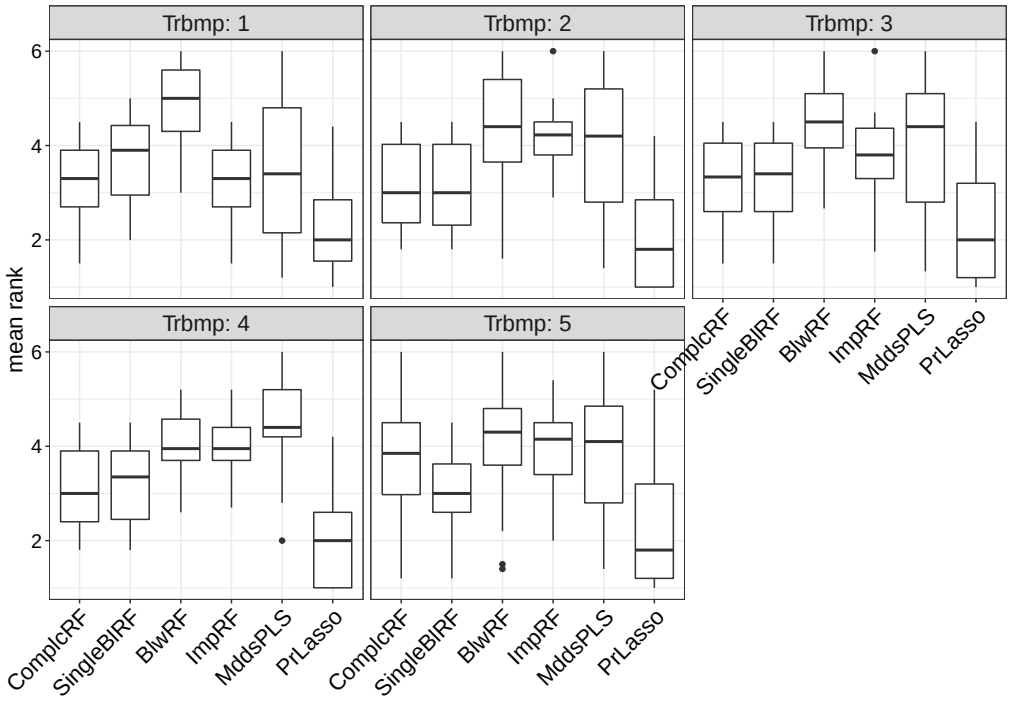

Fig. S11: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the Brier score – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

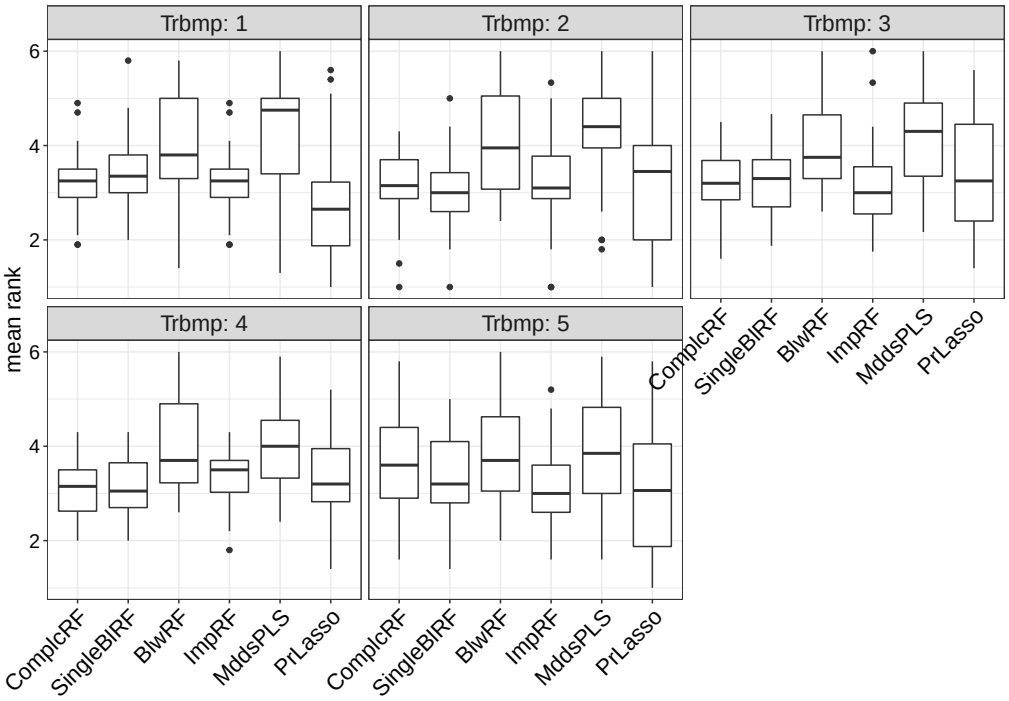

Fig. S12: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the AUC – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

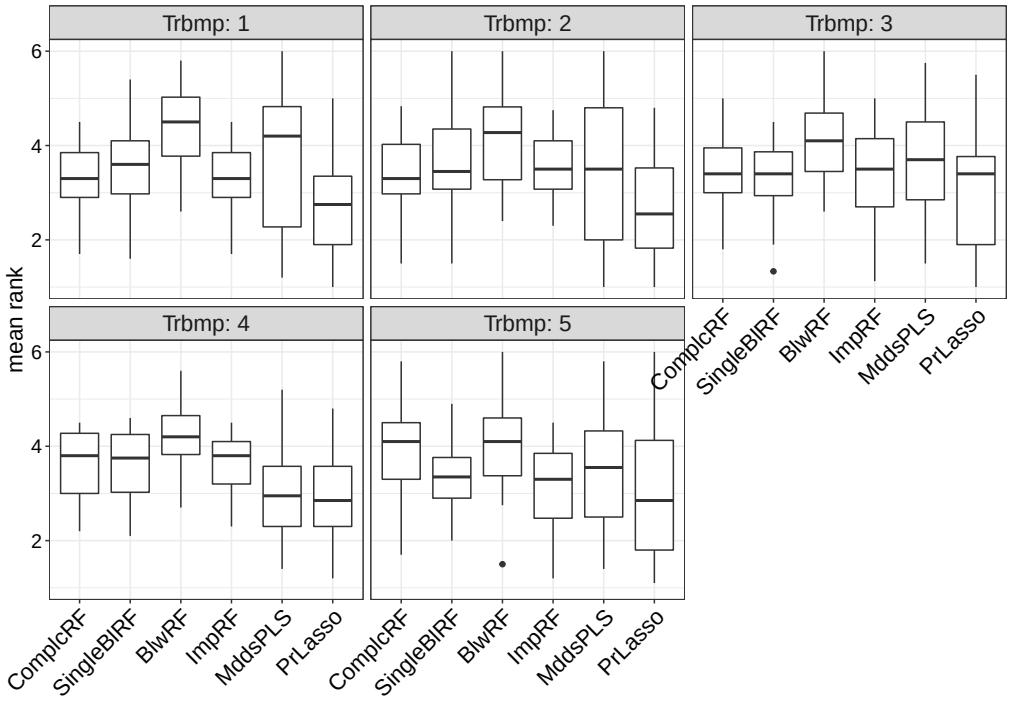

Fig. S13: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the accuracy – separately by trbmp. The ranks were computed for each combination of trbmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

C.2 Performance metric values

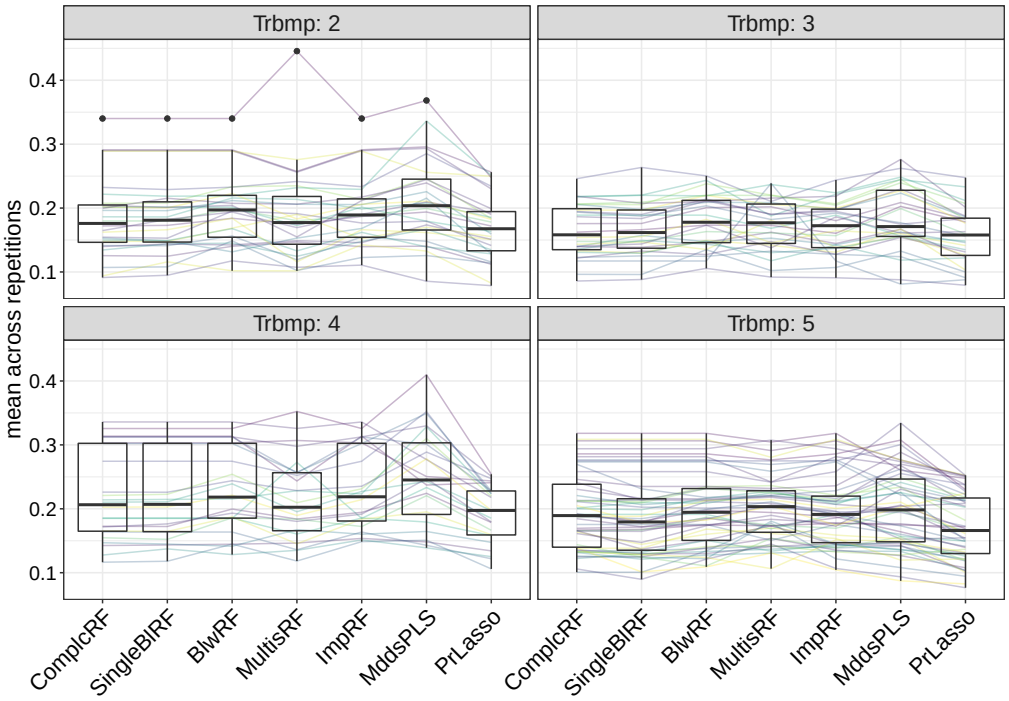

Fig. S14: Brier score values – separately by trbmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

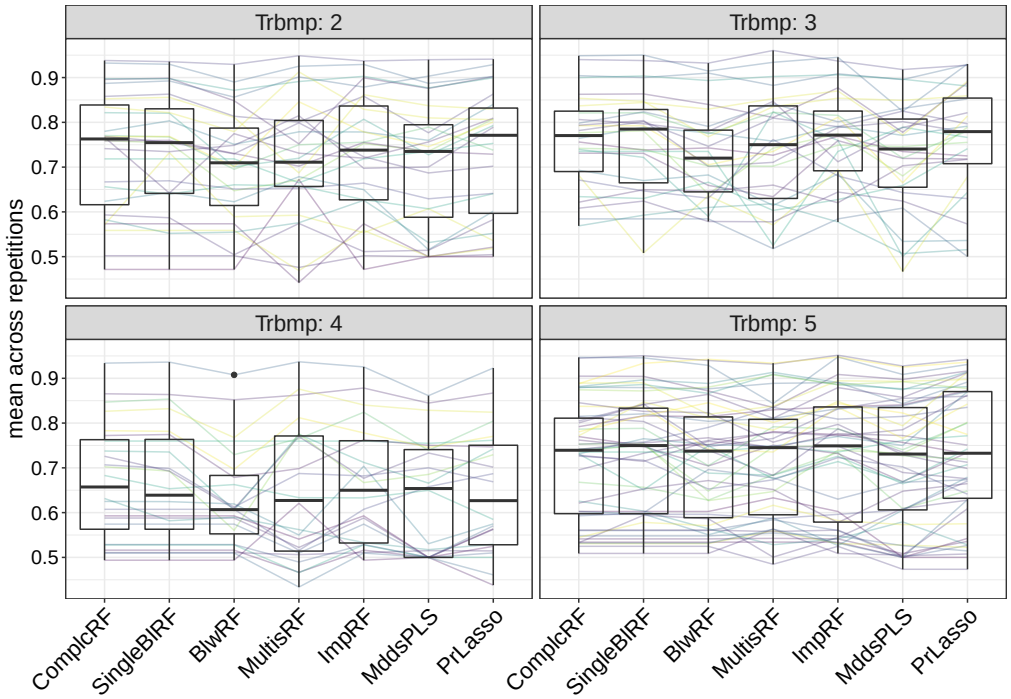

Fig. S15: AUC values – separately by trbmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

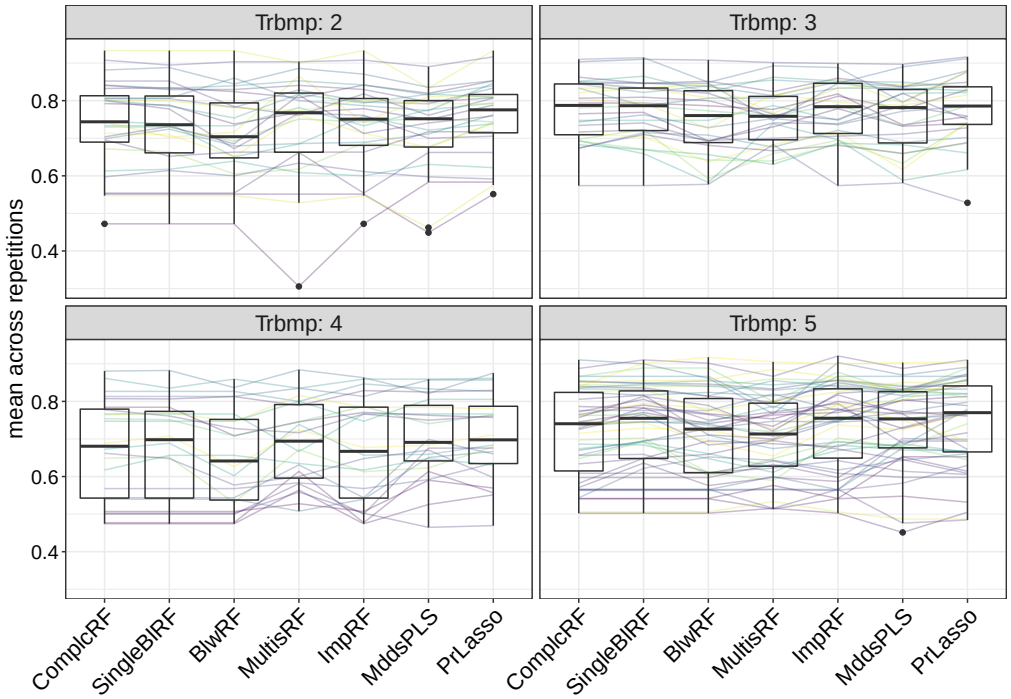

Fig. S16: Accuracy values – separately by trbmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

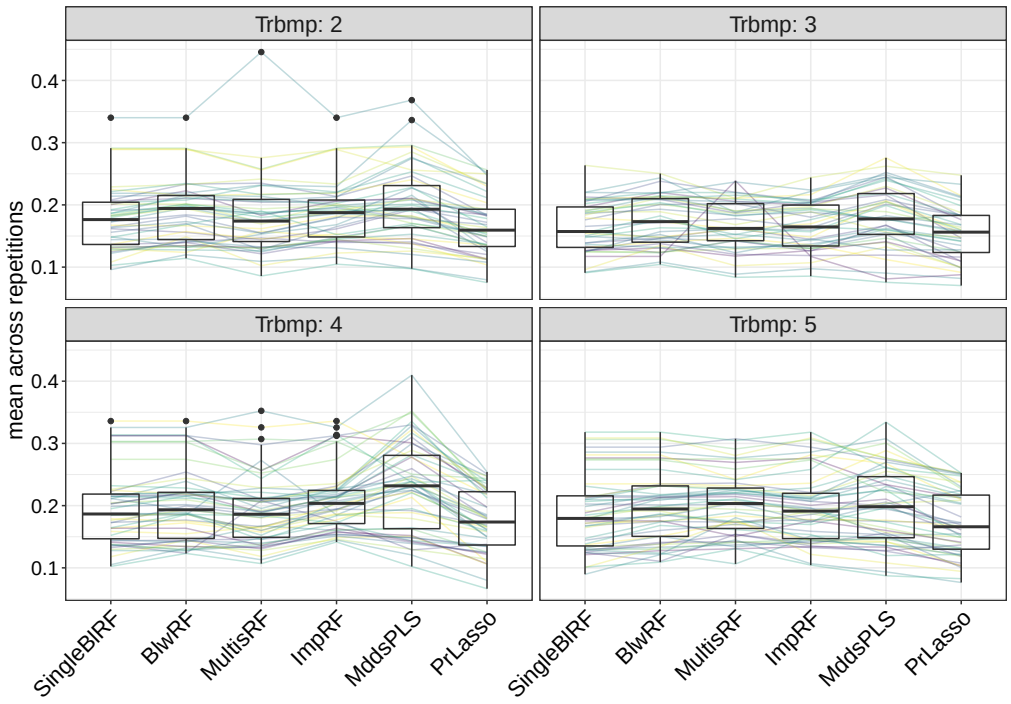

Fig. S17: Brier score values without `ComplcRF` – separately by trbmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

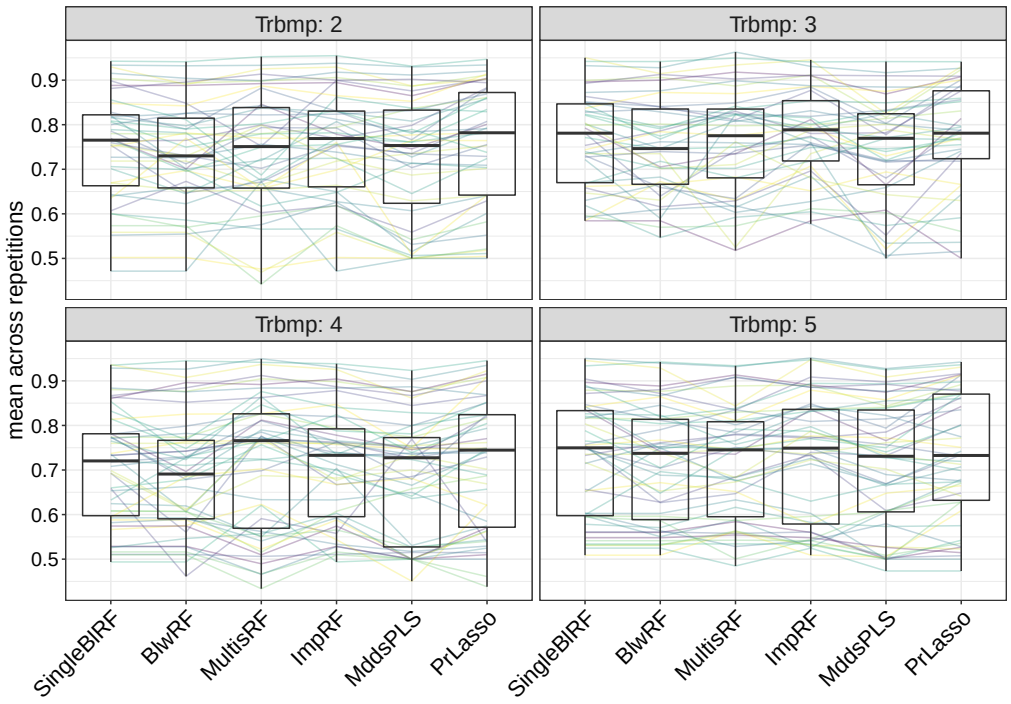

Fig. S18: AUC values without `ComplcRF` – separately by trbmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

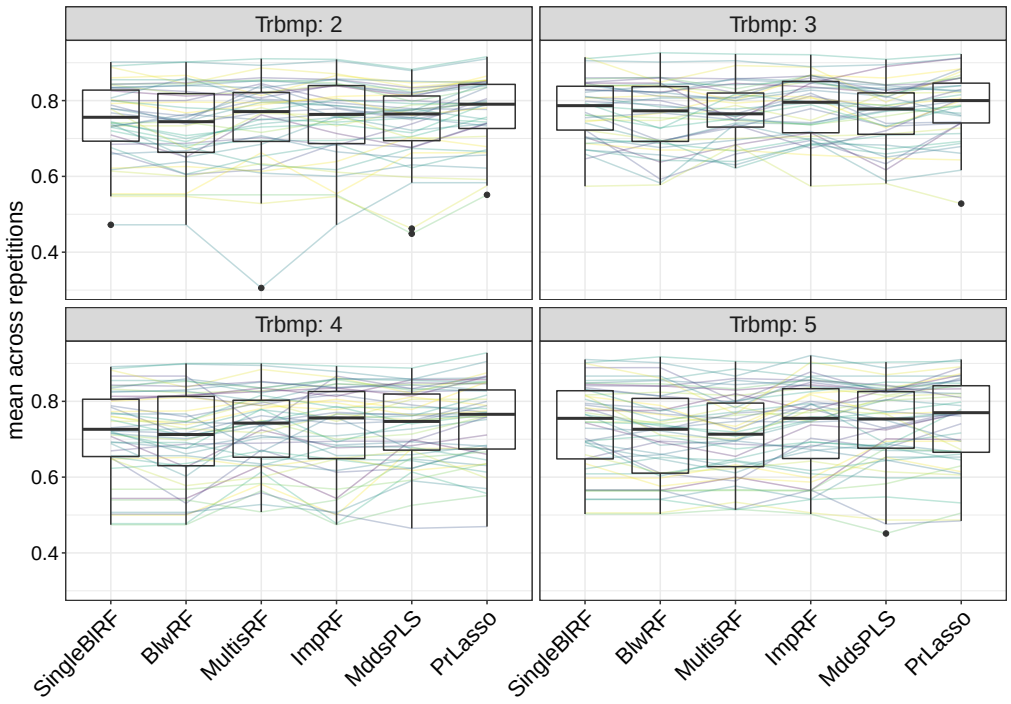

Fig. S19: Accuracy values without `ComplcRF` – separately by trbmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

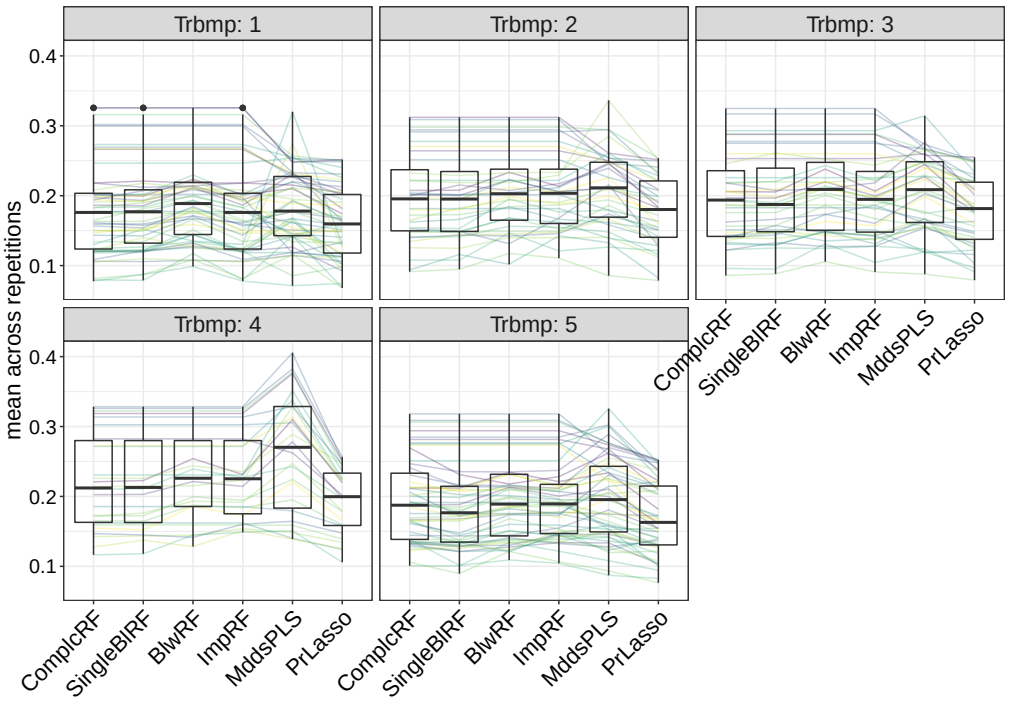

Fig. S20: Brier score values without `MultisRF` – separately by trbmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

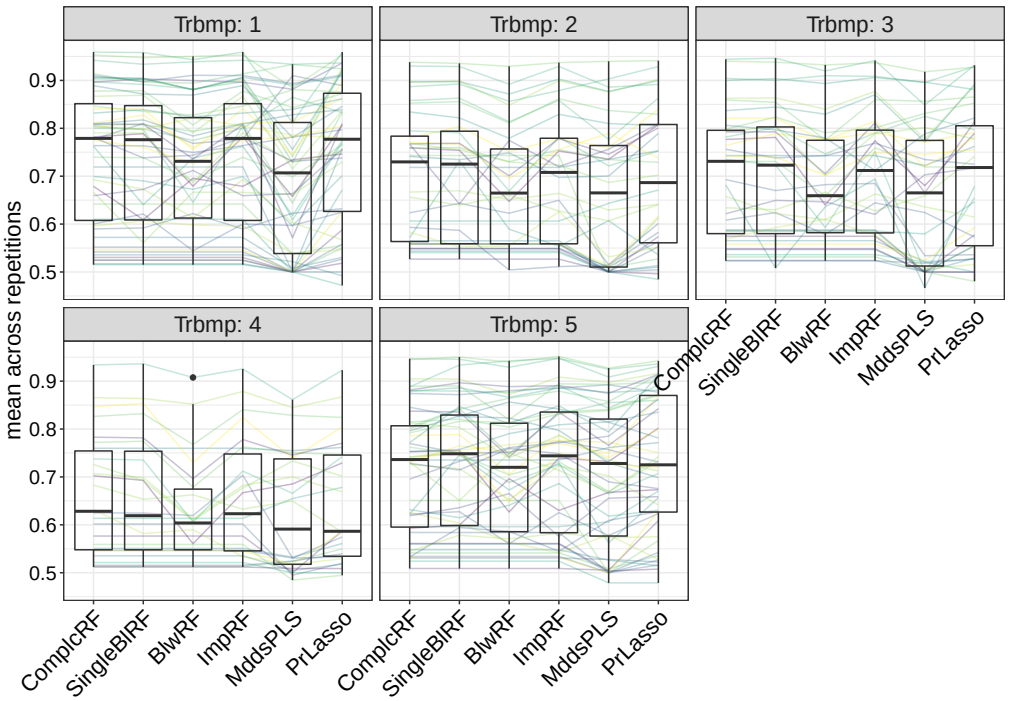

Fig. S21: AUC values without `MultisRF` – separately by trbmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

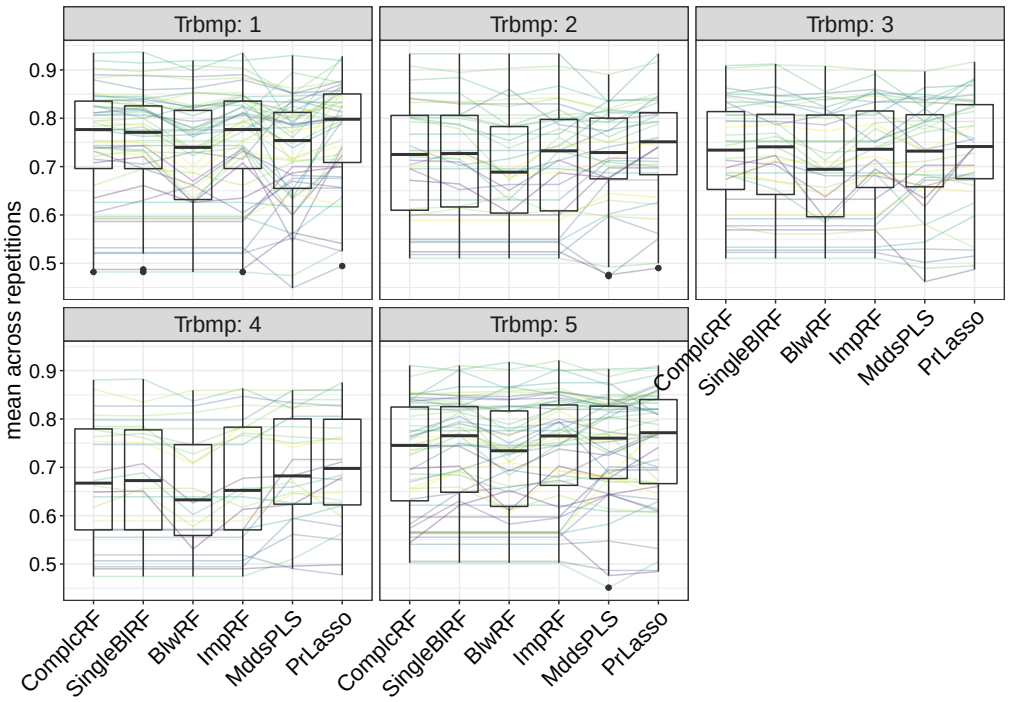

Fig. S22: Accuracy values without `MultisRF` – separately by trbmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

D Performance separately by tebmp

D.1 Rank values

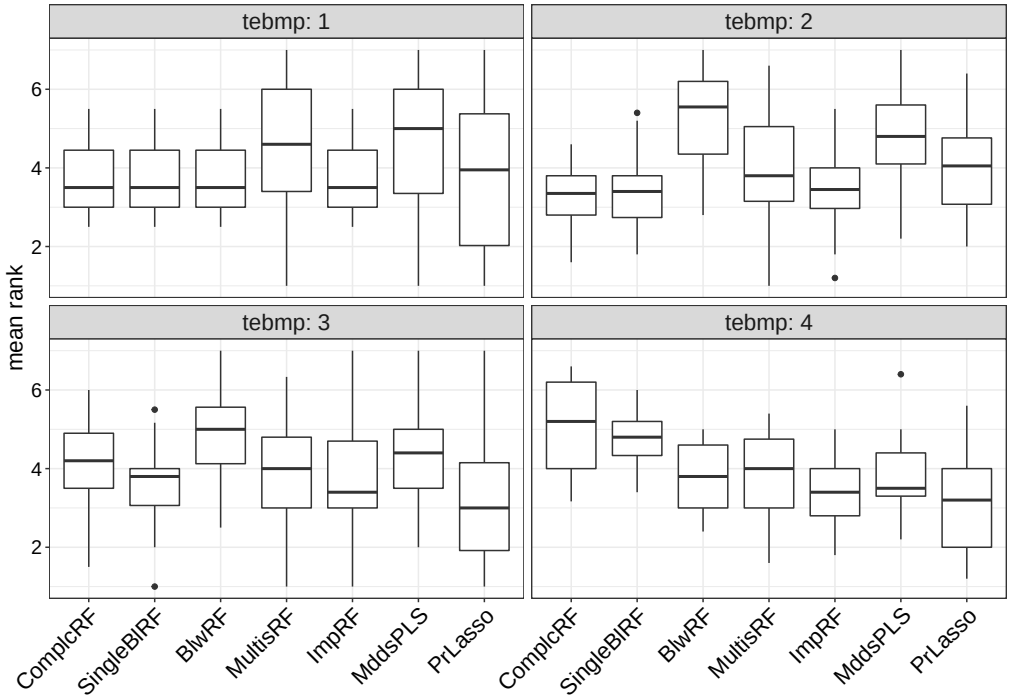

Fig. S23: Ranks each method achieved among the other methods in terms of the AUC – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all seven methods. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

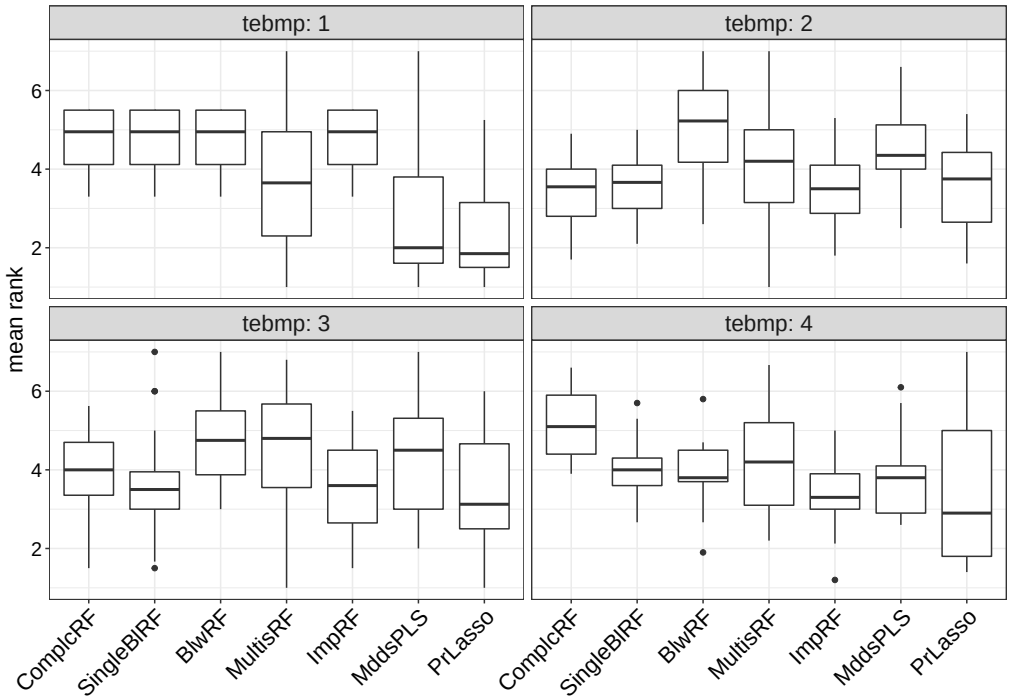

Fig. S24: Ranks each method achieved among the other methods in terms of the accuracy – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all seven methods. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

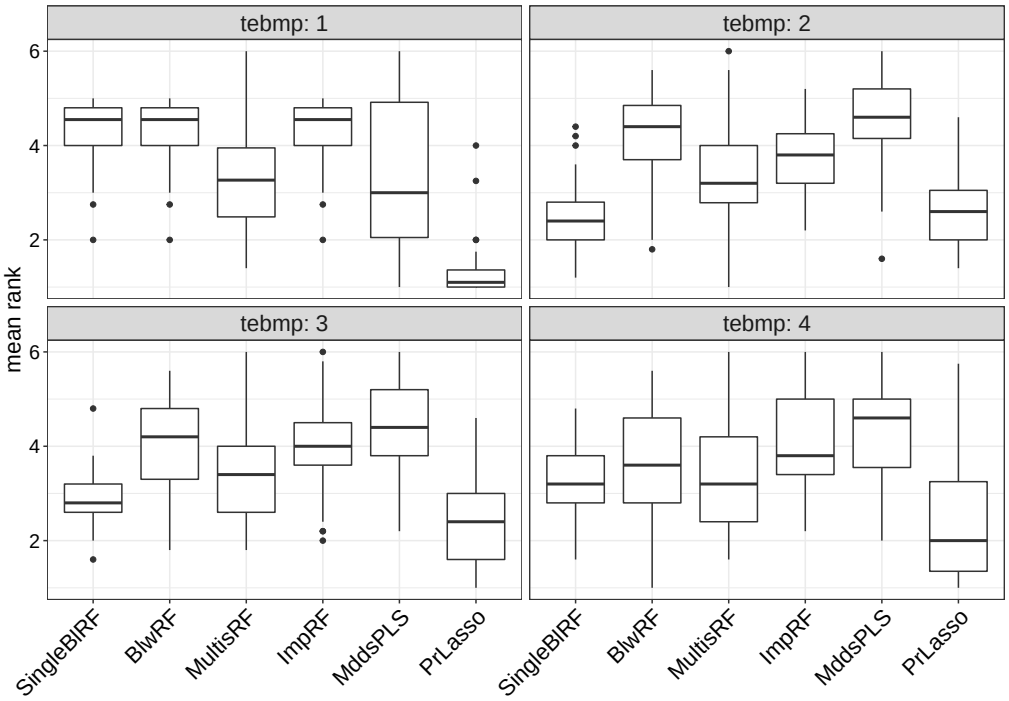

Fig. S25: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the Brier score – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

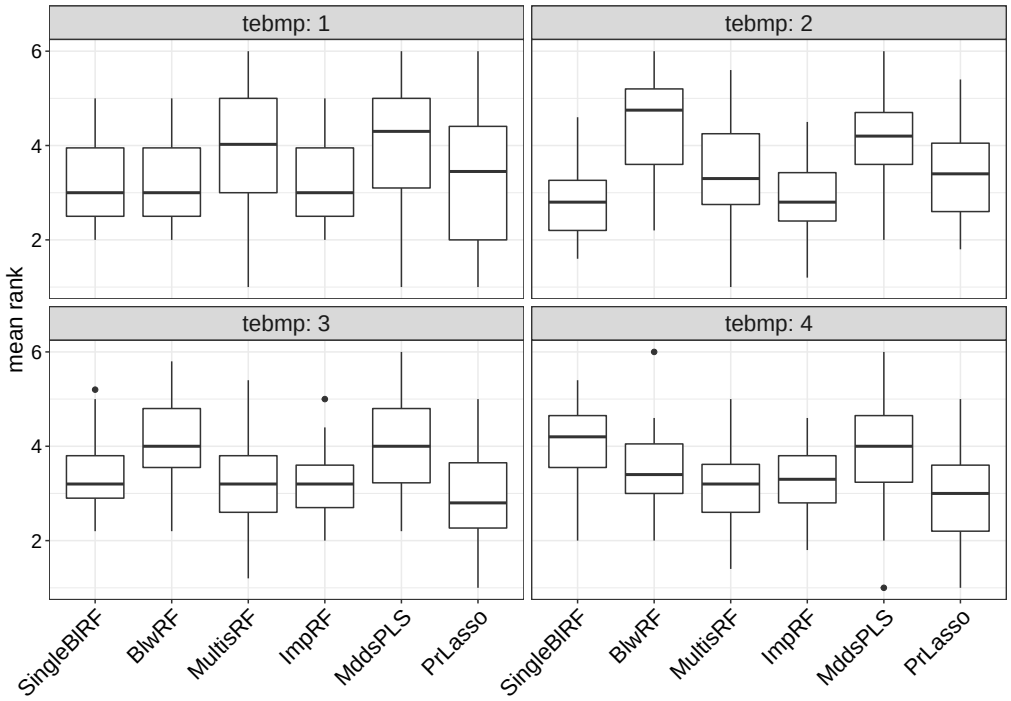

Fig. S26: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the AUC – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

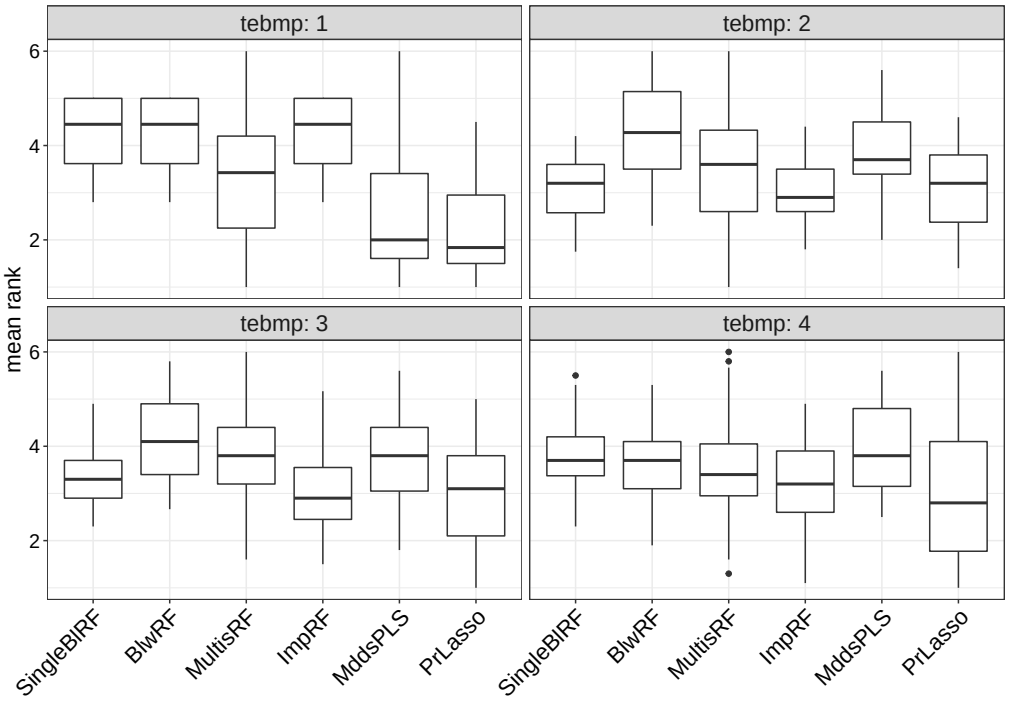

Fig. S27: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the accuracy – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

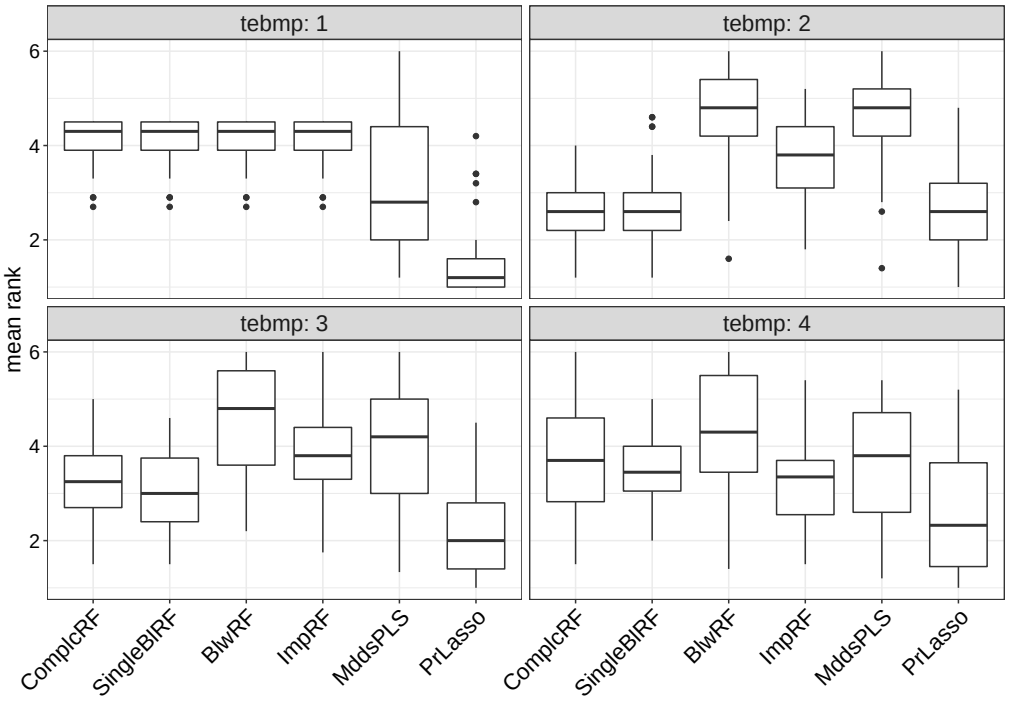

Fig. S28: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the Brier score – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

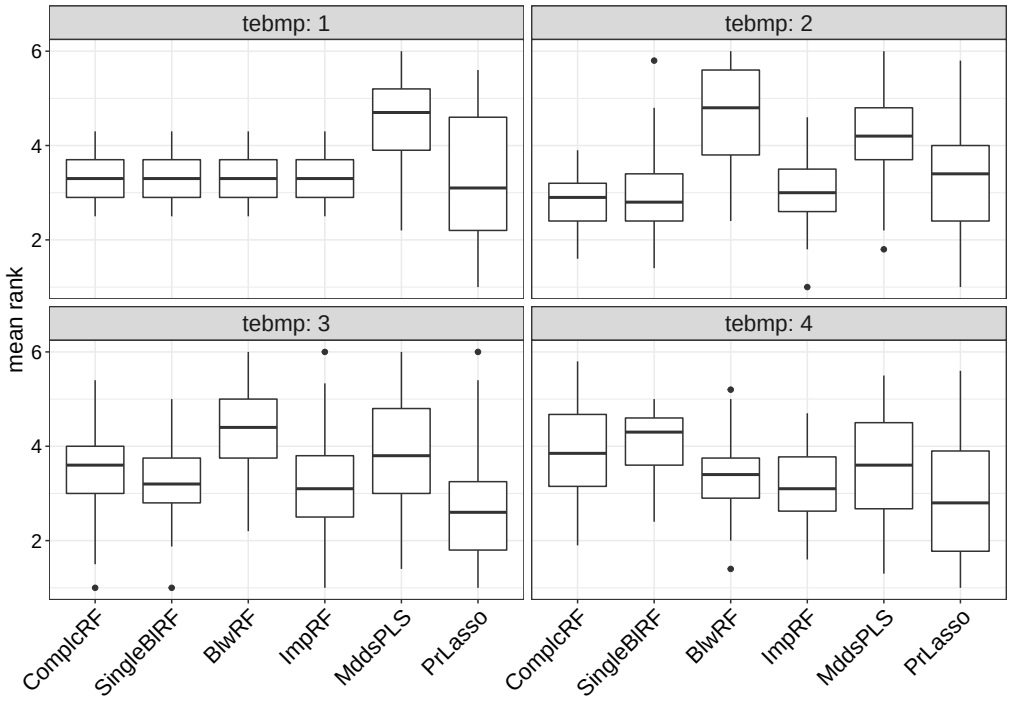

Fig. S29: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the AUC – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

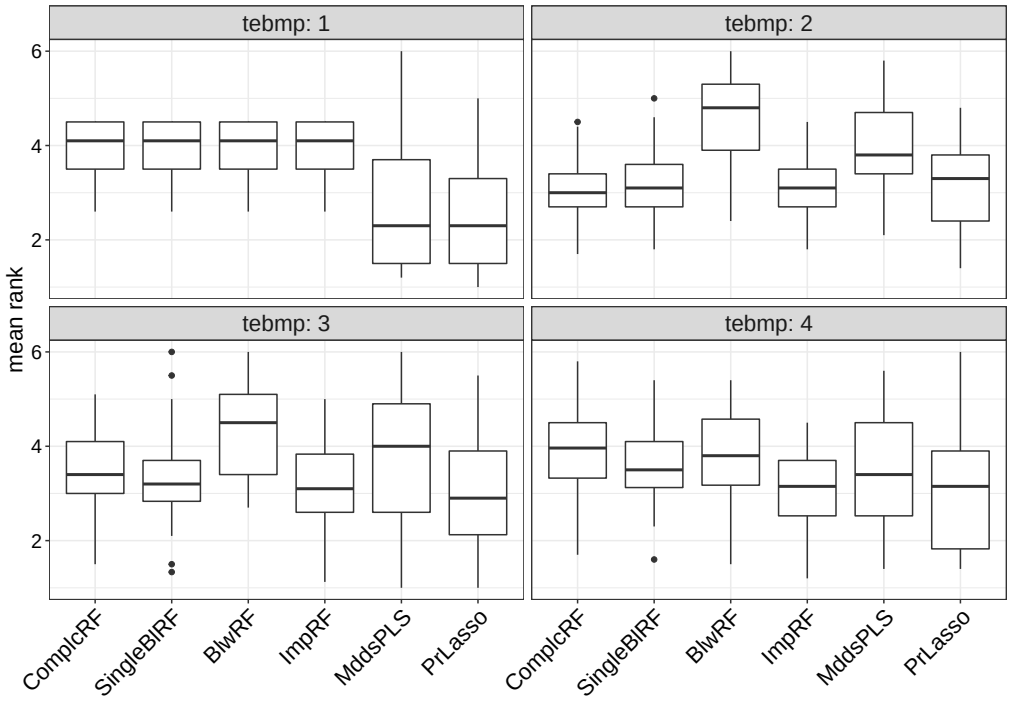

Fig. S30: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the accuracy – separately by tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

D.2 Performance metric values

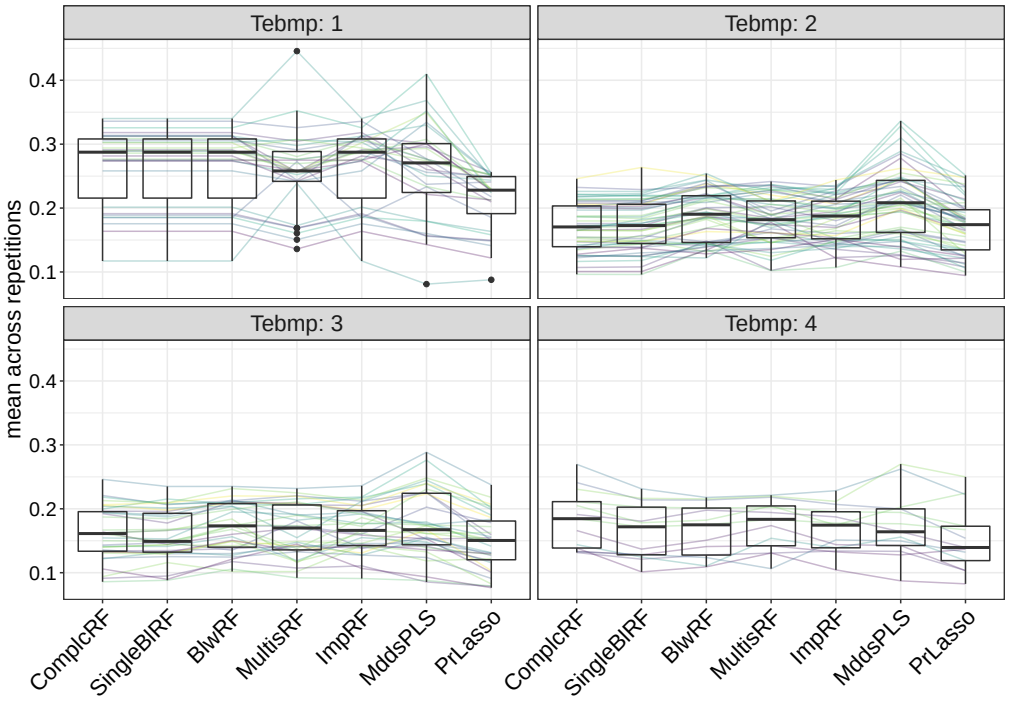

Fig. S31: Brier score values – separately by tebmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

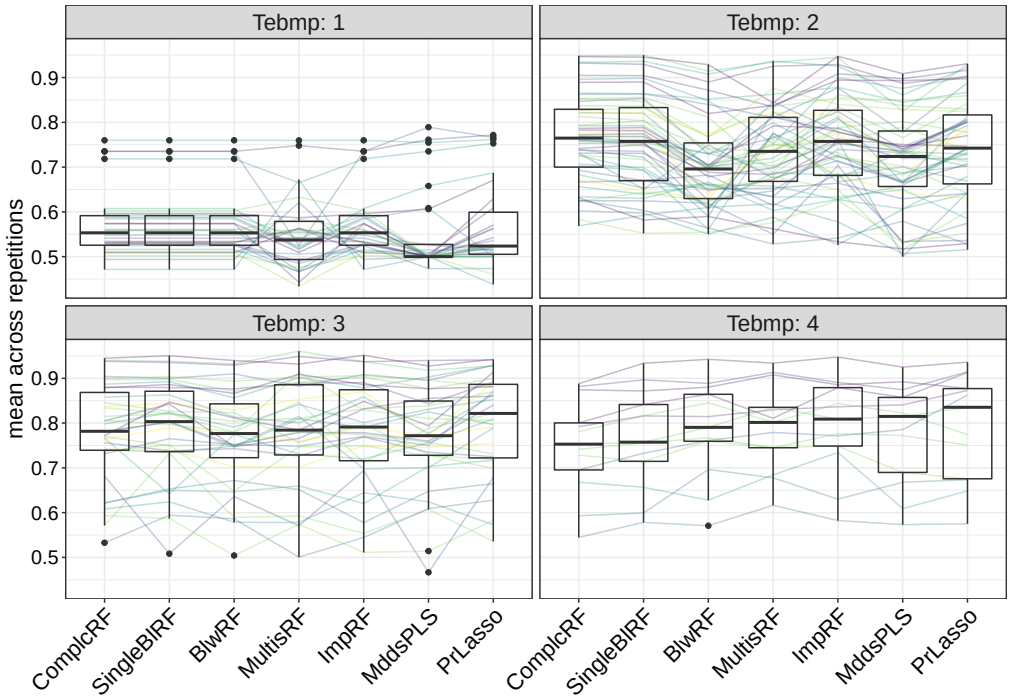

Fig. S32: AUC values – separately by tebmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

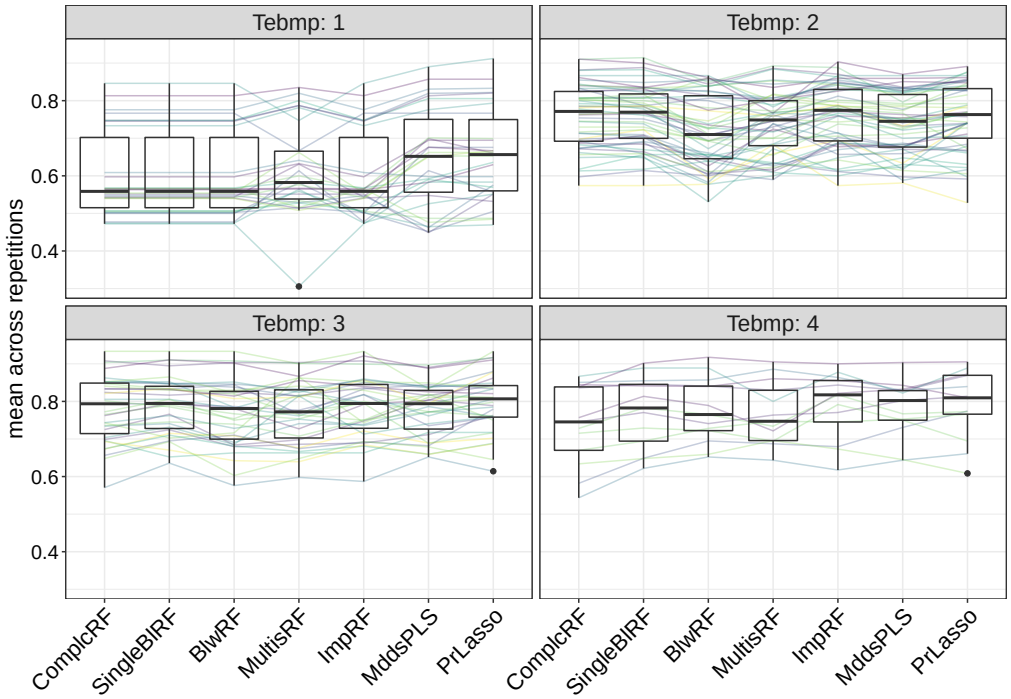

Fig. S33: Accuracy values – separately by tebmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

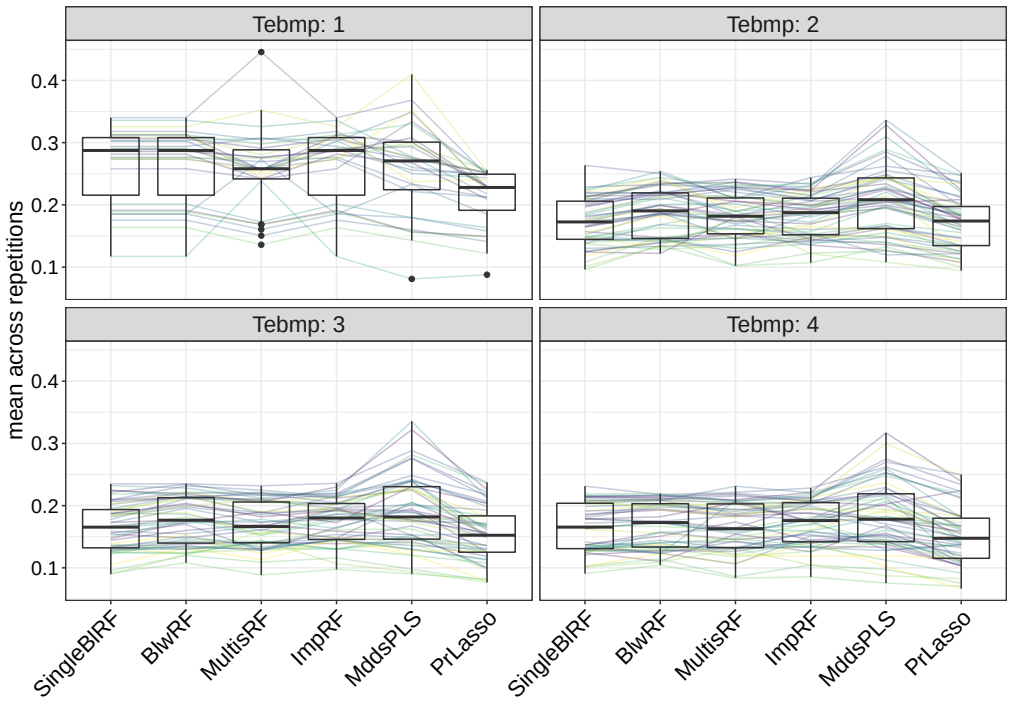

Fig. S34: Brier score values without `ComplcRF` – separately by tebmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

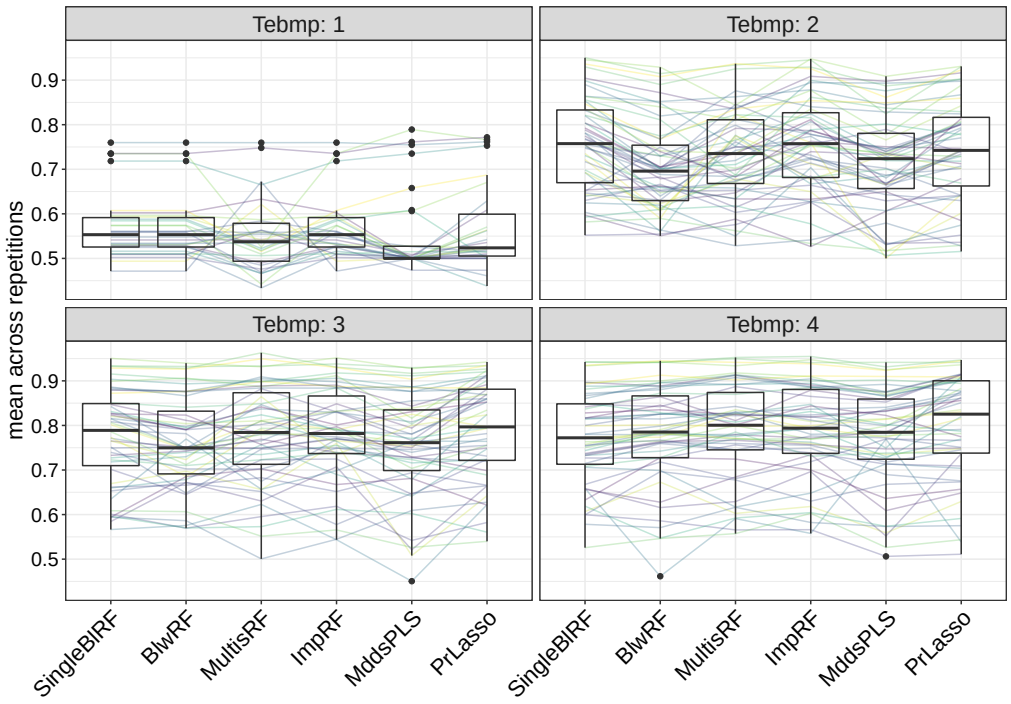

Fig. S35: AUC values without `ComplcRF` – separately by tebmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

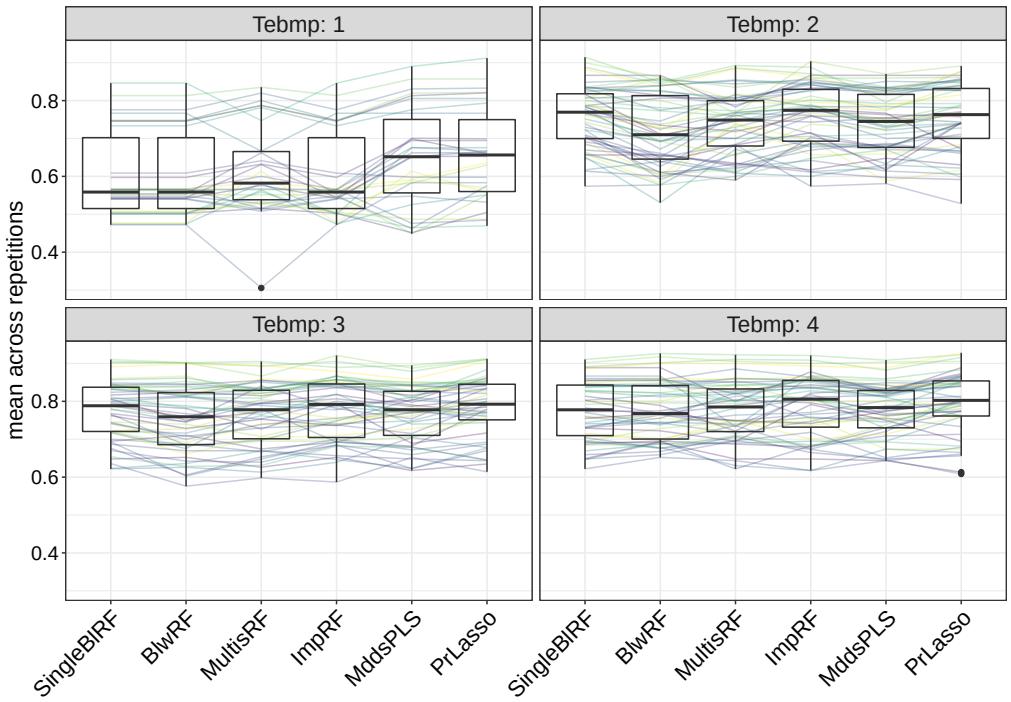

Fig. S36: Accuracy values without `ComplcRF` – separately by tebmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

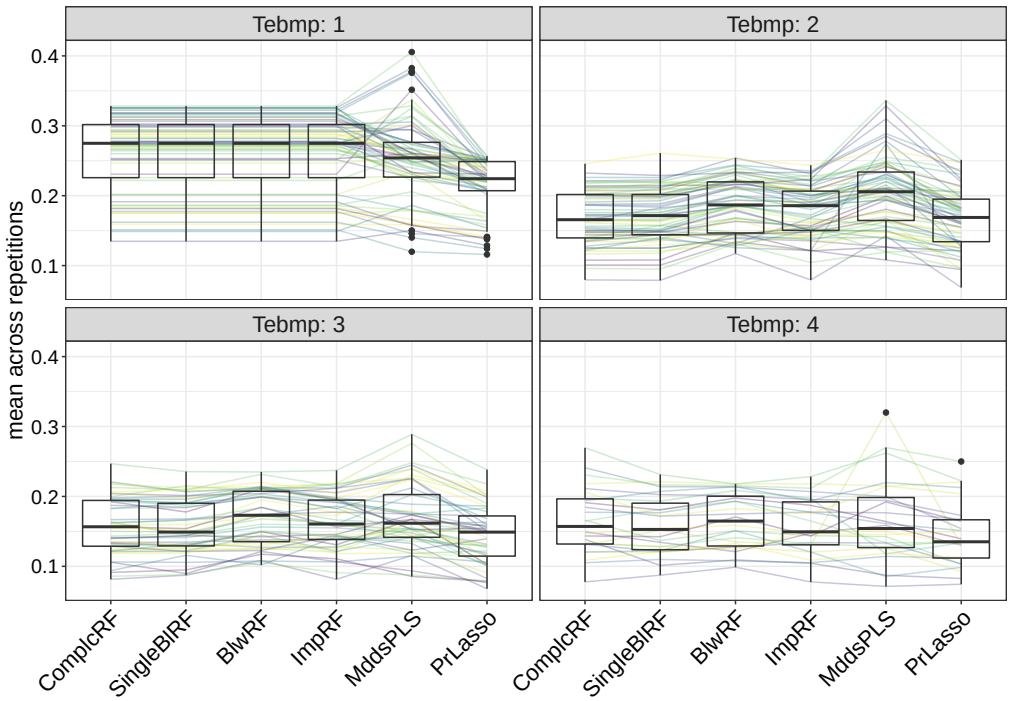

Fig. S37: Brier score values without `MultisRF` – separately by tebmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

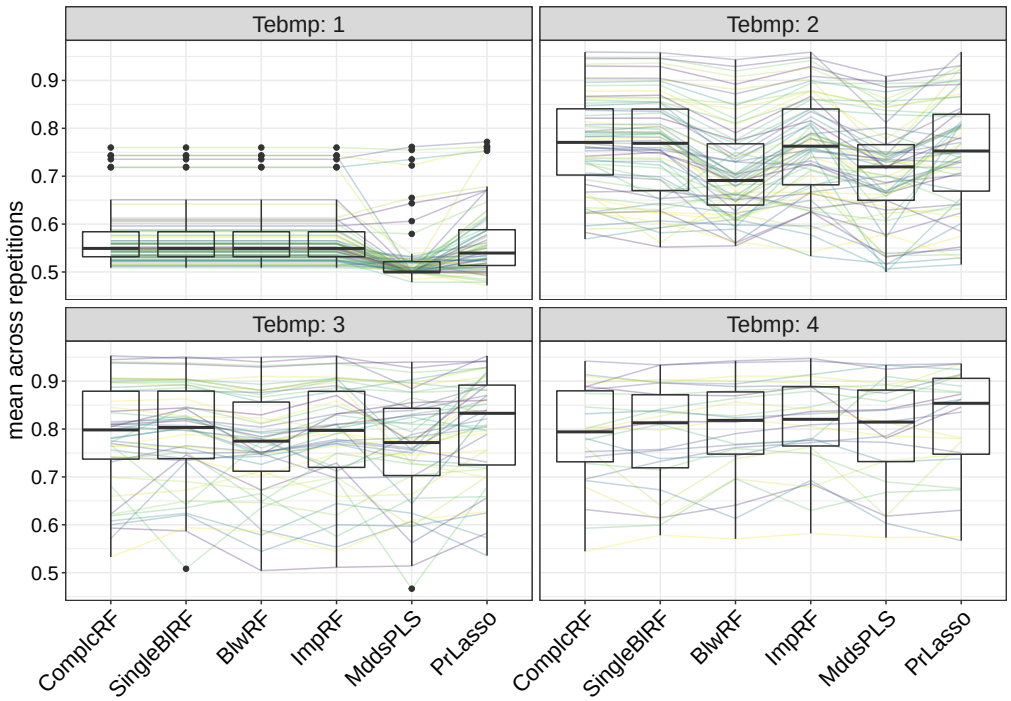

Fig. S38: AUC values without `MultisRF` – separately by tebmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

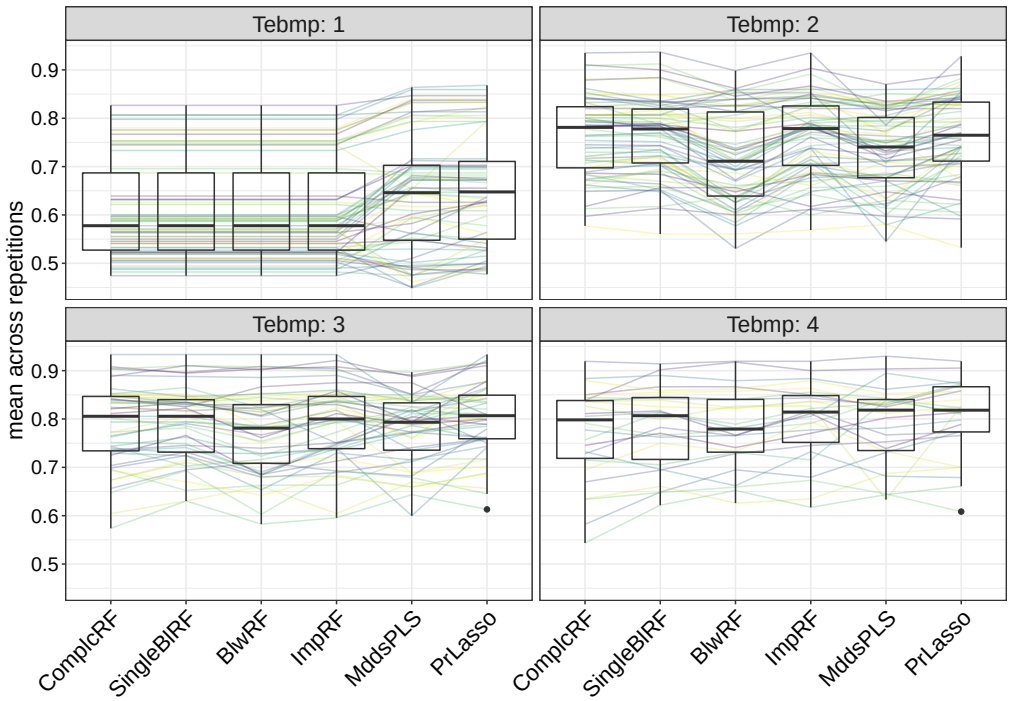

Fig. S39: Accuracy values without `MultisRF` – separately by tebmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

E Performance separately by each combination of trbmp and tebmp

E.1 Rank values

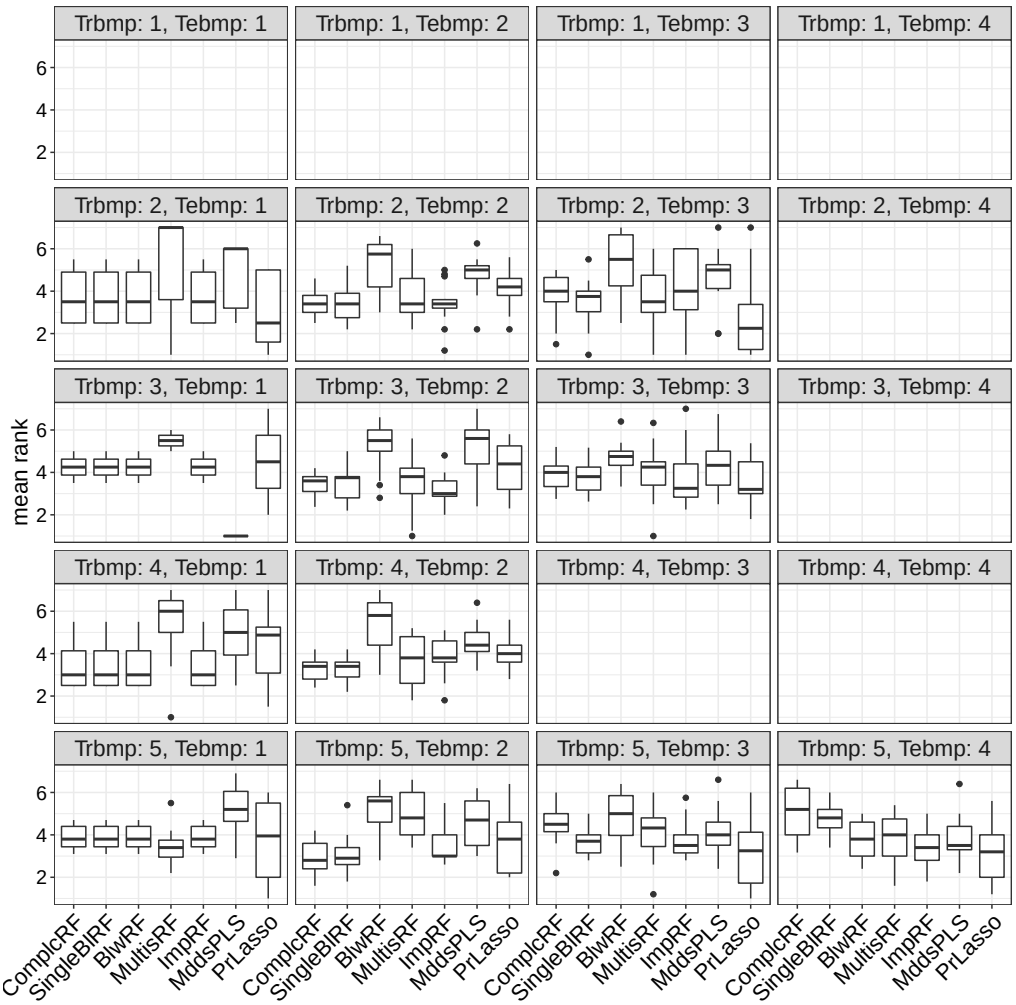

Fig. S40: Ranks each method achieved among the other methods in terms of the AUC – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all seven methods. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

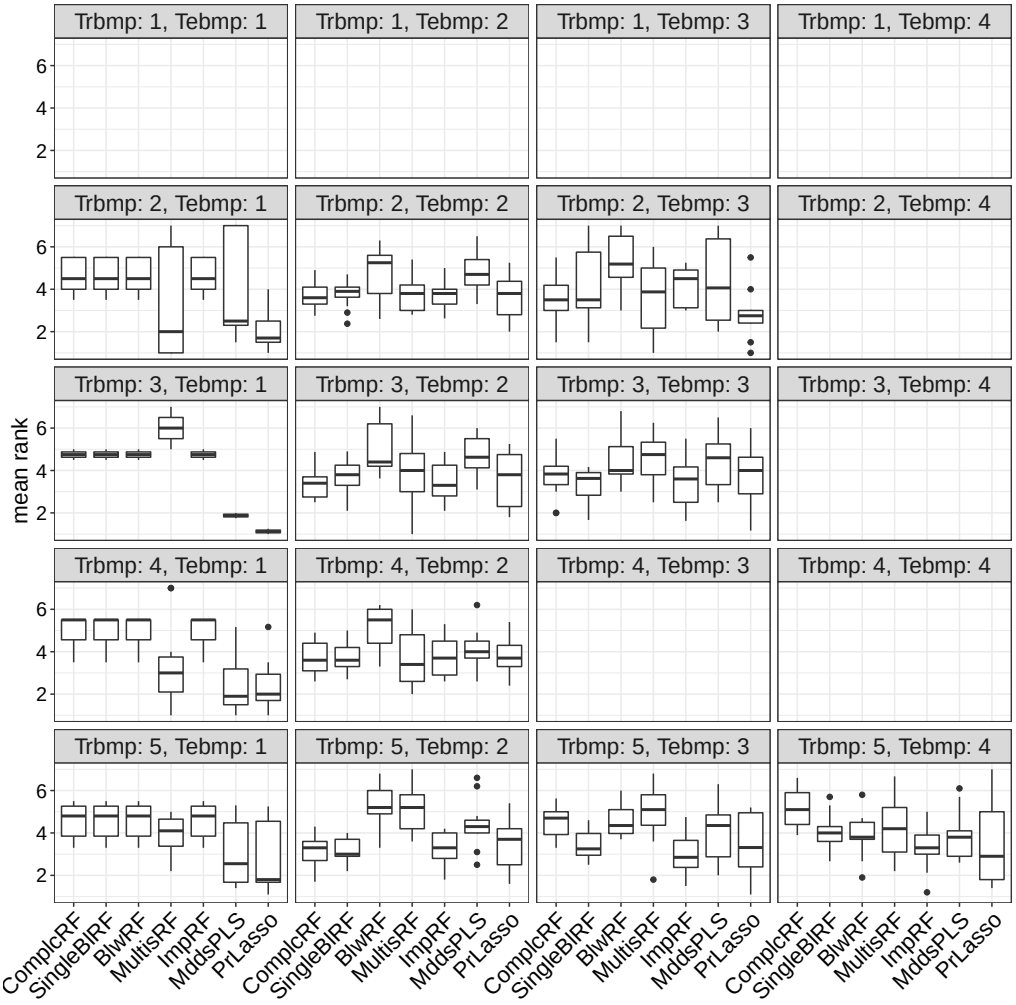

Fig. S41: Ranks each method achieved among the other methods in terms of the accuracy – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all seven methods. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

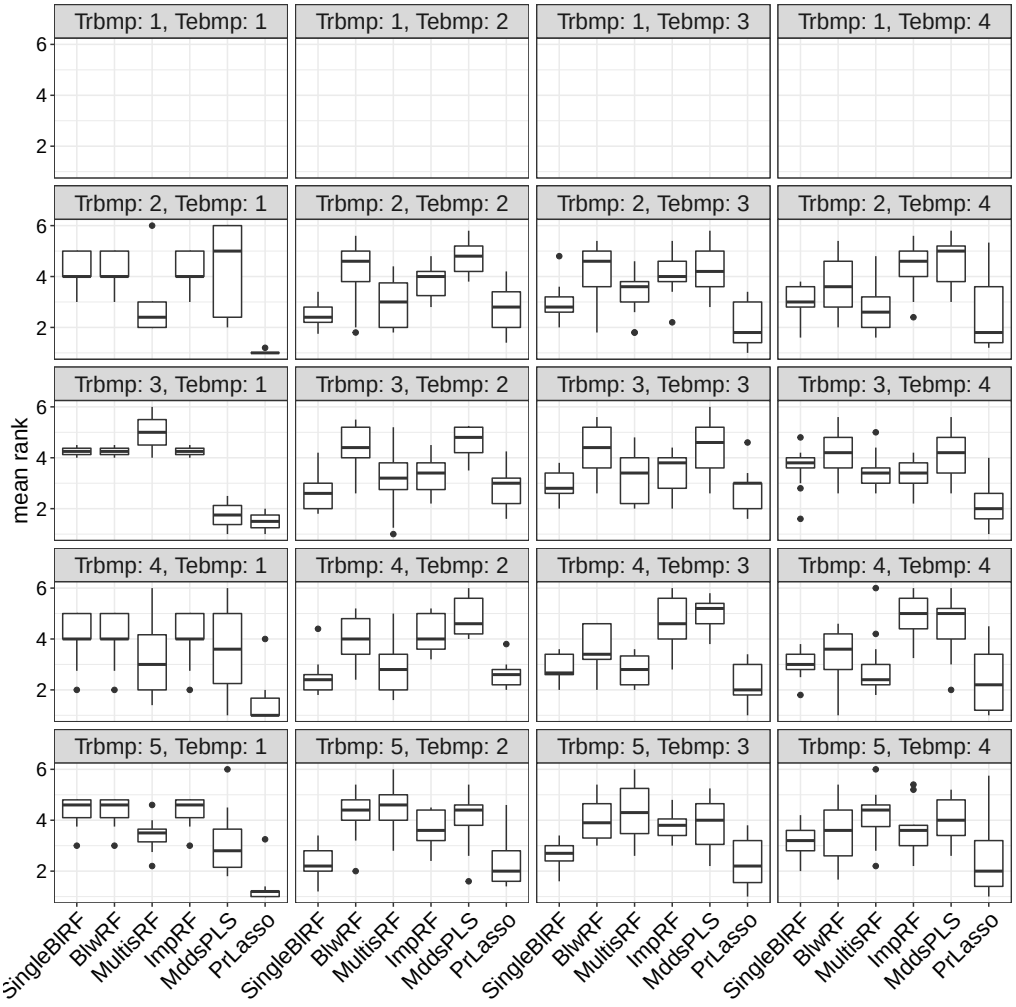

Fig. S42: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the Brier score – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

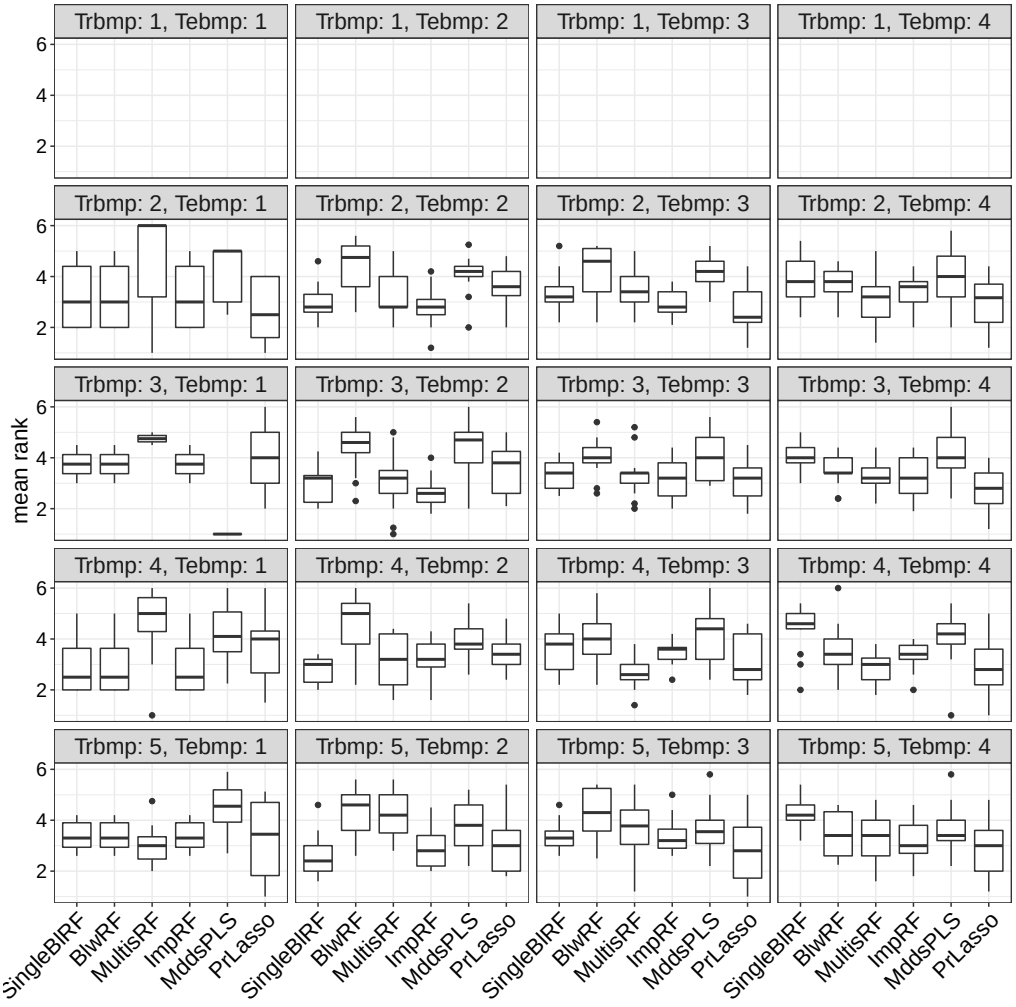

Fig. S43: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the AUC – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

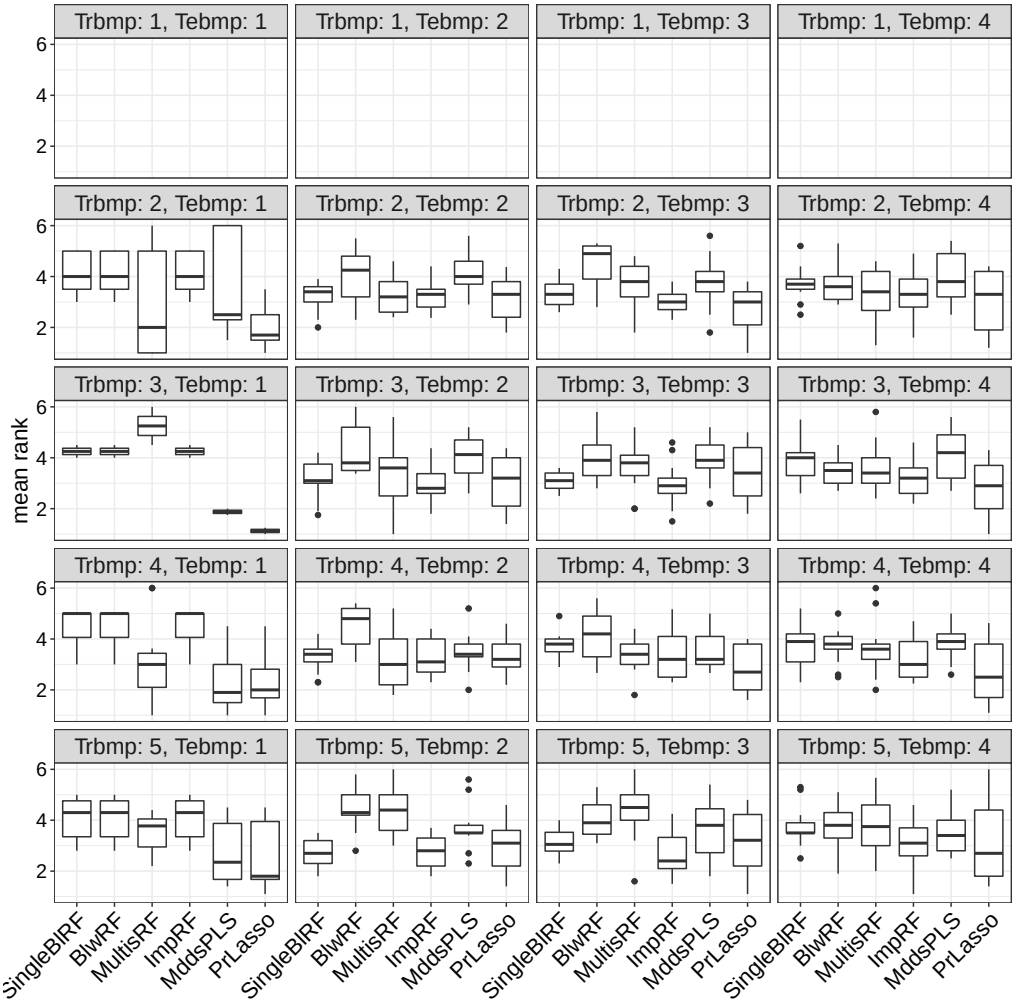

Fig. S44: Ranks each method achieved among the other methods excluding `ComplcRF` in terms of the accuracy – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

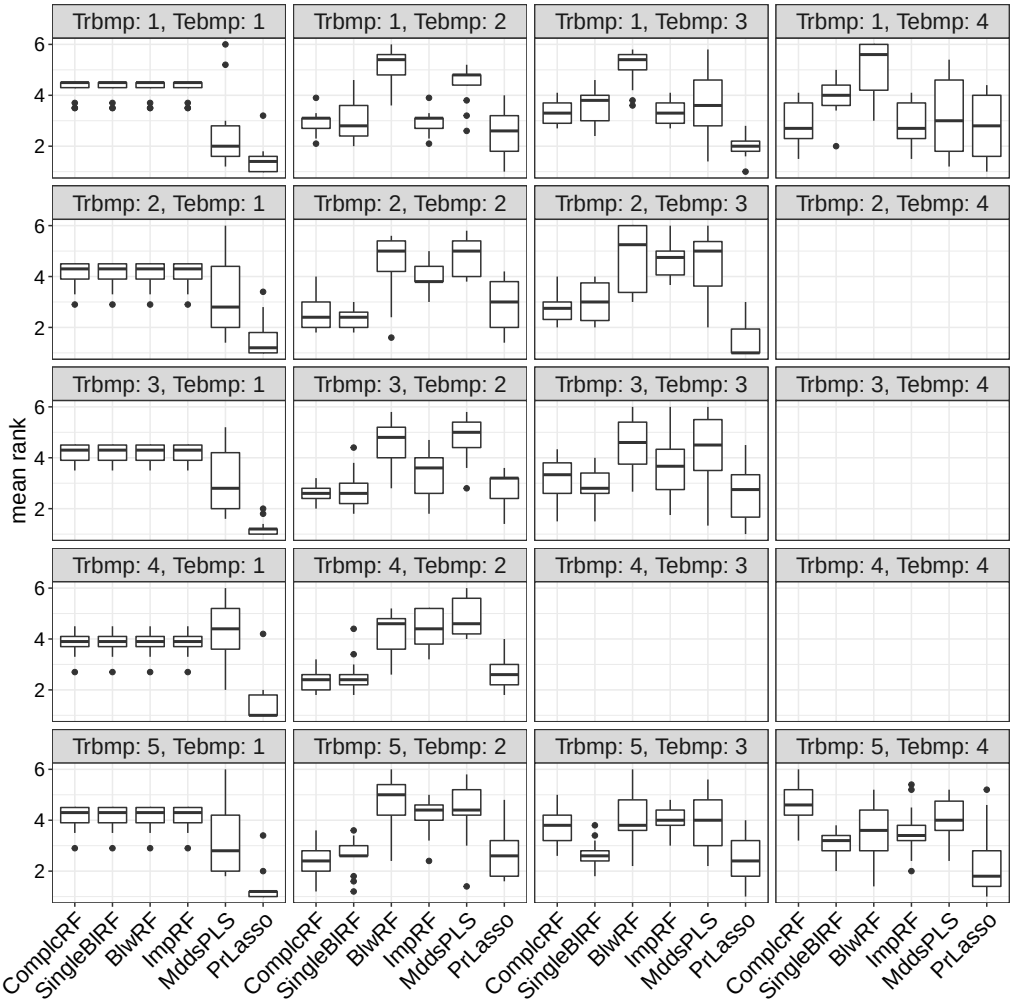

Fig. S45: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the Brier score – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

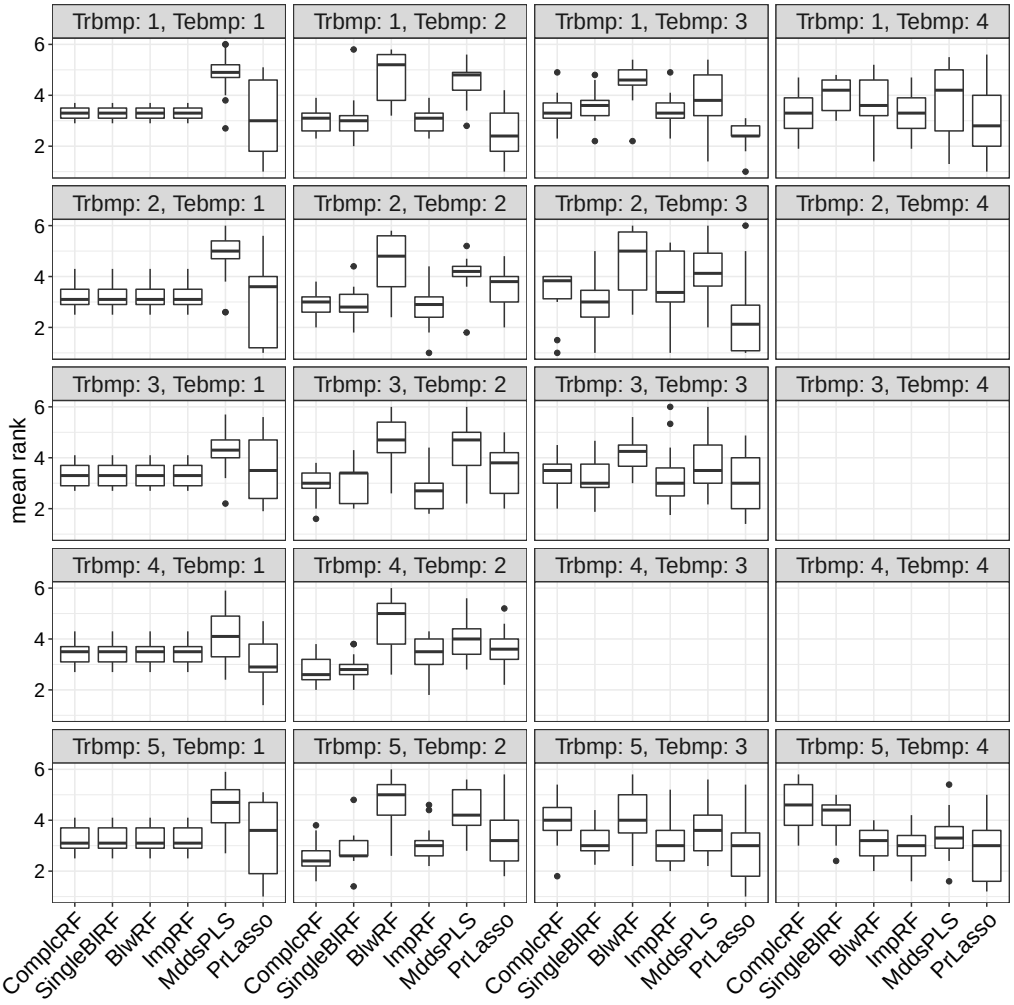

Fig. S46: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the AUC – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

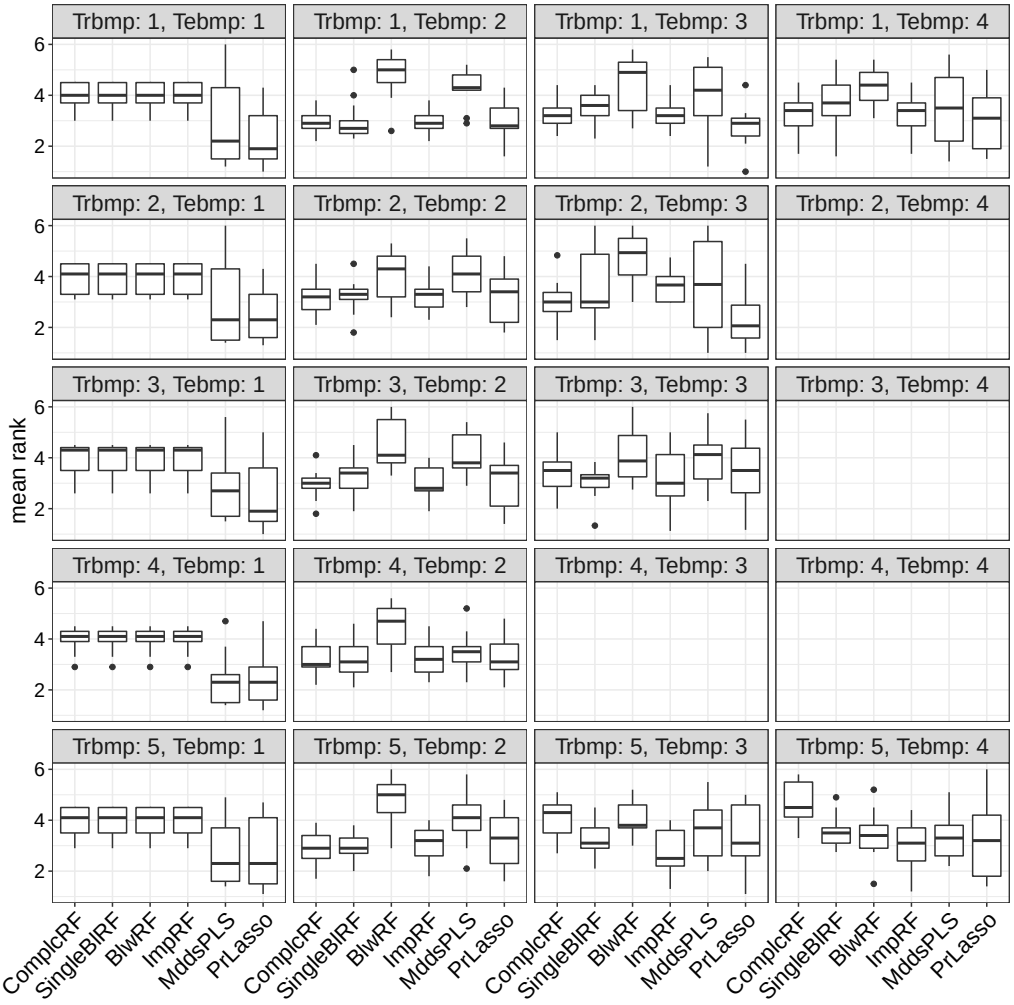

Fig. S47: Ranks each method achieved among the other methods excluding `MultisRF` in terms of the accuracy – separately by each combination of trbmp and tebmp. The ranks were computed for each combination of tebmp, tebmp, dataset, and repetition, where only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The ranks were then averaged across the repetitions. Smaller ranks indicate a better performance.

E.2 Performance metric values

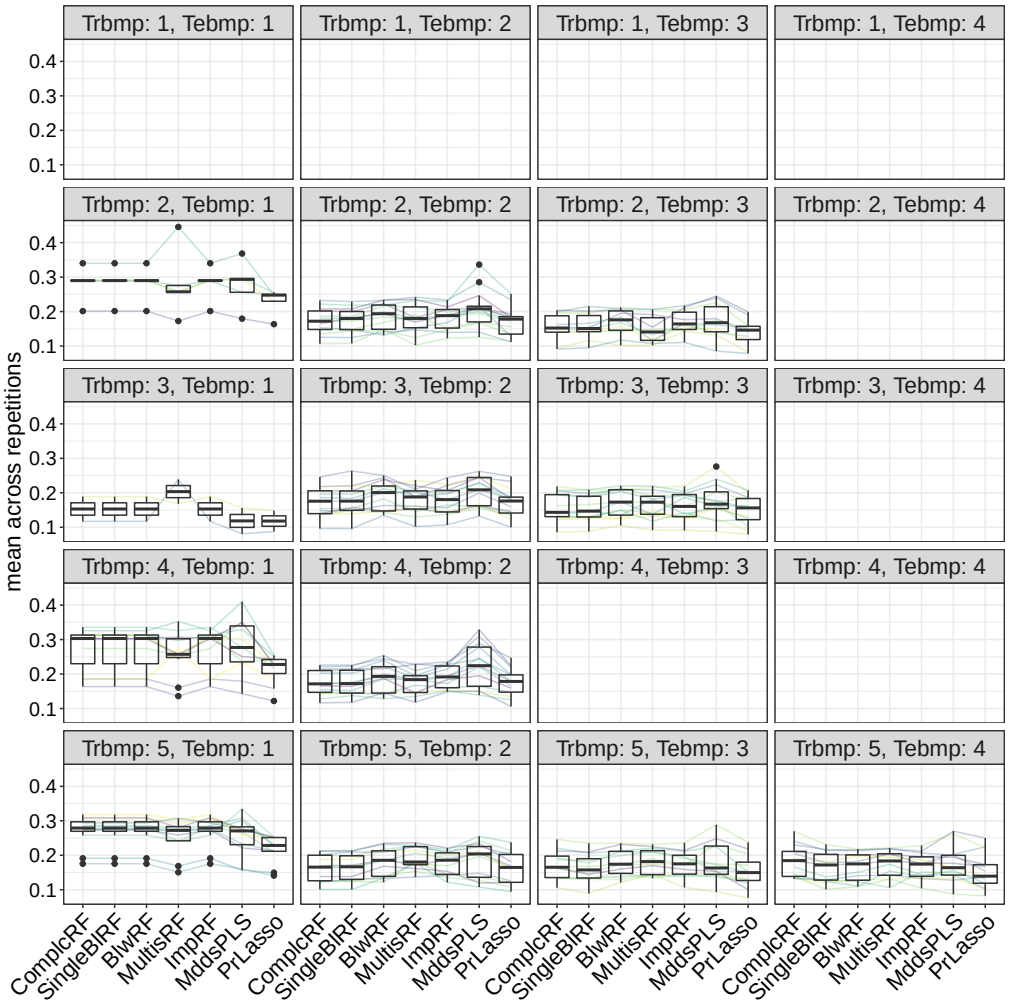

Fig. S48: Brier score values – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

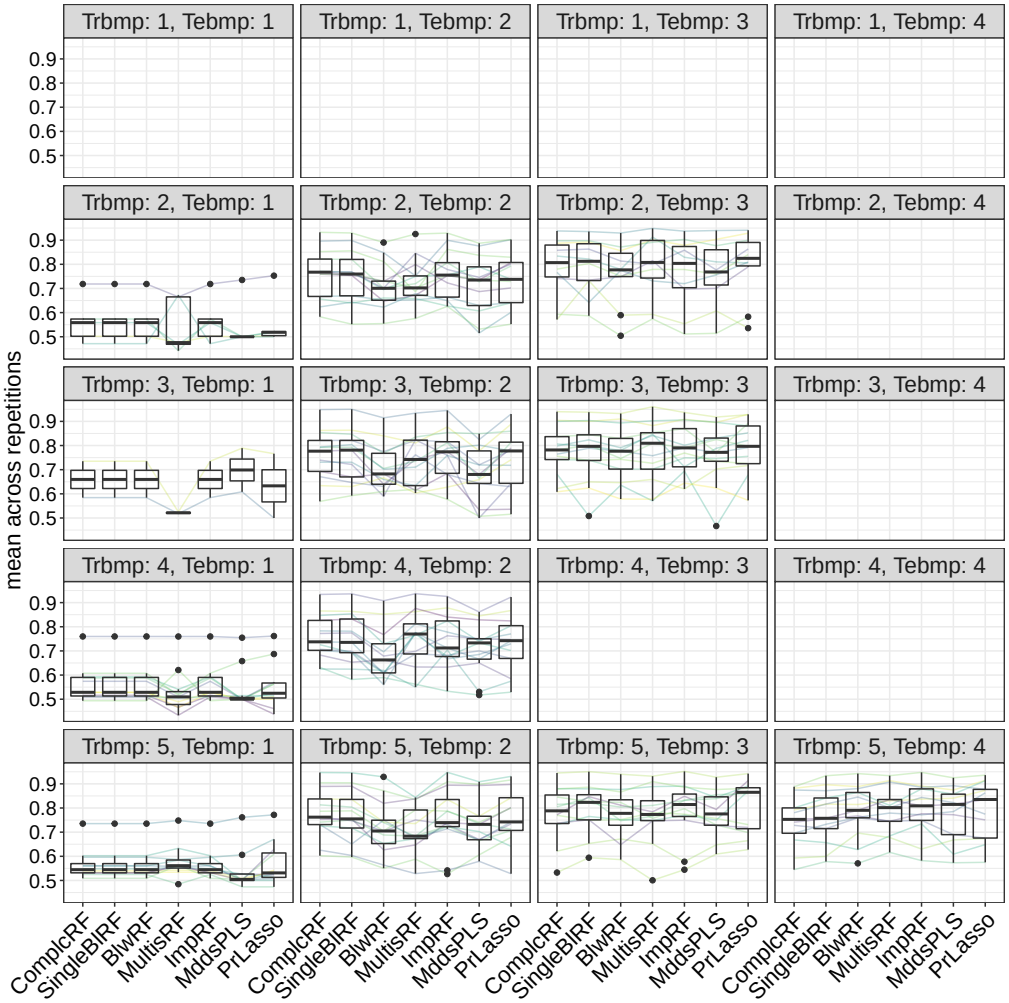

Fig. S49: AUC values – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

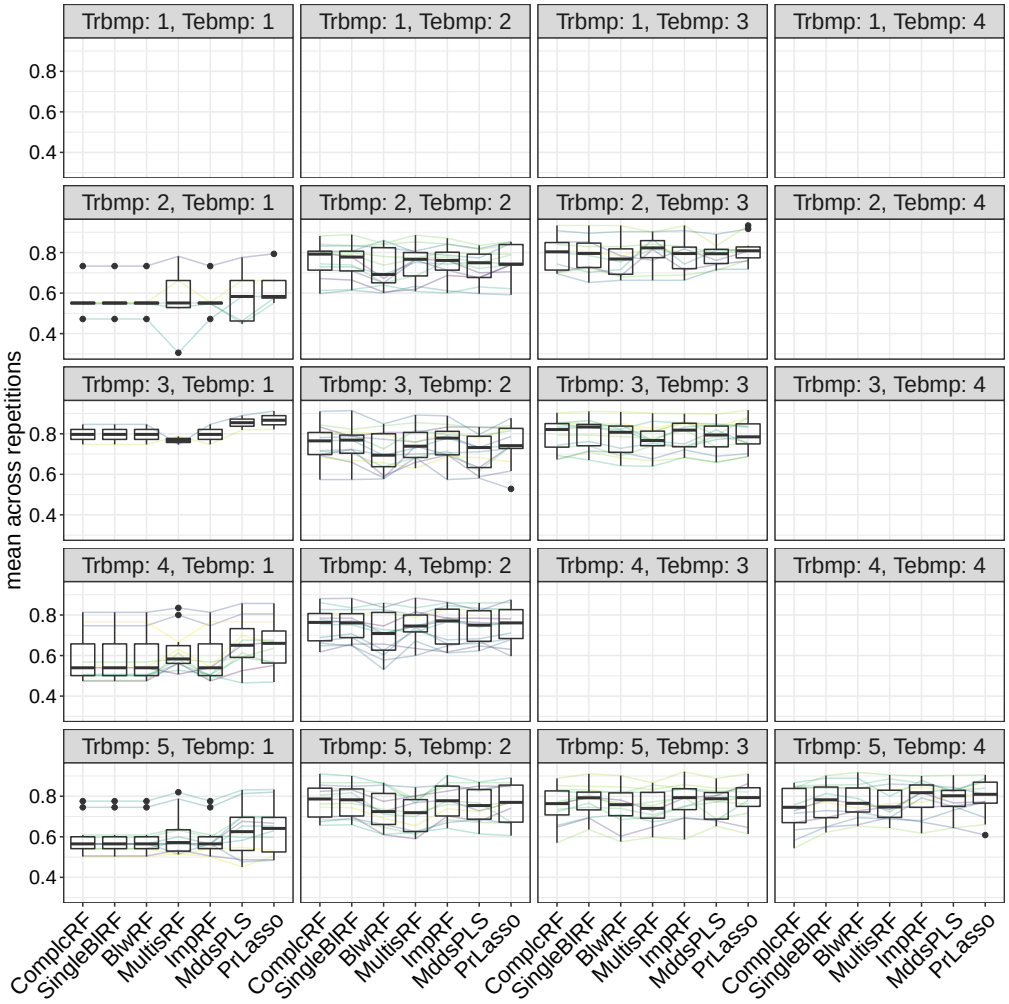

Fig. S50: Accuracy values – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all seven methods. The values were averaged across the repetitions.

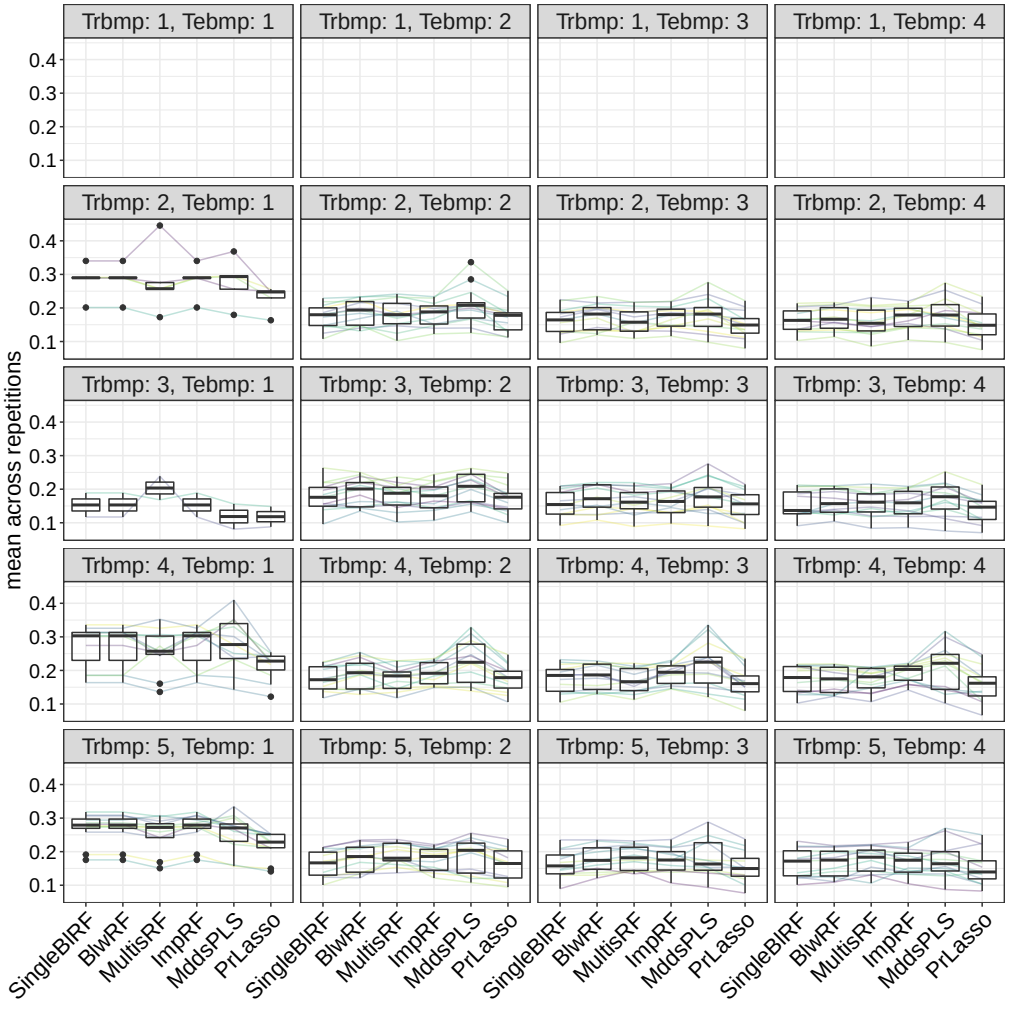

Fig. S51: Brier score values without `ComplcRF` – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

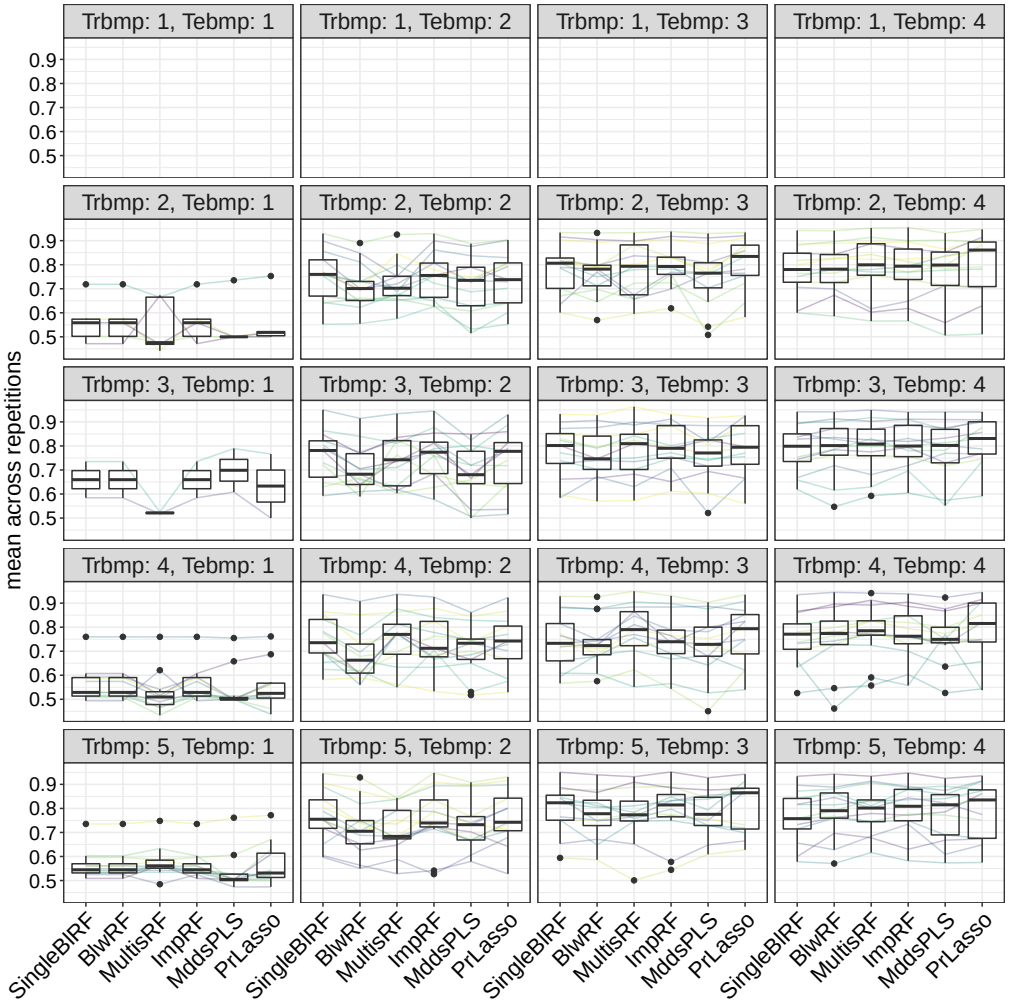

Fig. S52: AUC values without `ComplcRF` – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

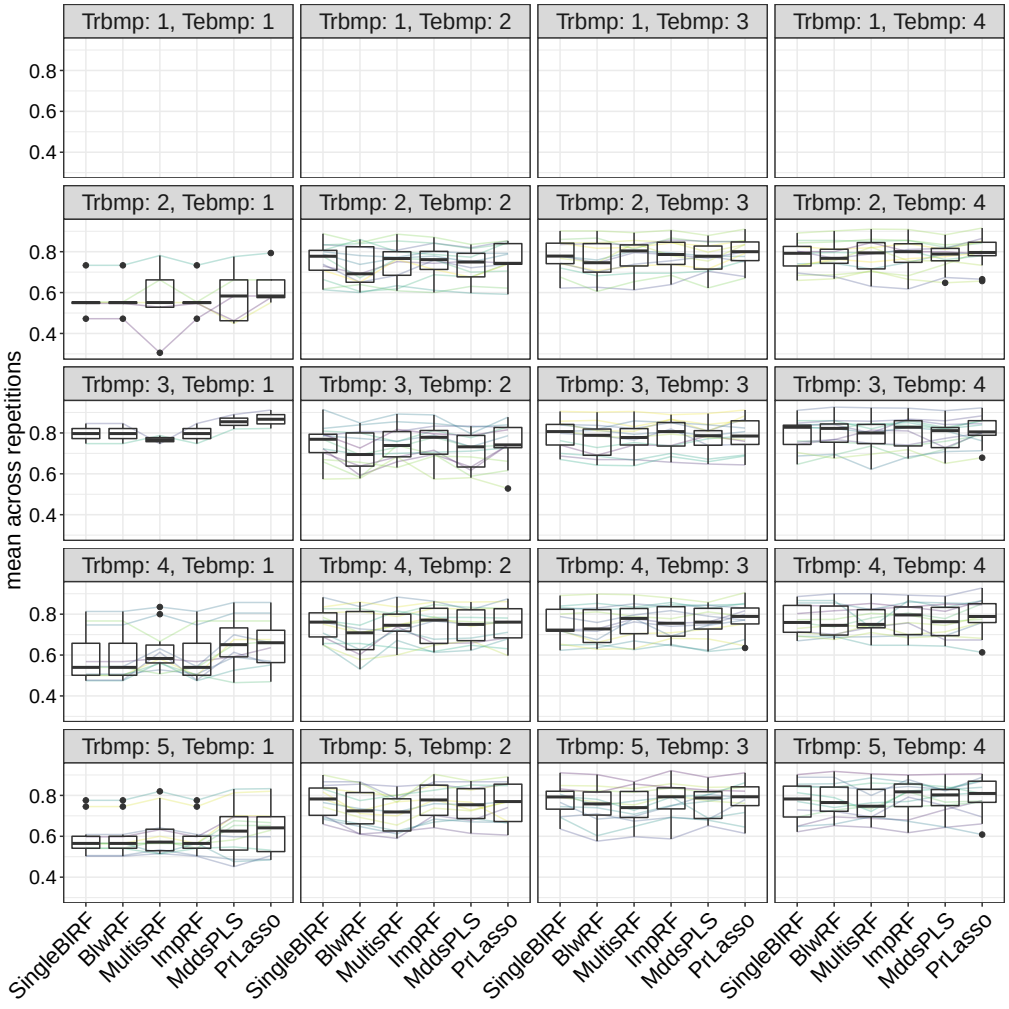

Fig. S53: Accuracy values without `ComplcRF` – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all methods excluding `ComplcRF`. The values were averaged across the repetitions.

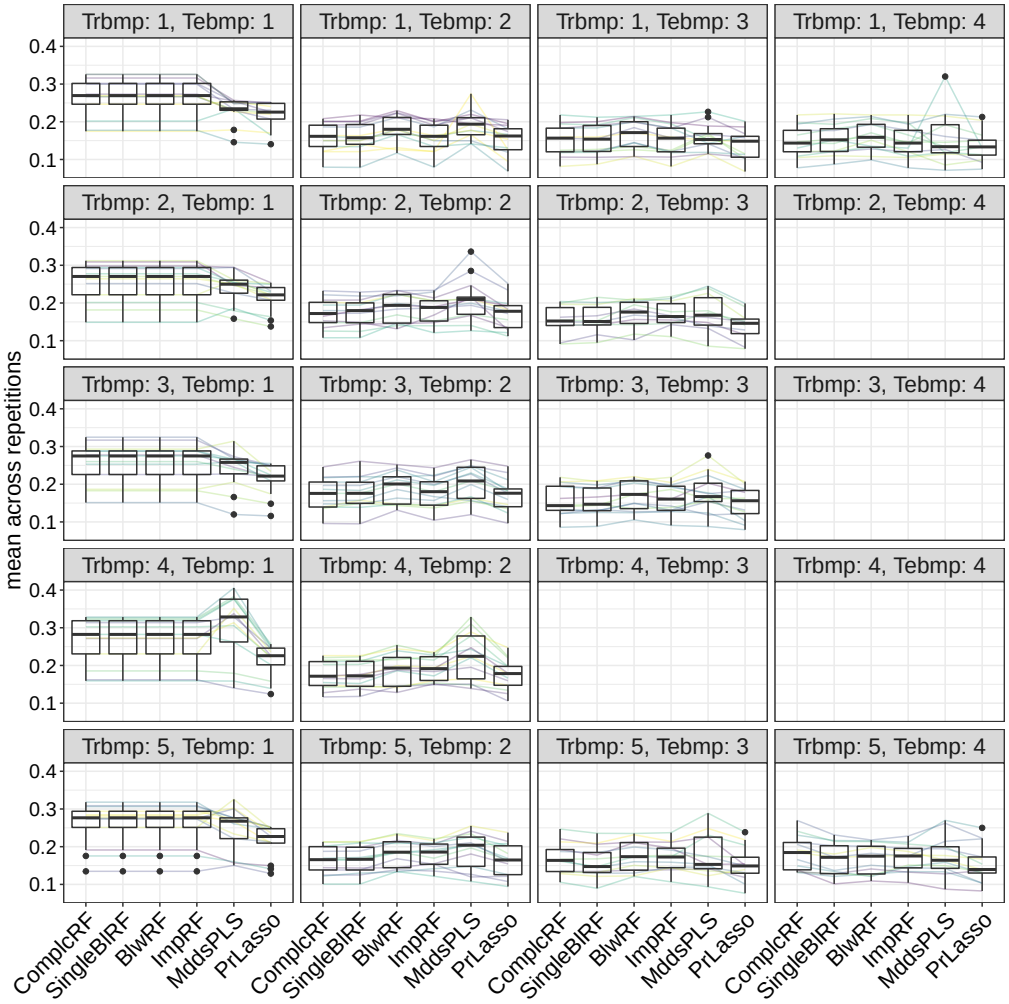

Fig. S54: Brier score values without `MultisRF` – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

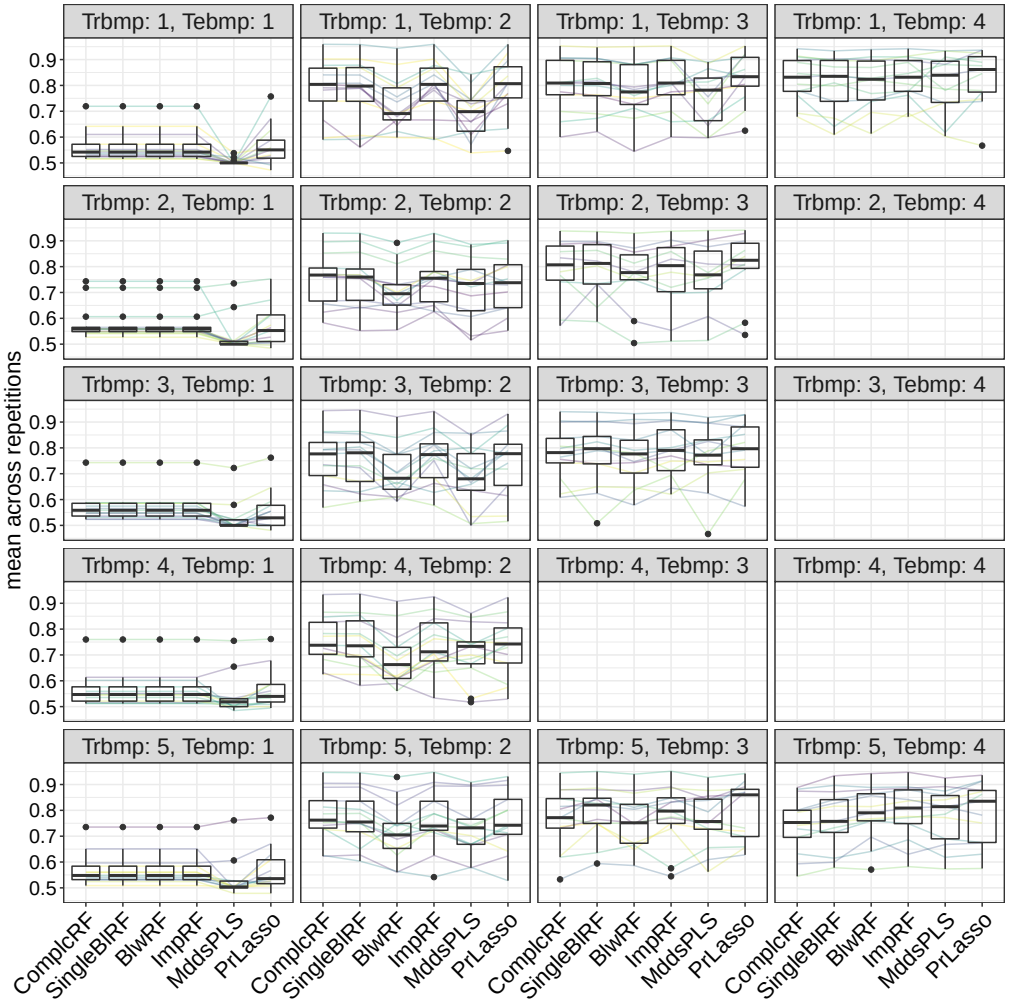

Fig. S55: AUC values without `MultisRF` – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.

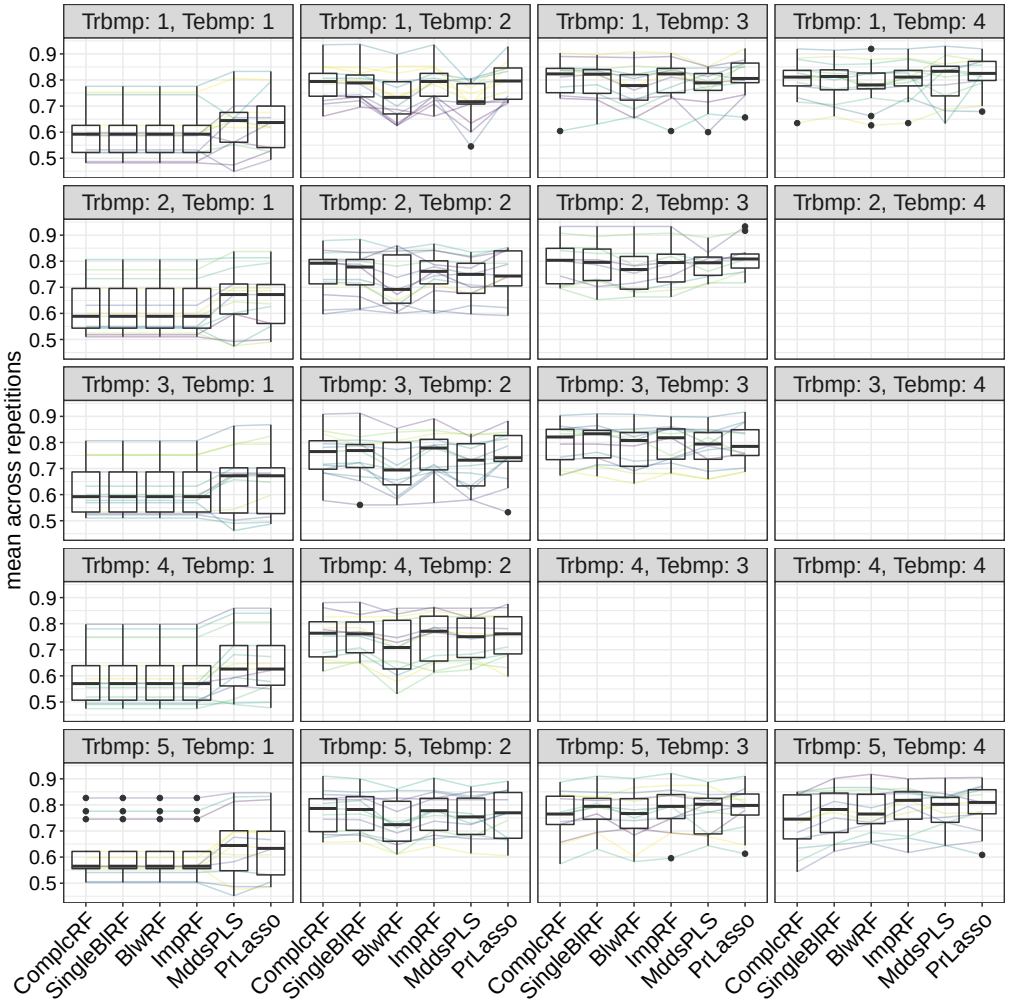

Fig. S56: Accuracy values without `MultisRF` – separately by each combination of trbmp and tebmp. Only those repetitions were considered for which the results were available for all methods excluding `MultisRF`. The values were averaged across the repetitions.