



Figure 15: Combined Accuracy of Uncertainty Estimation and Predictive Performance. Comparing the ENCE of all aleatoric uncertainty estimating models (top) and all epistemic uncertainty estimating models (bottom), aggregated over 10 experiments. For each dataset, the best model in terms of average ENCE is marked with a star together with any other models that are not statistically worse based on a one-sided Mann-Whitney-Wilcoxon test. Apart from the Random Forest and Evidential models, all other models are trained with censored labels.