Ensemble loss (w. hyperparameters as in Wenzel et al. (2020)) 0.400 Mean loss $\pm 1\sigma$ 0.375 Best individual model Categorical cross-entropy Wenzel et al. (2020) 0.350 0.325 0.300 0.275 0.250 0.225 0.200 2.0 2.5 3.0 3.5 4.0

Ensemble size