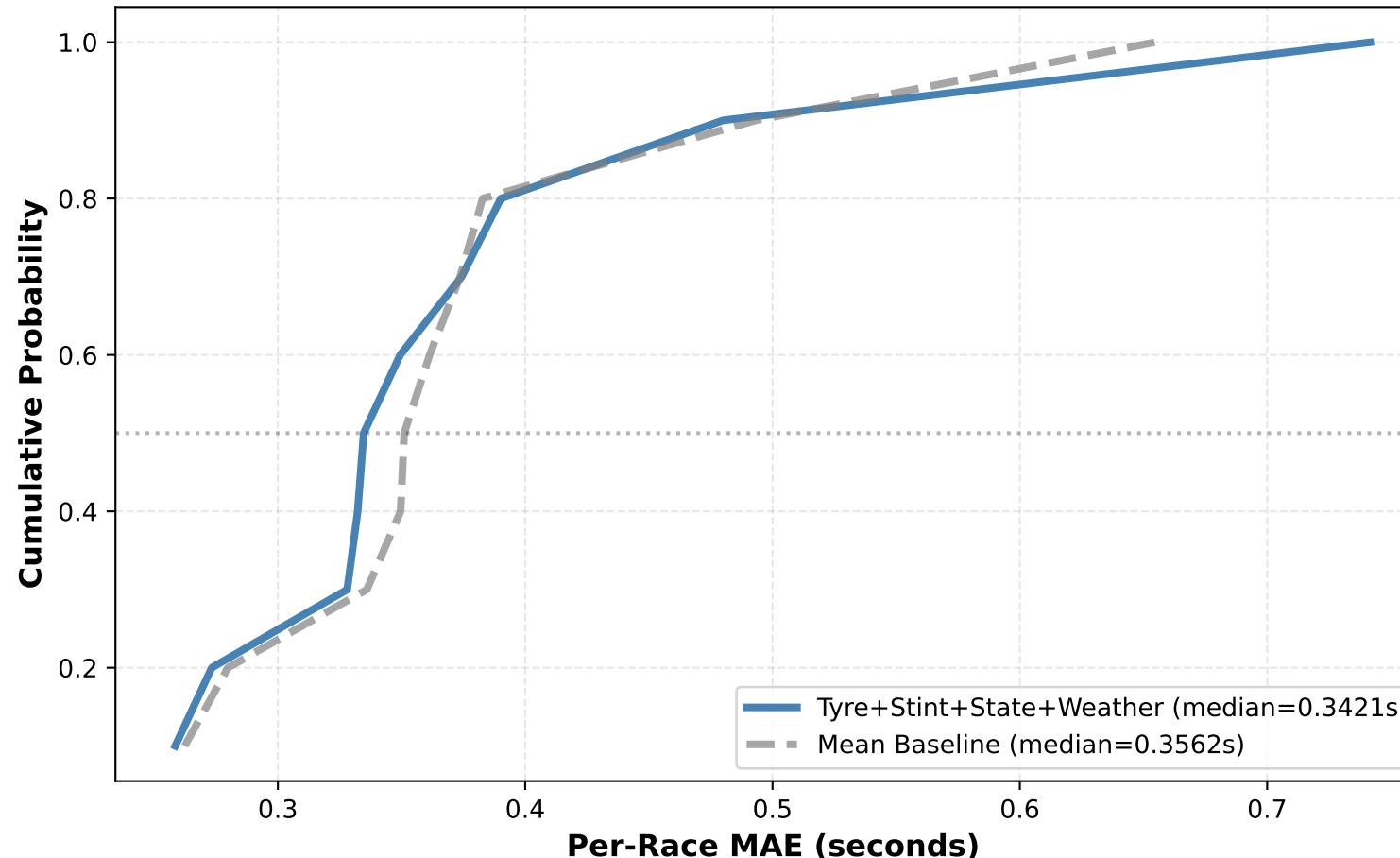


Model Improvement Across Races
→ Clear separation confirms generalization beyond just average metrics

Validation: Per-Race MAE Comparison
 $n=10$ races | Leftward shift = improvement



Test: Per-Race MAE Comparison
 $n=10$ races | Leftward shift = improvement

