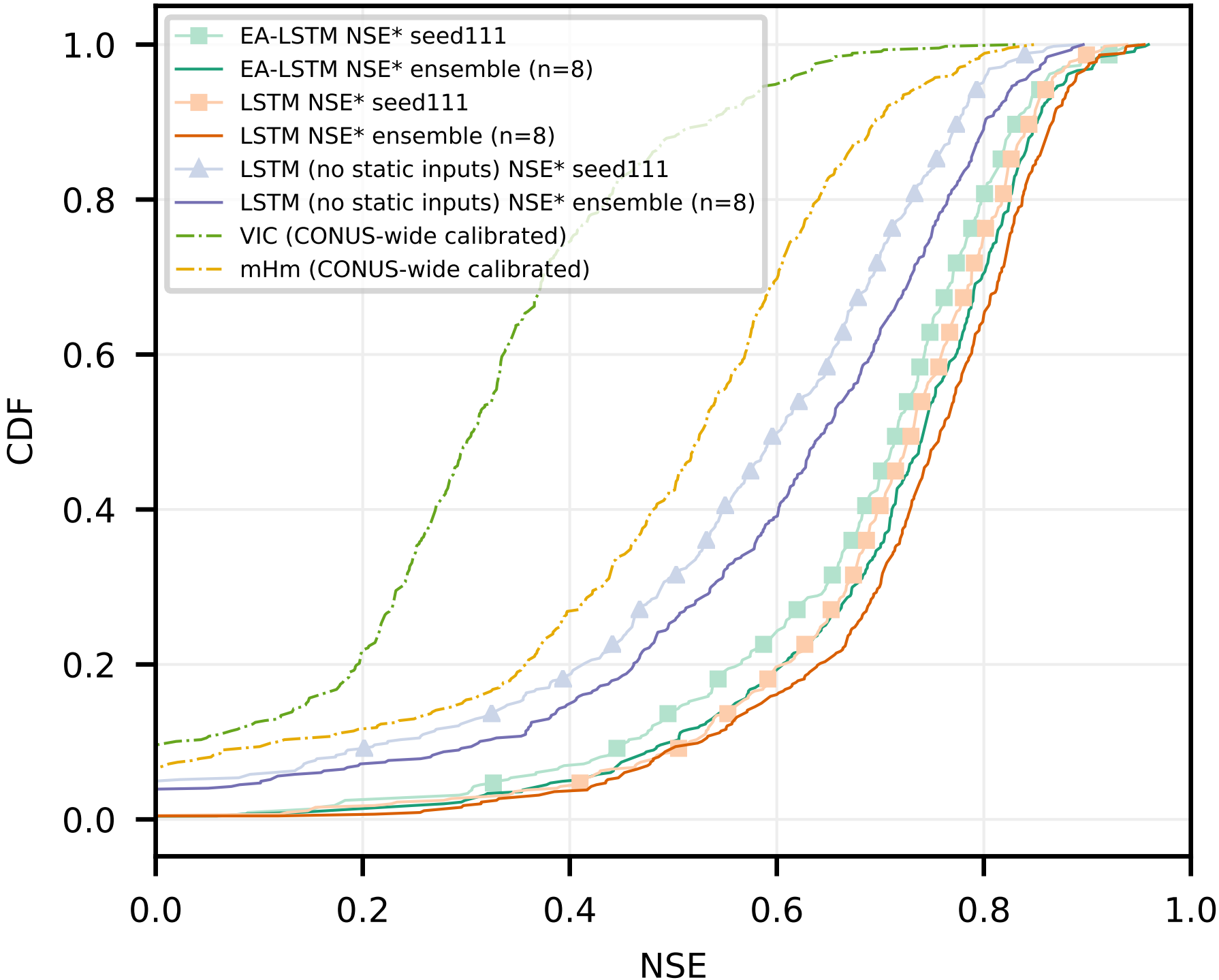Benchmarking vs CONUS-wide calibrated models