RMSE comparison across preprocessing scenarios