Sea surface temperature (SST) is a fundamental quantity to understand weather and climate dynamics. Although sensors aboard satellites provide global and repeated SST coverage, a *characterization of SST retrieval error, in terms of precision and bias is* necessary for determining the suitability of SST retrievals in various applications. Guidance on how to derive meaningful error estimates, however, is still being developed. Previous methods estimated retrieval uncertainty based purely on geophysical factors, such as season or “wet or “dry” atmospheres, but the discrete nature of these bins led to obvious spatial discontinuities in SST maps. Recently, a new approach was tried where retrievals were clustered based on the terms (excluding offset) in the statistical algorithm used to estimate SST. However, this method resulted in over 600 clusters, too many to understand the geophysical conditions that influence SST retrieval error. Using MODIS and buoy SST matchups (2012 - 2016), we explore the use of various machine learning algorithms (recursive and conditional trees, random forests)to gain insight into geophysical conditions leading to the different signs and magnitudes of MODIS SST residuals (satellite SST estimates minus buoy SSTs)

estimate the error range of SST retrievals.

MODIS residuals were first split into three categories: <  0.4 C, -0.4 C <= residual <= 0.4 C, and > 0.4 C. A first finding is that these categories are heavily unbalanced, with positive residuals being much less frequent . Performance of classification algorithms is affected by imbalance, we therefore rebalance classes testing a variety of algorithms, such as SMOTE (Synthetic Minority Over Sampling Technique). We consider a variety of features for the decision tree algorithms and start with regressors from the MODIS SST algorithm, proxies for temperature deficit, and measures of spatial homogeneity in brightness temperatures (BTs) such as the range of values in the 11 um channel over a 25 km^2 area centered on the buoy/retrieval area. These features and a rebalancing of classes led to a decision tree that estimated the error range classified the SST residuals into the three categories with an accuracy of 67.2%. ~~To the best of our knowledge this is the first numerical estimate of the error.~~ Furthermore, spatial homogeneity in BTs consistently appear among the most important variables for classification of SST residuals, suggesting that unidentified cloud contamination still is one of the causes leading to larger SST residuals.

~~are commonly seen as the first splitting variable in anked as some of the most important features, giving physical insight into the geophysical conditions that affect SST retrieval accuracy~~. We use this knowledge to enhance the precision and accuracy of error estimates from our decision tree classifier.