

# Outlier Detection for RS-Derived Surface Water Temperature Estimates Using A Variational Autoencoder Method

B. Steele

December 13, 2023

[GH Repo](#)

## Scientific Motivation and Problem Statement

We use the Landsat Collection 2 surface temperature (ST) product (Cook et al. 2014) to create a more robust dataset for our analyses, supplementing the manually-measured surface temperature at lakes monitored by Northern Water (see Figure 1). The product performs particularly poorly when clouds are present or even near the area of interest - instances of this tend to result in lower-than-expected (or measured) temperature values, even though the associated metadata does not indicate that there are clouds near or at the location of interest. Additionally, the thermal sensor on the Landsat constellation has a native resolution of 100m, and we mask the data at 30m (the resolution of the optical sensors, where we determine water extent). Because the reservoirs we are interested in obtaining surface temperature for are dynamic, it is possible we are inadvertently including land-contaminated pixels in the summary of the area of interest. Since the final goal of modeling daily surface temperature is to launch a real-time decision support system, it is imperative that we don't include contaminated data into our pipeline since these data are the foundational building blocks for other optically-derived parameter estimates we will be modeling.

In an attempt to identify outliers/novel data points from the remote sensing data, I will use a variational autoencoder (VAE) trained on the *in situ* temperature measurements and associated data to establish thresholds for outlier detection for the ST product at lakes where we have measured, *in situ* data. While this method is typically used on raster-type data in conjunction with convolutional neural networks, I'm hopeful that implementation could be similar on vector data since the input data, at some point in the VAE architecture are flattened to vector-like data.

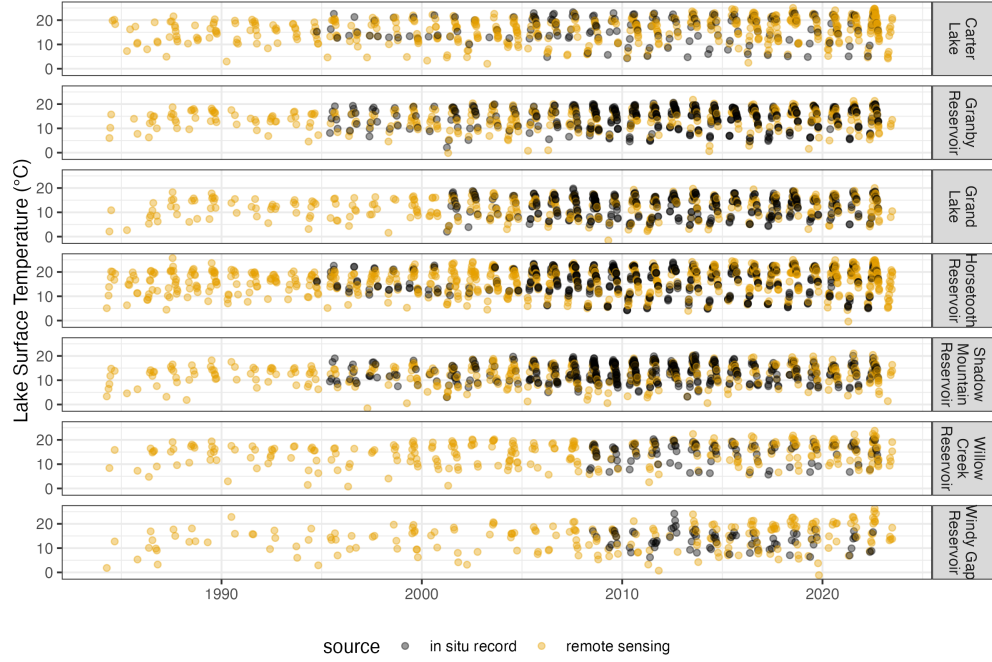


Figure 1: Remote sensing-enhanced lake surface temperature record for the seven manually-monitored Northern Water waterbodies. Manually-measured surface temperature represented by black translucent dots, remote-sensing derived values are in gold translucent dots.

The baseline for this implementation is the results of the previous neural network framework, where the daily model performance had an MAE of 1.8 and a RMSE of 2.2 degrees Celsius. If we feed only the remote sensing data which passes the variational autencoder threshold test and the MAE and RMSE decrease, we will consider this method to perform better than baseline.

## Description of the Data

Northern Water, the municipal subdistrict that delivers drinking water to approximately 1 million people in northern Colorado and irrigation water for ~600,000 acres of land has an exhaustive manual measurement record for many of the lakes within their network. The manual sampling record contains 1125 surface temperature measurements from 7 lakes. ‘Surface temperature’ in the manual sampling record is defined as any measured temperature at 1m depth. I retain only the top-most value when multiple measurements were available. The data from Landsat 4-9’s surface temperature product contains 2896 observations for these same lakes. While these thermal estimates are well-aligned with the manual monitoring data and have been bias-corrected (RMSE 1.31°C, MAE 1.06°C., Figure 2), there are still outliers when comparing these data with same-day measured surface temperature. Furthermore, these only represent 63 of the 2896 remotely-sensed values within these seven reservoirs, and 11 have residuals greater than the MAE for this subset.

The features that I will be using to train and validate this VAE for outlier detection are surface temperature, summarized NLDAS meteorological data (air temperature, precipitation, solar radiation, and wind), and static values for each of the reservoirs (elevation, surface area, maximum depth, volume, and shoreline distance). The NLDAS data have been summarized for the previous day’s weather, 3 days prior, and 5 days prior - meaning, the model does not use *today’s* weather for prediction. To capture annual warming and seasonal warm-up/cool-down, which are not always consistent between annual cycles, I’ve implemented an annual cumulative sum for both temperature and solar radiation and the day of year within the feature set. Because static features are incomplete for Windy Gap Reservoir, data from that reservoir is not included in this analysis.

The training set will include all 2896 measurements from the NW record in the six reservoirs with complete feature sets. These data are reliable and should adequately train the VAE to reconstruct the data from patterns in the latent space that are inherent to the feature set. The validation set to calculate a threshold for outliers will be the 63, where instances where the residual is less than MAE are defined as *inliers* and those greater than MAE are *outliers*. An analysis of the reconstruction error will be used to define a threshold for *inliers* and *outliers* as described in Sinha et al. (2021).

## Northern Water Waterbody/Remote Sensing Preliminary Temperature Product

Grey dashed line is 1:1

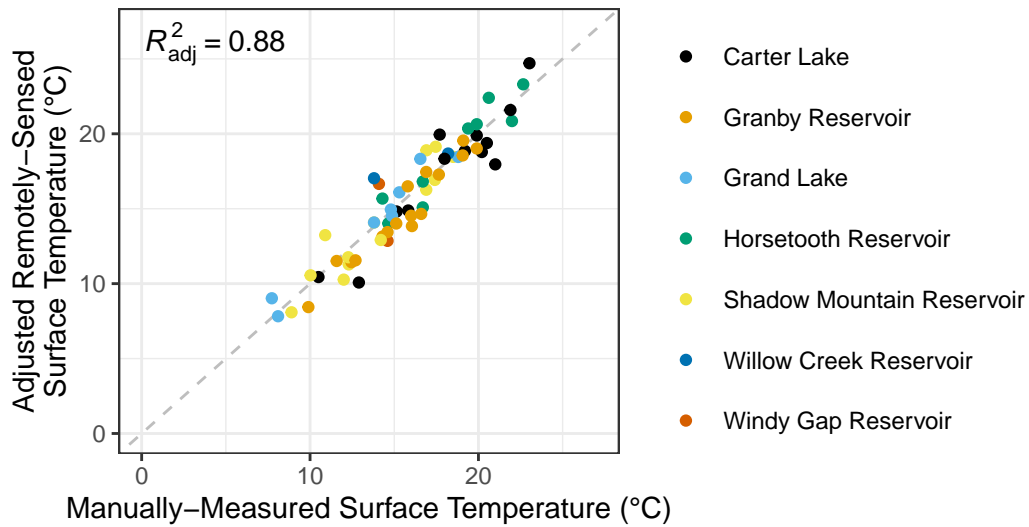


Figure 2: Preliminary calibrated Landsat-derived surface temperature where same-day *in situ* measurements available ( $n = 63$ ). Note obvious outliers at Willow Creek Reservoir (dark blue, Carter Lake(black), Windy Gap Reservoir (red) that require further investigation into QA/QC before integration into decision support system pipeline.

## Data preprocessing

All NLDAS precipitation data are right skewed heavily biased to low precip values including zero, to account for this and make the distribution more normal, I added 0.0001 to each value and applied a log-10 transformation to this subset. The wind data were left skewed and to transform the distribution, I used a square root transformation. All features and labels were then scaled using the mean and standard deviation from the training dataset to get the values closer around zero, which are preferable for neural networks. This is the same preprocessing that was used in the neural networks assignment.

## Training-Validation-Test Splits

For the autoencoder implementation, I used a leave-one-out training/validation split, where each lake was iteratively left out of the model development, creating 6 models. The test data are those from 2021 forward.

For development of the threshold cutoff, I used the instances of same-day *in situ* measurements and remote sensing values shown in Figure 2. For training-validation, I'll use the same leave-one-out technique, and if the threshold is similar across training/validation datasets, I will unify the data and create a more robust threshold model with all data. There will be no 'test' set in this instance because of the small amount of data. Testing for this example would be through implementation of this VAE-defined threshold method.

## Autoencoder Setup

The set up of this VAE was very conservative, using two fully connected dense layers for both the encoder and decoder with three codings for the latent space. Batch size was set to 32. The dense layers for the encoder were comprised of 20 then 10 nodes, the decoder comprised of 10 then 20 nodes all with *relu* activation. Input and reconstruction (output) layers contained 33 nodes, one for each standardized variable value. Linear activation was used for the reconstruction layer. To reduce overfitting, I also added a dropout layer with a dropout rate of 20% between the input and first dense layer.

Loss function for this VAE was the sum of the variational autoencoder's latent loss:

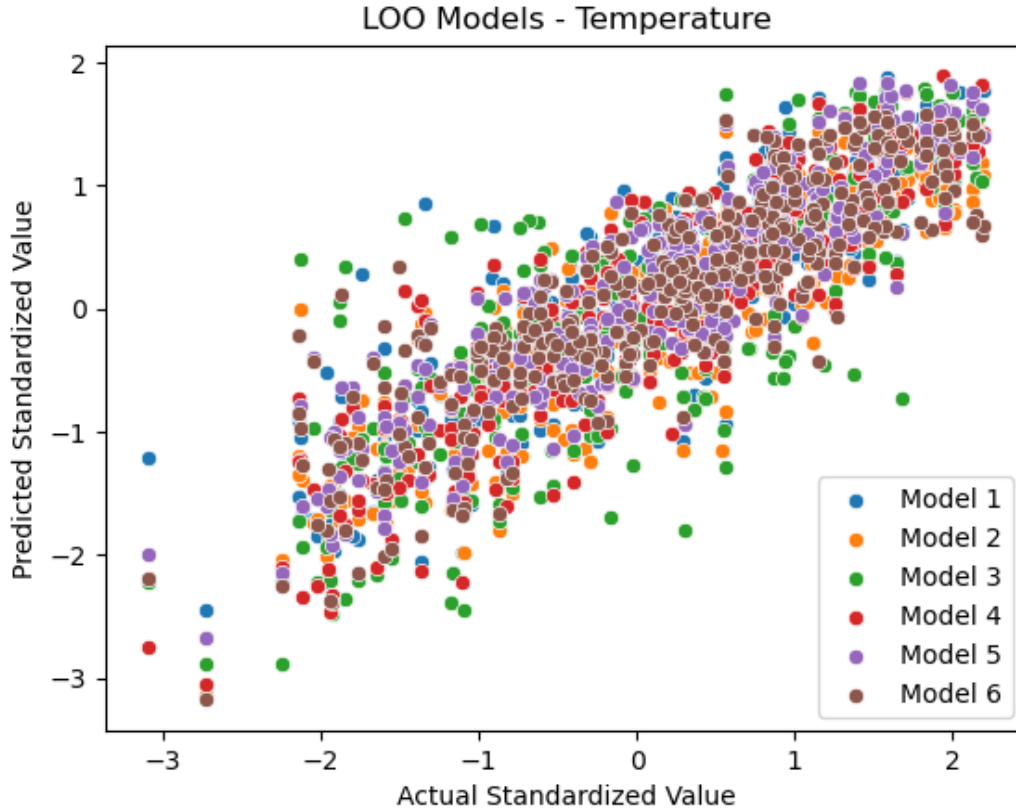
$$\mathcal{L} = -\frac{1}{2} \sum_{i=1}^n \left[ 1 + \log(\sigma_i^2) - \sigma_i^2 - \mu_i^2 \right]$$

And reconstruction loss, which was mean squared error (MSE):

$$MSE = \sum_{i=1}^n (x_i - \hat{x}_i)^2$$

## Results

To determine if the model was performing as expected, I will examine the temperature values in the reconstruction for the test dataset. While the other parameters are important to the modeling effort, determining temperature outliers is the purpose of the implementation of the VAE. The reconstructed temperature values should be strongly linearly related to the input temperature values. Figure ?@fig-VAEtemp shows the resulting reconstructed, standardized temperature value (y) plotted against the actual standardized temperature value.



## Conclusions

In hindsight, I think adding a weight to the reconstruction loss to add more weight divergence in temperature would have been particularly useful. Sadly, I've completely run out of time

to make thoughtful conclusions! I am particularly interested to see how the VAE handles the outlier/inlier data and whether or not it does make a difference in the NN models. While I don't have time to add that here, I will send an update if/when I get to the point where I have actually gone through the entire application of the VAE for outlier detection.

## Citations

- Cook, Monica, John R. Schott, John Mandel, and Nina Raqueno. 2014. "Development of an Operational Calibration Methodology for the Landsat Thermal Data Archive and Initial Testing of the Atmospheric Compensation Component of a Land Surface Temperature (LST) Product from the Archive." *Remote Sensing* 6 (11): 11244-11266. <https://doi.org/10.3390/rs6111244>.
- Sinha, Saumya, Sophie Giffard-Roisin, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, and Claire Monteleoni. 2021. "Variational Autoencoder Anomaly-Detection of Avalanche Deposits in Satellite SAR Imagery." In, 113119. CI2020. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3429309.3429326>.