Appendix S2: Landsat Filters

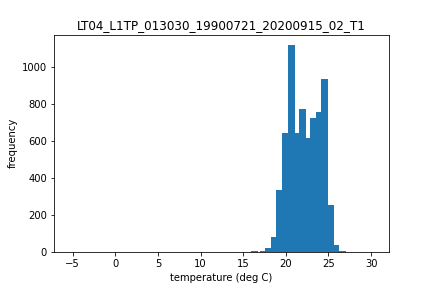
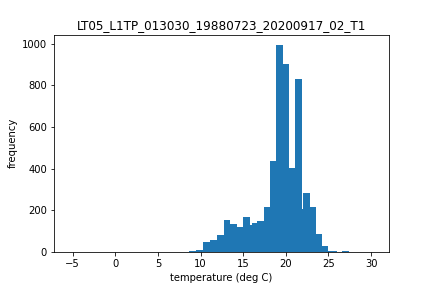
lakeCoSTR: An open-source, interactive retrieval tool to facilitate use of the Landsat Collection 2 surface temperature product to estimate lake surface water temperatures

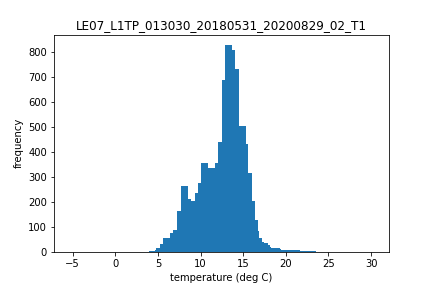
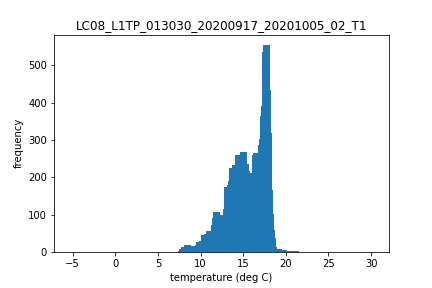
*Herrick, C, Steele, BG*, Brentrup, JA, Cook, B, Cottingham, KL, Ducey, M, Johnson, K, Lutz, DA, Palace, M, Sullivan, F, Thompson, M, Trout-Haney, JV, Weathers, KC

Ecosphere

# Background:

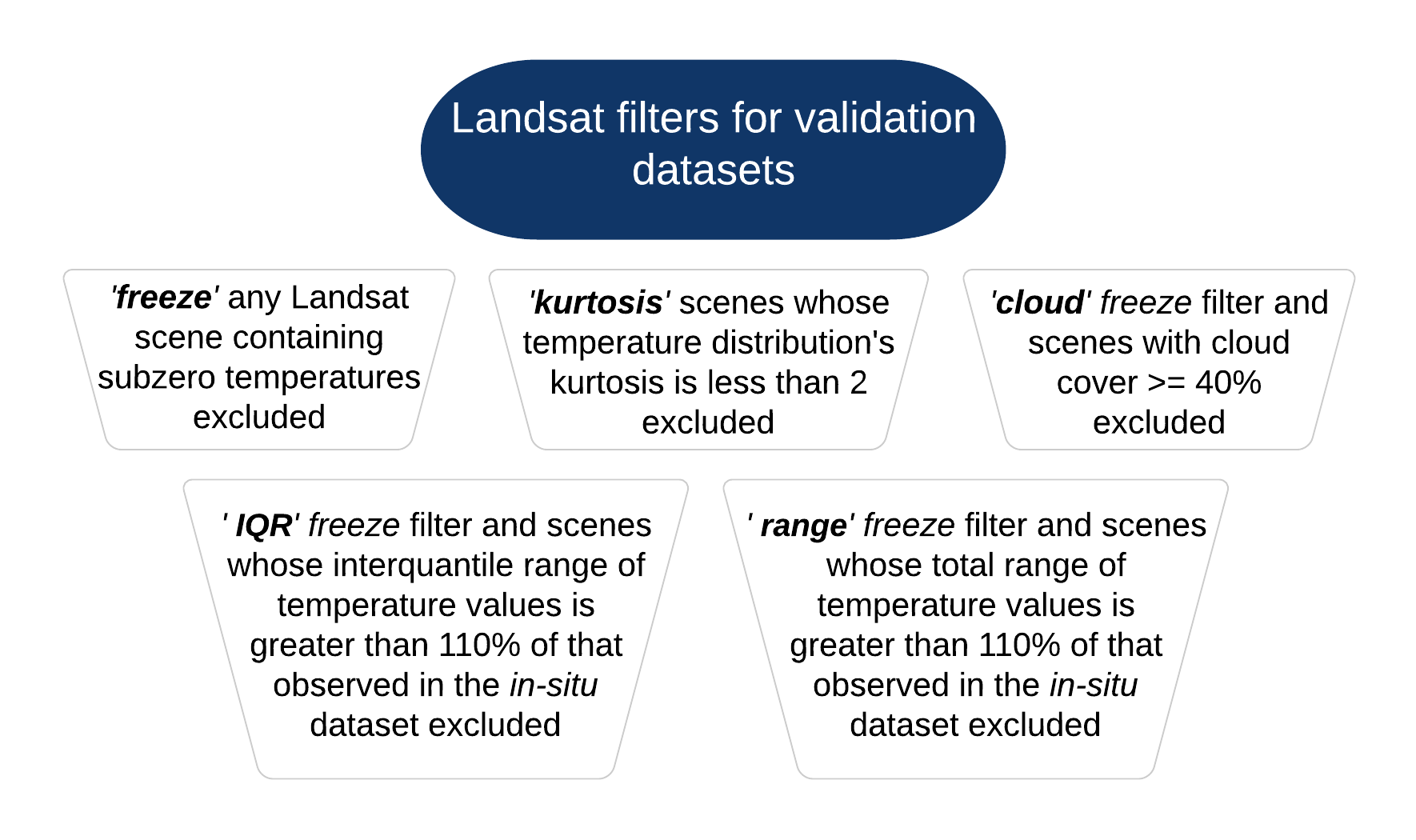
During our preliminary work to evaluate the Landsat Collection 2 surface water temperature data relative to in situ measurements for Lake Sunapee, NH, we looked carefully at the histograms of pixel-level data exported by lakeCoSTR for each Landsat scene. It was clear that some data distributions seemed ecologically unfeasible, displaying a range of 20 degrees Celsius or more over the surface of Lake Sunapee on a single flyover date (Figure S1). We assumed that these types of spread were indicative of lingering atmospheric correction issues or cloud effects, and that additional quality assurance filters were needed. In this appendix, we describe some of the filters we tried that may be helpful for others using this tool.

**Figure S1. Four histograms, one from each Landsat Mission, exhibiting unusual frequency distributions as exported from the lakeCoSTR tool.**

In addition to the possible range-of-value issues shown in Figure S1, there is also evidence of multi-modal distributions. Because we were interested in using a single median value to describe the lake surface temperature in a scene, we determined these scenes were not suitable for our analysis. Note that bimodal distributions in surface temperature across a lake surface are certainly ecologically possible outcomes and may be true for some systems or useful for some analyses. Finally, there were also scenes with negative temperatures even though pixels classified as snow or ice were filtered out in the Colab script.

 **Figure S2. The five filters we applied to the Landsat Collection 2 surface water temperature data to improve the Deming regression models relating Landsat - in situ data.**

Our desire was to create a filter that would remove the Landsat scenes with presumed atmospheric interference (indicated by grossly large estimated temperature ranges) or that were otherwise not suitable for our analysis. We used measurements from the extensive \*in situ() data network at Lake Sunapee to define some of these filters, as well as statistical measures of distribution, including quartile values, spread, and kurtosis (the ‘tailedness’ of the distribution). While we filter for clouds and cloud shadows in Google Earth Engine using the bitmask metadata at a pixel-level, there is documented interference of clouds and cloud shadows with the surface temperature product (Cook et al. 2014). Because the potential for pixels with cloud contamination is higher in scenes with greater cloud cover, we also tried a cloud filter to eliminate data where the scene-level metadata indicated cloud cover greater than 40%.

# Methods

The filters we explored, listed in order of increased stringency based on the number of scenes eliminated from analysis, were:

* **freeze**: removing all scenes whose minimum temperature was below 0 degrees Celsius
* **IQR**: removing scenes that reported interquartile temperature ranges greater than 110% in a summary of the high-frequency *in-situ* temperature record used in the main text
* **kurtosis**: removing any scenes whose histogram has a kurtosis value less than 2
* **cloud**: removing scenes with cloud cover greater than 40%
* **range**: removing scenes that reported temperature ranges greater than 110% observed in a summary of the high-frequency *in-situ* temperature record used in the main text

## Load, summarize, and filter data

To define some filters, we used the validation dataset described in section 3.1.2 of the main text, filtered to those temperatures measured between the months of May and October and the hours of 9 and 11 am (the approximate time of Landsat flyover). These values were aggregated to daily values of range, interquartile range, and number of locations contributing the ranges. From these daily values, we calculated the maximum range and interquartile range from the observed temperature data.

## [1] "Maximium spread observed is:"

## [1] 9.19

## [1] "Maximium interquartile range observed is:"

## [1] 2.6575

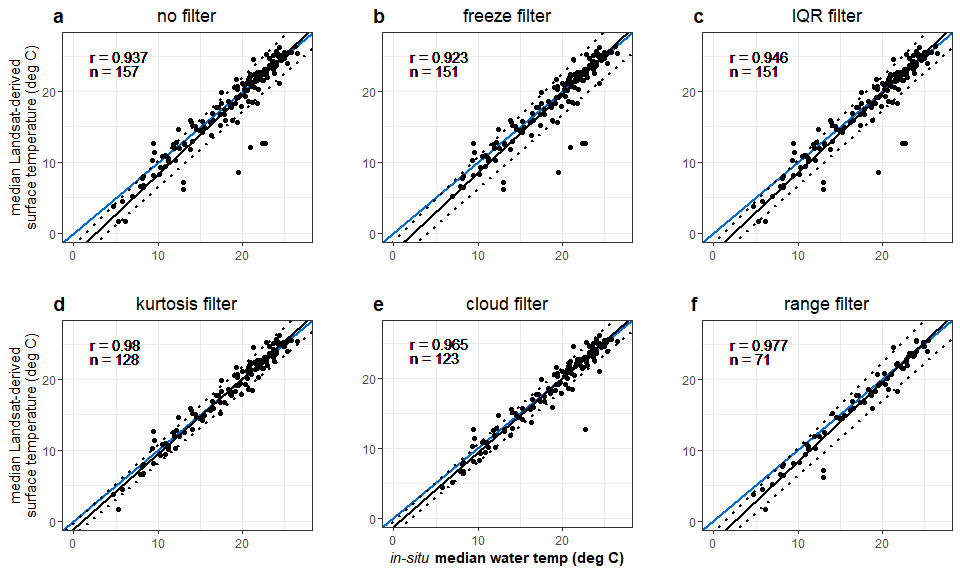
# Presentation and discussion of filter performance

To analyze the performance of each of the filters, we performed a Deming regression (Deming, 1943) on each of the Landsat *in situ* pair datasets filtered, as described above, from the output of the lakeCoSTR tool.

The considerations we made when comparing filter performance were:

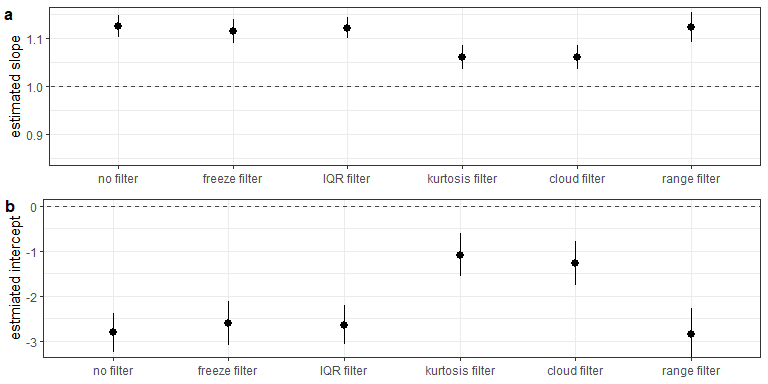
1. number of valid scenes for validation
2. Pearson correlation coefficient
3. presence/absence of outliers
4. slope and intercept of the regression line

Scenes for validation ranged from 157 (no additional QAQC) to 71 (*in situ* range filter). All datasets, including the unfiltered Collection 2 dataset, had acceptable Pearson correlation coefficients above 0.9 (Figure S3). The filters that reduced the outliers were the kurtosis filter, cloud filter, and the range filter. Of those, the kurtosis filter performed the best at reducing outliers. Of note, most outliers lie below the 1:1 line (Figure S3) - the presence of clouds and cloud shadows lowers the median temperature for each scene by reporting surface temperature values that are much cooler than in actuality.



**Figure S3. Deming regressions for 6 differently filtered datasets from the Landsat Collection 2 surface temperature product, arranged from least stringent to most stringent based on the number of scenes eliminated. Within each panel, the blue line is the 1:1 line, the black line shows the Deming error-in-variables regression, and the dotted black lines indicate the 95% prediction intervals for the Deming regression. The Pearson correlation (r) and sample size (n) are indicated in the upper left corner.**

While increasing the stringency of Landsat scene filters generally led to increased agreement between the Landsat-derived surface temperature product and the *in situ* water temperature (Figure S4), it is notable that the filter that eliminated the most scenes from the analysis (“range”) has an estimated slope and intercept furthest from 1 and 0, respectively. However, the kurtosis filter and the cloud filter performance was very similar, which may indicate that both filters perform similar functions.

 **Figure S4. A simple comparison of estimated slope (a) and intercept (b) with upper and lower 95% confidence intervals for each of the Collection 2 (raw + filtered) datasets presented in Figure C indicates that the cloud filter dataset’s Deming regression provided a slope closest to 1 and an intercept closest to 0 (indicated by the dashed line on each panel). Datasets are ordered by increasing stringency, as measured by the number of scenes in the validation dataset, from left to right.**

Because the two filters that seemed the most promising were the kurtosis and cloud filters, we carried out all analyses described in the main text with both filters. We chose the kurtosis filter after this exercise, because it removed a number of outliers in the long-term monthly temperature analysis that the cloud filter did not. We suspect that kurtosis, a measure of tailedness, was a successful filter because it incorporates shape and range in a single value, and we suspect these are the symptoms of cloud cover for this system. Additionally, the kurtosis filter resulted in a dataset that did not require calibration for Lake Sunapee. Importantly, though, the kurtosis filter threshold (we used a value of 2) may have to be changed for other systems - analyses at other lakes that have *in situ* data for ground truthing will be needed to determine if the value we chose is transferable.

# Literature Citations

Cook M, Schott JR, Mandel J, Raqueno N. 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. 2014; 6(11):11244-11266. <https://doi.org/10.3390/rs61111244>

Deming, W. E. 1943. Statistical adjustment of data. Dover Publications, 1985.