*Supplement of manuscript on*

# A global seamless 1 km resolution daily land surface temperature dataset (2003-2020)

**Tao Zhang, Yuyu Zhou, Zhengyuan Zhu, Xiaoma Li, Ghassem R. Asrar**

Correspondence to: Yuyu Zhou (yuyuzhou@iastate.edu)

## S1 Interpolation based on Correlation Weighting (ICW)

An Interpolation based on Correlation Weighting (ICW) technique was used to interpolate the residual of land surface temperature (LST) in each day of a year. This method was inspired by the Inverse Distance Weighting (IDW) interpolation method. The IDW method uses the weighted average values of neighboring sites to estimate the missing value, where the weight was calculated based on the inverse distances between target site and its neighboring sites. In the ICW method, the weight between target site and one of its neighboring sites was calculated based on the correlation between the time series of the pairs (days of a year) LSTs of the two locations.

The missing value of the target site $V_{S_0}$ at the time *t* was estimated based on values of the neighboring sites with Eq. (S1).

$$V_{S_0}(t) = \sum_{i=1}^{i=n} w(S_0, S_i) \cdot V_{S_0}(S_i, t) \quad \text{(S1)}$$

where $V_{S_0}(t)$ is the estimated value of target site at the time *t*; $w(S_0, S_i)$ is the weight of the *i*-th neighboring site $S_i$, which can be calculated with Eq. (S2); $V_{S_0}(S_i, t)$ is the estimated value of the target site at the time *t* based on the *i*-th neighboring site, which can be estimated with Eq. (S3); and *n* is the number of neighboring sites.

$$w(S_0, S_i) = \frac{cor(S_0, S_i)}{\sum_{i=1}^{i=n} cor(S_0, S_i)} \quad \text{(S2)}$$

where $w(S_0, S_i)$ and *n* are the same with those in Eq. (1); $cor(S_0, S_i)$ is the Pearson's correlation coefficient between $S_0$ and $S_i$.

$$V_{S_0}(S_i, t) = \alpha_i + \beta_i \cdot V_{S_i}(t) \quad \text{(S3)}$$

where $\alpha_i$ and $\beta_i$ are the intercept and slope of the linear function between target site and the *i*-th neighboring site, which was fitted using ordinary least square (OLS) method based on the matched time series (days of a year) of LST in the two locations; $V_{S_i}(t)$ is the value of the *i*-th neighboring sites at time *t*.

## S2 Implementation of the ICW method

The ICW method was implemented as follows. First, in order to improve the efficiency, each MODIS tile was divided into blocks with a size of 10 by 10 pixels, and the block center pixels were used as neighboring pixels for interpolating missing residuals. That is, missing residuals in a block can be interpolated based on the values from the 8 neighboring block center pixels. Second, in order to ensure that all the block center pixels have valid (good quality) data for the estimation of other pixels, the missing values in the block center pixels were interpolated using the IDW method. The steps used in this process are: (a) computing the average value of each block; (b) resampling the original MODIS tile of 1200 by 1200 to 120 by 120 and the value of each pixel in the new image is the average value of a block in the original MODIS tile; (c) interpolating missing values in the resampled image based on

the IDW method; and (d) assigning interpolated values to the block center pixels without valid values in the original MODIS tile. Third, in order to reduce the possible boundary effects of the interpolated residuals between neighboring blocks, for each pixel of a block, one of the neighboring block center pixels, which has the largest correlation coefficient with the target pixel was used for estimation. This process can avoid systematic deviation in the boundary pixels from different blocks that were estimated based on different combination of block center pixels, because all the pixels in a block were interpolated based on eight center pixels of neighboring blocks. Eqs. (S1) and (S3) can be simplified to Eq. (S4).

$$V_{S_0}(t) = \alpha_m + \beta_m \cdot V_m(t) \qquad \text{(S4)}$$

where $\alpha_m$ and $\beta_m$ are the intercept and slope of the linear function between target pixel and its neighboring pixel with the corresponding maximum correlation coefficient, which was fitted using the ordinary least square (OLS) method based on the matched time series (days of a year) values of the two locations; and $V_m(t)$ is the value of the neighboring pixel with the maximum correlation coefficient at the time *t*.

Finally, in order to mitigate boundary effects between neighboring tiles, multiple neighboring tiles were mosaicked as a region, and residual of the block center pixels in the region were interpolated at the same time. The overlapped areas between the two neighboring regions were also considered to avoid possible boundary effects. The interpolation was conducted following the order of the region ids. For example, as the id of North America is 1, which was the first region to be processed.

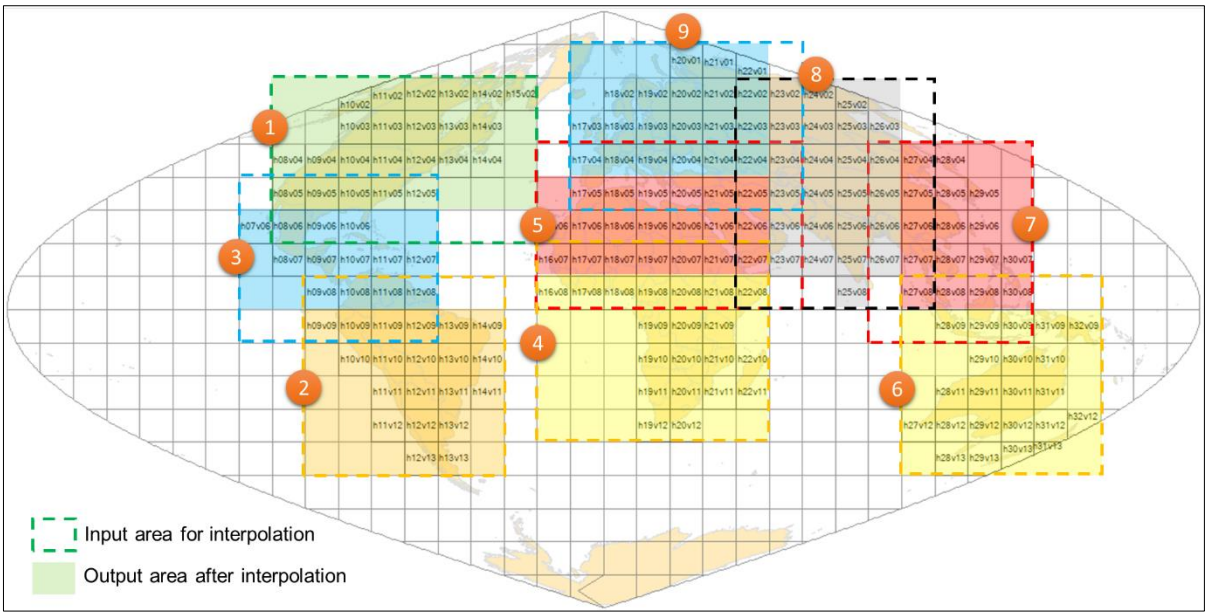


**Figure S1: Division of global regions. Dashed and shaded rectangles indicate the extent of input data and output data, respectively.**

## S3 Spatial patterns of gap-filled LST between neighboring regions

There are no significant boundary effects between neighboring regions as shown in the example data of gap-filled LST (Fig. S2), indicating that the division of regions with overlapped areas is useful (Sect. S2). For example, the spatial pattern of gap-filled LST based on the proposed gap-filling method has no sudden changes on the boundary of region 1 and region 3 (subregion A in Fig. S2), and region 2 and region 3 (subregion B in Fig. S2) (Fig. S1). This is because when interpolating the residual of center blocks near the boundary (Sect. S2), the valid residual values from the two neighboring regions were used, mitigating boundary effects in the finally gap-filled LST data.

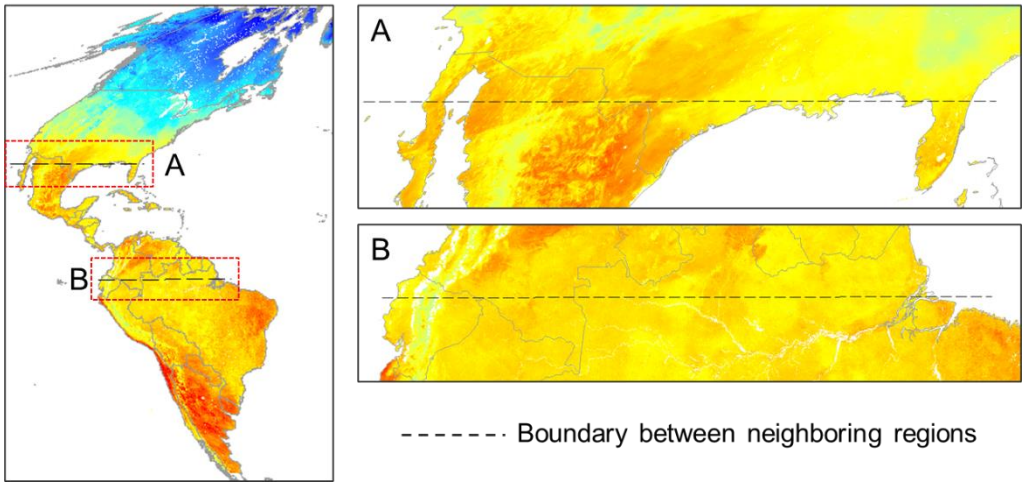


**Figure S2: Spatial pattern of gap-filled daytime LST around the boundary of two neighboring regions in America.**

## S4 Spatial patterns of the gap-filled LST between neighboring tiles

There are no significant boundary effects between neighboring tiles as shown in the example data of gap-filled LST (Fig. S3), indicating the interpolation of residuals for block center pixels in a region with multiple tiles is useful (Sect. S2). There are obvious boundary effects between tiles h10v04, h11v04, h10v05, and h11v05 (Fig. S3(a)) with these tiles gap-filled independently; on the contrary, there are no boundary effects when the residuals of the whole region were interpolated at the same time (Fig. S3(b)). When the residuals for block center pixels were calculated independently for each tile, systematic differences may occur near the boundary of the two neighboring tiles especially when there are a lot of missing values, leading to boundary effects in the interpolated residuals and gap-filled LST data.

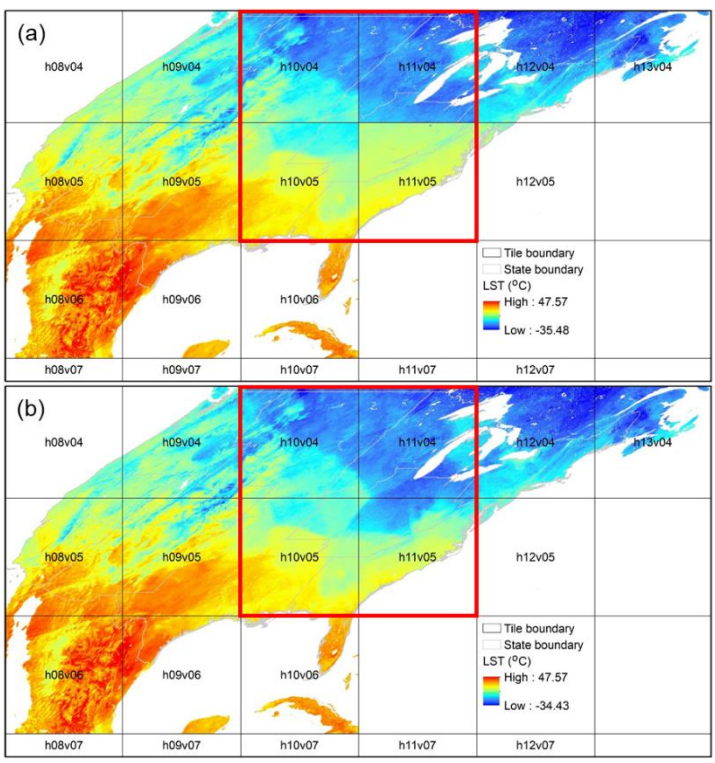


**Figure S3: Spatial patterns of gap-filled daytime LST without (a) and with (b) interpolating center pixels of multiple tiles on 20th day of 2012.**