

**Assessing gap-filled Landsat land surface temperature time series data using different observational datasets**

Journal:	<i>International Journal of Remote Sensing</i>
Manuscript ID	TRES-PAP-2024-1355
Manuscript Type:	IJRS Research Paper
Date Submitted by the Author:	09-Aug-2024
Complete List of Authors:	Shi, Hua; ASRC Federal Data Solutions - Contractor to USGS EROS Xian, George; USGS EROS
Keywords:	surface temperature, data fusion, landsat
Keywords (user defined):	temporal frequency, accuracy assessment, uncertainty analysis

SCHOLARONE™  
Manuscripts

1  
2  
3 **August 08, 2024**  
4

5 **Dear Dr. Peng,**  
6

7 **We have revised the manuscript, “Assessing gap-filled Landsat land surface temperature**  
8 **time series data using different observational datasets” (original title is “Assessing the**  
9 **accuracy of gap-filled land surface temperature time series for surface urban heat island**  
10 **study”) in accordance with your guidance on April 02, 2024. We reorganized our**  
11 **manuscript by following the IJRS’s instruction for authors. We responded to comments**  
12 **from two reviewers, providing point by point responses, and made the significant changes**  
13 **in text (see below). Since the topic of this manuscript is a high-profile topic within the U.S**  
14 **Department of Interior, we will need to know whether the manuscript will be accepted by**  
15 **your journal. Please ensure we are notified as soon as possible of your decision. If you have**  
16 **any questions about our revised submission, please contact us.**  
17  
18

19 **Thanks,**  
20

21  
22 **Hua Shi PhD,**  
23

24 **Research Ecologist/Geographer**  
25

26 **AFDS, Contractor to the USGS EROS Center**  
27

28 **47914 252nd Street | Sioux Falls, SD, USA 57198-0001**  
29

30 **605.594.6050 | [hshi@contractor.usgs.gov](mailto:hshi@contractor.usgs.gov)**  
31

32 **ORCID iD: <https://orcid.org/0000-0001-7013-1565>**  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## The response to referee (s). “R:” is our responses.

Referee(s)' Comments to Author:

Referee: 1

Comments to the Author

In this manuscript, titled ‘Assessing the accuracy of gap-filled land surface temperature time series for surface urban heat island study’, the authors presented the validation of a gap-filling method for Landsat LST data. The validation is based on several datasets, considering the NOAA Global Historical Climate Network (GHCN) air temperatures as the main reference. Although the evaluation of a well established method is always of interest to the community, since it reinforces the quality of the method and, thus, justify its use and its quality, the methods used in this research paper are not the most appropriate for the assessment of an LST dataset. Although some previous papers with validated LST products temperature data, that is not a good practice since the reference dataset is based on a completely air different magnitude which only shows similar values to those of LST under specific atmospheric conditions and times (see Best Practice protocol for LST validation; Guillevic et al. 2018: [https://gcc02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpvs.gsfc.nasa.gov%2FPDF%2FCEOS\\_LST\\_PROTOCOL\\_Feb2018\\_v1.1.0\\_light.pdf&data=05%7C02%7Cshsi%40contractor.usgs.gov%7C1eabf08df27d4ebbd42408dc530289ce%7C0693b5ba4b184d7b9341f32f400a5494%7C0%7C0%7C638476517568471560%7CUnknown%7CTWFpbGZsb3d8eyJWljiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTil6Iik1haWwiLCJXVCi6Mn0%3D%7C0%7C%7C%7C&sdata=w31x5kZhlmWL8IH0EfnYztjFV4kHAir7GMAIZEpNyQ%3D&reserved=0](https://gcc02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpvs.gsfc.nasa.gov%2FPDF%2FCEOS_LST_PROTOCOL_Feb2018_v1.1.0_light.pdf&data=05%7C02%7Cshsi%40contractor.usgs.gov%7C1eabf08df27d4ebbd42408dc530289ce%7C0693b5ba4b184d7b9341f32f400a5494%7C0%7C0%7C638476517568471560%7CUnknown%7CTWFpbGZsb3d8eyJWljiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTil6Iik1haWwiLCJXVCi6Mn0%3D%7C0%7C%7C%7C&sdata=w31x5kZhlmWL8IH0EfnYztjFV4kHAir7GMAIZEpNyQ%3D&reserved=0)).

**R: Thanks. However, we disagree with Referee 1’s comments regarding the use of validated LST products temperature data. Referee 1 stated, “Although some previous papers have validated LST products with temperature data, that is not a good practice since the reference dataset is based on a completely different magnitude of air temperature, which only shows similar values to those of LST under specific atmospheric conditions and times.” We believe Referee 1 may have misunderstood why we use GHCN station observation air temperature and how we evaluated the gap-filled LST. Here is our explanation:**

1. There are no field observation LST data available for evaluating gap-filled LST in our research, so using GHCN station data is a good option.
2. There is always debate about using air temperature to validate LST. While it may not be the most appropriate to use air temperature to assess Landsat-derived LST directly, but there is a strong relationship and spatial pattern between air temperature and LST at the same location and date.
3. We do NOT use GHCN station data to assess Landsat derived LST directly. we used GHCN air temperature as a baseline to compare its relationship with original clear Landsat LST and gap-filled LST to evaluate the gap-filled LST products whether could be used for SUHI research. For instance, the results of the statistical analysis show that the original LST and gap-filled LST are comparable or even better. This indicates that

1  
2  
3 **gap-filled LST products can be used to further SHUI research. Our manuscript stated**  
4 **this conclusion.**

- 5  
6 4. **For comparison, gap-filled LST was evaluated against existing remote sensing-derived**  
7 **LST from MODIS, VIIRS, and ECOSTRESS, which have different spatial resolutions. We**  
8 **used GHCN station data as a reference, averaging at 9 Landsat pixels (30 meters) to**  
9 **match the spatial resolution, to assess the performance of the gap-filling models.**

10  
11  
12  
13  
14 On the other hand, the manuscript is confusing since its title. It is mentioned in the title and several  
15 times during the manuscript, especially at the end of each section, that the assessed method is for  
16 SUHI studies. Although the authors are right that these data could be used for SUHI analysis, it is not  
17 as relevant as to be mentioned constantly, and in any case make sense to include it in the title and  
18 the abstract. I would suggest to remove that references to SUHI or, at least, include a small study  
19 case of the implementation of the data to SUHI, what in my opinion is out of the scope of this  
20 manuscript.  
21

22  
23 **R: Thanks. We changed title to “Assessing gap-filled Landsat land surface temperature**  
24 **time series data using different observational datasets”. We made some change in text. Our**  
25 **gap-filled LST products will be used for further SUHI research.**

26  
27  
28  
29 The hypothesis considered in the introduction should be deeply thought, they sound quite superfluous,  
30 e.g., what is ‘expert opinion’?, or in reference to hypothesis 3, the accuracy of LST is often assessed  
31 comparing with ground temperatures. Hypothesis 4 is also not clearly answered in the manuscript. The  
32 objectives are also unclear or not fully related with the overall manuscript.  
33

34 **R: Thanks. We deleted hypothesis 3 and made some changes.**

35  
36  
37  
38 The method section should be more detailed. The gap-filling method is not clearly explained. A  
39 suggestion would be to add a flowchart which helps to understand the processing flow. In page 13 I  
40 cannot understand how temperature and SUHI intensity can be estimated from digital camera  
41 photographs. Please, add more details.  
42

43  
44 **R: This manuscript focuses on the accuracy assessment and uncertainty analysis of gap-**  
45 **filled LST products generated by gap-filling models. We have previously published a paper**  
46 **that provides a detailed discussion of gap-filling methods, which is referenced in Section**  
47 **2.3. Therefore, we believe it is unnecessary to explain the gap-filling methods again in this**  
48 **manuscript.**

49  
50 **Published paper:**

51  
52 **Zhou, Q., Xian, G., & Shi, H. (2020). Gap fill of land surface temperature and reflectance products in**  
53 **Landsat analysis ready data. *Remote Sensing*, 12**

1  
2  
3 Other minor comments:  
4

5 Several times is mentioned that MODIS / VIIRS weekly composites have better quality than daily data. I  
6 can see your point, but that really depends on your purpose and the use of your data. E.g., on a strict  
7 validation of daily and weekly LST data, the former would provide better results. Thus, I would suggest  
8 to change that expression and clarify that it is preferred to use weekly composites to avoid or minimize  
9 the influence of potential cloud effects in the image.  
10

11 **R: Thanks. We made some improvement in text.**

12 I am also wondering which is the effect of under-cloud gap-filling. In Figure 3, the gap-filling LSTs are  
13 always above the other sources LSTs. Why they are not lower if the gap is due to a cloud effect. I would  
14 expect to have cooler surface temperatures for those situations.  
15

16 **R: Good point. As mentioned above, we have a detailed discussion about gap-filling models  
17 and their products. This seasonal training strategy uses clear observations as training data  
18 and employs linear regression models to predict missing (or cloud/snow-covered) LST  
19 pixels. However, we stated the limitations of our gap-filling models. We plan to improve  
20 these models with better training strategies in future work.**  
21

22 In Page 20, line 70: What do you mean with the spectral behavior of LST? LST is not a spectral  
23 magnitude.  
24

25 **R: thanks. There is no line 70 in page 20.**

26 Please, review criteria when referring to figures: Figure 1. or Fig. 1.  
27

28 **R: thanks. Fixed.**

29 Add units in tables where required, most of them are missing.  
30

31 **R: thanks. Fixed. Note: R2 and RMSE don't have units.**

32 The labels in figures are not readable on paper and hardly readable on digital format.  
33

34 **R: thanks. We reproduced figures with readable labels.**  
35

36 State if you are using MOD11 or MOD21 LST data. If it is the former, then the reference Hulley and Hook  
37 (2021) is not the most appropriate for page 10 line 22. I would rather suggest the original paper for the  
38 improvements in MOD11 C6 for bare soil.  
39

40 **R: We added new reference:**

41 **Duan, S.-B., Li, Z.-L., Cheng, J., & Leng, P. (2017). Cross-satellite comparison of operational  
42 land surface temperature products derived from MODIS and ASTER data over bare soil  
43 surfaces. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126, 1-10**  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Referee: 2

4  
5 Comments to the Author

6  
7 It is an interesting topic and fall within the scope of IJRS.

8  
9 But this paper is hard to follow, and needs to be improved to meet the standard of IJRS. My comments  
10 are listed below,

11  
12 Major comments

13  
14 1. Although the introduction is very detailed, its excessive length may hinder readability. It would be  
15 beneficial to streamline this section to ensure a more concise and reader-friendly presentation.

16  
17 **R: Thanks. We improved the introduction. This version retains the key points and**  
18 **objectives of the study while making the introduction more concise and reader friendly.**

19  
20 2. There is significant content repetition between lines 55 of page 5 to line 20 of page 6, and lines 37 to  
21 52 of page 5.

22  
23 **R: Thanks. Reorganized and removed repetition sentences.**

24  
25 3. It's important to note that air temperature and LST represent two distinct parameters, and it is not  
26 accurate to use near-surface air temperature as a proxy for LST due to inherent differences between  
27 these two measurements. The statement "Another motivation for implementing the extracting strategy  
28 is that GHCN is field observation and air temperature but more accurate than other existing remote  
29 sensing derived LST." in section 2.2 suggests a direct comparison of accuracy between GHCN air  
30 temperature data and remote sensing-derived LST data. However, considering the fundamental  
31 distinction in what they measure, this comparison seems inappropriate.

32  
33  
34 **R: Here are the explanation: Why we use GHCN station observation air temperature and**  
35 **how we evaluated the gap-filled LST. Here is our explanation:**

- 36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- 1. There are no field observation LST data available for evaluating gap-filled LST in our research, so using GHCN station data is a good option.**
  - 2. There is always debate about using air temperature to validate LST. While it may not be the most appropriate to use air temperature to assess Landsat-derived LST directly, but there is a strong relationship and spatial pattern between air temperature and LST at the same location and date.**
  - 3. We do NOT use GHCN station data to assess Landsat derived LST directly. we used GHCN air temperature as a baseline to compare its relationship with original clear Landsat LST and gap-filled LST to evaluate the gap-filled LST products whether could be used for SUHI research. For instance, the results of the statistical analysis show that the original LST and gap-filled LST are comparable or even better. This indicates that gap-filled LST products can be used to further SHUI research. Our manuscript stated this conclusion.**
  - 4. For comparison, gap-filled LST was evaluated against existing remote sensing-derived LST from MODIS, VIIRS, and ECOSTRESS, which have**

1  
2  
3 **different spatial resolutions. We used GHCN station data as a reference,**  
4 **averaging at 9 Landsat pixels (30 meters) to match the spatial resolution, to**  
5 **assess the performance of the gap-filling models.**  
6

7  
8 As acknowledged in section 4 regarding the limitations of this study, "Additionally, NOAA GHCN records  
9 air temperatures that are different from gap-filled and other LST datasets." Given that both Landsat LST  
10 and LST datasets used in the comparison analysis have their accuracy quantified based on GHCN air  
11 temperature, the rationale for using GHCN data as a cornerstone reference for remote sensing-derived  
12 LST requires further clarification to justify its appropriateness.  
13

14 **R: Thanks. Good points. We added sentences to justify its appropriateness.**  
15

16  
17 4. The length of section 4 is extensive, with some content repetition observed. It is advisable to  
18 streamline this section for conciseness. Consider separating the discussion and conclusion into distinct  
19 sections. Additionally, incorporating "3.3 Uncertainty analysis" as part of the experimental content  
20 within the discussion section could make the narrative more cohesive and logically structured.  
21

22 **R: Thanks for your suggestions. We separated discussion and conclusion into distinct**  
23 **sections.**  
24

25  
26  
27 Minor comments

28  
29 1. It is recommended to add latitude and longitude grids to Figures 1 and 2, and to supplement the color  
30 bar legends of Figures 1 and 2 with the unit "°C" for clarity.  
31

32 **R: Thanks. We reproduced Figure 1 by using Land cover classes and GHCN station**  
33 **locations with legend, the detailed information about each GHCN station see the Appendix**  
34 **1. We also added latitude and longitude grids for Figure 1 and 2.**  
35

36  
37 2. Please supplement the selected ground station points with their latitude and longitude information to  
38 enhance the geographical context.

39 **R: We reproduced Figure 1 with Land cover types and NOAA GHCN station's location**  
40 **with ID. Added appendix 1 for these NOAA GHCN stations detailed information including**  
41 **Latitude and longitude.**  
42

43  
44 3. Figures 3.1, 3.2, and 3.3 reveal some exceptionally low outliers in the original Landsat LST data. It is  
45 recommended that the authors clarify whether these outliers have a non-negligible impact on the Gap-  
46 Filled LST results.  
47

48 **R: Thank you. Figures 3.1, 3.2, and 3.3 (currently Figure 5, 6, and 7) reveal exceptionally**  
49 **low outliers in the original Landsat LST data. These data points predominantly occur**  
50 **during winter, as evidenced by snow/ice cover and their limited occurrence. Their impact**  
51 **on the Gap-Filled LST results is non-negligible due to our use of a seasonal training**  
52 **strategy.**  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 **Assessing gap-filled Landsat land surface temperature time series data**  
4 **using different observational datasets**  
5  
6  
7

8 Hua Shi<sup>a\*</sup> and George Xian<sup>b</sup>  
9

10  
11 *<sup>a</sup>ASRC Federal Data Solutions (AFDS), Contractor to the USGS Earth Resources*  
12 *Observation and Science Centre (EROS), Sioux Falls, SD 57198, U.S.A.; <sup>b</sup>USGS Earth*  
13 *Resources Observation and Science Centre (EROS), Sioux Falls, SD 57198, U.S.A*  
14  
15

16  
17 \*corresponding author: [hshi@contractor.usgs.gov](mailto:hshi@contractor.usgs.gov); Tel.: +01-605-594-6050  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Word count: 6338

## Assessing gap-filled Landsat land surface temperature time series data using different observational datasets

Landsat Analysis Ready Data (ARD)-based time series present challenges in monitoring surface urban heat islands (SUHI) due to rapid changes in land surface temperature (LST) compared to cloud-free satellite observations. This research is to investigate the use of a spatiotemporal gap-filling model as a feasible and cost-effective solution to produce Landsat time-series LST products with both high spatial resolution and temporal frequency. The study identified and filled Landsat ARD thermal times-series data gaps due to missing data, cloud and shadow effects, and data quality. The accuracy of Landsat gap-filled products was assessed using randomly selected clear observations of Landsat and uncertainty products from the gap-filling model, and was evaluated using various existing temperature datasets, including climate data from NOAA Global Historical Climate Network (GHCN) station observations, Daily Surface Weather and Climatological Summaries (DAYMET), and land surface temperature including MODIS, VIIRS, and ECOSTRESS. The result suggests that the gap-filled Landsat LST has significant correlations with existing datasets including field observation and remote sensing data derived from other sensors that have similar monthly and seasonal variation patterns. The uncertainty maps show spatial distributions of uncertainty for gap-filled pixels that have high or low uncertainties. The Landsat gap-filled time-series datasets can be used to measure annual, seasonal, or even monthly landscape thermal conditions, which are useful for SUHI and relevant research, and to perform multi-decade time series LST change analysis under climate change conditions.

Keywords: temporal frequency; gap-filling; Landsat surface temperature; uncertainty analysis, accuracy assessment

## 1. Introduction

Satellite images provide valuable geospatial data for characterizing thermal conditions and support ecosystem and environmental change monitoring, but standard satellite missions always have to live with the trade-offs between spatial resolution and temporal frequency (Luo, Guan, and Peng 2018; Roy and Yan 2020). Remote sensing techniques are increasingly used to combine data from multiple sensors and platforms to create high-quality data products (Bauer 2020). However, these data products may contain gaps due to missing or incomplete data (cloud cover or sensor malfunctions) that can reduce their accuracy and usefulness for scientific research and applications (Wulder et al. 2011). To overcome these limitations, gap-filling (or data fusion) models are widely used in environmental change monitoring, where continuous and accurate observations of environmental variables are crucial (Gao et al. 2006; Roy et al. 2008; Roy and Yan 2020; Zhou, Xian, and Shi 2020).

The growth of multi-sensor integrated datasets provides the opportunity to investigate land surface temperature (LST) dynamics and environmental changes at both high spatial resolution and temporal frequency, but also urges approaches to reduce the inconsistency of data availability. Many gap-filling approaches were developed for predicting missing values related to cloud contamination and Landsat 7 Scan Line Corrector (SLC)-off data (Chen et al. 2011; Yan and Roy 2018; Zhu et al. 2022). The main purpose of these models is to fill the gaps in time-series data caused by missing values, sensor failure, or other factors including clouds and cloud shadows, and to estimate the missing values with reasonable accuracy. The accuracy of gap-filling models is essential for environmental change monitoring, as inaccurate or biased estimates (Roy and Yan 2020) may lead to incorrect conclusions and decisions (Stehman et al. 2018; Stehman and Foody 2019; Wulder et al. 2022). To achieve

1  
2  
3 accurate gap-filling, several methods have been proposed, including interpolation,  
4 regression, and machine learning algorithms (Zhu et al. 2010; Zhu et al. 2016; Zhou et  
5 al. 2022). Despite the progress made in this field, gap-filling models still face  
6 challenges, such as uncertainty analysis and accuracy assessment, which need to be  
7 addressed to improve their performance for gap-filled products (Wang et al. 2022; Zhu  
8 et al. 2022).

9  
10 The accuracy of these gap-filling algorithms is crucial for ensuring the reliability and  
11 usefulness of the resulting data products (Nicolòs et al. 2011; Foody 2020). Accuracy  
12 assessment is the process of evaluating the agreement between the estimated values  
13 from gap-filling techniques and the true values. Uncertainty analysis is the process of  
14 quantifying the variability in the estimated values due to uncertainties in the gap-filling  
15 process (Rocchini et al. 2013). The accuracy assessment and uncertainty analysis of  
16 gap-filled products are necessary to understand the limitations of the data and to ensure  
17 that the data is suitable for use in environmental research and management (Rounsevell  
18 et al. 2021; Zhu et al. 2022).

19  
20 Uncertainty analysis of gap-filled LST products can also provide insights into the  
21 reliability of observed data and model outputs in urban environmental change  
22 monitoring, which is crucial for decision making and policy development (Rocchini et  
23 al. 2013). Accuracy assessment ensures that environmental change monitoring data is  
24 reliable and can be used for decision making purposes (Leyk et al. 2018).

25  
26 One of important environmental change monitoring is SUHI research. In urban heat  
27 island research, gap-filling models can be used to complete missing data points in  
28 temperature data, which is important for understanding the effects of urbanization on  
29 temperature patterns in temporally (Zhou, Xian, and Shi 2020). Uncertainty analysis can  
30

1  
2  
3 also help to quantify the uncertainty associated with temperature data and model  
4  
5 outputs, which is important for understanding the reliability of temperature data and for  
6  
7 making informed decisions about the impacts of urban heat islands on human health and  
8  
9 the environment (Rocchini et al. 2013). Accuracy assessment of gap-filled data is useful  
10  
11 for SUHI intensity and heat wave related analysis so that reliable SUHI information can  
12  
13 be used to develop effective public health strategies to mitigate the impacts of climate  
14  
15 change on human health. The main research problem associated with these models is  
16  
17 the input data and models' uncertainty, which can significantly affect the accuracy of  
18  
19 gap-filled data.  
20  
21  
22  
23  
24

25 In this study, we answered questions including what are the current methods used to fill  
26  
27 gaps in surface temperature data, what is the uncertainty of input data and models, what  
28  
29 is the accuracy of gap-filled products, and what are the limitations of our gap-filling  
30  
31 models and how can they be improved. We hypothesize that 1) the use of gap-filling  
32  
33 algorithms is a more effective method for filling gaps in Landsat thermal conditions  
34  
35 than traditional interpolation methods; 2) the uncertainty in surface temperature data can  
36  
37 be quantified using statistical models by incorporating prior knowledge; and 3) the use  
38  
39 of gap-filling models and uncertainty analysis can improve the accuracy of urban heat  
40  
41 island studies models inputs by reducing the bias and variance in surface temperature  
42  
43 datasets.  
44  
45  
46  
47  
48

49 With the hypotheses, our objectives were 1) to identify and analyse uncertainty of input  
50  
51 datasets and gap-filling models for surface temperature time series data; 2) to  
52  
53 investigate the impact of gap-filling on the surface temperature time series data; 3) to  
54  
55 compare the performance of the gap-filled surface temperature time series data with  
56  
57 field observation and other existing LST data for the validation of the accuracy and  
58  
59 reliability of the gap-filling technique(s); 4) to provide recommendations for the  
60

1  
2  
3 selection and application of gap-filling techniques for improved urban heat island study.  
4  
5  
6

## 7 **2. Materials and methods**

8

9  
10 We carried out this research through several steps. We first collected reference datasets  
11  
12 from various existing sources with multiple spatial resolutions and temporal frequencies  
13  
14 (Section 2.3). For each date of the time series within selected years, reference  
15  
16 temperature for each date was taken the same date (or a close date if data were missing)  
17  
18 as the gap-filled Landsat LST date (Section 2.4). These reference datasets provided the  
19  
20 basis for the accuracy estimates. Then, we conducted the accuracy assessment (Section  
21  
22 2.5) following protocols of consistent estimation required for a statistically rigorous  
23  
24 analysis. The statistical parameters of R-Square ( $R^2$ ) and Root Mean Square Error  
25  
26 (RMSE) were used. Finally, we analysed the uncertainty from gap-filling models with  
27  
28 input Landsat data and the uncertainty from comparison datasets by estimating standard  
29  
30 errors using gap-filling models (Section 2.6) and reporting the uncertainty of the users,  
31  
32 reference data, and overall accuracies.  
33  
34  
35  
36  
37  
38

### 39 **2.1. Study area**

40  
41 We selected three study areas in the conterminous United States (CONUS) (Section  
42  
43 2.1). These areas are Atlanta, GA; Phoenix, AZ; and Sioux Falls, SD during selected  
44  
45 years (1991, 2000, 2016, and 2020). The Atlanta area, one of the largest and most  
46  
47 populated urban centres in the U.S. covers four ARD tiles with a total area of 90,000  
48  
49 km<sup>2</sup> (Figure 1). It is located in northern Georgia, near the Blue Ridge Mountains. The  
50  
51 area has a population of 6,220,106 in 2022 according to the U.S. Census Bureau  
52  
53 (<https://www.census.gov/>). The rural landscapes surrounding the city comprise forest,  
54  
55 croplands, pastures, hayfields, and water bodies. The area has a humid subtropical  
56  
57 climate and monthly mean air temperatures of 6.1 °C in January and 26.8 °C in July  
58  
59  
60

1  
2  
3 (NWS 2021). The first six months of 2022 were Atlanta's fourth hottest on record,  
4  
5 according to the National Oceanic and Atmospheric Administration (NOAA). The area  
6  
7 receives abundant rainfall with an annual average of 1260 mm.  
8  
9

10  
11 The second study area is the Sioux Falls metropolitan area and surrounding rural areas  
12  
13 in South Dakota, United States. The area is within one ARD tile and has a spatial extent  
14  
15 of 22,500 km<sup>2</sup> (Figure 1). The city of Sioux Falls has grown at a rapid pace since the  
16  
17 late 1970s, with the city's population increasing from 81,000 in 1980 to 208,884 in  
18  
19 2022 (<https://www.census.gov/>). It is the 130<sup>th</sup> largest city in the US but the largest city  
20  
21 in South Dakota. The rural landscapes surrounding the city comprise croplands,  
22  
23 pastures, and hayfields, with patches of forests concentrated in parks, bottomlands,  
24  
25 shelterbelts, and farmsteads. Within the sub-humid continental temperate climate zone,  
26  
27 Sioux Falls has warm, humid summers and cold winters with most precipitation  
28  
29 occurring between April and September (yearly average about 840 mm). The monthly  
30  
31 mean air temperatures vary from -16.7 to -4.4 °C in winter (December– February) and  
32  
33 from 16.1 to 30.0 °C in summer (June– August). The area is also known for its strong  
34  
35 winds which can reach up to 56 km per hour.  
36  
37  
38  
39

40  
41 *Figure 1 near here*  
42  
43

44 The Phoenix area is within two Landsat ARD tiles with a spatial extent of 45,000 km<sup>2</sup>  
45  
46 (Figure 1). Phoenix is the most populous city of Arizona, with 1,644,409 residents as of  
47  
48 2022 (<https://www.census.gov/>). It is the fifth most populous city and the most  
49  
50 populated state capital in the country and the only U.S. state capital with a population of  
51  
52 more than one million residents. Phoenix lies near the confluence of the Gila and Salt  
53  
54 rivers and is situated at the northern edge of the Sonoran Desert, an arid ecological zone  
55  
56 whose characteristic plant is the nationally protected saguaro cactus. The area has a  
57  
58 typical arid subtropical climate. The metropolitan area is known as the "Valley of the  
59  
60

1  
2  
3 Sun" due to its location in the Salt River Valley. The area has monthly mean air  
4  
5 temperatures of 35 °C in July and 12 °C in January, has a large temperature difference  
6  
7 between day and night, and receives only 185 mm annual average rainfall (NWS 2021).  
8  
9

## 11 **2.2. Reference data and extracting strategy**

13  
14 We selected four existing reference datasets for the study (Table 1). The first one is  
15  
16 NOAA Global Historical Climatology Network daily (GHCN) (Figure 1 and Table 2),  
17  
18 an integrated database of daily climate summaries from land surface stations across the  
19  
20 globe (Menne et al. 2017). GHCN is made up of daily climate records from numerous  
21  
22 sources that have been integrated and subjected to a common suite of quality assurance  
23  
24 reviews. NOAA National Centres for Environmental Information (NCEI) provides  
25  
26 numerous daily variables, including maximum and minimum temperature, total daily  
27  
28 precipitation, snowfall, and snow depth. About half the stations only report  
29  
30 precipitation. Both record length and period of record vary by station and cover  
31  
32 intervals ranging from less than a year to more than 175 years. The second one is  
33  
34 Moderate Resolution Imaging Spectroradiometer (MODIS), onboard the NASA Terra  
35  
36 and Aqua Earth Observing System satellites, which provides daily multiple LST  
37  
38 products. The most recently Collection 6 (C6) MODIS LST includes three refinements  
39  
40 over bare soil surfaces compared to the Collection 5 (C5) MODIS LST product (Duan  
41  
42 et al. 2017; Hulley and Hook 2021). The third one is VIIRS-derived data products that  
43  
44 are used to measure cloud and aerosol properties, ocean colour, ocean and LST, ice  
45  
46 movement and temperature, fires, and Earth's albedo. Climatologists use VIIRS data to  
47  
48 improve our understanding of global climate change archived and distributed through  
49  
50 the Oak Ridge National Laboratory (ORNL) (Hulley and Hook. 2018). The fourth  
51  
52 reference dataset is the Daily Surface Weather and Climatological Summaries  
53  
54 (DAYMET) dataset (Thornton et al. 2021), which provides gridded estimates of daily  
55  
56  
57  
58  
59  
60

1  
2  
3 weather parameters for North America, including daily continuous surfaces of minimum  
4 and maximum temperature, precipitation occurrence and amount, humidity, shortwave  
5 radiation, snow water equivalent, and day length.  
6  
7  
8

9  
10 *Table 1 and Table 2 near here.*  
11  
12

13 We selected climate data from all available GHCN stations (Table 2), which have full  
14 temperature records for 1991, 2000, 2016, and 2020, as extracting points and converted  
15 the points to Landsat resolution (30 m × 30 m) from all pixels in the CONUS ARD grid  
16 system as a mask to spatially match the other existing LST datasets. The total sample  
17 consisted of 180 points within seven CONUS ARD tiles in three selected study areas.  
18 We chose this extracting strategy to prioritize four desirable strategy criteria: (1)  
19 probability based; (2) simple to implement; (3) easy to compare for multiple spatial  
20 resolution LST datasets; (4) extracting the same locations to GHCN stations. Another  
21 motivation for implementing the extracting strategy is that GHCN is field observation  
22 and air temperature but more accurate than other existing remote sensing derived LST.  
23 Also, we employed a weekly composite of MODIS and VIIRS data, which offers  
24 improved quality due to the composite algorithm mitigating cloud impact. For example,  
25 the daily MODIS and VIIRS data often contain large portion of missing values,  
26 potentially leading to misleading comparison results. Additionally, we used GHCN  
27 station locations to create 3\*3 Landsat pixel (30 m) masks to get average values from  
28 gap-filled Landsat LST and extract DAYMET (1000 m), MODIS and VIIRS (1000 m),  
29 and ECOSTRESS (70 m) to be comparable within the same land cover class. The gap-  
30 filled uncertainty layer is used for analysis.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

### 56 **2.3. Summary of gap-filling method**

57  
58

59 As summarized in the introduction, we have developed a new method of time series  
60

gap-filling that is designed for multi-sensor and multi-time data harmonization. This method uses pixels from the orbit overlap region to fill data gaps based on time series similarity, which retains the observation variation. Model assembling procedures were used to estimate stable predictions that are not only robust to occasionally cloud-contaminated training data but also allowed us to estimate the uncertainty of the predictions (Zhou, Xian, and Shi 2020). In this study all overlap regions in tree study areas have more than 21 clear observations, while non-overlap regions often have insufficient data. Thus, the 7-parameter linear harmonic model is used to replace those cloud-contaminated data, which maintains the details of seasonality in the training data (Equation (1)). The procedure for each target pixel that includes clustering the training data, stratifying random selection for the target pixel, and predict the full time series LST via linear regression (equation (2)). For detailed information see (Zhou, Xian, and Shi 2020).

$$f(t) = a_0 + \sum_{m=1}^M \left( a_m \cos \frac{2\pi t}{L} + b_m \sin \frac{2\pi t}{L} \right) \quad (1)$$

where  $f(t)$  is the modeled time series for a single pixel location in the overlap region;  $a_0$  describes the mean of  $f(t)$  over the time series;  $a_m$  and  $b_m$  are coefficients for harmonic component  $m$ ;  $t$  is day of year; and  $L$  is the length of the time period ( $L = 365.25$ ). Parameter  $M$  ( $M = 3$ ) determines the highest frequency used for modelling.

$$y = \beta_0 + \sum (\beta_i X_i) + \epsilon \quad (2)$$

where  $\beta_0$  and  $\beta_i$  represent the linear parameter to be estimated and  $\epsilon$  represents the error terms.  $X_i$  and  $y$  are sampled training data, and the target pixel time series at  $y$  are clear observation dates.

#### 2.4. Accuracy assessment

Accuracy assessment is the procedure used to quantify product quality. Attempts have been made to quantify limiting factors resulting from the Landsat low temporal availability of data used for generating high frequency LST information at regional level. Sub-pixel fractional error matrices are introduced as a more appropriate way for assessing the accuracy of mixed pixels. For classification with coarse spatial resolution data, limitations of the classification method produce a maximum achievable accuracy defined as the average percent fraction of dominant land cover of all pixels in the mapped area. We used a combination of station data, climate data (DAYMET), MODIS and VIIRS LST data, and Landsat data to validate the accuracy of the gap-filled products. Specifically, we used the following methods. (1) Station validation: We selected NOAA GHCN station observation data on temperature and intensity at 20 randomly selected sites in the study areas. We used a high-resolution digital camera to take photographs of the sites, and then used the photographs to estimate temperature and SUHI intensity using image analysis software. We compared these data to the corresponding data in the gap-filled product to assess the accuracy of the gap-filling techniques. (2) Landsat validation: We chose raw Landsat data to validate the accuracy of the gap-filled products over time. We compared the Landsat data from multiple time periods to the gap-filled product to Landsat data. We estimated temperature and UHI intensity using image analysis software and compared the results to the gap-filled product. (3) DAYMET LST, MODIS and VIIRS LST data: We also used these data to evaluate the gap-filled products. The GHCN station data is a point dataset, and DAYMET data is rasterized based on GHCN with 1 km resolution. The MODIS and VIIRS dataset have a spatial resolution of 1 km with 8-day composites. ECOSTRESS is a new dataset with limited products.

We compared the gap-filled data with the original data to assess the accuracy of the gap-filled products using two metrics to evaluate the accuracy: root mean square error (RMSE) and the coefficient of determination ( $R^2$ ). The uncertainty analysis was conducted to quantify the variability in the gap-filled data. We used the bootstrap method to simulate the variability in the gap-filled data and calculated the 95% confidence intervals for the gap-filled data. The following equations were used to calculate statistical parameters of RMSE and  $R^2$ :

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (G_i - R_i)^2}{N}} \quad (3)$$

where:  $(G_i - R_i)^2$  is differences, squared, and N is sample size.

$$r^2 = \left( \frac{\sum_i^N (X_i - \mu_x)(Y_i - \mu_y)}{(\sqrt{\sum_i^N (X_i - \mu_x)^2})(\sqrt{\sum_i^N (Y_i - \mu_y)^2})} \right)^2 \quad (4)$$

where: N is samples, X is the predictor variable, and Y is the response variable in this regression model.

## 2.5. Uncertainty analysis

There are many sources of uncertainty in Landsat gap-filling processing, such as clear observation training collection, QA band issues, and the modelling approach issues. The gap-filled Landsat LST is generally based on clustering of spatial entities within a spectral space. One major concern is the use of seasonal models to predict the variability of LST into several discrete dates within seasons. This type of approach is often inappropriate given the continuous values by regression model, which usually provides overestimated prediction in the high end and underestimated prediction in the low end. This leads to uncertainty in the products resulting from the use of remote

1  
2  
3 sensing data. Based on the assumption that any gap-filled Landsat LST has an  
4 associated error and/or uncertainty of unknown magnitude, the statistical quantification  
5 of uncertainty analysis should be a core part of scientific research. In this study we  
6 analysed uncertainty layers from Landsat gap-filling models and reviewed recent  
7 attempts to take explicitly into uncertainty when mapping LST. We used the Landsat  
8 gap-filling uncurtaining layers that calculated the standard deviation (SD) of iterations  
9 for each prediction as an indicator of uncertainty. Standard deviation is often used to  
10 quantify the level of uncertainty in a set of measurements. In the context of gap-filling,  
11 it is used to indicate the degree of uncertainty of gap-filled LST for the pixel. The  
12 magnitude of the standard deviation is directly proportional to the level of uncertainty of  
13 the pixel. In other words, the larger the standard deviation is, the more uncertainties of  
14 the pixel exist. The results have the same structure as ARD tiles with all missing data  
15 filled, using the following formula:  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

$$\sigma = \sqrt{1/n \sum_{i=1}^n (x - \mu)^2} \quad (5)$$

34  
35  
36  
37 where:  $\sigma$  is SD,  $n$  is the number of filled LST,  $x$  is filled LST, and  $\mu$  is mean of filled  
38 LST.  
39  
40  
41  
42  
43

## 44 **2.6. Comparison analysis**

45  
46 The synthesis and comparison of spatial data collected or derived at different spatial  
47 resolutions with various uncertainty factors is challenging work in geospatial statistics  
48 (Gotway and Young 2002). The geostatistical concept of support belongs to one of  
49 spatial scales, in which it is a property of a variable used in comparison analysis. The  
50 support can be as small as a point or as large as the full extent of the study area. In this  
51 study, the spatial unit, which is based on GHCN station with Landsat 9 pixels mask,  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 used to sample the multiple LST and climate datasets have also been considered in the  
4 support of the data. Comparison results (RSME, standard deviation, spatial  
5 autocorrelation, and others) can sometimes be estimated using geostatistical support-  
6 effect models. We present the case studies of LST illustrating some of these differences  
7 and how conclusions about zonal mean of temperature may be affected. We identify the  
8 influence of GHCN observational air temperature on statistical results as a local zonal  
9 mean temperature. The way to avoid the station air temperature is by careful  
10 construction of sampling design and analysis.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

### 23 **3. Results**

#### 24 **3.1. Landsat ARD gap-filled LST**

25  
26  
27  
28  
29 The study conducted in the three selected cities in 1991, 2000, 2016, and 2020  
30 demonstrated that the Landsat ARD gap-filled products can better differentiate the  
31 performances of the spatiotemporal gap-filling model with improved training data  
32 strategy. The top row of Figures 2 (Atlanta), 3 (Sioux Falls), and 4 (Phoenix) shows the  
33 examples of the gap-filled Landsat LST from different dates. The original Landsat LST  
34 only covered part of the tiles (bottom row) and was contaminated by scattered clouds.  
35  
36 By closely comparing these pair maps, the gap-filled Landsat LST showed a similar  
37 value range to the Landsat LST observations in clear areas. Both observations and gap-  
38 filled showed high LST in urban (or bare) areas and low LST along forest or water  
39 areas. The spatial distribution pattern of gap-filled Landsat LST is associated well with  
40 urban land and most of high values are located within the city limits in Atlanta and  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

1  
2  
3 other LST data because these products, including MODIS/VIIRS and ECOSTRESS are  
4  
5 at 500 m (daily) and 1000 m (8-day composite) while the gap-filling procedure is at 30-  
6  
7 m scale. Temporally, the gap-filled products revealed more detailed seasonal and  
8  
9 monthly variations, which can potentially be used to study seasonal SUHI change.  
10  
11

12 *Figures 2, 3, and 4 near here.*  
13  
14  
15

### 16 **3.2. Accuracy assessment**

17 We used randomly selected GHCN observation data as the validation datasets, in which  
18  
19 over 350 days per year of air temperature records are available. By comparing the 3x3  
20  
21 pixel average LST with weather station records, we found that the gap-filled Landsat  
22  
23 LST data had good agreement with the data from GHCN in all three study areas.  
24  
25

26  
27 Figures 5, 6, and 7 represent the time series of observed and gap-filled Landsat LST in  
28  
29 36 GHCN field stations from 1991, 2000, 2016, and 2020 with different Landsat  
30  
31 sensors within three study areas. The GHCN air temperature is also displayed as a  
32  
33 reference of seasonal temperature pattern in the three selected study areas. The gap-  
34  
35 filled values followed well with observations except for several scattered outliers in  
36  
37 specific dates and stations, while both observed and gap-filled summer LSTs are higher  
38  
39 than the station air temperatures. The outliers in Figure 5 with very low LST came from  
40  
41 Landsat 8 images that were mostly covered by clouds, but pixel QA only flagged part of  
42  
43 the images as cloud or cloud shadow. Similarly, the low LST at the peak of the summer  
44  
45 in Figure 5 also came from cloud-contaminated observations (Landsat 5 and 7 image).  
46  
47

48 Figure 6 shows that the LST gap-filled model didn't perform well in winter months  
49  
50 because of snow cover and limited clear observation for training. To assess the accuracy  
51  
52 quantitatively, the RMSE and  $R^2$  of observed LST with GHCN and gap-filled Landsat  
53  
54 LST with GHCN were calculated (Tables 3, 4, and 5). Generally, the gap-filled Landsat  
55  
56 LST has a better range of RMSE than the Landsat LST product accuracy, indicating the  
57  
58  
59  
60

1  
2  
3 algorithms have successfully filled the LST. The RMSE of the gap-filled Landsat LST  
4 data ranged from 1.5 to 4.2°C (see Table 3) in Atlanta, 1 to 5 °C in Sioux Falls (Table  
5 4), and 0.5 to 3°C in Phoenix (Table 5). The gap-filled data has a significant  
6  
7  
8 relationship with GHCN validation data. We compared both original clear Landsat LST  
9  
10  
11 (in brackets is the number of clear observations in Tables 3, 4, and 5 and gap-filled  
12  
13 Landsat LST with GHCN air temperature. The correlation between GHCN and gap-  
14  
15 filled LST is similar with (or lower than) GHCN and the original Landsat clear  
16  
17 observation if the number of gap-filled LST close or equal the number of original clear  
18  
19 observation, the average  $R^2$  values of all stations is about 0.81 and 0.68 ( $p < 0.05$ ) in  
20  
21 Atlanta (Table 3) , 0.74 and 0.81 ( $p < 0.05$ ) in Sioux Falls (Table 4), 0.82 and 0.74  
22  
23 ( $p < 0.05$ ) in Phoenix (Table 5). Some of the gap-filled Landsat LST show large  
24  
25 variations, while observations are consistent at low or high values. As the GHCN station  
26  
27 data were randomly selected and expanded to 9 Landsat pixels, we went through most  
28  
29 gap-filled images and found that some cloud or cloud shadow pixels were labelled as  
30  
31 clear observations by the Landsat pixel QA band, which were not filled and could lead  
32  
33 to the consistently high or low values of gap-filled LST.  
34  
35  
36  
37  
38  
39  
40

41 *Tables 3, 4, and 5 near here*

42  
43 *Figures 5, 6, and 7 near here*

44  
45  
46 The gap-filled Landsat LST products, on average, slightly overestimate LSTs in urban  
47  
48 and bare areas and underestimate LSTs in other land cover in all three study areas.

49  
50 Atlanta (Figure 5), located in a subtropical humid climate condition, has relatively small  
51  
52 seasonal or annual temperature variations compared with other study areas. Phoenix  
53  
54 (Figure 7), in a subtropical arid land climate condition, has more clear observations for  
55  
56 training, resulting in better performance of the gap-filling model than in Sioux Falls  
57  
58 (Figure 6), which is located in a moderate temperate continental climate condition with  
59  
60

1  
2  
3 snow cover in winter and has less training data available.  
4  
5

### 6 7 **3.3. Uncertainty analysis** 8

9  
10 We also analysed the uncertainty of the gap-filled Landsat surface temperature data  
11 using a Monte Carlo simulation. The simulation randomly sampled the input data from  
12 their respective probability distributions and calculated the resulting distribution of the  
13 gap-filled data. The simulation showed that the uncertainty of the gap-filled Landsat  
14 LST data was primarily determined by the uncertainty in the input Landsat data,  
15 followed by the uncertainty in the auxiliary data and model parameters. Figure 8 shows  
16 the spatial distribution of the yearly number of clear observations of Landsat (top) and  
17 annual mean of gap-filling models uncertainty map (bottom) in 2016 and 2020 in the  
18 three study areas. Table 6 gives the maximum, mean, and minimum number of clear  
19 observations and uncertainty values for the same times. Table 6 also shows that 1) in  
20 Atlanta, 55% of pixels have prediction uncertainties of  $\pm 0.65$  °C in 2016 and  $\pm 0.97$  °C  
21 in 2020 with annual mean LST of 19.2 °C in 2016 and 18.4 °C in 2020; 2) in Sioux  
22 Falls, over 50% of prediction uncertainties are around  $\pm 0.94$  °C in 2016 and  $\pm 2.08$  °C in  
23 2020 with annual mean LST of 13.9 °C in 2016 and 13.6 °C in 2020; 3) in Phoenix,  
24 about 60% of prediction uncertainties are about  $\pm 0.76$  °C in 2016 and  $\pm 0.68$  °C in 2020  
25 with annual mean LST of 29.2 °C in 2016 and 29.9 °C in 2020. The gap-filled products  
26 from the gap-filling model have better results in Atlanta and Phoenix, except for  
27 individual outliers. The uncertainty is high in Sioux Falls because it has fewer clear  
28 observation for training data due to snow/ice cover in wintertime. The uncertainty is  
29 low in Phoenix because it has more observations for training data than the other two  
30 areas.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Table 6 near here,*

1  
2  
3  
4  
5  
6 We used 9-pixel masks to calculate the mean of prediction uncertainty by using GHCN  
7 data (Figure 8).  
8  
9

10  
11  
12 *Figures 8 and 9 near here.*  
13

14 The factors affecting gap-filled Landsat LST product accuracy could come from several  
15 sources including violations of the fundamental assumptions underlying the approach,  
16 Landsat swath overlap areas, seasonal linear regression algorithm, errors in the  
17 collection training dataset by using the QA band, and Landsat thermal data quantity and  
18 quality (Figure 9). For example, the gap-filled products have higher value of accuracy  
19 assessment in Phoenix due to high availability of clear observations for training data  
20 (Figure 9C). The gap-filled products have lower value of accuracy assessment in Sioux  
21 Falls due to snow/ice cover even it is a clear observation, because of the lack of clear  
22 observation, the gap-filling modelling approach assumes that the spectral behaviour of  
23 LST through time can only be represented with seasonal harmonic models that are used  
24 for training all dates during periods of time with limited training data. Therefore, the  
25 model output is expected to be better in winter when training data is extracted from  
26 warm months (dates) and the regression model that has a high uncertainty in snow cover  
27 dates. Figures 9B show high uncertainty values in the early dates of the year (mostly in  
28 January and February in Sioux Falls). We also checked the original Landsat images for  
29 these pixels having high uncertainty value by dates and location in Figures 8 and 9.  
30  
31 Most of these pixels have cloud cover or no clear observation available. All these  
32 factors, such as snow cover and limited training data with seasonal model, depend on  
33 the regional geography and vary across different climate zones. For example, the  
34 quantity of available cloud- and snow-free Landsat data varies considerably in these  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 three study areas and is shown by the presence of swath overlap zones of Landsat  
4 acquisition footprints within ARD tiles. How the algorithm responds to these variations  
5 in input data can impact the accuracy of the gap-filled products. For the gap-filling  
6 model, the selection of training data and accuracy of training data labels may both exert  
7 considerable influence on accuracy. Decisions made with respect to the filtering and  
8 balance of the pool of seasonal training data points may have different effects on  
9 accuracy of LST. Finally, using overlapped and clear observations from adjacent ARD  
10 tiles in training models improved the models and removed a source of error that arose  
11 from the gap-filling model design.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

#### 25 **3.4. Comparison analysis**

26  
27 We compared the gap-filled Landsat LST with three other large-scale time series LST  
28 products including MODIS, VIIRS, ECOSTRESS, and rasterized DAYMET climate  
29 data. We also compared these LST datasets with GHCN station observations. We  
30 created 3x3 Landsat pixels based on GHCN points as a centre pixel to match coarse  
31 LST datasets. The comparison shows that the gap-filled Landsat LST matches the  
32 GHCN air temperature better than the others, except for DAYMET data. We did  
33 compare DAYMET with GHCN, and they are well matched with  $R^2 = 0.99$  in most  
34 cases. We utilized MODIS and VIIRS 8-day composites due to their superior data  
35 quality compared to daily data, primarily because of the impact of cloud cover.  
36 Specifically, the daily MODIS and VIIRS data often contain numerous missing values,  
37 even on clear-sky days. This makes it challenging to compare them with gap-filled Land  
38 Surface Temperature (LST) data at a 3x3 30 m pixel size based on GHCN locations. By  
39 filtering out cloud pixels from the daily ECOSTRESS data using pixel QA information,  
40 we found that there isn't enough usable data available. Additionally, ECOSTRESS data  
41 is only available for the year 2020. Table 7 shows  $R^2$  and RMSE from the selected  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 GHCN station locations. In general, gap-filled LST attains the accuracy that is broadly  
4 consistent with accuracy of these other existing LST datasets. For example, in the daily  
5 data comparison, gap-filled has low or similar RMSE to VIIRS in 2020 and  
6 ECOSTRESS has the highest RMSE. One reason is the gap-filled and ECOSTRESS  
7 LSTs have daily data, but VIIRS and MODIS LSTs are 8-day composite data that have  
8 better results with cloud and bad pixels fixed. For monthly and seasonal, all datasets  
9 have similar patterns with GHCN except ECOSTRESS.

10  
11  
12  
13  
14  
15  
16  
17  
18  
19 *Table 7. near here*

#### 20 21 22 23 **4. Discussion**

24  
25 This study presented a comprehensive accuracy assessment, uncertainty analysis, and  
26 multi-dataset comparison for evaluating the gap-filled Landsat LST data. The multi-  
27 disciplinary history provided conveys the complexity of the issues encountered in vali-  
28 dating and comparing multiple source spatial data and the widespread interest in solu-  
29 tions that have been developing over several decades. This is now an active area of re-  
30 motely sensed data fusion and gap-filling research, and much novel research work has  
31 recently been developed (Roy and Yan 2020; Zhu et al. 2022).

32  
33 Our results show that the product using gap-filling techniques has high accuracy for  
34 estimating annual, seasonal, and even monthly thermal condition for UHI and trends  
35 analysis. It also shows that the gap-filling algorithm is effective in filling missing values  
36 in remote sensing data. However, the accuracy of the gap-filled data varied depending  
37 on the test site with type of climate zones. The application of models was more accurate  
38 in arid areas (Arizona) than in humid regions (Georgia). However, the model outcomes  
39 in South Dakota are similar with these in other two locations in summer but are not  
40 good in winter months due to snow/ice cover. The RMSE values were lower in the  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 urban site than in the other land covers, while there was no significant difference of  
4  
5 RMSE in different geographical regions. The correlation coefficient was higher in the  
6  
7 arid region than in the humid region. The uncertainty analysis showed that the  
8  
9 variability in the gap-filled data was higher in the semi-arid region of Arizona than in  
10  
11 the humid region of Georgia. The uncertainty affecting gap-filled Landsat LST product  
12  
13 accuracy could arise from several sources including modelling approach, seasonal linear  
14  
15 regression algorithm, errors in the collection training dataset by using the QA band, and  
16  
17 Landsat thermal data quantity and quality. The comparison analysis reveals that gap-  
18  
19 filled Landsat LSTs are more accurate in monthly and seasonal estimates.  
20  
21  
22

23  
24 Several limitations are found to the current study. First, the validation data sources used  
25  
26 in our study have their own uncertainties, and these uncertainties may propagate into  
27  
28 our assessment of the accuracy and uncertainty of the gap-filled Landsat LST.  
29  
30

31  
32 Additionally, NOAA GHCN records air temperatures that are different from gap-filled  
33  
34 and other LST datasets. Second, we only evaluated our own gap-filling techniques.  
35

36  
37 Third, we only used Landsat data and our results may not be applicable to other remote  
38  
39 sensing datasets. Fourth, our study only evaluated the accuracy and uncertainty of the  
40  
41 gap-filled products at the 9 pixel (30x30 m) and two Landsat ARD tile levels  
42  
43 (5000x5000 m). Future studies may need to investigate the accuracy and uncertainty at  
44  
45 the continental or global scale.  
46  
47

48  
49 Gap-filling can introduce uncertainty into the final products because of combining  
50  
51 factors including input data and models selection (Friedl et al. 1995; Murphy et al.  
52  
53 2004; Zhou, Xian, and Shi 2020; Rounsevell et al. 2021). Depending on the method  
54  
55 used, gap-filling can be a complex and computationally intensive process. Gap-filling  
56  
57 requires access to multiple dates of data, and the availability of these data may be  
58  
59 limited in some regions or for certain time periods because of missing data, cloud and  
60

1  
2  
3 shadow, and snow and ice cover (Gao et al. 2006; Zhu et al. 2022). The accuracy of  
4  
5 gap-filling can be affected by spatial and temporal variability in the data. For example,  
6  
7 filling in missing data in a forested area may be more difficult than in an area with  
8  
9 lower temperature. Different methods may be more or less suitable for different types of  
10  
11 data or for different applications (Rocchini et al. 2013; Leyk et al. 2018; Stehman and  
12  
13 Foody 2019). Gap-filling relies on the quality of the surrounding data to estimate  
14  
15 missing values (Zhang et al. 2020; Zhou, Xian, and Shi 2020). If the surrounding data is  
16  
17 poor quality or affected by noise, the accuracy of the gap-filled data may be  
18  
19 compromised. Validating the accuracy and reliability of the gap-filling data can be  
20  
21 challenging, as there may not be ground-based measurements or other independent  
22  
23 sources of data available for comparison except air temperature from weather station  
24  
25 observations. Also, as it is difficult to determine the true value of missing data, it can  
26  
27 make it difficult to evaluate the accuracy of the gap-filling method.  
28  
29  
30  
31  
32  
33

34 Future work may need to address some of the limitations of this study by focusing on  
35  
36 investigating the accuracy and uncertainty of gap-filled products using other remote  
37  
38 sensing datasets and at different spatial and temporal scales. Additionally, the impact of  
39  
40 the gap-filling techniques on the accuracy of downstream analyses, such as thermal  
41  
42 condition, vegetation indices and land cover classifications, should be investigated.  
43  
44

45 Future studies may need to investigate the accuracy and uncertainty at the regional or  
46  
47 global scale. Other topics may need further attention: 1) Machine learning techniques,  
48  
49 such as deep learning and artificial neural networks, which could offer better accuracy  
50  
51 and un-certainty estimates for gap-filled products. 2) Investigation of the impact of gap-  
52  
53 filling techniques on downstream analyses for other landscapes in addition to urban. 3)  
54  
55 Evaluation of the accuracy and uncertainty of gap-filling products over longer time  
56  
57 periods and larger spatial scales. 4) Development of auto-standardized methods for  
58  
59  
60

1  
2  
3 evaluating gap-filled products by using other existing remote sensing derived LST  
4 products. Standardized methods for evaluating the accuracy and uncertainty of gap-  
5 filled products would facilitate comparison and benchmarking of different techniques  
6 and products.  
7  
8  
9  
10  
11  
12

## 13 **5. Conclusion**

14  
15  
16 Gap-filling accuracy assessment, uncertainty analysis, and comparison analysis  
17 are essential to ensure the reliability of gap-filled Landsat LST products. This  
18 paper presents the results of accuracy assessment, uncertainty analysis, and  
19 comparison analysis of a new time series gap-filled Landsat LST that is modelled  
20 from multi-sensor and multi-time Landsat data harmonization. Landsat LST  
21 observations within the ARD tiles without gap filling are not adequate to represent  
22 temporal frequency of surface thermal conditions in a time series, resulting in  
23 either overestimates or underestimates of their seasonal or annual temporal means.  
24  
25 The Landsat LST with gap-filling substantially added temporal density for daily  
26 Landsat LST records and can be used to calculate monthly and seasonal Landsat  
27 LST. This increased frequency Landsat time-series LST provides an optional  
28 temperature source for SUHI monitoring, assessment, and trend analysis. The  
29 gap-filled Landsat LST has significant correlations with air temperature recorded  
30 from gridded weather records, suggesting similar daily, monthly, and seasonal  
31 variation patterns between the two datasets. The data can be used in longtime  
32 time-series SUHI and intensity annual, seasonal, even monthly change analysis.  
33  
34 Furthermore, we demonstrate that widespread uncertainty is occurring across our  
35 study area and this uncertainty is influencing the gap-filled Land-sat LST.  
36  
37 Our study provides important insights into the accuracy of gap-filling techniques  
38 for gap-filled products derived from remote sensing data. By assessing the  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 accuracy of the techniques, we can provide the information of reliability and  
4  
5 usefulness of remote sensing data products for various applications. In conclusion,  
6  
7 our study shows that gap-filling techniques are effective in filling missing values  
8  
9 in remote sensing data. However, the accuracy of the gap-filled data varied  
10  
11 depending on the test site and the type of gap-filling models used. The accuracy  
12  
13 assessment showed that the models performed better in arid regions than in humid  
14  
15 regions. Also, the model results are similar in in summer months in all regions.  
16  
17 The uncertainty analysis indicates that the variability in the gap-filled data is  
18  
19 higher in the arid and semi-arid regions than in humid region. The variability is  
20  
21 larger in the cold region than the warm region during winter months.  
22  
23  
24  
25  
26  
27

## 28 **Acknowledgments**

29  
30  
31 The authors thank Dr. Heather J. Tollerud and Mr. Roger Auch for reviewing this manuscript.  
32  
33 The manuscript contents do not constitute a statement of endorsement, policy, decision, or  
34  
35 position on behalf of USGS. The views expressed in this article are those of the author and do  
36  
37 not necessarily represent the views or policies of the USGS. Any use of trade, firm, or product  
38  
39 names is for descriptive purposes only and does not imply endorsement by the U.S.  
40  
41 Government. Hua Shi's work was performed under Work performed under USGS contract  
42  
43 140G0119C0001.  
44  
45  
46

## 47 **Disclosure statement**

48  
49  
50 No potential conflict of interest was reported by the author(s).  
51  
52  
53

## 54 **References**

55  
56 Bauer, Marvin E. 2020. "Remote Sensing of Environment: History, Philosophy, Approach and  
57  
58 Contributions, 1969 –2019." *Remote Sensing of Environment* 237:111522. doi:  
59  
60 <https://doi.org/10.1016/j.rse.2019.111522>.

- 1  
2  
3 Chen, J., X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin. 2011. "A simple and effective method for  
4 filling gaps in Landsat ETM+ SLC-off images." *Remote Sensing of Environment* 115  
5 (4):1053-64. doi: 10.1016/j.rse.2010.12.010.
- 6 Duan, Si-Bo, Zhao-Liang Li, Jie Cheng, and Pei Leng. 2017. "Cross-satellite comparison of  
7 operational land surface temperature products derived from MODIS and ASTER data  
8 over bare soil surfaces." *ISPRS Journal of Photogrammetry and Remote Sensing* 126:1-  
9 10. doi: <https://doi.org/10.1016/j.isprsjprs.2017.02.003>.
- 10 Foody, Giles M. 2020. "Explaining the unsuitability of the kappa coefficient in the assessment  
11 and comparison of the accuracy of thematic maps obtained by image classification."  
12 *Remote Sensing of Environment* 239:111630. doi:  
13 <https://doi.org/10.1016/j.rse.2019.111630>.
- 14 Friedl, M. A., F. W. Davis, J. Michaelsen, and M. A. Moritz. 1995. "Scaling and uncertainty in the  
15 relationship between the NDVI and land surface biophysical variables: An analysis  
16 using a scene simulation model and data from FIFE." *Remote Sensing of Environment*  
17 54 (3):233-46. doi: [http://dx.doi.org/10.1016/0034-4257\(95\)00156-5](http://dx.doi.org/10.1016/0034-4257(95)00156-5).
- 18 Gao, F., J. Masek, M. Schwaller, and F. Hall. 2006. "On the blending of the landsat and MODIS  
19 surface reflectance: Predicting daily landsat surface reflectance." *IEEE Transactions on*  
20 *Geoscience and Remote Sensing* 44 (8):2207-18. doi: 10.1109/TGRS.2006.872081.
- 21 Gotway, C. A., and L. J. Young. 2002. "Combining incompatible spatial data." *Journal of the*  
22 *American Statistical Association* 97 (458):632-48. doi: 10.1198/016214502760047140.
- 23 Hulley, G., and S. Hook. 2021. "MOD21A2 MODIS/Terra Land Surface Temperature/3-Band  
24 Emissivity 8-Day L3 Global 1km SIN Grid V061." In, edited by NASA EOSDIS Land  
25 Processes DAAC.
- 26 Hulley, G., and S. Hook. 2018. "VIIRS/NPP Land Surface Temperature and Emissivity 8-Day L3  
27 Global 1km SIN Grid V001." In, edited by NASA EOSDIS Land Processes DAAC.
- 28 Leyk, S., J. H. Uhl, D. Balk, and B. Jones. 2018. "Assessing the accuracy of multi-temporal built-  
29 up land layers across rural-urban trajectories in the United States." *Remote Sensing of*  
30 *Environment* 204:898-917. doi: 10.1016/j.rse.2017.08.035.
- 31 Luo, Y., K. Guan, and J. Peng. 2018. "STAIR: A generic and fully-automated method to fuse  
32 multiple sources of optical satellite data to generate a high-resolution, daily and cloud-  
33 /gap-free surface reflectance product." *Remote Sensing of Environment* 214:87-99.  
34 doi: 10.1016/j.rse.2018.04.042.
- 35 Menne, Matthew J., Byron E. Gleason, Jay Lawrimore, Jared Rennie, and Claude N Williams.  
36 2017. "Global Historical Climatology Network - Monthly Temperature, Version 4 BETA  
37 " In, edited by NOAA National Centers for Environmental Information. Washington D.C.
- 38 Murphy, J. M., D. M. H. Sexton, D. H. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A.  
39 Stainforth. 2004. "Quantification of modelling uncertainties in a large ensemble of  
40 climate change simulations." *Nature* 430 (7001):768-72. doi: 10.1038/nature02771.
- 41 Niclòs, Raquel, Joan M. Galve, Jose A. Valiente, Maria J. Estrela, and César Coll. 2011.  
42 "Accuracy assessment of land surface temperature retrievals from MSG2-SEVIRI data."  
43 *Remote Sensing of Environment* 115 (8):2126-40. doi:  
44 <https://doi.org/10.1016/j.rse.2011.04.017>.
- 45 NWS. "National Weather Service." NWS, NOAA, Accessed May, 01, 2022.  
46 <https://www.weather.gov/>.
- 47 Rocchini, D., G. M. Foody, H. Nagendra, C. Ricotta, M. Anand, K. S. He, V. Amici, et al. 2013.  
48 "Uncertainty in ecosystem mapping by remote sensing." *Computers and Geosciences*  
49 50:128-35. doi: 10.1016/j.cageo.2012.05.022.
- 50 Rounsevell, M. D. A., A. Arneth, C. Brown, W. W. L. Cheung, O. Gimenez, I. Holman, P. Leadley,  
51 et al. 2021. "Identifying uncertainties in scenarios and models of socio-ecological  
52 systems in support of decision-making." *One Earth* 4 (7):967-85. doi:  
53 10.1016/j.oneear.2021.06.003.
- 54  
55  
56  
57  
58  
59  
60

- Roy, D. P., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen, and E. Lindquist. 2008. "Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data." *Remote Sensing of Environment* 112 (6):3112-30. doi: 10.1016/j.rse.2008.03.009.
- Roy, D. P., and L. Yan. 2020. "Robust Landsat-based crop time series modelling." *Remote Sensing of Environment* 238:110810. doi: <https://doi.org/10.1016/j.rse.2018.06.038>.
- Stehman, Stephen V., Cidália C. Fonte, Giles M. Foody, and Linda See. 2018. "Using volunteered geographic information (VGI) in design-based statistical inference for area estimation and accuracy assessment of land cover." *Remote Sensing of Environment* 212:47-59. doi: <https://doi.org/10.1016/j.rse.2018.04.014>.
- Stehman, Stephen V., and Giles M. Foody. 2019. "Key issues in rigorous accuracy assessment of land cover products." *Remote Sensing of Environment* 231:111199. doi: <https://doi.org/10.1016/j.rse.2019.05.018>.
- Thornton, Peter E., Rupesh Shrestha, Michele Thornton, Shih-Chieh Kao, Yaxing Wei, and Bruce E. Wilson. 2021. "Gridded daily weather data for North America with comprehensive uncertainty quantification." *Scientific Data* 8 (1):190. doi: 10.1038/s41597-021-00973-0.
- Wang, Jing, Calvin K. F. Lee, Xiaolin Zhu, Ruyin Cao, Yating Gu, Shengbiao Wu, and Jin Wu. 2022. "A new object-class based gap-filling method for PlanetScope satellite image time series." *Remote Sensing of Environment* 280:113136. doi: <https://doi.org/10.1016/j.rse.2022.113136>.
- Wulder, M. A., J. C. White, J. G. Masek, J. Dwyer, and D. P. Roy. 2011. "Continuity of Landsat observations: Short term considerations." *Remote Sensing of Environment* 115 (2):747-51. doi: 10.1016/j.rse.2010.11.002.
- Wulder, Michael A., David P. Roy, Volker C. Radeloff, Thomas R. Loveland, Martha C. Anderson, David M. Johnson, Sean Healey, et al. 2022. "Fifty years of Landsat science and impacts." *Remote Sensing of Environment* 280:113195. doi: <https://doi.org/10.1016/j.rse.2022.113195>.
- Yan, L., and D. P. Roy. 2018. "Large-area gap filling of landsat reflectance time series by spectral-angle-mapper based spatio-temporal similarity (SAMSTS)." *Remote Sensing* 10 (4). doi: 10.3390/rs10040609.
- Zhang, Xiaoyang, Jianmin Wang, Geoffrey M. Henebry, and Feng Gao. 2020. "Development and evaluation of a new algorithm for detecting 30 m land surface phenology from VIIRS and HLS time series." *ISPRS Journal of Photogrammetry and Remote Sensing* 161:37-51. doi: <https://doi.org/10.1016/j.isprsjprs.2020.01.012>.
- Zhou, Q., G. Xian, and H. Shi. 2020. "Gap fill of land surface temperature and reflectance products in landsat analysis ready data." *Remote Sensing* 12 (7). doi: 10.3390/rs12071192.
- Zhou, Qiang, Zhe Zhu, George Xian, and Congcong Li. 2022. "A novel regression method for harmonic analysis of time series." *ISPRS Journal of Photogrammetry and Remote Sensing* 185:48-61. doi: <https://doi.org/10.1016/j.isprsjprs.2022.01.006>.
- Zhu, X., J. Chen, F. Gao, X. Chen, and J. G. Masek. 2010. "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions." *Remote Sensing of Environment* 114 (11):2610-23. doi: 10.1016/j.rse.2010.05.032.
- Zhu, X., E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky. 2016. "A flexible spatiotemporal method for fusing satellite images with different resolutions." *Remote Sensing of Environment* 172:165-77. doi: 10.1016/j.rse.2015.11.016.
- Zhu, Xiaolin, Wenfeng Zhan, Junxiong Zhou, Xuehong Chen, Zicong Liang, Shuai Xu, and Jin Chen. 2022. "A novel framework to assess all-round performances of spatiotemporal fusion models." *Remote Sensing of Environment* 274:113002. doi: <https://doi.org/10.1016/j.rse.2022.113002>.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11 Table 1. Main data sources used in the study.

12  
13 Table 2, NOAA GHCN Station ID (Figure 1), Latitude (Lat.) and Longitude (Lon.),  
14 Name, and Land Cover Class (LC, 2020) in three study areas.

15  
16 Table 3. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Table 3. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA  
GHCN stations for annual accuracy assessment, Atlanta, GA.

Table 4. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA  
GHCN stations for annual accuracy assessment, Sioux Falls, SD.

Table 5. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA  
GHCN stations for annual accuracy assessment, Phoenix, AZ.

Table 6. The number of Landsat clear observations and gap-filled uncertainty ( $^{\circ}\text{C}$ ) in  
2016 and 2020.

Table 7. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE), selected NOAA  
GHCN stations for comparison analysis among GHCN, gap-filled LST, MODIS LST,  
VIIRS LST, and ECOSTRESS LST by daily, monthly, and seasonal for three study  
areas in year 2016 and 2020.

1  
2  
3  
4  
5  
6  
7  
8 Figure 1. The zoom in 2020 land cover map for three selected study areas within the  
9 ARD tiles, GHCN station location are white-purple boxes associate with the ID number  
10 in black (the ID, name, land cover class, and detailed information of each station see  
11 Table 2). These Land cover map of individual urban centre are not at the same scale.

12  
13  
14  
15 Figure 2. Gap-filled Landsat LST (top) and Original Landsat (5, 7, and 8) LST (bottom)  
16 in Atlanta, GA in 1991, 2000, 2016 and 2020 from left to right.

17  
18  
19  
20 Figure 3. Gap-filled Landsat LST (top) and Original Landsat (5, 7, and 8) LST (bottom)  
21 in Sioux Falls, SD in 1991, 2000, 2016 and 2020 from left to right.

22  
23  
24  
25 Figure 4. Gap-filled Landsat LST (top) and Original Landsat (5, 7, and 8) LST (bottom)  
26 in Phoenix, AZ in 1991, 2000, 2016 and 2020 from left to right.

27  
28  
29  
30 Figure 5. Time series of land surface temperature and air temperature at 12 stations in  
31 Atlanta, GA. The blue boxes are valid Landsat observations, and red dots are gap-filled  
32 values at each Landsat acquisition date. The black triangles are station air temperature  
33 that match all Landsat acquisition dates within an ARD tile in Atlanta, GA  
34 (<https://www.ncdc.noaa.gov/ghcn-daily-description>). Information of GHCN stations see  
35 the Figure 1 and Table 2.

36  
37  
38  
39  
40 Figure 6. Time series of land surface temperature and air temperature at 12 stations in  
41 Sioux Falls, SD. The blue boxes are valid Landsat observations, and red dots are gap-  
42 filled values at each Landsat acquisition date. The black triangles are station air  
43 temperature that match all Landsat acquisition dates within an ARD tile in Atlanta, GA  
44 (<https://www.ncdc.noaa.gov/ghcn-daily-description>). Information of GHCN stations see  
45 the Figure 1 and Table 2.

46  
47  
48  
49  
50  
51 Figure 7. Time series of land surface temperature and air temperature at 12 stations in  
52 Phoenix AZ. The blue boxes are valid Landsat observations, and red dots are gap-filled  
53 values at each Landsat acquisition date. The black triangles are station air temperature  
54 that match all Landsat acquisition dates within an ARD tile in Atlanta, GA  
55 (<https://www.ncdc.noaa.gov/ghcn-daily-description>). Information of GHCN stations see  
56 the Figure 1 and Table 2.

1  
2  
3 Figure 8. Annual clear observation (top) and annual mean of gap-filled uncertainty  
4 (bottom) for three study areas in year 2016 and 2020.  
5  
6

7 Figure 9. Uncertainty of gap-filling models in the three study areas: Atlanta (A), Sioux  
8 Falls (B), and Phoenix (C) by selected NOAA GHCN validation stations in 2020 (top)  
9 and 2016 (bottom). X axis is dates of year and Y axis is uncertainty values from the  
10 gap-filling model. The information of GHCN stations see Figure 1 and Table 2.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

Table 1. Main data sources used in the study.

<b>Name</b>	<b>Type</b>	<b>Resolution</b>	<b>Temporal</b>	<b>Spectral accuracy</b>	<b>Source</b>
Landsat ARD LST Collection 1	LST	30 m	7 days	~0.5 kelvin (vary by pixel)	USGS
GHCN	Air Temp.	points	Daily	-	NOAA
MODIS LST	LST	1000 m	Weekly	1.5~2.5 kelvin	NASA
VIIRS LST	LST	1000 m	Weekly	1.5~2.5 kelvin	NASA
ECOSTRESS	LST	70 m	Daily	1~2 kelvin	NASA
DAYMET	LST	1000 m	monthly	-	ORNL

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

Table 2, NOAA GHCN Station ID (Figure 1), Latitude (Lat.) and Longitude (Lon.), Name, and Land Cover Class (LC, 2020) in three study areas.

Atlanta, GA					Sioux Falls, SD					Phoenix, AZ				
ID	Name	Lat.	Lon.	LC**	ID	Station Name	Lat.	Lon.	LC	ID	Station Name	Lat.	Lon.	LC
1	USC00091640	33.60	-85.08	90	22	USC00211263	43.50	-96.70	81	50	USC00020632	33.81	-111.65	52
2	USC00092485	34.00	-84.75	21	23	USC00216565	43.46	-96.73	81	50	USW00093139	33.82	-111.90	52
3	USC00099486	34.30	-83.86	81	25	USC00390128	43.48	-96.76	21	51	USC00021282	33.60	-111.71	52
4	USC00098740	33.33	-83.70	22	26	USC00391032	43.52	-96.67	22	52	USC00023190	33.55	-111.44	22
5	USW00013874	33.63	-84.44	24	27	USC00391076	43.54	-96.68	81	53	USC00025700	33.56	-111.54	52
6	USC00091665	34.20	-84.79	21	28	USC00392302	43.53	-96.81	23	54	USC00028214	33.43	-111.92	90
7	USC00092180	34.26	-83.49	22	33	USC00392984	43.54	-96.84	21	55	USC00028499	33.60	-112.30	21
8	USC00096335	33.40	-84.91	71	35	USC00394037	43.58	-96.80	21	56	USC00029634	33.43	-112.00	24
9	USC00092006	34.17	-84.73	43	36	USW00014944	43.52	-96.75	24	57	USW00023183	33.34	-112.15	22
10	USC00098950	33.87	-83.54	81	37	USC00395090	43.50	-96.79	81	58	USC00027281	33.46	-111.48	52
11	USC00092318	33.60	-83.84	21	40	USC00399042	43.53	-96.75	21	59	USC00020288	33.50	-112.36	21
12	USC00094700	33.53	-84.35	24	41	USC00390422	43.50	-96.67	71	60	USC00025521	33.11	-112.03	22
13	USW00053819	33.36	-84.57	21	44	USW00094950	43.55	-96.66	82	61	USC00025270	33.07	-111.77	71
14	USC00099466	33.93	-83.73	81	45	USC00390281	43.59	-96.73	21	62	USC00027370	33.21	-111.68	24
15	USW00003888	33.78	-84.52	21	45	USC00397666	43.51	-96.67	81	63	USC00021514	33.38	-112.07	21
16	USW00053838	34.27	-83.83	22	46	USC00395671	43.55	-96.63	95	64	USC00028112	33.69	-112.08	22
17	USW00053863	33.88	-84.30	23						65	USW00003184	33.78	-112.52	21
18	USC00093271	33.26	-84.28	23						66	USC00029464	33.62	-111.91	21
19	USW00053873	34.12	-84.85	23						67	USW00003192	33.48	-111.93	24
20	USC00092283	34.23	-84.13	23										

Table 3. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA GHCN stations for annual accuracy assessment, Atlanta, GA.

Station name	Year	LC*	No. Obs. #		$R^2$		RMSE	
			$G^+$ ( $O^{\wedge}$ )	G	O vs. $V^{\ddagger}$	G vs. V	O vs. V	G vs. V
USC00092283		22	23 (19)	31	0.70	0.79	6.31	6.41
USC00099466	1991	81	22 (18)	31	0.84	0.86	4.57	3.92
USC00098740		21	14 (12)	31	0.77	0.78	6.44	6.96
USC00098950		81	30 (28)	79	0.77	0.84	7.57	5.98
USC00092180	2000	22	38 (38)	79	0.63	0.76	7.77	6.13
SW00053863		24	28 (23)	79	0.83	0.81	9.33	11.01
USC00099486		81	24 (26)	86	0.74	0.81	5.61	2.88
USW00053838	2016	23	45 (45)	86	0.88	0.82	5.64	5.68
USC00094700		22	24 (24)	86	0.70	0.85	6.16	8.56
USC00092006		41	60 (38)	83	0.48	0.72	7.81	3.49
USC00092283	2020	23	56 (29)	83	0.44	0.80	9.84	5.81
USC00091965		23	62 (38)	83	0.36	0.82	12.47	6.09

Table 4. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA  
GHCN stations for annual accuracy assessment, Sioux Falls, SD.

Station name	Year	LC	No. Obs.			$R^2$		RMSE	
			G (O)	G	O vs. V	G vs. V	O vs. V	G vs. V	
USW00014944		24	12 (9)	38	0.94	0.79	4.34	7.19	
USC00395090	1991	81	9 (9)	31	0.88	0.75	6.04	9.74	
USC00390128		21	17 (17)	38	0.98	0.86	3.26	6.24	
USC00216565		81	43 (32)	96	0.89	0.58	6.14	10.44	
USC00390281	2000	21	34 (25)	96	0.88	0.74	5.09	5.29	
USC00392984		23	18 (12)	96	0.84	0.56	6.37	10.74	
USW00014944		24	25 (14)	117	0.87	0.76	6.43	8.14	
USC00211263	2016	81	66 (22)	117	0.73	0.59	5.48	10.31	
USC00391076		23	66 (16)	117	0.73	0.78	4.27	8.37	
USC00391032		22	58 (25)	125	0.59	0.74	11.44	9.85	
USC00395090	2020	81	26 (13)	125	0.57	0.68	11.48	8.48	
USW00014944		24	30 (13)	125	0.77	0.70	9.23	8.31	

Table 5. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE) in selected NOAA GHCN stations for annual accuracy assessment, Phoenix, AZ.

Station name	Year	LC	No. Obs.			$R^2$		RMSE	
			G (O)	G	O vs. V	G vs. V	O vs. V	G vs. V	
USW00023183		22	16 (8)	31	0.85	0.71	9.07	9.37	
USW00093139	1991	52	27 (13)	31	0.41	0.81	9.28	7.52	
USC00025270		71	26 (16)	31	0.77	0.72	10.22	9.85	
USC00021282		52	34 (33)	83	0.87	0.88	10.43	11.38	
USC00023190	2000	22	61 (60)	78	0.84	0.85	5.86	5.82	
USC00025512		23	33 (31)	83	0.80	0.79	8.70	8.79	
USW00093139		52	66 (54)	89	0.73	0.79	5.32	5.89	
USW00023183	2016	22	66 (30)	89	0.73	0.79	7.15	6.13	
USW00003192		24	46 (41)	89	0.81	0.85	12.15	12.52	
USC00020288		24	64 (53)	88	0.91	0.89	6.94	9.28	
USC00027281	2020	52	61 (37)	88	0.59	0.89	13.38	7.59	
USC00028499		24	74 (60)	88	0.59	0.85	9.67	7.75	

Table 6. The number of Landsat clear observations and gap-filled uncertainty ( $^{\circ}\text{C}$ ) in 2016 and 2020.

<b>Atlanta, GA (H24V13) Sioux Falls, SD (H16V06) Phoenix, AZ (H07V13)</b>							
<b>Year</b>	<b>Class*</b>	<b>Clear Obs.</b>	<b>Uncertainty</b>	<b>Clear Obs.</b>	<b>Uncertainty</b>	<b>Clear Obs.</b>	<b>Uncertainty</b>
2016	H	1	2.25	1	4.65	4	2.34
	M	30	0.65	25	0.94	41	0.76
	L	61	0.3	46	0.3	70	0.03
2020	H	1	4.95	1	4.71	1	3.68
	M	21	0.97	22	1.08	45	0.68
	L	48	0.3	40	0.3	79	0.3

Table 7. The correlation ( $R^2$ ) and Root Mean Square Error (RMSE), selected NOAA GHCN stations for comparison analysis among GHCN, gap-filled LST, MODIS LST, VIIRS LST, and ECOSTRESS LST by daily, monthly, and seasonal for three study areas in year 2016 and 2020.

Station	Frequency	No. Sample	GHCN vs. Gap-filled		GHCN vs. VIIRS (MODIS)		GHCN vs. ECOSTRESS	
			RMSE	$R^2$	RMSE	$R^2$	RMSE	$R^2$
USC00094648	Daily	84	5.77	0.76	6.35	0.64	10.34	0.42
Atlanta, GA	Monthly	12	7.0	0.97	5.31	0.92	8.72	0.61
2020 (Urban)	Seasonal	4	6.05	0.98	4.89	0.96	8.49	0.74
USC00097827	Daily	84	5.27	0.77	4.88	0.66	4.51	0.30
Atlanta, GA	Monthly	12	4.69	0.98	2.25	0.94	4.87	0.49
2020 (Forest)	Seasonal	4	3.68	0.99	1.91	0.94	5.05	0.46
USC00098740	Daily	87	6.35	0.88	7.51	0.87	-	-
Atlanta, GA	Monthly	12	7.60	0.92	7.68	0.91	-	-
2016 (Urban)	Season	4	7.79	0.94	10.79	0.95	-	-
USW00014944	Daily	125	8.45	0.69	7.57	0.87	10.96	0.57
Sioux Falls, SD	Monthly	12	6.07	0.87	7.16	0.94	6.45	0.80
2020 (Urban)	Season	4	4.02	0.99	6.66	0.97	13.53	0.87
USC00391076	Daily	125	9.25	0.67	8.05	0.78	10.75	0.54
Sioux Falls, SD	Monthly	12	9.25	0.85	8.05	0.90	10.75	0.77
2020 (Agri.)	Season	4	5.8	0.93	6.71	0.93	12.55	0.87
USW00094950	Daily	117	6.7	0.80	5.53	0.84	-	-
Sioux Falls, SD	Monthly	12	4.93	0.95	2.94	0.96	-	-
2016 (Urban)	Seasonal	4	2.70	0.99	1.03	0.98	-	-
USC00023190	Daily	89	4.93	0.87	11.38	0.92	11.05	0.67
Phoenix, AZ	Monthly	12	4.43	0.98	11.16	0.94	13.60	0.43
2020 (Urban)	Season	4	2.50	0.98	6.31	0.96	13.70	0.18
USC00027281	Daily	89	7.04	0.90	9.31	0.89	9.93	0.66
Phoenix, AZ	Monthly	12	7.62	0.99	8.63	0.96	10.02	0.88
2020 (Shrub)	Season	4	3.97	0.99	4.93	0.98	9.59	0.82
USW00003192	Daily	89	6.14	0.89	8.92	0.82	-	-
Phoenix, AZ	Monthly	12	5.85	0.97	7.87	0.95	-	-
2016 (Urban)	Season	4	2.86	0.88	3.34	0.87	-	-

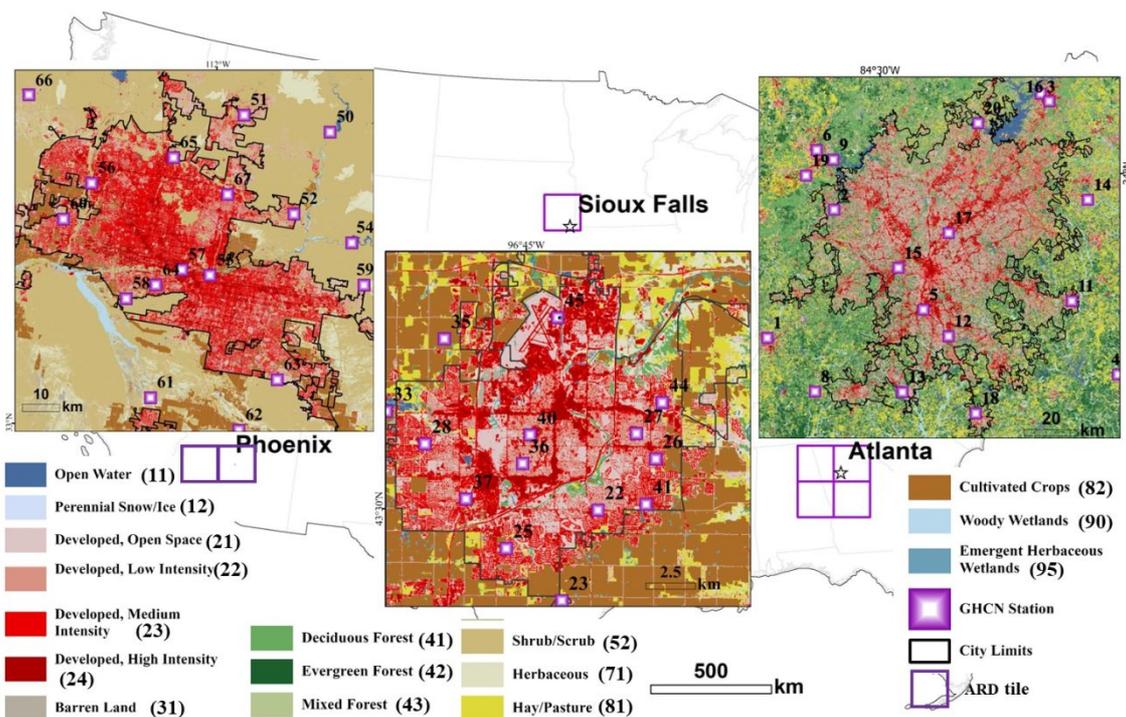


Figure 1. The zoom in 2020 land cover map for three selected study areas within the ARD tiles, GHCN station location are white-purple boxes associate with the ID number in black (the ID, name, land cover class, and detailed information of each station see Table 2). These Land cover map of individual urban centre are not at the same scale.

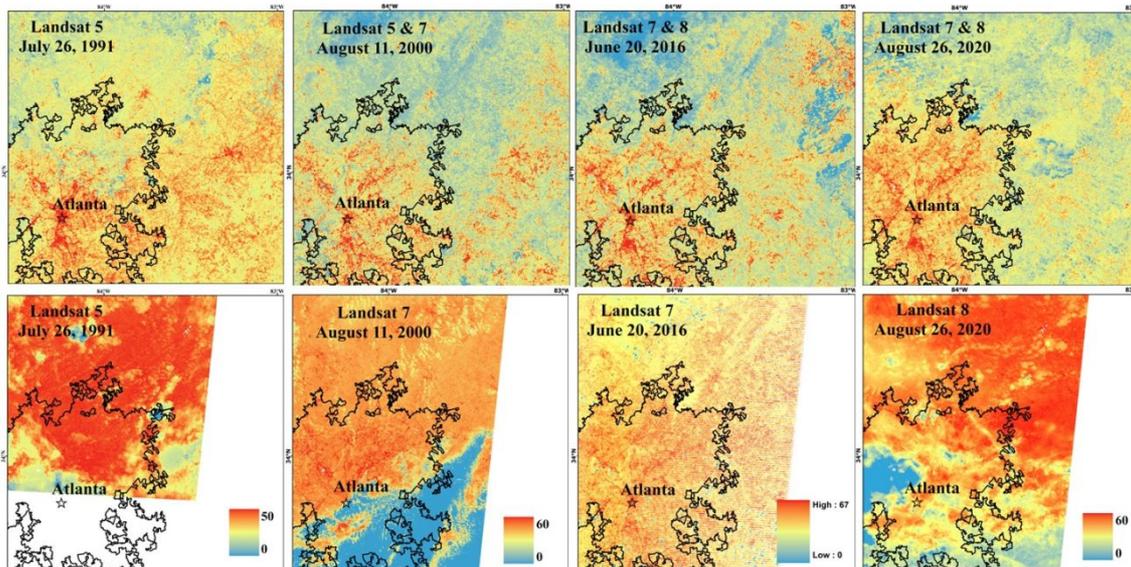


Figure 2. Gap-filled Landsat LST (top) and Original Landsat (5, 7, and 8) LST (bottom) in Atlanta, GA on the selected dates of 1991, 2000, 2016 and 2020 from left to right.

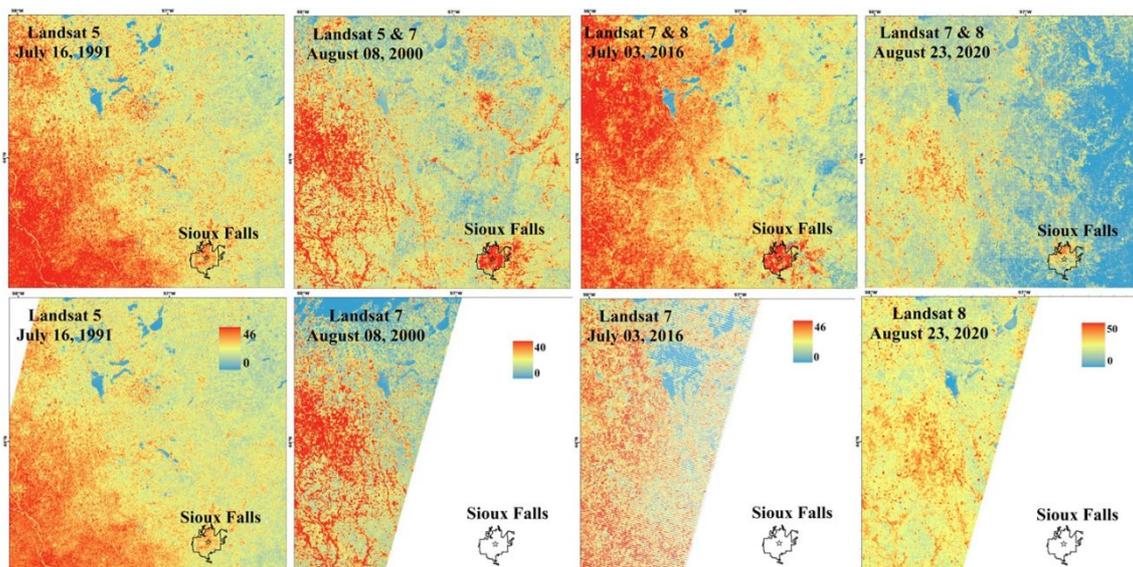


Figure 3. Gap-filled Landsat LST (top) and Original Landsat (5, 7, and 8) LST (bottom) in Sioux Falls, SD on the selected dates of 1991, 2000, 2016 and 2020 from left to right.

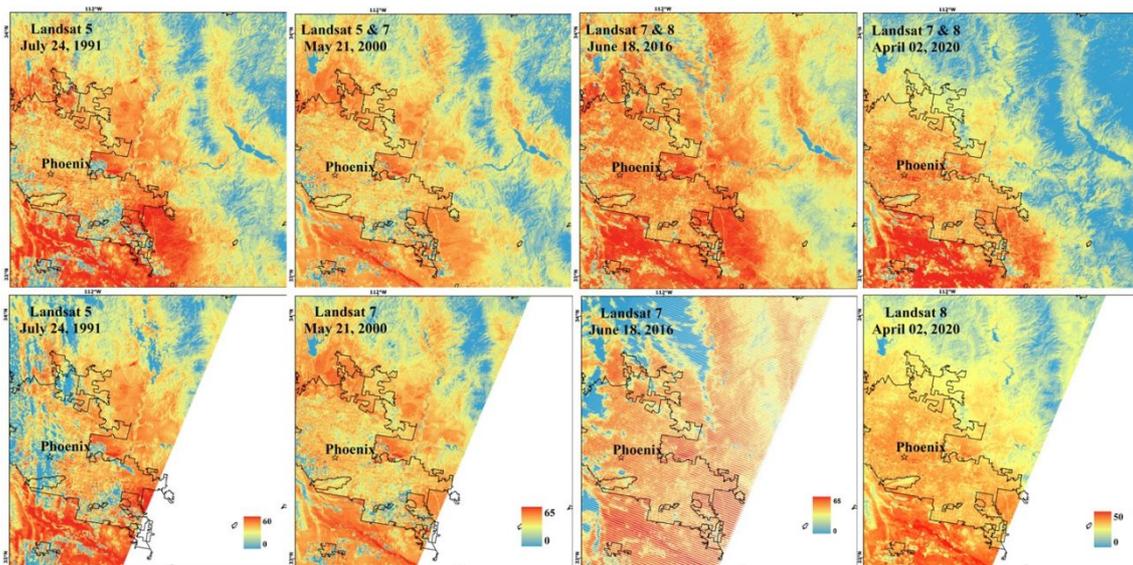


Figure 4. Gap-filled Landsat LST (top) and Original Landsat (5, 7, and 8) LST (bottom) in Phoenix, AZ on the selected dates of 1991, 2000, 2016 and 2020 from left to right.

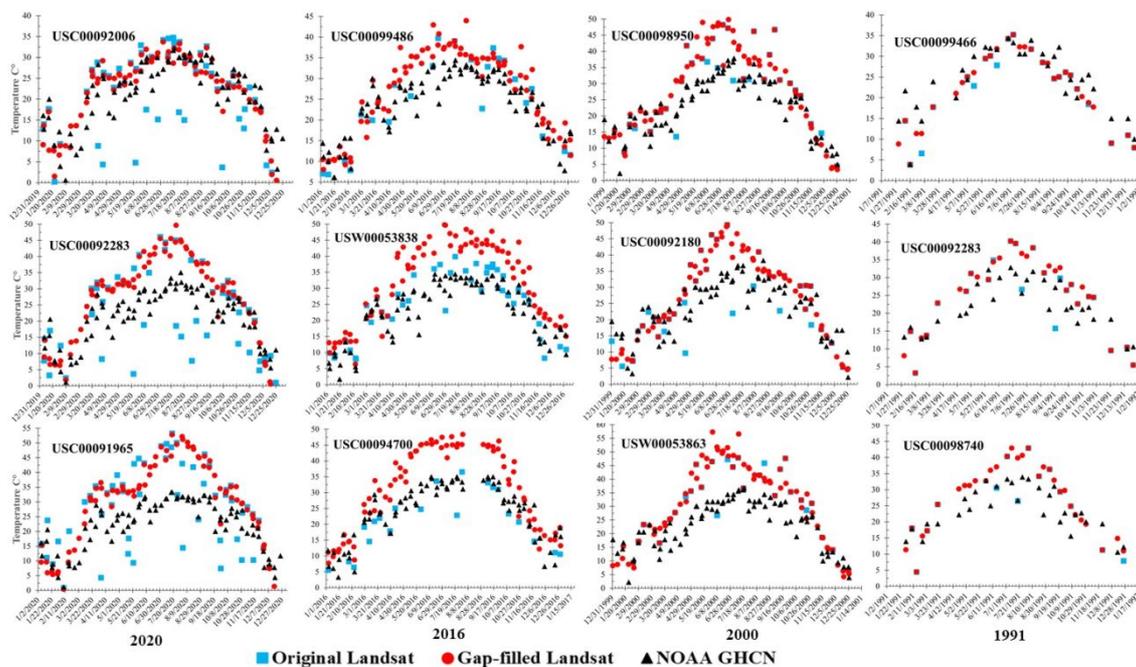


Figure 5. Time series of land surface temperature and air temperature at 12 stations in Atlanta, GA. The blue boxes are valid Landsat observations, and red dots are gap-filled values at each Landsat acquisition date. The black triangles are station air temperature that match all Landsat acquisition dates within an ARD tile in Atlanta, GA (<https://www.ncdc.noaa.gov/ghcn-daily-description>). Information of GHCN stations see the Figure 1 and Table 2.

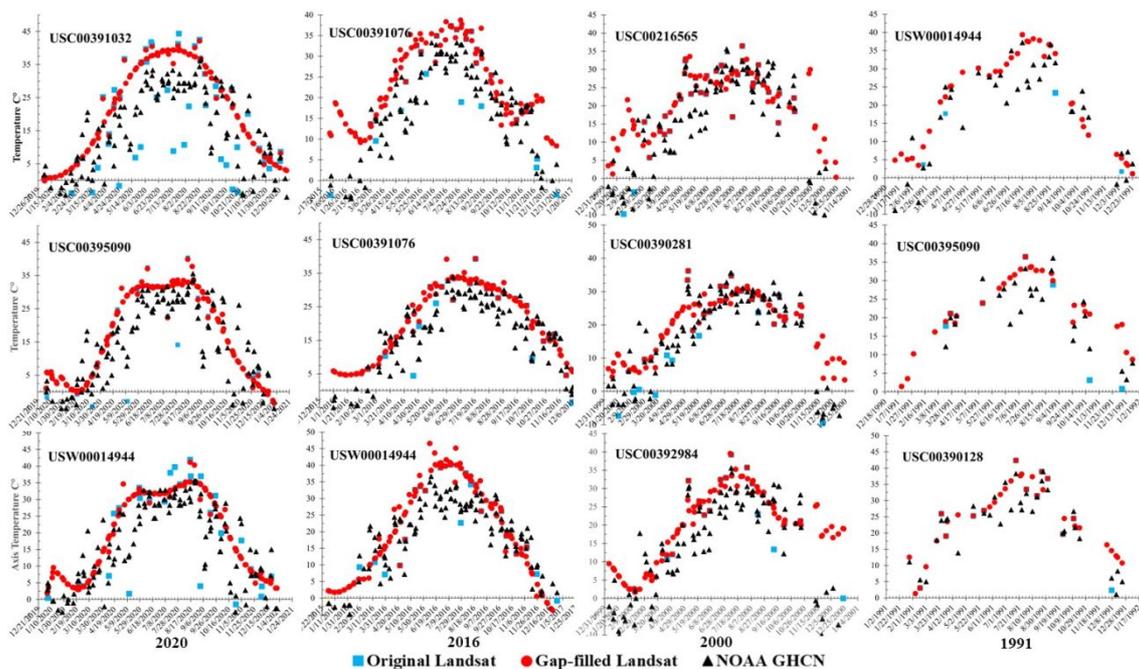


Figure 6. Time series of land surface temperature and air temperature at 12 stations in Sioux Falls, SD. The blue boxes are valid Landsat observations, and red dots are gap-filled values at each Landsat acquisition date. The black triangles are station air temperature that match all Landsat acquisition dates within an ARD tile in Sioux Falls, SD (<https://www.ncdc.noaa.gov/ghcn-daily-description>). Information of GHCN stations see the Figure 1 and Table 2.

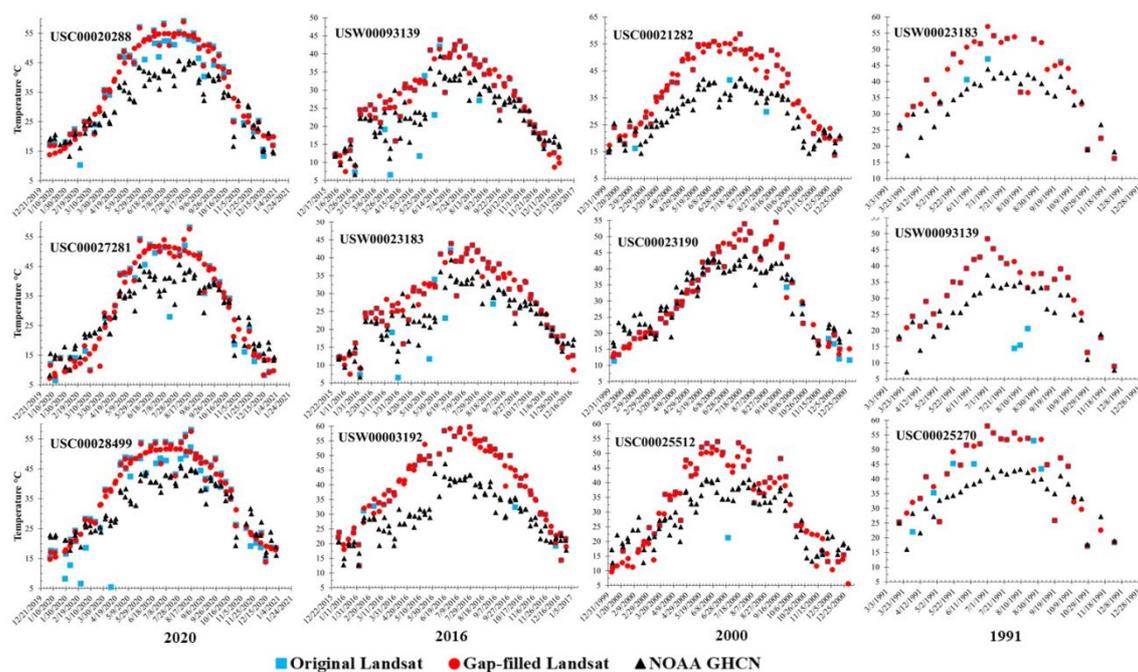


Figure 7. Time series of land surface temperature and air temperature at 12 stations in Phoenix AZ. The blue boxes are valid Landsat observations, and red dots are gap-filled values at each Landsat acquisition date. The black triangles are station air temperature values at each Landsat acquisition date that match all Landsat acquisition dates within an ARD tile in Phoenix, AZ (<https://www.ncdc.noaa.gov/ghcn-daily-description>). Information of GHCN stations see the Figure 1 and Table 2.

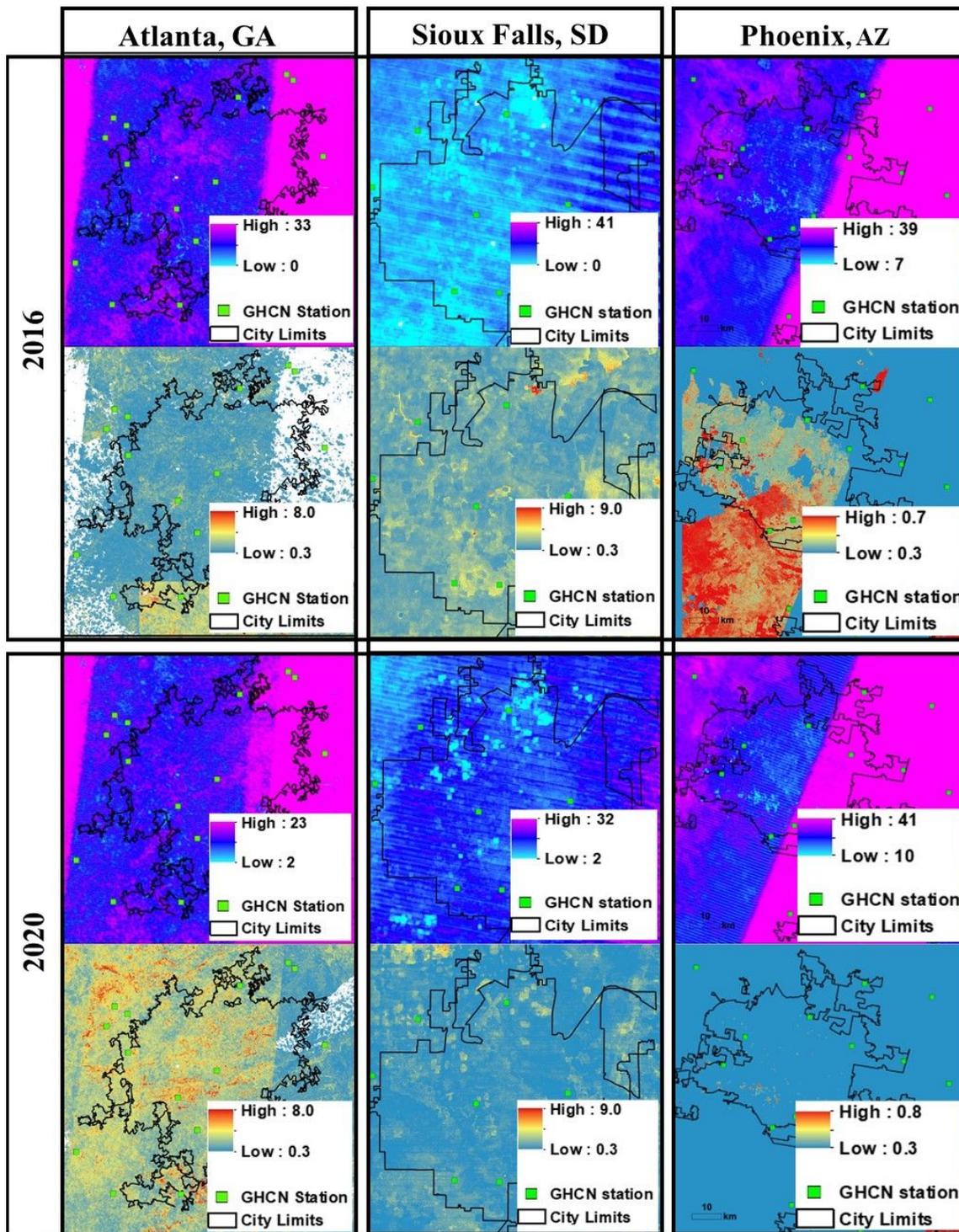


Figure 8. Annual clear observation (top) and annual mean of gap-filled uncertainty (bottom) for three study areas in year 2016 and 2020.

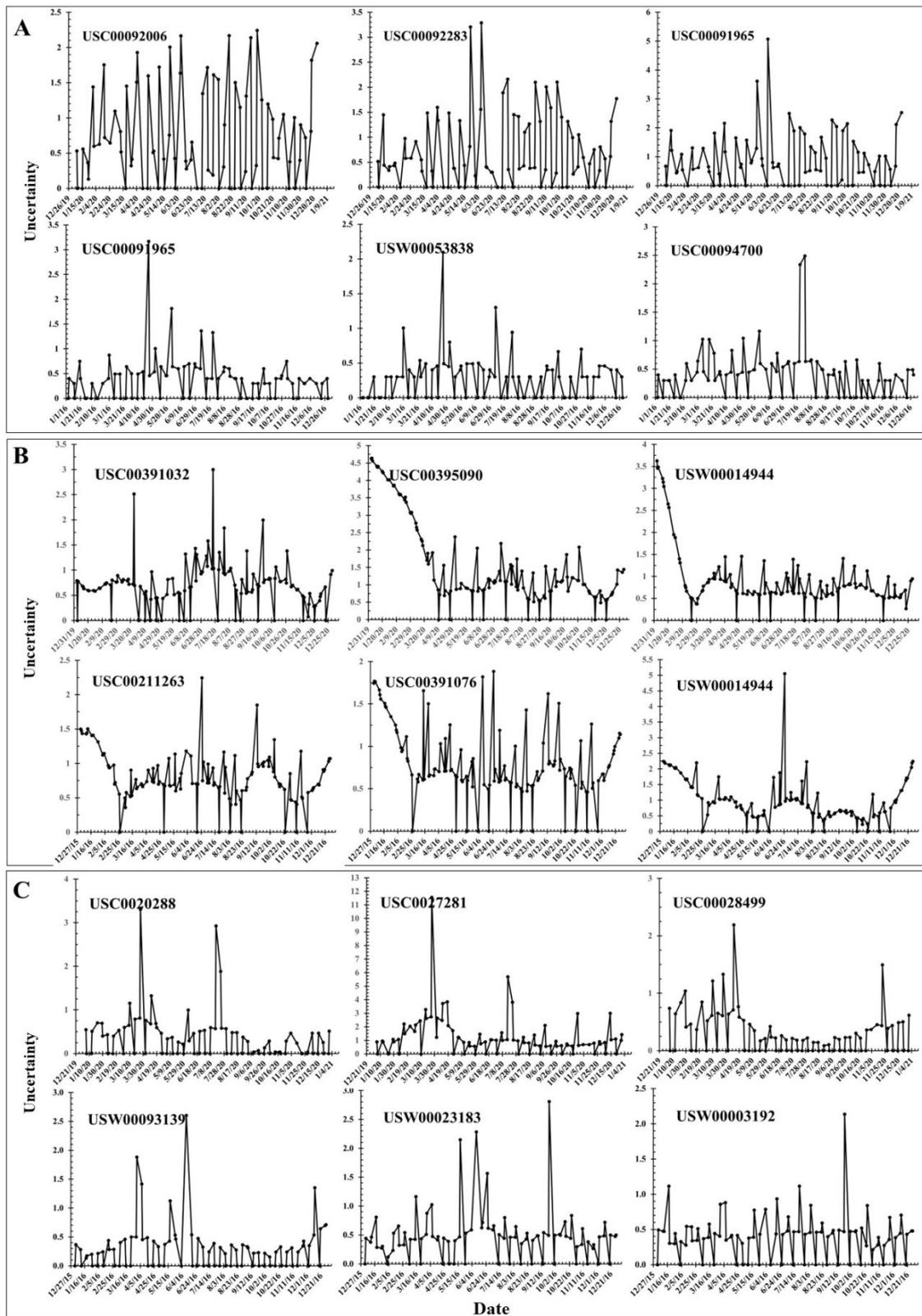


Figure 9. Uncertainty of gap-filling models in the three study areas: Atlanta (A), Sioux Falls (B), and Phoenix (C) by selected NOAA GHCN validation stations in 2020 (top)

1  
2  
3 and 2016 (bottom). X axis is dates of year and Y axis is uncertainty values from the  
4  
5 gap-filling model. The information of GHCN stations see Figure 1 and Table 2.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only