scientific **data**

Check for updates

OPEN

DATA DESCRIPTOR

Gap-free 16-year (2005–2020) sub-diurnal surface meteorological observations across Florida

Julie A. Peeling [1] ✉, Jasmeet Judge[1], Vasubandhu Misra[2], C. B. Jayasankar[2] & William R. Lusher[3]

The sub-tropical, flat, peninsular region of Florida is subject to a unique climate with extreme weather events that impact agriculture, public health, and management of natural resources. Meteorological data at high temporal resolutions especially in tropical latitudes are essential to understand diurnal and semi-diurnal variations of climate, which are considered as the fundamental modes of climate variations of our Earth system. However, many meteorological datasets contain gaps that limit their use for validation of models and further detailed observational analysis. The objective of this paper is to apply a set of data gap filling strategies to develop a gap-free dataset with 15-minute observations for the sub-tropical region of Florida. Using data from the Florida Automated Weather Network (FAWN), methods of linear interpolation, trend continuation, reference to external sources, and nearest station substitution were applied to fill the data gaps depending on the extent of the gap. The outcome of this study provides continuous, publicly accessible surface meteorological observations for 30 FAWN stations at 15-minute intervals for years 2005–2020.

Background & Summary

Data such as rainfall, temperature, wind patterns, and solar radiation are significant meteorological variables in determining climate variations and change. For example, high spatial and temporal resolution rainfall data is necessary for the development of hydrological models, flood risk assessment, land management, and climate model validation[1–4]. In sub-tropical, flat regions such as Florida, slight seasonal climate shifts can have drastic impacts on flooding, agricultural production, and public health[5–8]. Florida is a sub-tropical region with average air temperatures fairly stable in the summer across the state and varying from North to South (increasingly warm) in the winter[8]. During summer months, average temperatures are typically between 24 and 28 °C (297–301 K) across the state, and in winter, Northern Florida averages around 7–13 °C (280–286 K) while the southern part of the state tends to average around 15–19 °C (288–292 K)[9]. The elevation levels in Florida range from sea level to about 105 m above sea level[10]. Due to its peninsular geography in subtropical latitudes and interactions with relatively warm oceans, Florida has a unique climate to the rest of the United States[11,12]. Its wet season is heavily interconnected with fresh water availability and ecosystem functionality, and as population growth continues throughout the state, there is the further strain placed on its natural resources[11].

Within the climate system, the diurnal and semi-diurnal scale variations represent a fundamental mode of variability[13–16]. Diurnal variations are generated from diurnally varying solar heating that affects near the surface, through the depths of the troposphere, and in the stratosphere that manifests as pronounced oscillations with periods of approximately 24 h (diurnal) and 12 h (semi-diurnal). These periodic oscillations that appear in the upper atmosphere are also called atmospheric tides, which significantly impact the diurnal and semi-diurnal variations of many climatic variables[16]. Often, the fidelity of numerical climate and weather models and reanalyses is assessed in the ability to represent the diurnal scales owing to its feature of being a fundamental mode of variation of the climate system (e.g.[17–20]). However, a huge limitation of verifying these models to simulate the

[1]University of Florida, Center for Remote Sensing, Department of Agricultural and Biological Engineering, Gainesville, FL, 32611, USA. [2]Florida State University, Center for Ocean-Atmospheric Prediction Studies, Department of Earth, Ocean and Atmospheric Science, Tallahassee, FL, 32306, USA. [3]University of Florida IFAS Extension, Florida Automated Weather Network, Gainesville, FL, 32611, USA. ✉e-mail: juliepeeling@ufl.edu
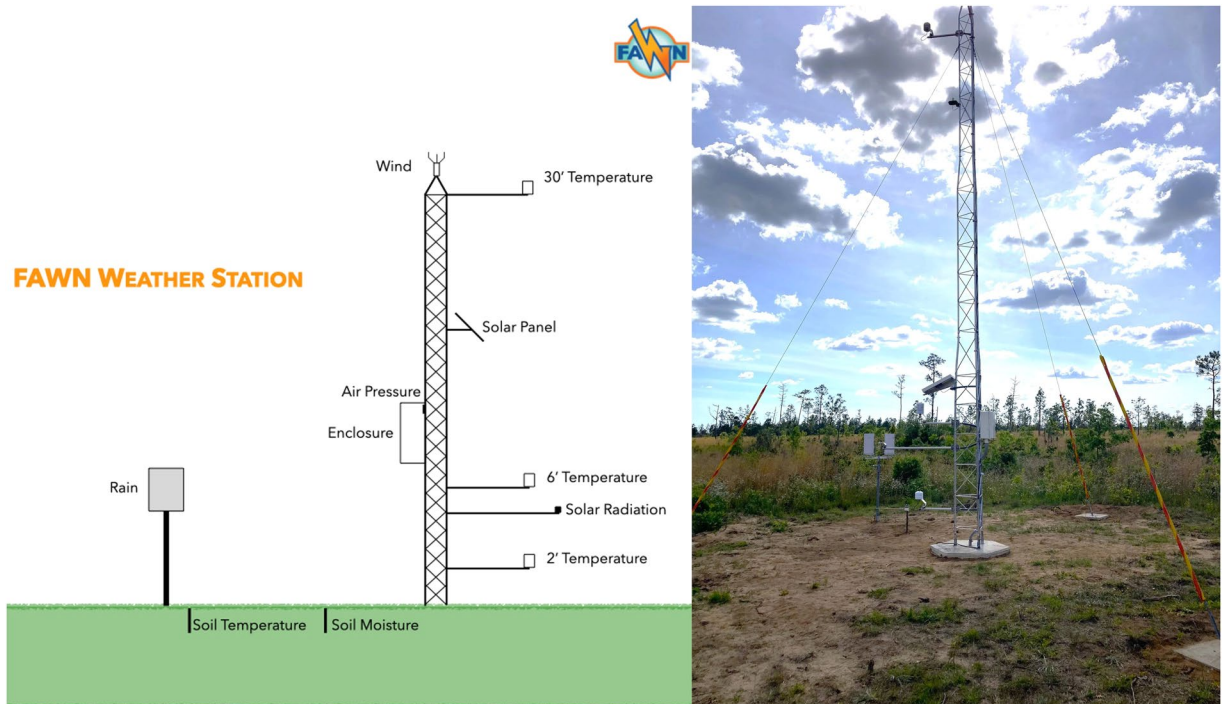
Fig. 1 Typical FAWN tower configuration (left) and photo of a FAWN tower (right).
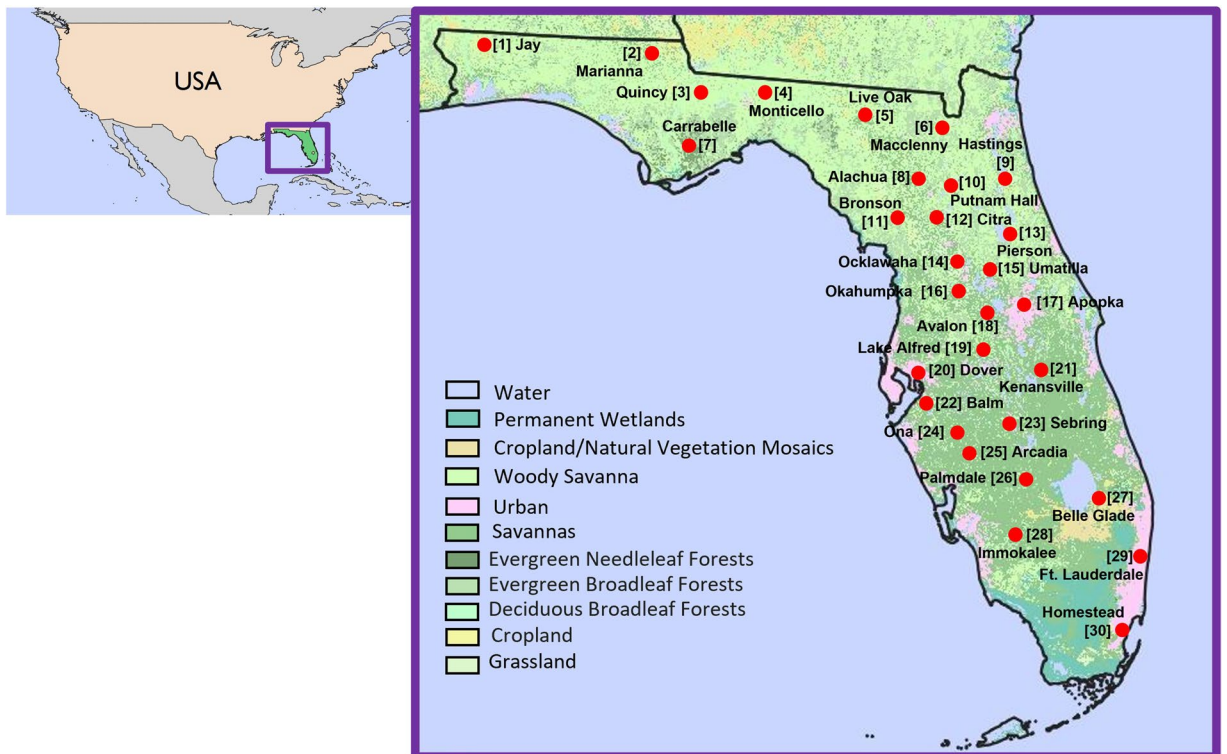

Fig. 2 Location of 30 FAWN stations selected for this study across Florida, along with their names and numbers. The base map is a 2019 Land Cover map from Moderate Resolution Imaging Spectroradiometer (MODIS).

diurnal cycles is the lack of data that robustly resolves the diurnal variations[21]. To examine these variations in sub-tropical areas such as the state of Florida, continuous data is needed at sub-diurnal (hourly or finer) temporal and spatial resolution.

Meteorological Variable	Minimum	Maximum
All Temperatures (°C)	−20	50
Relative Humidity (%)	0	100
Precipitation (in)	0	3
Solar Radiation (W/m²)	0	1200
Wind Speed (mph)	0	75
Wind Direction (°)	0	360

Table 1. Quality control measures implemented by FAWN, adapted from the FAWN measurement system specifications[47].
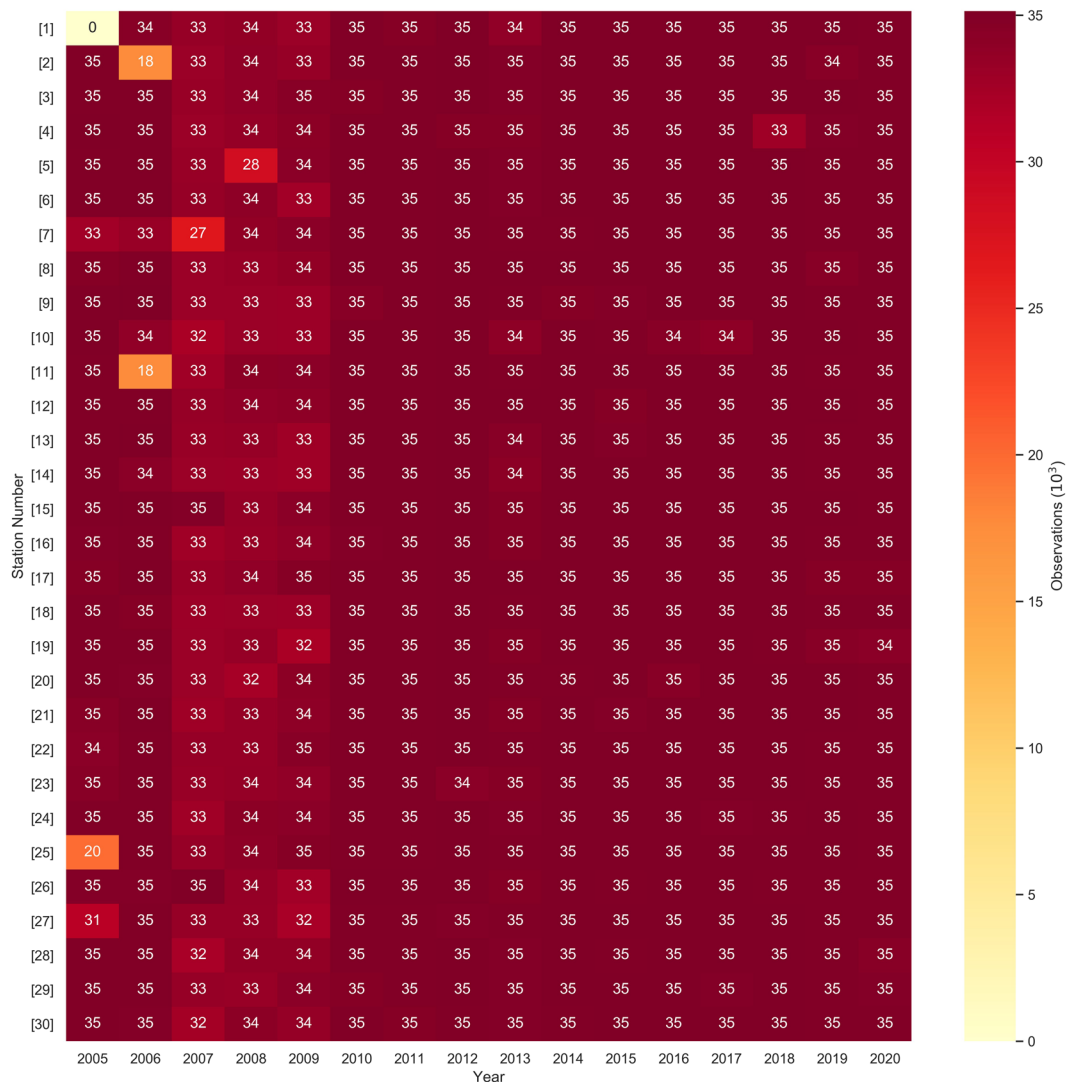

Fig. 3 Heat map representing the number of observations (in thousands) present in the raw data for each station in each year.

The Florida climate is representative of a trade wind regime for latitudes between about 25 degrees North and South of the equator, including monsoon regions such as India and Vietnam[11,22,23]. Areas such as these often lack a high-density observational network, but for the Florida region, many climate and weather datasets exist to provide information at varying spatio-temporal resolutions. For example, the Florida Climate Center, affiliated with the National Climatic Data Center (NCDC), provides daily precipitation and temperatures for approximately 100 stations across Florida[24], and hourly local climatological data (LCD) is available through the National Oceanic and Atmospheric Administration (NOAA)[25,26]. These datasets lack a sub-hourly temporal resolution, limiting their applicability. In addition, there are two precipitation-only data sources. Integrated Multi-Satellite Retrievals for Global Precipitation Mission (IMERG) data from the National Aeronautics and
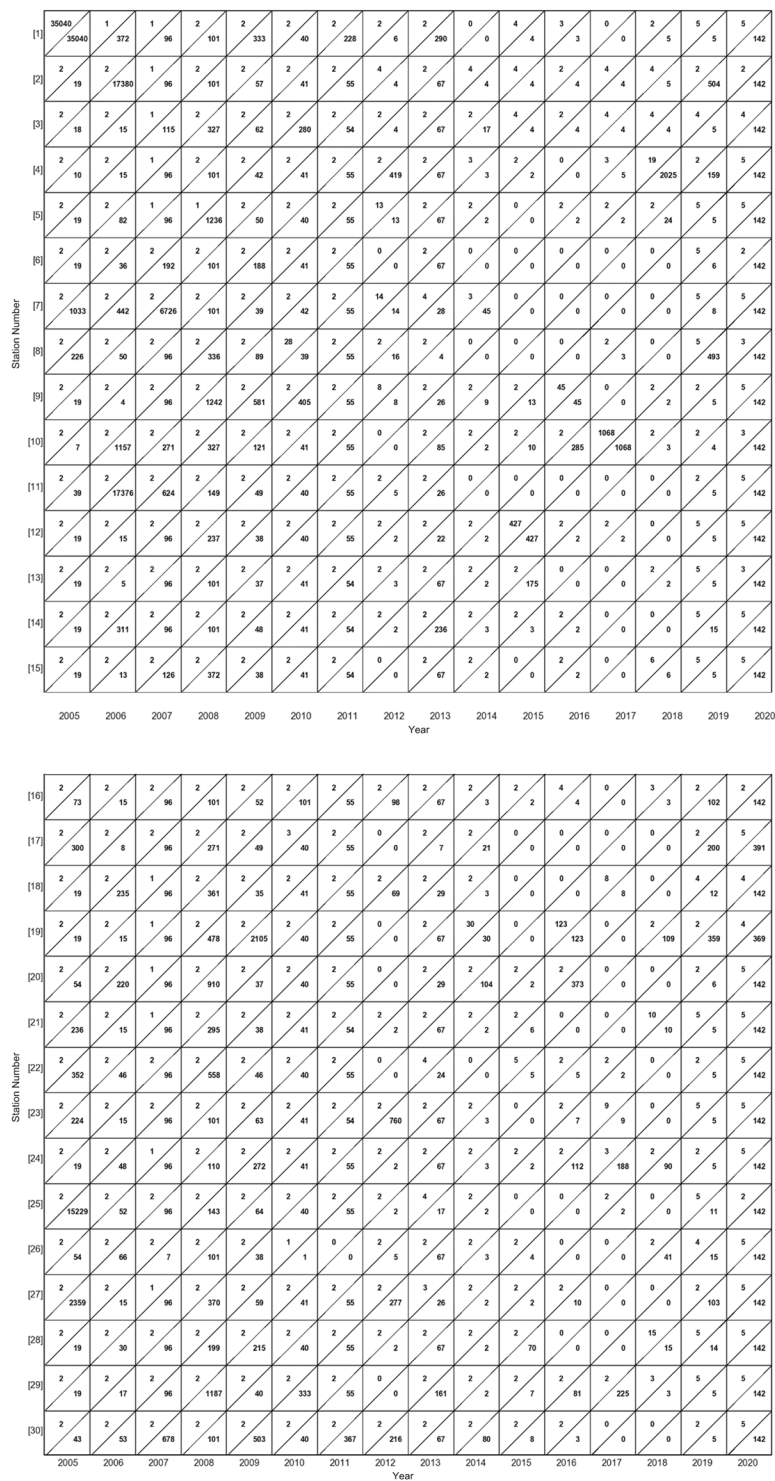
Fig. 4 The minimum (left triangle) and maximum (right triangle) number of 15-minute data gaps present at each station in each year for (**a**) station #s 1–15 and (**b**) station #s 16–30.

Space Administration (NASA) provides 30-minute precipitation data at a 10 km resolution over the period June 2000-present[12]. The second source, from the NCDC, offers 15-minute precipitation observations from stations that are sparsely located (31 stations throughout Florida)[27].

The Florida Automated Weather Network (FAWN) is the only network providing sub-hourly data for 10 meteorological variables[28]. Initiated in 1997 to provide climatic data to rural areas in Florida to inform growers, it is currently comprised of 42 stations[29]. The goal of FAWN is to provide accurate, reliable, and real-time weather data to users across Florida for applications such as cold weather protection strategies, irrigation scheduling, and extreme precipitation analysis[29–32]. Each FAWN tower, as shown in Fig. 1, is equipped with sensors that
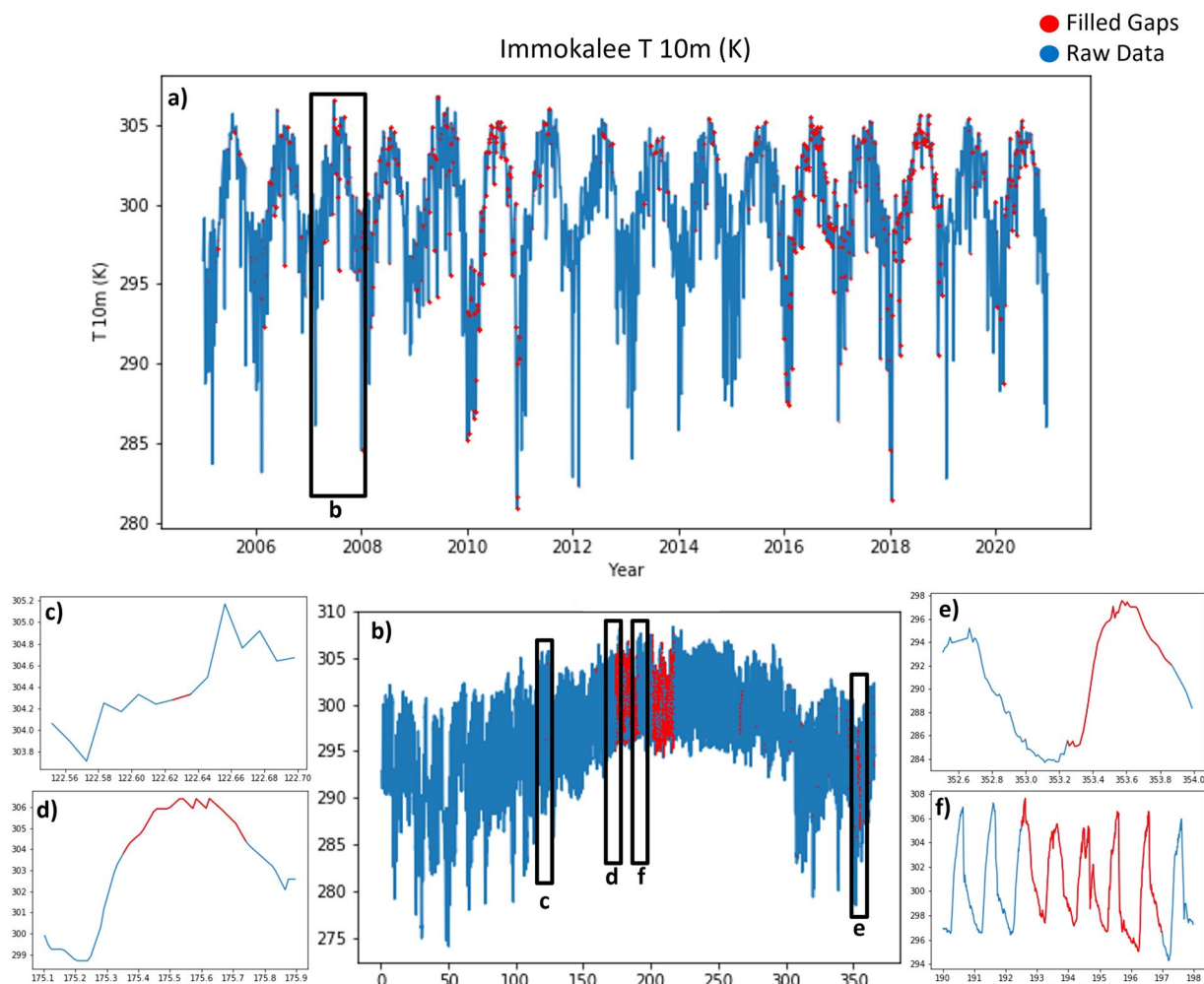
Fig. 5 Examples for the methods used for filling diurnal meteorological variable such as 10 m air temperature at the Immokalee station (#28): (**a**) 16-year time-series data; (**b**) Zoomed in time-series for 2007; (**c**) linear interpolation for gaps less than 6 hours; (**d**) trend continuation method for gaps between 6–12 hours; and (**e**) external data source and (**f**) nearest station for gaps larger than 12 hours.

Meteorological Variable	Description
T 60 cm	Air Temperature at 60 cm (°C)
T 2 m	Air Temperature at 2 m (°C)
T 10 m	Air Temperature at 10 m (°C)
T Dew	Dew Point Temperature (°C)
T Soil	Soil Temperature (°C)
RH	Relative Humidity (%)
PPT	Rainfall Amount (inches)
WS	Wind Speed (mph)
WD	Wind Direction (degrees)
Sol Rad	Solar Radiation (W/m²)

Table 2. Micrometeorological variables and their descriptions as available from FAWN on an annual basis for download.

measure air temperature at 60 cm (T 60 cm), air temperature at 2 m (T 2 m), air temperature at 10 m (T 10 m), soil temperature (T Soil), relative humidity (RH), precipitation (PPT), wind speed (WS), wind direction (WD), solar radiation (Sol Rad), and barometric pressure[29]. The FAWN dataset also includes derived parameters such as dew point temperature (T Dew), wet bulb temperature, and potential evapotranspiration[29]. However, barometric pressure, wet bulb temperature, and potential evapotranspiration are not included in annual datasets. FAWN data is offered at a fine temporal resolution of 15 minutes, but many of these datasets contain gaps of

Meteorological Variable	T-test p-value	F-test p-value	KS p-value
T 60 cm	0.104	0.483	0.149
T 2 m	0.216	0.466	0.289
T 10 m	0.274	0.438	0.374
T Dew	0.549	0.291	0.595
T Soil	0.818	0.460	0.855
RH	0.374	0.343	0.375
PPT	0.812	0.500	0.815
WS	0.683	0.389	0.700
WD	0.705	0.385	0.693
Sol Rad	0.750	0.496	0.750

Table 3. Minimum p-values for each meteorological variable from the statistical significance tests, including the T-test, F-test, and Kolmogorov-Smirnov (K-S) test, on each station.

various sizes, from 15 minutes to one year, due to operational issues that limit their use for applications that require continuous datasets[29]. To address this challenge, this paper details the methods utilized to fill gaps in FAWN meteorological observations to generate a continuous dataset over 2005–2020. Previous studies have employed various data homogenization and gap-filling methods for meteorological variables, such as WS and PPT, across Florida to improve prediction methods and better understand trends in extreme weather[32–34]. The value of the the dataset generated in this study lies in its fine temporal resolution and diverse set of meteorological variables. This study leverages the FAWN infrastructure in order to create a gap-free dataset for wider scientific applications in regions of similar characteristics. The publicly available dataset will provide a unique resource within a complex sub-tropical region for climate analysis and modeling.

Methods

Data acquisition and preprocessing.   Yearly observations at 15-minute intervals were obtained from FAWN for all active stations[28]. In this study, we examined the FAWN data available between 1997 and 2020 and selected the stations with data present across the longest period of time during which the most stations were available, resulting in the chosen 30 stations over 2005–2020, as shown in Fig. 2. In the northern part of the State, 16 stations were located in forested and woody environments, and in the South, nine stations were in areas classified as savanna. Four of the stations were positioned in urban areas, and one station was located in cropland.

FAWN implements initial quality control measures and filtering before publishing the raw data, details for which are given in Table 1. Annual tests are conducted to determine if repair or replacement of sensors is needed based upon EPA guidelines, and filtering of these potential operational incidents as well as power failures result in data gaps[35]. In this study, supplemental quality control mechanisms were implemented to enhance the data reliability. For all temperature measurements, if there was a difference >5 °C within one time step of 15 minutes, the data point was marked as a data gap that was filled as described below. For WS higher than 30 mph, the event was manually checked against nearby FAWN stations and LCD reports to confirm high WS. If the high WS value was confirmed, then it remained in the data set, and if it could not be verified, the value was marked as a data gap. Additionally, RH values of 0% were marked as data gaps.

Gap filling.   Data gaps occurred if the difference between two consecutive data points was greater than 15 minutes. The number of 15-minute observations in the raw data for each station at each year is given in Fig. 3 to provide insight into the amount of data points present and the extent of missing data. Figure 4a,b provide the minimum and maximum number of consecutive 15-minute data gaps missing for each station in each year, respectively, to demonstrate the distribution of data gap extents across space and time. Station #s 7 and 8 had a larger number of 15-minute gaps than other stations, with the most gaps occurring in 2007 and 2008, respectively. The years 2007–2009 had the most gaps for all stations, with over 1000 gaps for most stations during that period. Large data gaps such as these are primarily due to operational issues, generally from power failures. Gap filling of meteorological variables is inherently uncertain and challenging, with differing methodological approaches for different variables[36–41]. We applied several methods of data gap filling based upon gap size and the nature of the meteorological variables.

Datasets with diurnal cycles.   Gap filling for datasets with diurnal cycles such as temperatures, RH, and Sol Rad followed the same methodology. Figure 5a depicts an example of gap filling for T 10 m over the study period for station #28. The year 2007 had the most gaps for this station (see Fig. 5b), and various gap filling techniques were applied based on the gap size. For data gaps <6 hours (about 82% of gaps), *linear interpolation* was implemented using the slope between the two data points at either end of the gap to estimate the missing data points (see Fig. 5c)[38–40]. Such temporal interpolation is a reliable data gap filling method in continuous climate variables such as near-surface air temperature and solar radiation data[38–40]. For gaps between 6 and 12 hours (about 1% of gaps), *trend continuation* for the meteorological variables was implemented, similar to Tardivo and Berti[42] and Kemp *et al.*[43], by extracting the data values and trends from two days prior to and two days after gaps (see Fig. 5d). At each missing time step, the measurement was filled with the average values at that particular time from the surrounding days. When the data gaps were greater than 12 hours (about 17% of gaps), an outside data source was referenced. These large gaps mainly occurred in the years 2005–2009 for most stations (see Figs. 3, 4).
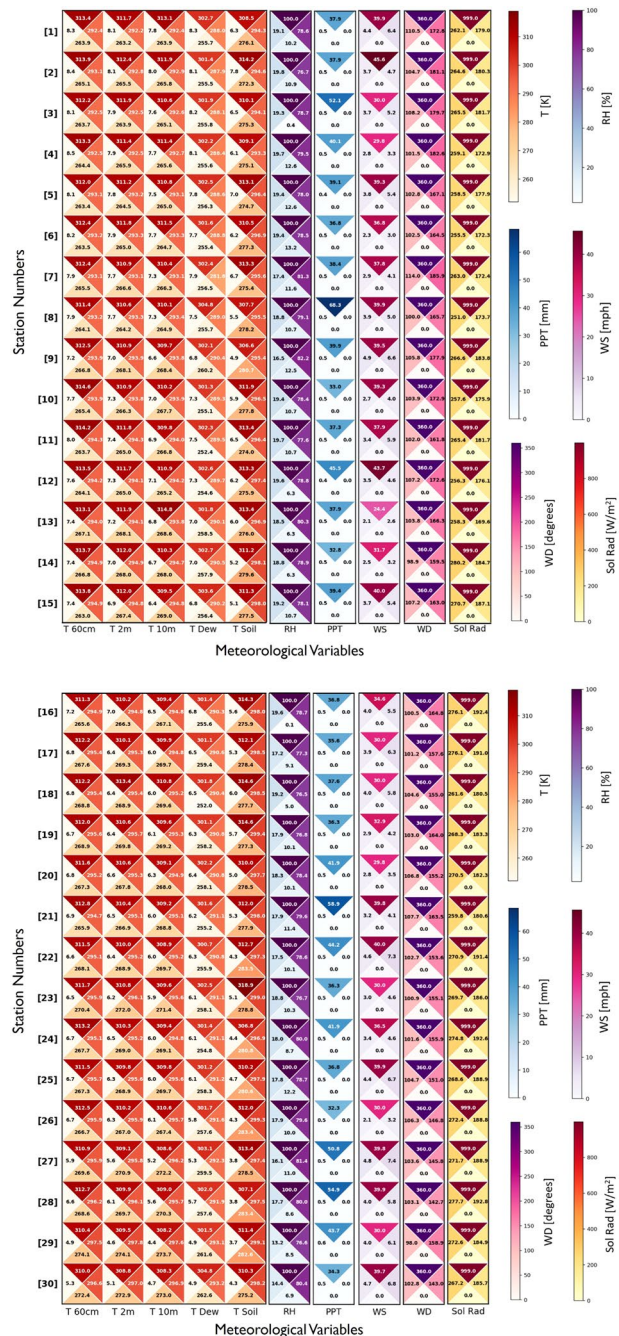
Fig. 6 Statistical description for the 10 meteorological variables at (**a**) station #s 1–15 and (**b**) station #s 16–30, provided in four triangles. In the clockwise direction from the top, each triangle provides the maximum, mean, minimum, and standard deviation of the gap-filled dataset over the 16-year period.

In this study, LCD from NOAA[25] was used as an *external data source* to fill these large data gaps with data from weather stations within the same city, or if not available then the same county (see Fig. 5e). Hourly LCD values were linearly interpolated to 15-minute intervals. Since LCD is available only for T 10 m, T Dew, and RH, larger data gaps in T 60 cm, T 2 m, T Soil and Sol Rad were filled using data from the nearest FAWN station (see Fig. 5f)[37,38,41]. The nearest station was determined through the smallest euclidean distance to a station with available data. The temporal correlations were high between the monthly means of these nearest stations. The nearest station method was also applied for any periods when there were gaps in the LCD, following Luedeling[38] and Graf[41]. In this study, the distance to the nearest station was typically around 32 km.

Discrete datasets. To fill gaps in the discrete datasets such as PPT, WS, and WD, LCD and nearest FAWN stations were used, similar to the larger data gaps mentioned above[37,38,41]. The NCDC PPT data could not be used due to large distances from FAWN stations and lack of observations consistent with the FAWN and LCD PPT

observations. Given the distribution of available FAWN stations, it was reasonable to assume that gradients of observed data for these meteorological variables were captured by filling the gaps using the nearest station.

Data Records

Gap-free data for 30 FAWN stations over the period 2005–2020 are available through Figshare, an open access repository, in CSV file format titled "*Florida Automated Weather Network Yearly CSV Data (Gap Free)*"[44]. The data is continuous over 16 years for each station listed in Fig. 2, and annual data within the given time period can be downloaded. There are 10 gap-filled meteorological variables provided in the datasets, the units and labels of which are given in Table 2.

Technical Validation

In addition to visual inspection of filled data such as comparing diurnal patterns with surrounding days, the validity of the data was assessed to ensure consistency between the filled data and the raw data for each station and meteorological variable. This was assessed by conducting differential statistics between the raw data and the filled data. A two-tailed T-test on the means of each meteorological variable at each station was conducted to determine whether the mean of the filled data differed significantly from the mean of the raw data[40]. This test was chosen as one source of validation in order to ensure that the gap filling process did not significantly alter the mean of the filled data as compared to the raw data. All p-values resulting from the T-test were >0.1, so there was no significant difference found between the filled data means and raw data means (see Table 3 for minimum p-values).

Figure 6a,b provide the mean, along with the standard deviation, minimum, and maximum values, for the 10 meteorological variables at each station. As expected, the mean air temperature values increase from station #1, at around 292 K, to station #30, at around 297 K (North to South). The maximum PPT was highest, between 52.1 mm and 68.3 mm, at station #s 3, 8, 21, and 28, providing information on the areas which received the highest intensity rainfall within a 15-minute period over the study period. The standard deviation of the temperature values tended to decrease from station 1 (around 8 K) to 30 (around 5 K), supporting higher temperature variability in the more northern stations.

To test the difference in standard deviations between the meteorological variables at each station in the filled dataset and raw dataset, an F-test was implemented. As we determined that the means of the filled and raw data were not significantly different, this test was conducted to reveal whether the dispersal of values around the averages of each dataset significantly varied. These p-values resulting from the F-test were also all >0.1, indicating no notable difference in standard deviations.

In order to test the statistical difference in distribution of the raw and filled datasets, the Kolmogorov-Smirnov (K-S) test was implemented. This test essentially checks whether two datasets come from the same distribution, and the test statistic can be interpreted to represent the greatest distance between the cumulative distribution function of each dataset[45]. Thus, the K-S test was chosen as a third validation metric to determine if there existed significant difference between the shape and spread of the filled and raw datasets. The resulting p-values from the K-S test showed no such difference, as they were all > 0.1.

Usage Notes

The gap-filled dataset generated through this work is unmatched in temporal resolution and spatial extent across the state of Florida. It provides information on 10 meteorological variables at 15-minute intervals, spanning 30 stations from as far north as Jay (latitude 31°N) to Homestead in the south (latitude 25.5°N). It also has potential applications in climate monitoring, agriculture, and hydrology. The gap free data can be applied to understand climate variability and verify numerical climate and weather models, which can be used to predict future weather conditions from current observation[46]. The continuous 16-year data product developed through the methods outlined above can serve as an important resource for climate research and forecasting in sub-tropical regions such as Florida[11,21].

Code availability

No custom software was used to process the data described in this paper. The open-source software used to conduct this study was Python version 3.7.6. The packages and libraries used included Numpy (V 1.18.1), Pandas (V 1.3.0), Matplotlib (V 3.1.3), and Scipy (V 1.4.1). Specific functions for the statistical analysis including the T-test, F-test, and Kolmogorov-Smirnov functions were conducted using the Scipy stats module.

References

1. Huang, Y.-F. *et al*. Hourly rainfall data from rain gauge networks and weather radar up to 2020 across the hawaiian islands. *Scientific Data* **9**, 334, https://doi.org/10.1038/s41597-022-01430-2 (2022).
2. Ochoa-Tocachi, B. F. *et al*. High-resolution hydrometeorological data from a network of headwater catchments in the tropical andes. *Scientific Data* **5**, 180080, https://doi.org/10.1038/sdata.2018.80 (2018).
3. Takhellambam, B. S. *et al*. Temporal disaggregation of hourly precipitation under changing climate over the southeast united states. *Scientific Data* **9**, 211, https://doi.org/10.1038/s41597-022-01304-7 (2022).
4. Dai, A. The diurnal cycle from observations and ERA5 in surface pressure, temperature, humidity, and winds. *Climate Dynamics* **61**, 2965–2990, https://doi.org/10.1007/s00382-023-06721-x (2023).
5. Dewan, A. *et al*. Developing a high-resolution gridded rainfall product for bangladesh during 1901–2018. *Scientific Data* **9**, 471, https://doi.org/10.1038/s41597-022-01568-z (2022).

6. Raimi, D., Keyes, A. & Kingdon, C. Florida climate outlook: Assessing physical and economic impacts through 2040. Tech. Rep., Resources for the Future (2020).

7. Misra, V. *et al.* The florida water and climate alliance (FloridaWCA): Developing a stakeholder-scientist partnership to create actionable science in climate adaptation and water resource management. *Bulletin of the American Meteorological Society* **102**, E367–E382, https://doi.org/10.1175/BAMS-D-19-0302.1 (2021).

8. Misra, V. & Bhardwaj, A. The impact of varying seasonal lengths of the rainy seasons of india on its teleconnections with tropical sea surface temperatures. *Atmospheric Science Letters* **21**, https://doi.org/10.1002/asl.959 (2020).

9. Shin, S., Her, Y., Zhang, G. & Lusher, W. What does florida weather during the past 20 years look like? florida weather represented by the florida automated weather network (fawn). *EDIS* https://doi.org/10.32473/edis-ae537-2020 (2020).

10. South Florida Regional Council. Assessment report of regional differences within florida: Within the context of the trainers and tools project: Building coastal flood hazard resiliency in florida's regional planning council communities (2016).

11. Misra, V., Bhardwaj, A. & Mishra, A. Characterizing the rainy season of peninsular florida. *Climate Dynamics* **51**, 2157–2167, https://doi.org/10.1007/s00382-017-4005-2 (2027).

12. Misra, V., Jayasankar, C. B., Beasley, P. & Bhardwaj, A. Operational monitoring of the evolution of rainy season over florida. *Frontiers in Climate* **4**, 793959, https://doi.org/10.3389/fclim.2022.793959 (2022).

13. Yang, G.-Y. & Slingo, J. The diurnal cycle in the tropics. *Monthly Weather Review* **129**, 784–801, 10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2 (2001).

14. Deser, C. & Smith, C. A. Diurnal and semidiurnal variations of the surface wind field over the tropical pacific ocean. *Journal of Climate* **11**, 1730–1748, 10.1175/1520-0442(1998)011<1730:DASVOT>2.0.CO;2 (1998).

15. Dai, A. & Deser, C. Diurnal and semidiurnal variations in global surface wind and divergence fields. *Journal of Geophysical Research: Atmospheres* **104**, 31109–31125, https://doi.org/10.1029/1999JD900927 (1999).

16. Chapman, S. & Lindzen, R. S. *Atmospheric Tides*, vol. 7 (Dordrecht, Holland: Reidel, 1970).

17. Lin, X., Randall, D. A. & Fowler, L. D. Diurnal variability of the hydrologic cycle and radiative fluxes: Comparisons between observations and a gcm. *Journal of Climate* **13**, 4159–4179, 10.1175/1520-0442(2000)013<4159:DVOTHC>2.0.CO;2 (2000).

18. Trenberth, K. E., Dai, A., Rasmusssen, R. M. & Parsons, D. B. The changing character of precipitation. *Bulletin of the American Meteorological Society* **84**, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205 (2003).

19. Dai, A. & Trenberth, K. E. The diurnal cycle and its depiction in the community climate system model. *Journal of Climate* **17**, 930–951, 10.1175/1520-0442(2004)017<0930:TDCAID>2.0.CO;2 (2004).

20. Lee, M.-I. *et al.* An analysis of the warm-season diurnal cycle over the continental united states and northern mexico in general circulation models. *Journal of Hydrometeorology* **8**, 344–366, https://doi.org/10.1175/JHM581.1 (2007).

21. Dai, A. Recent climatology, variability and trends in global surface humidity. *Journal of Climate* **19**, 3589–3606, https://doi.org/10.1175/JCLI3816.1 (2006).

22. Misra, V., Bhardwaj, A. & Noska, R. Understanding the variations of the length and the seasonal rainfall anomalies of the indian summer monsoon. *Journal of Climate* **30**, 1753–1763, https://doi.org/10.1175/JCLI-D-16-0501.1 (2017).

23. Vu, T. M., Raghavan, S. V., Liong, S.-Y. & Mishra, A. K. Uncertainties of gridded precipitation observations in characterizing spatio-temporal drought and wetness over vietnam. *International Journal of Climatology* **38**, 2067–2081, https://doi.org/10.1002/joc.5317 (2018).

24. *Climate Data Access Tools - Florida Climate Center*. https://climatecenter.fsu.edu/climate-data-access-tools (2023).

25. *Local Climatological Data (LCD)*. https://www.ncei.noaa.gov/products/land-based-station/local-climatological-data (2021).

26. Mukherjee, S. & Nateghi, R. Climate, weather, socio-economic and electricity usage data for the residential and commercial sectors in FL, u.s. *Data in Brief* **13**, 192–195, https://doi.org/10.1016/j.dib.2017.05.031 (2017).

27. *U.S. 15-Minute Precipitation Data*. https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00505 (2021).

28. *FAWN - Florida Automated Weather Network*. https://fawn.ifas.ufl.edu/ (2023).

29. Lusher, W., Jackson, J. & Morgan, K. The florida automated weather network (FAWN): Ten years of providing weather information to florida growers. *Proc. Fla. State Hort. Soc.* **121**, 69–74 (2008).

30. Jackson, J., Morgan, K. & Lusher, W. Citrus cold weather protection and irrigation scheduling tools using florida automated weather network data. *Proceedings of the Florida State Horticultural Society* (2008).

31. Morgan, K. T. & Lusher, W. R. Use of real-time florida automated weather network data in support of strawberry cold protection and irrigation scheduling. *Proceedings of the Florida State Horticultural Society* **124**, 174–176 (2011).

32. Zhang, C., Wu, S., Li, T., Yu, Z. & Bian, J. Interpreting the trends of extreme precipitation in florida through pressure change. *Remote Sensing* **14**, 1410, https://doi.org/10.3390/rs14061410 (2022).

33. Huang, W. & Xu, S. Neural network and harmonic analysis for recovering missing extreme water-level data during hurricanes in florida. *Journal of Coastal Research* **2009**, 417–426, https://doi.org/10.2112/07-0863.1 (2009).

34. Peng, G., Mooers, C. N. K. & Graber, H. C. Coastal winds in south florida. *Journal of Applied Meteorology and Climatology* **38**, 1740–1757, 10.1175/1520-0450(1999)038<1740:CWISF>2.0.CO;2 (1999).

35. United States Environmental Protection Agency. *Guidance for data quality assessment* (2000).

36. Longman, R. J. *et al.* Data Descriptor: Compilation of climate data from heterogeneous networks across the Hawaiian Islands. *Scientific Data* **5**, https://doi.org/10.1038/sdata.2018.12 (2018).

37. Richard, A. *et al.* Filling gaps in micro-meteorological data. In Dong, Y., Ifrim, G., Mladenić, D., Saunders, C. & Van Hoecke, S. (eds.) *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track, Lecture Notes in Computer Science*, 101–117, https://doi.org/10.1007/978-3-030-67670-4_7 (Springer International Publishing, 2021).

38. Luedeling, E. *Chapter 10 Filling gaps in temperature records | Tree phenology analysis with R* (Eike Luedeling, 2022).

39. Denhard, A., Bandyopadhyay, S., Habte, A. & Sengupta, M. Evaluation of time-series gap-filling methods for solar irradiance applications. *Tech. Rep. NREL/TP-5D00-79987, 1826664, MainId:41192, National Renewable Energy Laboratory*. https://doi.org/10.2172/1826664 (2021).

40. Henn, B., Raleigh, M. S., Fisher, A. & Lundquist, J. D. A comparison of methods for filling gaps in hourly near-surface air temperature data. *Journal of Hydrometeorology* **14**, 929–945, https://doi.org/10.1175/JHM-D-12-027.1 (2013).

41. Graf, A. Gap-filling meteorological variables with Empirical Orthogonal Functions. In *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, 8491 (2017).

42. Tardivo, G. & Berti, A. A dynamic method for gap filling in daily temperature datasets. *Journal of Applied Meteorology and Climatology* **51**, 1079–1086, https://doi.org/10.1175/JAMC-D-11-0117.1 (2012).

43. Kemp, W. P., Burnell, D. G., Everson, D. O. & Thomson, A. J. Estimating missing daily maximum and minimum temperatures. *Journal of Applied Meteorology and Climatology* **22**, 1587–1593, 10.1175/1520-0450(1983)022<1587:EMDMAM>2.0.CO;2 (1983).

44. Peeling, J. A., Judge, J., Misra, V., Jayasankar, C. B. & Lusher, W. R. Florida Automated Weather Network Yearly CSV Data (Gap Free). *figshare* https://doi.org/10.6084/m9.figshare.c.6899269 (2023).

45. Nguyen, C. D., Carlin, J. B. & Lee, K. J. Diagnosing problems with imputation models using the kolmogorov-smirnov test: a simulation study. *BMC Medical Research Methodology* **13**, 144, https://doi.org/10.1186/1471-2288-13-144 (2013).

46. Mitra, A. K. *Use of Remote Sensing in Weather and Climate Forecasts*, 77–96 (Springer Nature, 2023).

47. *FAWN Measurement System Specifications*. https://fawn.ifas.ufl.edu/tour/fawn_info/ (2023).

Acknowledgements

Author contributions

J.J., V.M. and J.P. were responsible for the conceptualization; R.L. and J.P. were responsible for data handling and quality control; J.P., J.J., V.M. and J.B. developed the methodology; J.B., J.J., V.M. and J.P. contributed to visualization and writing (the original draft); All authors contributed to reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that influenced the work reported in this paper.

Additional information

Correspondence and requests for materials should be addressed to J.A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.