Multi-Source Data Driven Forecasting of Extreme Heat Events Using an ARIMA–XGBoost Hybrid Framework

Dandi Liao, Lina Xu*, Pengfei Long, Yuxuan Wang, Qiying Dong, Siyu Liu, Xincai Chang.

School of Geophysics and Geomatics, China University of Geosciences, Wuhan, China.

Keywords: Extreme Heat Events (EHEs); Multi-source Data; ARIMA; XGBoost; Disaster Management.

Abstract

Extreme heat events (EHEs) pose growing risks to densely populated subtropical cities such as Hong Kong, yet there remains a need for lightweight, interpretable tools that can provide multi-day forecasts based on readily available observations. This study develops a multi-source data driven framework that integrates aerosol optical depth (AOD), land surface temperature (LST), precipitable water (PW), and precipitation (Precip), together with ARIMA-based anomaly features, to predict EHEs over Hong Kong. Using a seven-day sliding window, independent XGBoost classifiers are trained to forecast daily EHE occurrence probabilities for the next 1–5 days over ten climate years (March 2015–February 2025). A lead-specific threshold optimization on a validation subset is applied to maximize F1-score. Test results show that AUC values for Lead 1–Lead 5 remain between 0.935 and 0.883, with F1-scores between 0.738 and 0.639, indicating robust predictability up to five days in advance. A process-scale duration inference method based on the leading continuous segment of the predicted sequence achieves 67.08% exact-match accuracy, 77.69% accuracy within ±1 day, and a mean absolute error of 0.75 days. The proposed framework is computationally efficient and operationally relevant, offering practical support for urban heat early warning and risk management.

1. Introduction

In recent years, global climate change has led to more frequent and persistent extreme heat events (EHEs), posing substantial threats to densely populated urban regions (Perkins-Kirkpatrick & Lewis, 2020; IPCC, 2021; Zhang et al., 2023). Located in the subtropical monsoon zone, Hong Kong is strongly influenced by land–sea breeze circulations, humidity fluctuations, and atmospheric energy exchanges. As a result, high-temperature episodes often exhibit abrupt and unstable characteristics. Producing refined forecasts is therefore essential not only for urban meteorological monitoring, but also for risk management, public health protection, and energy system operation. In high-density urban environments, EHEs can trigger heat-related illnesses, increase building energy demand, and compromise the reliability of critical infrastructure. Scientific prediction of such events is thus of marked social and economic importance. The increasing availability of multi-source remote sensing and ground-based observations further enhances the feasibility of accurate EHE prediction (Chen et al., 2018; Wu et al., 2024; Yao et al., 2024).

Traditional extreme weather prediction approaches often rely on numerical weather prediction (NWP) products or computationally intensive deep learning architectures. Although these methods have established physical foundations and strong fitting capabilities, they typically suffer from high computational costs, strong dependence on input conditions, and limited interpretability (Wulff & Domeisen, 2019; Jia et al., 2024; Li et al., 2026). Motivated by these challenges, we develop a lightweight, multi-source data-driven EHE prediction framework tailored for Hong Kong. The model integrates four environmental factors, they are aerosol optical depth (AOD), land surface temperature (LST), precipitable water (PW), and precipitation (Precip). The model also incorporates ARIMA-based time series anomaly detection to construct a multi-dimensional feature set combining both raw and anomaly indicators. Long-term samples are extracted using a sliding-window mechanism, and multi-step independent XGBoost models are employed to forecast daily EHE occurrence probabilities for the next 1–5 days. Furthermore, considering that EHEs commonly manifest as consecutive multi-day processes, we propose a process-scale duration inference method based on the sequence of predicted daily probabilities to estimate the potential duration of each event, enabling dual predictions of both whether an EHE will occur and how long it may persist.

Compared with existing studies, this work makes several key contributions.

(1) It develops a prediction framework that achieves both high interpretability and computational efficiency, enabling stable multi-step forecasting without relying on complex deep models.

(2) An ARIMA-based anomaly detection module is introduced to enhance the signal expressiveness of environmental predictors, providing effective support for the early identification of impending heat events.

(3) A duration-inference method is proposed to estimate the potential number of consecutive hot days, addressing the limitation of daily binary classifiers that are unable to directly represent process-level characteristics.

(4) Through systematic experiments using daily multi-source remote sensing and meteorological data over Hong Kong from

https://doi.org/10.5194/isprs-annals-XI-3-2026-911-2026

2015 to 2025, this study yields a practically applicable forecasting scheme with strong predictability suitable for real-world operations.

These findings of this study hold potential applications for urban meteorological forecasting, natural hazard early warning, and city-scale safety management. Moreover, the proposed framework provides a scalable and generalizable technical pathway for future multi-source-data-based studies on extreme event prediction.

2. Study Area and Data

2.1 Study area

The study area encompasses the entirety of the Hong Kong Special Administrative Region, located along the southeastern coast of China (approximately 22°10′–22°36′N, 113°50′–114°24′E). Characterized by hot and humid summers and pronounced temperature variations during autumn and winter, Hong Kong experiences complex microclimatic conditions shaped by land–sea breeze circulations, heterogeneous topography, and a highly urbanized built environment. These factors collectively contribute to strong spatial variability in the occurrence and intensity of extreme heat events across the region.

The temporal scope of this study spans ten years, from 1 March 2015 to 28 February 2025, with a daily temporal resolution. To ensure scientific rigor and maintain consistency in model training, the analysis adopts a “climate-year” framework, in which each year is defined from March of a given year to February of the following year. Accordingly, this study covers ten consecutive climate years from 2015 to 2024.

2.2 Data

This study focuses on four environmental factors: aerosol optical depth (AOD), land surface temperature (LST), precipitable water (PW), and precipitation (Precip).

Aerosol optical depth is used to characterize aerosol loading. AOD quantifies the attenuation of solar radiation by aerosols within the atmospheric column, with larger values indicating higher aerosol concentrations. Because aerosol measurements are highly sensitive to weather conditions, single-source datasets often suffer from substantial data gaps. To mitigate this issue, we integrate the Himawari-8/9 Level-2 aerosol retrieval product (ARP030) and Level-1.5 AERONET observations (Holben et al., 1998), covering July 2015–February 2025 and March 2015–February 2025, respectively. For Himawari-8/9, daily imagery at 11:00 local time (UTC+8) was downloaded, and all grid cells within the whole Hong Kong region were spatially averaged to obtain a regional AOD value. For AERONET, we used in-situ measurements from the Hong Kong PolyU and Hong Kong Sheung Shui stations. After filtering and quality assurance, the Himawari dataset contains 2,083 valid daily records and the AERONET dataset contains 929 valid records, corresponding to 57.02% and 25.43% of the total number of days within the study period, respectively. The merged dataset contains 2,259 valid AOD records, accounting for 61.84% of all days from March 2015 to February 2025.

Land surface temperature (LST) is represented using the daily LST parameter from the MODIS MOD11A1 product (Wan et al., 2015). Following the same procedure used for Himawari-based AOD, all grid cells within Hong Kong were extracted and averaged to obtain a domain-representative LST. Within the ten-year study period, the MOD11A1 product provides 3,653 daily entries; however, only 2,050 of these contain valid LST observations over the Hong Kong region, corresponding to 56.12% of all days.

Precipitable water (PW) is derived from the Integrated Global Radiosonde Archive (IGRA) and represents the total atmospheric water vapor content from the land surface to the 500 hPa level. A total of 3,644 valid PW records are available from the Kowloon IGRA station during the study period, accounting for 99.75% of all days.

Precipitation (Precip) is obtained from the NASA Climate Prediction Center (CPC) Global Unified Gauge-Based Analysis of Daily Precipitation dataset (Chen et al., 2008), representing the actual daily rainfall amount. Following the procedure used for the other factors, all grid cells within the Hong Kong region were extracted and averaged to obtain a domain-level precipitation estimate. Although the dataset provides valid precipitation records for all 3,653 days within the ten-year study period, 1,960 of these days report zero rainfall. Thus, only 1,693 days (46.35%) contain measurable precipitation.

Table 1 summarizes the four environmental factors compiled for this study.

Factors	Days with valid record	Percentage
AOD	2,259	61.84%
LST	2,050	56.12%
PW	3,644	99.75%
Precip	3,653	100%

Table 1. Number and percentage of valid data.

The definition of Extreme Heat Events (EHEs) follows the criterion used by the Hong Kong Observatory (HKO) for “Very Hot Weather”, which is issued when the maximum air temperature recorded at the HKO Headquarters reaches 33°C or above. The EHE occurrence dataset used in this study is based on the official Very Hot Weather Warnings issued by HKO. Over the ten-year study period, a total of 635 EHE days were recorded, representing 17.38% of all days.

For subsequent model training, all environmental factors were normalized to the range of 0–1. The EHE labels were binarized, where days with an issued Very Hot Weather warning were assigned a value of 1, and other days were assigned a value of 0.

https://doi.org/10.5194/isprs-annals-XI-3-2026-911-2026 | © Author(s) 2026. CC BY 4.0 License.

3. Methodology

The framework of this study consists of four main components: environmental-factor anomaly detection, time-series sample construction, multi-step daily prediction modelling, and process-scale duration inference. The overall methodological workflow is illustrated in Figure 1.

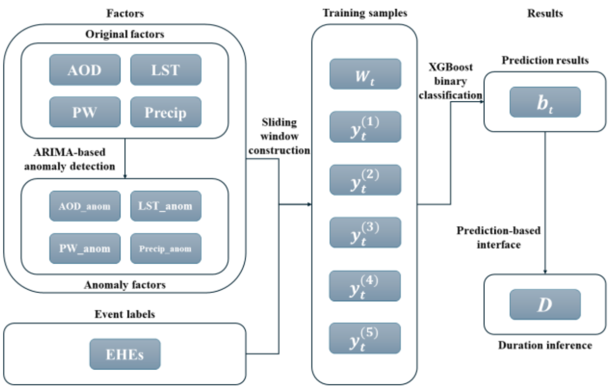

Figure 1. Methodological workflow.

3.1 ARIMA-based Anomaly Detection

The time series of environmental variables (AOD, LST, PW, and Precip) contain long-term trends, seasonal components, and short-term fluctuations. To enhance the model's sensitivity to abrupt changes—such as rapid warming or sudden humidity shifts—this study applies an ARIMA model to each factor on a daily basis and derives anomaly features from the prediction residuals.

Let the original factor sequence be denoted as x_t, and the ARIMA-based prediction as $\hat{x}_t$. The anomaly series is defined as the residual:

$$a_t = x_t - \hat{x}_t \quad (1)$$

To quantify the magnitude of anomalies, the residuals are normalized using their standard deviation σ_a:

$$z_t = \frac{a_t}{\sigma_a} \quad (2)$$

After normalization, values satisfying $|z_t| > 2$ are treated as significant anomalies, and a corresponding binary anomaly indicator is generated for each factor. For example, the anomaly component of AOD is represented as *AOD_anom*. The four original factors and their anomaly counterparts together form an eight-dimensional feature vector [*AOD, LST, PW, Precip, AOD_anom, LST_anom, PW_anom, Precip_anom*], which is used in the subsequent modelling stages. These anomaly features effectively enhance the model's ability to capture rapid precursor signals leading up to extreme heat events.

3.2 Sliding Window Construction

Extreme heat events typically exhibit a distinct evolution pattern. The pre-event accumulation of thermal energy, variations in atmospheric moisture, and land-surface heating processes often determine whether an event will occur. To enable the model to learn these precursor signals, this study adopts a sliding-window approach to construct input samples from the preceding time series.

Let the 8-dimensional feature vector on day t be denoted as $X_t = [x_{t,1}, x_{t,2}, \dots, x_{t,8}]$. The input window covers the previous seven days including day t, forming $W_t = [X_{t-6}, X_{t-5}, \dots, X_t]$.

To support multi-step forecasting and obtain predictions for the subsequent five days, five lead times (Lead 1–5) are used as the corresponding event labels:

$$y_t^{(k)} = EHE_{t+k} \quad (3)$$

where $EHE_t \in \{0,1\}$ denotes the extreme heat event flag issued by HKO, and k =1,2,3,4,5 represents the 1–5-day lead.

The resulting training samples take the form of $(W_t, y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(5)})$.

Missing values (NaN) present in the time series are preserved and accompanied by a separate missing-value mask matrix during model training, enabling the model to learn the realistic missing-data patterns inherent in the observational records.

The dataset is divided based on 1 March 2023, with the first eight climate years used for training and the subsequent two climate years used for testing.

3.3 Multi-step Daily Prediction with XGBoost

To determine whether an extreme heat event will occur on each of the next 1–5 days, an independent XGBoost binary classifier $P(y_t^{(k)} = 1 \mid W_t)$ is trained for each lead time. Rather than applying a fixed threshold of 0.5, the model outputs daily occurrence probabilities that are subsequently post-processed to determine lead-specific optimal decision thresholds.

After model training, a validation set extracted from the tail of the training set (i.e., the last two climate years within the training range) is used to evaluate threshold performance. By scanning through a range of probability thresholds and computing the corresponding F1-scores, the threshold that maximizes the F1-score is selected as the optimal decision threshold for each lead.

3.4 Process-scale Duration Inference

Extreme heat events often occur as multi-day episodes, and their duration is an important metric in operational forecasting (Ren et al., 2021). Directly training a multi-class model to predict event duration poses challenges due to severe class imbalance and the scarcity of long-duration cases. To address this, we develop a process-scale duration inference method that estimates event length based on the sequence of daily occurrence probabilities.

Given the five daily probabilities $\boldsymbol{p} = [p_1, p_2, p_3, p_4, p_5]$ produced by the lead-specific models, each probability is compared with its lead-dependent optimal threshold to generate a binary prediction sequence $\boldsymbol{B} = [b_1, b_2, b_3, b_4, b_5]$.

The inferred duration corresponds to the length of the first continuous block of predicted occurrences, computed as:

$$D = \max(n \mid b_1 = b_2 = \cdots = b_n = 1). \tag{4}$$

For example, the sequences [11100] and [11011] yield inferred durations of 3 and 2 days respectively, while [00011] corresponds to a duration of 0 days.

This focus on the leading continuous segment is strategically chosen to align with the requirements of real-time early warning. In operational disaster management, the primary objective of a five-day forecast is to characterize the immediate upcoming heat event. While subsequent predicted blocks may represent potential recurring events, they are treated as distinct meteorological processes that would be more accurately captured and updated in the following days' rolling forecasts. Focusing on the first continuous block ensures that the duration inference provides the most urgent and reliable temporal guidance for immediate emergency response.

4. Results

After model training, we evaluated the proposed framework using the test set and obtained a series of performance outcomes. This section presents the results from two perspectives: (1) multi-step daily prediction performance and (2) duration inference performance.

4.1 Multi-step Daily Prediction Performance

The evaluation of the multi-step daily prediction models shows that predictive skill decreases gradually with lead time, yet remains overall stable (Figure 2). The Lead-1 model achieves the highest accuracy, with an AUC of 0.935 and an F1-score of 0.738. These results suggest that the seven-day sliding window provides sufficient precursor information for identifying EHE onset, as this duration effectively captures the characteristic synoptic-scale evolution of atmospheric and land-surface anomalies. Preliminary sensitivity tests also indicated that shorter windows (e.g., 3–5 days) were insufficient to captured precursor signals, while longer windows introduced redundant noise with limited gains in F1-scores.

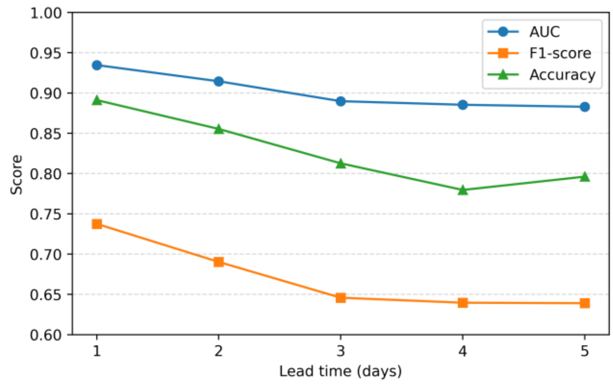

Figure 2. Multi-lead EHEs prediction scores.

Further analysis of the intermediate lead times (Lead-2 to Lead-4) reveals a gradual and consistent decline in predictive performance. This attenuation is consistent with the stochastic nature of atmospheric perturbations, where the accumulation of non-linear errors typically leads to a reduction in predictability over time. Despite this trend, the performance metrics do not exhibit a sharp drop; even at Lead-5, the model maintains an AUC of 0.883 and an F1-score of 0.639. The persistence of meaningful predictive capability at longer lead times indicates that the integration of multi-source features successfully captures large-scale signals, which possess higher inertia and longer memory than localized, short-term fluctuations. This ensures that the framework provides reliable early warning information even five days in advance.

To further illustrate how the seven-day sliding window captures these precursor signals, Figure 3 displays the daily variations of the four normalized environmental factors during a representative period preceding a five-day EHE in 2016. In this visualization, data points enclosed in black circles indicate significant anomalies identified by the ARIMA-based detection.

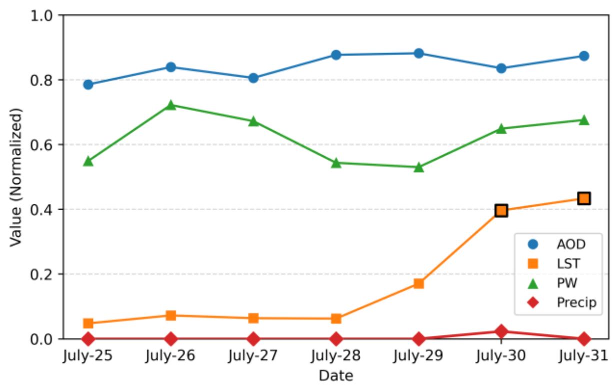

Figure 3. Seven-day temporal evolution of multi-source environmental factors and ARIMA-based anomalies preceding a typical EHE in 2016.

As illustrated in the figure, each factor exhibits a distinct evolutionary trajectory within this seven-day window. AOD fluctuates slightly at a sustained high level (normalized values > 0.8) without triggering any anomaly flags, suggesting a stable presence of aerosol loading. Similarly, PW shows relatively larger fluctuations at high levels, yet these variations remain within the expected range and do not result in significant anomalies. In contrast, LST displays a sharp and monotonic increase throughout the period; the magnitude of this rise is so pronounced that the last two days (July 30 and July 31) are identified as significant anomalies by the ARIMA module. Meanwhile, Precip remains at an absolute low level, with values of zero for six out of the seven days. Although an extremely small non-zero value is observed on July 30, it does not constitute an anomaly. These multidimensional precursors, particularly the rapid intensification of LST and the persistent lack of precipitation, collectively provide the model with a clear signal of the impending extreme heat event.

The temporal variations captured in the seven-day window provide a physical basis for the observed multi-step predictive performance. EHEs in Hong Kong are closely associated with the sustained accumulation of thermal energy, characterized by persistent aerosol loading and a lack of significant precipitation. As shown in the case study, AOD and PW maintain consistently high levels throughout the pre-event phase, providing steady precursor signals that the XGBoost models can effectively utilize even at longer lead times (Lead-4 and Lead-5). Meanwhile, the rapid and monotonic intensification of LST, particularly the significant anomalies detected towards the end of the window (July 30–31), offers a more immediate and decisive indicator of the impending event. The model's ability to integrate these multi-day evolutionary trends with the ARIMA-based anomaly detection contributes to the higher precision observed at shorter lead times (Lead-1 and Lead-2).

Overall, the proposed multi-step daily prediction framework achieves stable and operationally meaningful performance across the entire five-day forecasting horizon. By combining raw environmental factors with ARIMA-derived anomaly indicators, the framework demonstrates that a lightweight machine learning approach can successfully capture the critical transition from background environmental loading to extreme heat onset. This predictive capability holds significant practical value for urban heat-health early warning, energy demand forecasting, and city-scale emergency operations management, providing a computationally efficient tool for proactive disaster risk reduction.

4.2 Duration Interface Performance

The evaluation results of the duration inference model, derived from lead-specific daily probabilities, are summarized in Table 2. The framework demonstrates robust performance within the five-day predictability horizon, achieving an overall exact-match accuracy of 67.08%. This high level of precision is significantly supported by the model's exceptional ability to identify non-EHE periods (Duration = 0), where it reaches an accuracy of 84.19%, reflecting strong stability in excluding non-event days.

Duration (Days)	True Count	Prediction Count	Hit Count	Recall
0	487	420	410	84.19%
1	65	46	12	18.46%
2	54	48	9	16.67%
3	29	50	5	17.24%
4	26	37	2	7.69%
5	65	125	49	75.38%
Overall	726	726	487	67.08%

Table 2. Performance of process-scale duration inference across different EHE durations.

However, it is observed that the exact-match accuracy for specific short-term EHE durations (1–4 days) remains relatively low. This indicates a current limitation of the model in precisely distinguishing between different short-duration categories, which serves as a key direction for future optimization. Potential improvements may involve refining the feature engineering process to better capture the rapid evolution of short-lived thermal anomalies.

Despite these challenges, the model maintains significant practical utility. When a tolerance of ±1 day is allowed, the overall accuracy increases to 77.69%, and the mean absolute error (MAE) is kept at a low level of 0.75 days. These metrics suggest that the predicted duration typically deviates by less than one day from the observations.

It is important to note that the inference strategy adopted in this study does not penalize cases where the real event continues beyond the model's forecasting window. This design choice aligns well with real operational requirements: early-warning systems prioritize the expected duration in the upcoming days, rather than the full lifecycle of the event.

Overall, the duration inference results confirm that the proposed framework is capable not only of determining whether an extreme heat event will occur, but also of estimating how long it may persist. This dual capability enhances the model's practical utility for government agencies and industry sectors in emergency response planning and resource allocation.

5. Conclusions

This study develops an efficient, interpretable, and multi-step forecasting framework for extreme heat events (EHEs) in Hong Kong. By integrating multi-source remote sensing and ground-based observations, incorporating an ARIMA-based anomaly detection module, constructing temporal samples via a sliding-window mechanism, and training a set of multi-step XGBoost classifiers to predict the likelihood of EHEs in the next 1–5 days, we establish a lightweight system capable of determining both the occurrence and the potential duration of extreme heat. Using ten years of daily data from March 2015 to February 2025, we conduct a series of systematic experiments, and the main findings are summarized below.

First, in terms of multi-source data integration, this study combines four key environmental factors derived from multiple datasets including AERONET, Himawari-8/9, MODIS, IGRA, and CPC. Through the construction of an ARIMA residual–based anomaly detection mechanism, we further generate anomaly counterparts for each factor in addition to their original values. This design enables the model to capture short-term perturbations that often precede the onset of extreme heat, thereby enriching the expressiveness of the input features.

Second, the proposed sliding-window strategy effectively leverages temporal dynamics while maintaining a lightweight model structure. It allows the model to learn physical processes such as accumulated surface heating, moisture evolution, and radiative perturbations. Based on these temporal samples, the multi-step independent XGBoost classifiers demonstrate stable and operationally meaningful predictive skill across the entire 5-day forecasting range. The AUC values for Lead 1–Lead 5 remain between 0.935 and 0.883, and the corresponding F1-scores fall between 0.738 and 0.639, indicating that the seven-day sliding window effectively captures both the rapid thermal intensification and persistent background signals. These results highlight not only the influence of relatively slow-varying atmospheric variables in driving extreme heat events but also the potential of lightweight models in multi-step forecasting applications.

In addition, the proposed process-scale duration inference method effectively avoids the challenges associated with direct multi-class modeling, which is often hindered by severe class imbalance in event durations. By estimating the duration based on the leading continuous segment of the daily probability sequence, the model achieves an exact match accuracy of 67.08% and a ±1-day tolerance accuracy of 77.69% in the test set, with a mean absolute error of 0.75 days. Although distinguishing specific short-duration categories (1–4 days) remains a limitation to be addressed, the model shows exceptional stability in identifying non-EHE periods with an accuracy of 84.19%, effectively excluding non-event days. These results demonstrate that the framework is capable not only of identifying the occurrence of extreme heat events but also of providing process-level predictive information useful for emergency management, thereby enhancing the precision of early warnings and resource allocation.

Overall, the lightweight extreme heat forecasting framework proposed in this study shows strong potential in terms of data requirements, computational efficiency, and interpretability, making it a practical complement to operational urban meteorological forecasting and hazard early-warning systems. Future work may extend this study in several directions. First, incorporating higher-resolution reanalysis datasets and urban-attribute information—such as building density and land-cover types—may improve spatial forecasting accuracy. Second, the precision of short-duration event predictions could be further improved through more granular feature engineering. Third, the framework can be generalized to other hazards, including extreme rainfall and compound heat–humidity risks. Fourth, integrating causal inference approaches and explainable artificial intelligence techniques may further enhance the physical consistency and decision transparency of the model. Through these possible extensions, the proposed framework has the potential to provide more comprehensive and reliable technical support for urban extreme-weather monitoring, early warning, and risk management.

Acknowledgements

The authors would like to acknowledge the support from the 2025 Undergraduate Independent Innovation Funding Program of China University of Geosciences (Grant No. 2025XLA177) and the 2025 National Undergraduate Innovation Training Program of China (Grant No. S202510491216).

The authors also express their gratitude to the data providers whose high-quality and long-term observational products made this study possible, including AERONET for aerosol optical depth observations, Himawari-8/9 for Level-2 aerosol retrievals, MODIS (MOD11A1) for land surface temperature products, the IGRA radiosonde archive for precipitable water measurements, the CPC Global Unified Gauge-Based Daily Precipitation dataset for rainfall information, and the Hong Kong Observatory for providing the official extreme heat warnings used as ground-truth labels. Their continued efforts in maintaining open, reliable, and accessible datasets provided essential support for the multi-source analysis and model development conducted in this study.

References

Chen, M., P. Xie, and Co-authors, 2008. CPC Unified Gauge-based Analysis of Global Daily Precipitation, Western Pacific Geophysics Meeting, Cairns, Australia, 29 July - 1 August, 2008.

Chen, Q., Ding, M., Yang, X., Hu, K., & Qi, J., 2018. Spatially explicit assessment of heat health risk by using multi-sensor remote sensing images and socioeconomic data in Yangtze River Delta, China. *International Journal of Health Geographics,* 17. https://doi.org/10.1186/s12942-018-0135-y

Holben, B. N., T. F. Eck, I. a. Slutsker, D. Tanré, J. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. Kaufman and T. Nakajima, 1998. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment,* 66(1): 1-16.

Intergovernmental Panel on Climate Change (IPCC), 2021. *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://doi.org/10.1017/9781009157896

Jia, Z., Feng, G., Zhang, Z., Zhang, H., Zang, N., & Zheng, Z., 2024. The abrupt rise of midsummer high-temperature days and surface air temperature in Southern China around the early 2000s and it's influences on climate forecasts. *Climate Dynamics,* 62, 3605–3619. https://doi.org/10.1007/s00382-023-07087-w

Li, J., Zhu, R., Guo, Y., Li, Z., Yang, S., Deng, K., Liu, Z., & Zhang, T., 2026. Long-lead seasonal prediction of compound extreme heat and drought events in China using machine learning methods. *Climate Dynamics,* 64, 110. https://doi.org/10.1007/s00382-026-08090-7

Perkins-Kirkpatrick, S. E., & Lewis, S. C., 2020. Increasing trends in regional heatwaves. *Nature Communications,* 11, 3357. https://doi.org/10.1038/s41467-020-16970-7

Ren, J., Huang, G., Li, Y., Zhou, X., Xu, J., Yang, Z., Tian, C. & Wang, F., 2021. A Stepwise-Clustered Simulation Approach for Projecting Future Heat Wave Over Guangdong Province. *Frontiers in Ecology and Evolution,* 9: 761251. https://doi.org/10.3389/fevo.2021.761251

Wan, Z., Hook, S., & Hulley, G., 2015. MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Dataset]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD11A1.006

Wu, H., Xu, Y., Zhang, M., Su, L., Wang, Y., & Zhu, S., 2024. Spatially explicit assessment of the heat-related health risk in the Yangtze River Delta, China, using multisource remote sensing and socioeconomic data. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2024.105300

Wulff, C. O., & Domeisen, D. I., 2019. Higher subseasonal predictability of extreme hot European summer temperatures as compared to average summers. *Geophysical Research Letters*, *46*(20), 11520-11529. https://doi.org/10.1029/2019GL084314

Yao, Y., Lu, L., Guo, J., Zhang, S., Cheng, J., Tariq, A., Liang, D., Hu, Y., & Li, Q., 2024. Spatially explicit assessments of heat-related health risks: A literature review. *Remote Sensing,* 16(23), 4500. https://doi.org/10.3390/rs16234500

Zhang, L., Yu, X., Zhou, T., Zhang, W., Hu, S., & Clark, R., 2023. Understanding and attribution of extreme heat and drought events in 2022: Current situation and future challenges. *Advances in Atmospheric Sciences*, 40(11), 1941−1951. https://doi.org/10.1007/s00376-023-3171-x