HiTIC-Monthly: A High Spatial Resolution (1 km×1 km) Monthly Human Thermal Index Collection over China from 2003 to 2020

Hui Zhang¹, Ming Luo¹,²*, Yongquan Zhao³*, Lijie Lin⁴, Erjia Ge⁴, Yuanjian Yang⁶, Guicai Ning¹,², Jing Cong⁷, Zhaoliang Zeng⁸, Ke Gui⁹, Jing Li¹⁰, Ting On Chen¹, Xiang Li⁴, Sijia Wu¹, Peng Wang¹, Xiaoyu Wang¹

¹School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, China.
²Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Hong Kong SAR, China.
³School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 519082, China.
⁴School of Management, Guangdong University of Technology, Guangzhou 510520, China.
⁵Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario M5T 3M7, Canada.
⁶School of Atmospheric Physics, Nanjing University of Information Science & Technology, Nanjing 210044, China.
⁷Tianjin Municipal Meteorological Observatory, Tianjin 300074, China.
⁸State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China.
⁹State Key Laboratory of Severe Weather (LASW) and Key Laboratory of Atmospheric Chemistry (LAC), Chinese Academy of Meteorological Sciences, Beijing 100081, China.
¹⁰College of Resources and Environment, Fujian Agriculture and Forest University, Fuzhou 35002, China.

*Correspondence to: Ming Luo (luom38@mail.sysu.edu.cn) and Yongquan Zhao (yqzhao@link.cuhk.edu.hk)
Abstract

Human thermal comfort measures the combined effects of temperature, humidity, and wind speed, etc., and can be aggravated under the influences of global warming and local human activities. With the most rapid urbanization and the largest population, China is being severely threatened by aggravating human thermal stress. However, the variations of thermal stress in China at a fine scale have not been fully understood. This gap is mainly due to the lack of a high-resolution gridded dataset of human thermal indices. Here, we generate the first high spatial resolution (1 km × 1 km) dataset of monthly human thermal index collection (HiTIC-Monthly) over China from 2003 to 2020. In this collection, 12 commonly used thermal indicators are generated by the LGBM machine learning algorithm from multi-source gridded data, including MODIS land surface temperature, topography, land cover and land use, population density, and impervious surface fraction. Their accuracies were comprehensively assessed based on observations at 2419 weather stations across the mainland of China. The results show that our dataset has desirable performance, with mean $R^2$, root mean square error, mean absolute error, and bias of 0.996, 0.693°C, 0.512°C, and 0.003°C, respectively, by averaging the 12 indicators. Moreover, the predictions exhibit high agreements with observations across spatial and temporal dimensions, demonstrating the broad applicability of our dataset. The comparison with two existing datasets also suggests that our high-resolution dataset can describe a more explicit spatial distribution of the thermal information, showing great potentials in fine-scale (e.g., intra-urban) study. Further investigation reveals that nearly all indicators exhibit increasing trends in most parts of China during the year 2003~2020. The increase is especially stronger in North China, Southwest China, the Tibetan Plateau, and parts of Northwest China, and in the spring and summer seasons. The HiTIC-Monthly dataset is publicly available via https://zenodo.org/record/6895533 (Zhang et al., 2022a).
1 Introduction

Global climate change has brought significant challenges to human society and natural systems (Arias et al., 2021; Haines and Ebi, 2019), by inducing higher air temperature and more frequent extreme weather and climate events around the world (Arias et al., 2021; Schwingshackl et al., 2021). Heat-related disasters, e.g., heatwaves, droughts, and wildfires, are occurring more frequently and becoming more intense (Tong et al., 2021; Arias et al., 2021; Luo et al., 2022), exacerbating the thermal environment and threatening the tolerance limits of humans, animals, and plants (Raymond et al., 2020). Substantial warming and increasing extreme weather and climate events aggravate human thermal comfort and lead to increased exposures to uncomfortable thermal environments (Brimicombe et al., 2021), thus posing adverse impacts on public health, socio-economy development, and agricultural productivities (Budhathoki and Zander, 2019; Moda et al., 2019; Tuholske et al., 2021; Sun et al., 2019; Zhao et al., 2017).

The thermal stress that human beings actually perceive is not only related to air temperature, but also jointly influenced by other environmental variables such as humidity, wind, and/or direct sunlight (Mistry, 2020; Djongyang et al., 2010). These climatic variables alter the heat balance that maintains the core temperature of human bodies by influencing the heat exchange (e.g., radiation, convection, conduction, and evaporation) between humans and the surrounding environment (Periard et al., 2021; Stolwijk, 1975). High atmospheric humidity can exacerbate the thermal stress on human bodies by reducing evaporation from the skin through sweating when the air temperature is high (Li et al., 2018; Rogers et al., 2021; Luo and Lau, 2021). Furthermore, abnormal weather with a combination of extremely high air temperature, humidity, and/or wind can reduce labor capacity and human performance (Roghanchi and Koçsis, 2018; Lazaro and Momayez, 2020; Enander and Hygge, 1990), leading to temperature-related discomfort, stress, morbidity, and even death (Di Napoli et al., 2018; Kuchcik, 2021; Nastos and Matzarakis, 2011), particularly during heatwaves. For example, in the summer of 2017, 2018, and 2019, there were 1489, 1700, and 161 heatwave-related deaths, respectively, in the United Kingdom (Rustemeyer and Howells, 2021). Additionally, vulnerable groups, such as children, the elderly, chronic patients, and poor communities are at higher risk of being affected by thermal stress (Patz et al., 2005;
Wang et al., 2019), which is likely to be further exacerbated as population aging and warming globally (United Nations, 2017).

The changes and impacts in human thermal stress have attracted increasing attention in recent years (Schwingshackl et al., 2021; Krzysztof et al., 2021; Li et al., 2018; Rahman et al., 2022; Ren et al., 2022; Luo and Lau, 2021). For instance, Szer et al. (2022) estimated the impact of heat stress on construction workers based on the universal thermal climate index (UTCI). Ren et al. (2022) and Luo and Lau (2021) explored the contribution of urbanization and climate change to the urban human thermal comfort in China. Schwingshackl et al. (2021) assessed the severity and trend of future global heat stress based on Coupled Model Intercomparison Project phase 6 (CMIP6). These studies were mainly based on meteorological stations or coarse-gridded data. However, the meteorological stations are sparsely distributed (Peng et al., 2019), particularly in undeveloped, mountainous, and oceanic areas, which cannot reveal continuously spatial distributions of air temperature and thermal stress conditions (He et al., 2021). Additionally, existing low spatial resolution image products (Mistry, 2020; Di Napoli et al., 2020) cannot be applied to fine-scale studies because they cannot provide information with spatial details and variations. However, the changes in human thermal stress at a fine scale (e.g., 1 km × 1 km) remain much less understood. This research gap is mainly inhabited by the unavailability of a high spatial resolution (high-resolution) gridded dataset of human thermal stress.

Limited is high-resolution multiple human thermal stress indices dataset, compared with extensive studies in producing land surface temperature (LST) or near-surface air temperatures (SAT). In particular, numerous LST datasets, such as Land Surface Temperature in China (LSTC) (Zhao et al., 2020) and the global seamless land surface temperature dataset (Zhang et al., 2022b; Hong et al., 2022), and near surface air temperature datasets such as ERA5 (ECMWF, 2017), TerraClimate (Abatzoglou et al., 2018), and GPRChinaTemp1km (He et al., 2021) have been produced. Few coarse-resolution human thermal stress datasets have been produced, such as ERA5-HEAT (Di Napoli et al., 2020), HDI_0p25_1970_2018 (hereafter, HDI) (Mistry, 2020), and HiTiSEA (Yan et al., 2021). ERA5-HEAT was derived from ERA5, which includes two global hourly human thermal stress indices (UTCI and mean radiant temperature (MRT)) from January 1979 to the present (Di Napoli et al., 2020) (Table S1).
The HDI dataset was generated using 3-hourly climate variables of the global land data assimilation system (GLDAS), and it contains ten daily indices with a spatial resolution of 0.25°×0.25°, covering 90°N–60°S from 1970 to 2018 (Mistry, 2020). HiTiSEA contains ten daily human thermal stress indices from 1981 to 2017, with a spatial resolution of 0.1°×0.1° over South and East Asia (Yan et al., 2021), which was derived from ERA5-Land and ERA5 reanalysis products. However, these existing thermal index datasets have very coarse spatial resolutions. There is an urgent need for high-resolution (e.g., 1 km) dataset of human thermal stress.

Various indicators have been proposed to measure human thermal stress, and currently there is no universal or superior thermal stress indicator in all climate zones (Schwingshackl et al., 2021; Brake and Bates, 2002; Roghanchi and Kocsis, 2018; Luo and Lau, 2021). Existing human thermal stress indices have considered different climate conditions, direct or indirect exposures to weather elements, human metabolism, and the local working environment (Di Napoli et al., 2020). These indices have been designed to evaluate or quantify the comprehensive environmental pressure of temperature, humidity, wind, and other meteorological factors on human bodies (Epstein and Moran, 2006). They are based on the thermal exchange between the human and surrounding environments or empirical relationships gained by studying human responses to various environmental factors, varying in complexity, applicability, and capacity (Staiger et al., 2019). For example, the heat index (HI) is used for meteorological service (NWS, 2011); wet-bulb temperature (WBT) is used to measure the upper physiological limit of human beings (Raymond et al., 2020); physiologically equivalent temperature (PET) and UTCI are used to estimate human thermal comfort (Varentsov et al., 2020). A high-resolution dataset of different commonly used human thermal stress indicators is urgently called in global and regional studies, particularly for those with complex climate conditions (e.g., China).

China has been threatened by deteriorating thermal environments under global climate change and rapid local urbanization over the past decades (Ren et al., 2022; Luo and Lau, 2019). The changes and characteristics of human thermal stress across China have attracted extensive attention in recent years (Yan, 2013; Tian et al., 2022; Li et al., 2022). Wang et al. (2021) found that the frequency of extreme human-perceived temperature events increases in summer and decreases in winter in most urban...
agglomerations (UAs) of China. Li et al. (2022) found that the frequency of thermal discomfort days in China exhibits a significant increasing trend from 1961 to 2014, and there will be more threats from thermal discomfort in the future. Therefore, a long-term and high-resolution dataset with multiple human thermal stress indices in China is of great importance for studying detailed spatial and temporal variations of human thermal stress across the country. Such a dataset has the potential to (1) assess population exposure to extreme thermal conditions and heat-related health risks, (2) study the spatiotemporal evolution of human thermal stress and its influence on public health, tourism, industries, military, epidemiology, and biometeorology at a fine scale, and (3) provide the policymakers with data in manipulating targeted strategies to mitigate heat stress and protect vulnerable people.

In this study, we produced a high-resolution (1 km × 1 km) thermal index collection at a monthly scale (HiTIC-Monthly) in China over a long period (2003 to 2020). This collection contains 12 widely used human thermal indices, including Surface Air Temperature (SAT), indoor Apparent Temperature (AT\text{\textsubscript{in}}), outdoor shaded Apparent Temperature (AT\text{\textsubscript{out}}), Discomfort Index (DI), Effective Temperature (ET), Heat Index (HI), Humidex (HMI), Modified Discomfort Index (MDI), Net Effective Temperature (NET), simplified Wet Bulb Globe Temperature (sWBGT), Wet-Bulb Temperature (WBT), and Wind Chill Temperature (WCT). The remainder of this paper is structured as follows. Sections 2 and 3 respectively introduce the data sources and describe the procedure for predicting the human thermal indices. Section 4 presents a comprehensive analysis of the accuracies and trends of the human thermal indices. Section 5 provides data availability. Section 6 compares our products with two existing datasets, and the main findings of this paper are summarized in Section 7.

2 Data

2.1 Meteorological dataset

Daily mean surface air temperature, relative humidity, and wind speed recorded at the 2419 weather stations across China (Figure 1) from 2003 to 2020 were collected from the China Meteorological Data Service Center (CMDC) at http://data.cma.cn/en. All station records were subjected to strict quality control and evaluation, including homogenization based on a statistical approach (Xu et al., 2013) and
evaluation of temporal inhomogeneity based on the Easterling-Peterson method (Li et al., 2004).

2.2 Covariates

Human thermal stress is related to temperature, topography, land cover and land use, population density, surface water, and vegetation (Wang et al., 2020; Rahman et al., 2022; Krzysztof et al., 2021). In this study, eight variables reflecting the changes and spatial distribution characteristics of temperature were used to predict human thermal indices (Table 1) in addition to the meteorological variables. As LST is one of the most essential parameters for predicting human thermal indices, the seamless LST dataset created by Zhang et al. (2022b) was introduced into model training. This LST dataset used a spatiotemporal gap-filling algorithm to fill the missing or invalid value caused by clouds in the Moderate Resolution Imaging Spectroradiometer (MODIS) LST dataset (MOD11A1 and MYD11A1). It includes daily mid-daytime (13:30) and mid-nighttime (01:30) LST with 1 km × 1 km spatial resolution. The mean root mean squared errors (RMSEs) of daytime and nighttime LST are 1.88 °C and 1.33 °C, respectively. The land cover and land use dataset developed by Sulla-Menashe and Friedl (2019) based on a supervised classification method was downloaded via Google Earth Engine (GEE). The Multi-Error-Removed Improved-Terrain (MERIT) elevation dataset developed by Yamazaki et al. (2017) was downloaded from GEE. This dataset was generated after removing the errors from existing Digital Elevation Models (DEMs), such as SRTM3 and AW3D-30m, based on multi-source satellite data and filtering algorithms. The spatial resolution of this dataset is 3″ (i.e., ~90 meters at the equator). In addition, the slope was also extracted from the elevation data to act as the topography predictor. As the artificial surface is closely related to human activities (Zhao and Zhu, 2022), the dataset of global artificial impervious area (GAIA) produced by Gong et al. (2020) from GEE was used to delineate human footprints. The overall accuracy of GAIA is greater than 90% (Gong et al., 2020). The population dataset was downloaded from the WorldPop Project (Gaughan et al., 2013). Then, the abovementioned eight datasets were pre-processed to have the same spatial extend, projection, and spatial resolution of 1 km × 1 km through image mosaicking, reprojection, resampling, clipping, aggregating, and monthly synthesizing. Moreover, year and month of the year were also used as covariates.
3 Methodology

3.1 Calculation of human thermal indices

In addition to SAT, the calculation of human thermal indices used in this study is described in Table 2. These indices are calculated based on SAT (also simply denoted as T), relative humidity (RH), wind speed (V), and actual vapor pressure (E_a). E_a is derived from T and RH, rather than directly observed at meteorological stations (Eqs. 1~2). Furthermore, monthly human thermal indices were derived by averaging daily values in each month.

\[
E_a = 6.112 \times \exp\left(\frac{17.67 \times T}{T + 243.5}\right) 
\]

(1)

\[
E_a = \frac{R_H \times 100}{T} \times E_s 
\]

(2)

Here E_s is saturation vapor pressure (hPa) near the surface, T (°C) is air temperature at 2 m above the ground, and RH (%) is relative humidity at 2 m above the ground.

3.2 Prediction of human thermal indices using LGBM

The Light Gradient Boosting Machine (LGBM) algorithm was employed to predict human thermal indices from 2003 to 2020. Developed by Microsoft Research (Ke et al., 2017), LGBM is one of the gradient boosting decision tree (GBDT) algorithms. This algorithm has become a very popular nonlinear machine learning algorithm due to its superior performance in machine learning competitions and efficiency (Candido et al., 2021). Its performance has been evaluated and shows desirable results in different applications, such as evapotranspiration estimation (Fan et al., 2019), land cover classification (Candido et al., 2021; Mccarty et al., 2020), air quality prediction (Su, 2020; Zeng et al., 2021; Tian et al., 2021), subsurface temperature reconstruction (Su et al., 2021), and above-ground biomass estimation (Tamiminia et al., 2021).

LGBM adopts gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) algorithms to improve the training speed (Su et al., 2021). GOSS is used to select data instances with larger gradients and to exclude a considerable proportion of small gradient data instances (Ke et al., 2017), and EFB is used to merge features (Ke et al., 2017). Compared with traditional GBDT algorithms including eXtreme gradient boosting (XGBoost) and stochastic gradient boosting (SGB), LGBM
effectively decreases the training time without reducing the accuracy (Los et al., 2021; Ke et al., 2017). We used the Python package Scikit-Learn to perform the LGBM training, and hyperparameters of LGBM were tuned based on Grid Search Methods. The observed monthly human thermal indices from 2003 to 2020 at the 2419 weather stations across mainland China were randomly classified into a training set (80%) for hyperparameters tuning and model training and a testing set (20%) for model evaluation.

3.3 Accuracy assessment

Four statistic metrics, namely, determination coefficient ($R^2$), Mean Absolute Error (MAE), RMSE, and Bias, were used to evaluate the prediction accuracy of the human thermal indices. Ranging from 0 to 1, $R^2$ measures the proportion of variance explained by the model, representing how well the human thermal indicators were predicted compared to the observations. MAE represents the average absolute error between the predictions and the observations. RMSE is the standard deviation of the residuals and is sensitive to outliers. Bias describes the differences between the predictions and the observations. These metrics are computed as follows.

$$MAE = \frac{1}{N} \times \sum_{i=1}^{N} |y_i - \bar{y}|$$  \hspace{1cm} (3)

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - \bar{y})^2}$$  \hspace{1cm} (4)

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$  \hspace{1cm} (5)

$$Bias = \frac{1}{N} \times \sum_{i=1}^{N} (y_i - \bar{y})$$  \hspace{1cm} (6)

where $\bar{y}$ is the predicted value of human thermal indices, $\bar{y}$ is the mean of the observed human thermal indices calculated from meteorological stations, and $N$ is the number of samples.

4 Results

4.1 Evaluation of the prediction of human thermal indices

4.1.1 Overall accuracy

The prediction accuracies of the 12 human thermal indices were evaluated based on the validation data introduced in Section 3.2. All predicted human thermal indices exhibit high accuracies. Figure 2 shows...
the scatter plots of the observed versus the predicted values of 12 human thermal indices. As the figure displays, the data points of all indices are concentrated around the corresponding 1:1 line, indicating a good consistency between the observed and the predicted values. Figure 3 and Table S2 present the $R^2$, MAE, RMSE, and Bias values of 12 thermal indices from 2003 to 2020. The $R^2$ values of the 12 indices are all higher than 0.99, and their RMSE, MAE, and Bias are lower than 0.9 °C, 0.7 °C, and 0.003 °C, respectively. Particularly, HMI has the highest RMSE and MAE values, i.e., 0.859 °C and 0.645 °C, respectively; while ET shows the lowest RMSE and MAE, i.e., 0.377 °C and 0.281 °C, respectively. The larger error in terms of relatively higher RMSE and MAE for NET is likely caused by the incorporation of wind speed in computing this thermal indicator (see Table 2). These overall accuracies demonstrate that 12 human thermal indices predicted in this study are of good quality.

The spatial distributions of $R^2$, MAE, RMSE, and Bias at individual stations across mainland China are depicted in Figure 4–7, respectively. The predicted indices have high $R^2$ values (i.e., >0.98, Figure 4) in almost all stations across China, demonstrating the superiority of LGBM. Better predictions in terms of higher $R^2$ are distributed in eastern China, particularly in the North China Plain (NCP) and the Yangtze River Delta (YRD), while southwestern China (e.g., the Yunnan-Guizhou Plateau (YGP)) has relatively lower R2 values (<0.98). For MAE and RMSE, all indices have small values <1 °C at most stations across China. MAE and RMSE show similar patterns, which exhibit the highest values in HMI (Figure 5g and 6g), followed by NET and WCT, and ET has the lowest MAE and RMSE values (i.e., < 0.4 °C, Figure 5e and 6e). The MAE and RMSE of NET and WCT decrease from northwestern to southeastern China (Figure 5i, 5l, 6i, 6l). For other indices, small MAE and RMSE values are mainly observed in plain regions including NCP, while high values tend to appear in regions with complex topography, such as arid Northwest China, mountainous Northeast and South China, and the Hengduan Mountains. This difference is related to the uneven distribution of weather stations, i.e., dense in plain areas and coarse in complex terrain areas. The Bias values range from -0.3 °C to 0.3 °C (Figure 7). Positive Bias values are mainly distributed in northern China, while negative Bias is seen in the south.
4.1.2 Annual and monthly accuracies

The accuracies in terms of $RMSE$, $MAE$, and $Bias$ of the 12 human thermal indices in individual years from 2003 to 2020 are assessed in Figure 8. $RMSE$s and $MAE$s of all indices in nearly all years are less than 1.0 °C (Figure 8a-b). Yearly $RMSE$ ($MAE$) of ET fluctuates around 0.3 °C (0.2 °C) from 2003 to 2020. $RMSE$s ($MAE$s) of other indices range from 0.5 to 1.1 °C (0.4–0.8 °C) with marginal variations from year to year. $Biases$ vary between -0.04 °C and 0.04 °C across all years. $Biases$ seem to be slightly positive in 2003, 2004, and 2014 and negative in 2012, 2017, and 2018. Moreover, Figure S1 shows the $RMSE$s, $MAE$s, and $Biases$ of all human thermal indices in different months. In terms of $RMSE$, all the indices in 12 months are lower than 1.4 °C, and their $MAE$s are less than 1 °C. HI and HMI have relatively higher $RMSE$ and $MAE$ values in summer than in other seasons; whereas, other indices tend to have larger errors in winter than in summer. Additionally, the magnitude of $Bias$ is smaller than 0.03 °C for all the indices in 12 months.

4.1.3 Accuracies in major urban agglomerations

More than half of the national population in China lives in cities, particularly in UAs (i.e., also known as city clusters). Here we assessed the prediction accuracies in 20 major UAs in China, which hold 62.83% and 80.57% of the total population and gross domestic product (GDP) of the country (Fang, 2016). These accuracy assessments are presented in Tables S3–S6. As shown in Table S3, all UAs have the $R^2$ values higher than 0.9837, with an average of 0.9947. Table S4 also shows that these UAs have small $RMSE$ values, most of which are smaller than 1 °C, except for the UA of North Tian Shan Mountain in arid Northwest China. As the biggest UA in China, YRD has the lowest $RMSE$ of 0.288 °C among all 20 UAs. The $MAE$s of the thermal indicators in all UAs are smaller than 1 °C and with an average value of 0.477 °C (Table S5). The $Biases$ in the 20 UAs range from -0.160 °C to 0.123 °C (Table S6). These results suggest that all predicted human thermal indices in different UAs across China are of good quality at the local scale. It implies that our prediction model and results have great potential in evaluating local thermal environment changes (e.g., in urban areas or cities).
4.2 Spatial variations of the human thermal indices

The abovementioned assessments show that our model based on LGBM can yield high-accuracy predictions at both national and local scales. Therefore, this model is employed to generate a high-resolution human thermal index collection at a monthly scale over China (HiTIC-Monthly) from 2003 to 2020. By taking monthly ET in 2020 as an example, we examined the monthly evolution of spatial patterns of the HiTIC-Monthly dataset in this subsection.

Figure 9 shows the monthly distribution of the predicted ET in 2020, which exhibits obvious seasonality with higher temperatures in summer and lower in winter. The temperature shows a significant zonal difference with colder temperature in northern than southern China. Temperature has a close relationship with topography and decreases with elevation, varying from plateaus to plains. The Qinghai-Tibet Plateau (TP) has the lowest temperature, while southern China, the Sichuan Basin, and the Gobi regions in Northwest China witness the highest temperature. The distribution of temperature exhibits different patterns among the four seasons, especially between winter (e.g., January) and summer (e.g., July). In winter, the temperature increases from northern to southern areas and is the coldest in Northeast and Northwest China and the warmest on the Hainan Island. In summer, the hottest temperature appears in the Tarim and Jungar Basins of Xinjiang. The NCP region also has a high temperature in summer, which might be related to local urbanization (Liu et al., 2008) and irrigation (Kang and Eltahir, 2018).

The spatial variations of the predicted human thermal indices in summer (which is often characterized by severe heat stress) are examined in Figure 10 by taking July 2020 as an example. As it shows, 12 indicators show similar distribution patterns. There are significant differences in temperature among northwest, northeast, and southeast. Generally, the temperature decreases from the southeast to the northwest, and the southeast and northwest parts have the highest and lowest temperature, respectively. HMI exhibits the highest temperature while NET shows the lowest in July 2020.

4.3 Temporal changes in the human thermal indices

The yearly evolutions of the annual mean human thermal indices during 2003–2020 are displayed in Figure 11. Despite the interannual fluctuation in the time series, all indicators exhibit upward trends
except for NET and WCT, of which the decreasing trends are mainly affected by the recovering wind speed in the recent decade (Zeng et al., 2019). The fastest warming appears in HMI (0.303 °C/decade), and the slowest is in ET (0.111 °C/decade). These warming trends are stronger than the rising rate of global mean near surface temperature (IPCC, 2021), demonstrating China as one of the severest hotspots suffering from dramatic climate warming under global change. The detailed spatial variations regarding the trends of the human thermal indices across China are further depicted in Figure 12. Most parts of China experience increases in nearly all the indicators during 2003–2020. These increases are especially more profound in North China, Southwest China, TP, and parts of Northwest China. The possible reasons for the prominent warming trends in North China are explained as follows. The urbanization process has been prevailing in this area, with rapid growth in the economy and population. This process is accompanied by dramatic increases in impervious surfaces and decreases in green spaces. These changes lead to warmer surface and near surface air temperature, known as urban heat islands (UHI), thus increasing thermal stress in this region. The urbanization effects on local heat stress have also been reported by (Luo and Lau, 2021). Moreover, North China has a large amount of croplands with prominent irrigation activities, which may increase air humidity near the surface and exacerbate the combined effects of temperature and humidity, leading to increased heat stress (Kang and Eltahir, 2018). In addition, this area has experienced a weakening of surface wind speed (Zhang et al., 2021), which also exacerbates thermal stress, especially in NET and WCT.

Furthermore, different indices have different degrees of increasing trends. HMI has the largest increasing magnitude (Figure 12h), and ET is seen with relatively slight increases across China (Figure 12f). The trends of NET and WCT have similar spatial distribution patterns, with large proportions having cooling trends since 2003 (Figure 12j&l). Most parts of Xinjiang, northeastern and southern China have obvious decreasing trends, and the Inner Mongolia Plateau (IMP), NCP, eastern TP, YRD, and YGP have slight increasing trends.

The temporal trends of the human thermal indices in different seasons were also examined (Figure 13). The fastest warming tendency is observed in the spring season. The rising trends of spring HMI, HI, MDI, AT$_{in}$, and AT$_{out}$ exceed 0.4 °C/decade, and the trends of other indices (except ET and NET) are larger
than 0.3 °C/decade (Figure S2). Summer also has been experiencing significant increasing trends in all indicators, i.e., at a rate of > 0.2 °C/decade (except ET and NET). The trends in summer HMI, HI, WBT, MDI, DI, sWBGT, \textit{AT}_{in}, and \textit{AT}_{out} exceed 0.3 °C/decade (Figure S3). Differ from spring and summer, the human thermal indicators (except WCT and NET) in the autumn season show slightly cooling trends (Figure S4). Autumn WCT and NET have significantly strong decreasing trends, i.e., -0.349 and -0.507 °C/decade, respectively. Similar strong cooling trends of WCT and NET appear in winter, i.e., -0.661 and -0.453 °C/decade, respectively, while other indicators experience marginal long-term changes (Figure S5).

Figure S6 maps the spatial patterns of the trends of summer mean human thermal indices over mainland China during 2003–2020. All indicators show warming trends in most parts of China, particularly in NCP and TP. As one of the most densely populated regions in China, the prominent increases in thermal indices in NCP indicate that people living there have been experiencing increasing threats of intensifying heat stress. Among the 12 indicators, \textit{AT}_{out}, HI, NET and WCT tend to have a slight cooling trend in southeastern China. This cooling trend is consistent with the corresponding summer SAT.

The spatial distributions of the changing trends in winter across mainland China during 2003–2020 are depicted in Figure S7. The trend patterns in winter are similar to that in summer to some degrees. The warming trends are concentrated in Southwest China, most parts of Northwest China, and parts of East China (e.g., YRD). The cooling trends are located in TP, parts of Northeast and South China. The cooling tendencies are especially profound in NET and WCT (Figures S7j&m), Northeast China, and most parts of Northwest and South China. Parts of central China are seen with even stronger cooling thermal comfort. In spring, increases in all thermal indicators are observed in most parts of China (Figure S8), particularly in northern regions, such as central Inner Mongolia, parts of NCP, and Northeast China, while parts of southern China have slight decreases. These decreases are noticeable in NET and WCT (Figures S8j&m).

In contrast to spring, the autumn season is observed with decreased thermal temperature in the north and increases in the south (e.g., Southwest China, Figure S9).
5 Discussion

5.1 Comparison with existing human thermal indices datasets

Our 1 km \( \times \) 1 km HITIC-Monthly product is compared with two existing datasets, HDI (Mistry, 2020) and HiTiSEA (Yan et al., 2021), which have coarser spatial resolutions of 0.25° \( \times \) 0.25° and 0.1° \( \times \) 0.1° (Figure S1), respectively. We derived monthly mean AT\(_{in}\) in July 2018 from the HDI and HiTiSEA, and compare them with our HITIC-Monthly over mainland China, with a particular highlight in the four largest UAs, including Beijing-Tianjin-Hebei (BTH), YRD, middle Yangtze River Valley (mYRV) and Pearl River Delta (PRD) (Figure 14). The summer of 2018 is selected by considering that it was included in all three datasets and frequent heat events occurred in this summer (Zhou et al., 2020). Generally, the three datasets depict similar spatial patterns. However, our HiTIC-Monthly dataset obviously provides more detailed and clearer spatial information on human thermal stress than the other two. Additionally, the observed AT\(_{in}\) values at individual weather stations are also compared in Figure 14. It can be seen that HDI and HiTiSEA overestimate AT\(_{in}\), and such overestimation is especially severe in HDI, whereas our dataset is in good agreement with the observed AT\(_{in}\) at individual weather stations. Therefore, our predicted temperature can describe the spatial variations in the city areas well, thereby providing fundamental support for fine-scale climate studies, such as urban climate research.

5.2 Limitations and future works

There are 12 commonly used human thermal indices in the HiTIC-Monthly dataset produced in this study. Nine of these indices were computed from temperature and humidity (or water vapor) and the other three (i.e., AT\(_{out}\), NET, and WCT) were derived from temperature, humidity, and wind speed. In addition, by considering the combined effect of environmental variables, some other indices have been proposed in the literature, such as sunlight (Blazejczyk, 1994; Fanger, 1970; Höppe, 1999; Yaglou and Minaed, 1957). These indices include wet bulb globe temperature (WBGT), predicted mean vote (PMV), UTCI, physiological equivalent temperature (PET), etc. These thermal indices are not included in our study due to the lack of sunshine and radiative flux data.

Since LST is the most important variable for predicting the human thermal indices, the uncertainty in the LST dataset may harm the accuracy of the human thermal indices. The LST variable in our prediction is
from a global seamless 1 km resolution daily LST dataset (Zhang et al., 2022b). This dataset was generated based on spatiotemporal gap-filling algorithms and the MODIS LST dataset. It may overestimate LST in some cases because the LST under cloudy weather was filled based on the data under clear sky conditions (Zhang et al., 2022b). A high-quality LST dataset would further improve the prediction accuracy of the human thermal indices.

The human thermal indices dataset is at a monthly scale, but the temporal resolution may not be sufficient for the research of extreme weather events (e.g., heatwaves) and related environmental health (e.g., heat-related mortality). A dataset of high spatial resolution daily human thermal index collection (HiTIC-Daily) will be produced and released in our future studies. In the current study, we provided the first national multiple monthly human thermal indices dataset over the mainland of China, and this dataset shows high prediction accuracies in all climate regimes across China. A global dataset of multiple human thermal indices dataset is also expected in the near future.

**6 Data availability**

The high spatial resolution monthly human thermal index collection (HiTIC-Monthly) generated in this study is freely available to the public in network common data form (NetCDF) at [https://zenodo.org/record/6895533](https://zenodo.org/record/6895533) (Zhang et al., 2022a). The human thermal indices include surface air temperature (SAT), indoor Apparent Temperature (AT\text{in}), outdoor shaded Apparent Temperature (AT\text{out}), Discomfort Index (DI), Effective Temperature (ET), Heat Index (HI), Humidex (HMI), Modified Discomfort Index (MDI), Net Effective Temperature (NET), simplified Wet Bulb Globe Temperature (sWBGT), Wet-Bulb Temperature (WBT), and Wind Chill Temperature (WCT). This dataset has a spatial resolution of 1 km × 1 km and covers mainland China from January 2003 to December 2020, stacking by year. Each stack is composed of 12 monthly images. The unit of the dataset is 0.01 degree Celsius (°C), and the projection coordinate system is Albers Equal Area Conic Projection. The naming rule and other detailed information can be found in “README.pdf”.
7 Conclusions

A long-term and high-resolution dataset of multiple human thermal indices is of great significance for studying detailed spatial and temporal changes of human thermal stress in different climate regions across China and assessing the health risks of people exposed to extreme heat at a fine scale. However, the current datasets of human thermal indices (e.g., HDI and HiTISEA) only have coarse spatial resolutions (> 0.1°). In this study, we generated a dataset of monthly human thermal index collection with a high spatial resolution of 1 km × 1 km over mainland China (HiTIC-Monthly). In this collection, twelve human thermal indicators were predicted, including SAT, AT_in, AT_out, DI, ET, HI, HMI, MDI, NET, sWBGT, WBT, and WCT, from January 2003 to December 2020.

The HiTIC-Monthly dataset was produced by LGBM based on multi-source grided data, including MODIS LST, DEM, land cover and land use, population density, and impervious surface fraction. This dataset shows a desirable performance, with mean $R^2$, RMSE, MAE, and Bias of 0.996, 0.693°C, 0.512°C, and 0.003°C, respectively. Our predictions also exhibit good agreements with the observations in both spatial and temporal dimensions, demonstrating the broad applicability of our dataset. Moreover, the comparison with two existing datasets (i.e., HDI and HiTISEA) suggests that HiTIC-Monthly has more detailed spatial information, indicating that our dataset can well support fine-scale studies. Further investigation shows that almost all the indicators show warming trends in most parts of China during 2003–2020, particularly for North China, Southwest China, TP, and parts of Northwest China. Additionally, the warming tendency is faster in the spring and summer seasons. WCT and NET show similar and strong cooling trends in autumn and winter, while other indicators exhibit slight long-term changes.

Author contribution

H.Z.: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft preparation; M.L.: Conceptualization, Investigation, Funding acquisition Methodology, Supervision Writing – review & editing; Y.Z.: Investigation, Supervision Writing – review & editing; L.J.: Investigation, Writing – review & editing; E.G.: Investigation, Writing – review & editing; Y.Y.: Investigation, Writing – review
& editing; G.N.: Investigation, Writing – review & editing; J.G.: Investigation, Writing – review & editing; Z.Z.: Investigation, Writing – review & editing; K.G.: Investigation, Writing – review & editing; J.L.: Investigation, Writing – review & editing; X.L.: Investigation, Writing – review & editing; S.W.: Investigation, Writing – review & editing; P.W.: Investigation, Writing – review & editing; X.W.: Investigation, Writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41871029), the National Key R&D Program of China (2019YFC1510400), the Natural Science Foundation of Guangdong Province, China (2019A1515011025), and the Pearl River Talent Recruitment Program of Guangdong Province (2017GC010634).

References

Budhathoki, N. K. and Zander, K. K.: Socio-Economic Impact of and Adaptation to Extreme Heat and


He, Q., Wang, M., Liu, K., Li, K., and Jiang, Z.: GPRChinaTemp1km: a high-resolution monthly air


Luo, M. and Lau, N. C.: Increasing Human-Perceived Heat Stress Risks Exacerbated by Urbanization in

Rothfusz, L. P. and Headquarters, N. S. R.: The heat index equation (or, more than you ever wanted to know about heat index), Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology, 9023, 1990.


Figures 1. Spatial distribution of meteorological stations in the mainland of China, with color shadings indicating the elevation in meters.
Figure 2. Scatter plots of predicted versus observed values of 12 human thermal indices over mainland China during 2003–2020: (a) SAT, (b) $AT_a$, (c) $AT_{out}$, (d) DI, (e) ET, (f) HI, (g) HMI, (h) MDI, (i) NET, (j) sWBGT, (k) WBT, and (l) WCT. The black straight line is the 1:1 line.
Figure 3. Overall prediction accuracies of 12 human thermal indices over mainland China during 2003–2020: 
(a) $R^2$, (b) MAE, (c) RMSE, (d) Bias.
Figure 4. Spatial distribution of $R^2$ of the human thermal index predictions at individual meteorological stations over mainland China during 2003–2020.
Figure 5. As Figure 4 but for MAE.
Figure 6. As Figure 4 but for RMSE.
Figure 7. As Figure 4 but for Bias.
Figure 8. Prediction accuracies of human thermal indices in individual years over mainland China during 2003–2020.
Figure 9. Spatial distributions of monthly mean ET over mainland China in 2020.
Figure 10. Spatial distributions of human thermal indices over mainland China in July 2020.
Figure 11. Temporal variations of the national average of annual mean human thermal indices over mainland China during 2003–2020. The straight line illustrates the linear trend, the number in square bracket means the corresponding trend per decade, and the asterisk next to the number indicates that the trends are significant at the 0.05 level.
Figure 12. Spatial distributions of the trends (unit: °C per decade) of annual mean human thermal indices over mainland China during 2003–2020.
Figure 13. Temporal trends of the national average of annual and seasonal mean human thermal indices over mainland China during 2003–2020. The number means linear trend per decade. The asterisk indicates that the trends are significant at the 0.05 level.
Figure 14. Comparison of the spatial patterns among HDI_0p25_1970_2018 (HDI), HiTiSEA, and HITIC-
Monthly for AT$_m$ in mainland China and four major UAs, i.e., Beijing-Tianjin-Hebei (BTH), Yangtze River
Delta (YRD), middle Yangtze River Valley (mYRV) and Pearl River Delta (PRD), in July 2018. Colored
circles indicate the observed AT$_m$ values at individual meteorological stations.
### Tables

**Table 1. Grided datasets and variables used in this study.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Dataset</th>
<th>Spatial Resolution</th>
<th>Temporal Resolution</th>
<th>Variables</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land surface temperature</td>
<td>A global seamless 1 km resolution daily land surface temperature dataset (2003-2020)</td>
<td>1 km×1 km</td>
<td>Daily</td>
<td>Land surface temperature</td>
<td>(Zhang et al., 2022b)</td>
</tr>
<tr>
<td>Land cover and land use</td>
<td>MCD12Q1.006</td>
<td>500 m×500 m</td>
<td>Annual</td>
<td>Main class of land cover and land use classes in 1 km×1 km pixels</td>
<td>(Sulla-Menashe and Friedl, 2019)</td>
</tr>
<tr>
<td>Elevation</td>
<td>MERIT DEM: Multi-Error-Removed Improved-Terrain DEM</td>
<td>90 m×90 m</td>
<td>/</td>
<td>Averaged Elevation in 1 km×1 km, averaged slope in 1 km×1 km pixels</td>
<td>(Yamazaki et al., 2017)</td>
</tr>
<tr>
<td>Impervious surface</td>
<td>Tsinghua/FROM-GLC/GAIA/v10</td>
<td>30 m×30 m</td>
<td>Annual</td>
<td>Proportion of impervious surface layer in 1 km×1 km pixels</td>
<td>(Gong et al., 2020)</td>
</tr>
<tr>
<td>Population density</td>
<td>WorldPop</td>
<td>1 km×1 km</td>
<td>Annual</td>
<td>Population density</td>
<td>(Gaughan et al., 2013)</td>
</tr>
<tr>
<td>Temporal variation</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>Year, Month</td>
<td>/</td>
</tr>
</tbody>
</table>
### Table 2. Computation of human thermal indices.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Human thermal index</th>
<th>Computation model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Apparent Temperature (indoors)</td>
<td>( AT_{\text{i}} = -1.3 + 0.92 \times T + 2.2 \times E_a )</td>
<td>Steadman (1979)</td>
</tr>
<tr>
<td>AT&lt;sub&gt;out&lt;/sub&gt;</td>
<td>Apparent Temperature (outdoors, in the shade)</td>
<td>( AT_{\text{out}} = -2.7 + 1.04 \times T + 2 \times E_a - 0.65 \times V )</td>
<td>Steadman, (1984)</td>
</tr>
<tr>
<td>DI</td>
<td>Discomfort Index</td>
<td>( DI = 0.5 \times WBT + 0.5 \times T )</td>
<td>Sohar et al., (1963)</td>
</tr>
<tr>
<td>ET</td>
<td>Effective Temperature</td>
<td>( ET = T - 0.4 \times \left( T - 10 \right) \times (1 - 0.001 \times RH) )</td>
<td>Gagge et al., (1972)</td>
</tr>
</tbody>
</table>
| HI | Heat Index* | \( HI^* = -0.804695 + 1.61139411 \times T - 2.338549 \times RH 
- 0.14611605 \times T \times RH 
- 1.238094 \times 10^{-2} \times T^2 
- 1.6424828 \times 10^{-2} \times RH^2 
+ 2.211732 \times 10^{-3} \times T^2 \times RH 
+ 7.2546 \times 10^{-4} \times T \times RH^2 
+ 3.582 \times 10^{-4} \times T^2 \times RH^2 \) | Rothfusz and Headquarters, (1990) |
| HMI | Humidex | \( HMI = T + 0.5555 \times (0.1 \times E_a - 10) \) | Masterton et al., (1979) |
| MDI | Modified discomfort index | \( MDI = 0.75 \times WBT + 0.38 \times T \) | Moran et al., (1998) |
| NET | Net Effective Temperature | \( NET = \frac{37 - T}{0.68 - 0.0014 \times RH + \frac{0.7}{1.12 + 1.4 \times V^{0.16}}} - 0.29 \times T \times (1 - 0.01 \times RH) \) | Houghton and Yaglou, (1923) |
| sWBGT | Simplified Wet Bulb Globe Temperature | \( sWBGT = 0.567 \times T + 0.0393 \times E_a + 3.94 \) | Gagge and Nishi, (1976) |
| WBT | Wet-bulb Temperature | \( WBT = T \times \left[ \text{atan} \left( 0.151977 \times (RH + 8.313659)^{0.5} \right) \right] + \text{atan} \left( T + RH \right) - \text{atan} \left( RH - 1.676331 \right) \) | Stull, (2011) |
| WCT | Wind Chill Temperature | \( WCT = 13.12 + 0.6215 \times T - 11.37 \times (V \times 3.6)^{0.16} \) | Osczewsiki and Bluestein, (2005) |

T is air temperature (°C), RH is relative humidity (%), V is wind speed (m/s), and \( E_a \) is actual water vapor pressure (kPa). Asterisk means that an adjustment is needed. All units of human thermal indices in this study are in degrees Celsius (°C).