Science of Remote Sensing 12 (2025) 100275

Contents lists available at ScienceDirect

Science of Remote Sensing

journal homepage: www.sciencedirect.com/journal/science-of-remote-sensing

Reliability of satellite, reanalysis and observation-based gridded temperature datasets for climate change impact studies in Bhutan

Nima Dorji [a], Joseph L. Awange [b], Ayalsew Zerihun [c,*]

[a] School of Earth and Planetary Sciences, Discipline of Spatial Science, Curtin University, WA, Australia
[b] Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[c] School of Molecular and Life Sciences, Curtin University, WA, Australia

ARTICLE INFO

Keywords:
Bhutan
Climate change
Complex topography
Global warming
MODIS LST
Reanalysis
Systematic bias

ABSTRACT

The impacts of global warming are pronounced in mountainous regions, yet a scarcity of long-term climate data hinders robust documentation. Reanalysis (ERA5, ERA5-Land, MERRA2), gridded observational (CRU TS), and satellite-derived (MODIS LST) datasets serve as alternatives, but their reliability for local-scale impact studies remains uncertain without rigorous evaluation. Here, we present the first comprehensive assessment of these datasets across Bhutan's complex topography, comparing them to *in-situ* observations (1996–2023) using systemic statistical metrics, which is a critical prerequisite for their applications. Results reveal that pre-corrected datasets contain severe systematic cold bias increasing with elevation at 3.1–4.2 °C/km, culminating to bias up to −19 °C in the high-altitude areas. The post-correction analysis reveals that elevation-corrected reanalyses data reduces mean bias by a maximum of 31 %. However, enhancement of spatial representativeness of temperature through dynamically estimated lapse rate on *in-situ* temperature markedly reduces mean bias across all datasets including MODIS-derived air temperature. The altitudinal bias gradient, depending on reanalyses data, is reduced to 0.1°C–0.8 °C/km. Despite these notable improvements in accuracy, MODIS LST and reanalyses/CRU datasets continue to exhibit over- and underestimation, respectively. These findings suggest that limitations of accuracy stem not only from model assimilation or interpolation, but also from limited spatial representativeness of station observations. Our findings underscore that the use of these datasets directly in climate impact studies is impractical without prior corrections. This work provides a framework for evaluating temperature products in mountainous regions, ensuring their utility for adaptation planning in Bhutan and analogous terrains globally.

1. Introduction

Anthropogenic activities have been the primary drivers of global warming, with global surface temperatures reaching 1.1 °C higher in 2011–2020 than in the pre-industrial period (IPCC, 2023). However, this warming, exhibits significant spatial variations with some regions experiencing higher temperature increase than the global average (IPCC, 2018) particularly in the mountainous regions where complex topography amplifies local climate impacts (Pepin et al., 2015). Bhutan's highly varied topography and climatic diversity are likely experiencing amplified impacts of climate change across key socio-economic sectors, particularly agriculture, which is sensitive to temperature shifts. Some reports indicate shifts in the distribution of agricultural crops towards higher altitudes, attributing the shift mainly to climate change (Kuensel, 2021; The Bhutanese, 2018). Despite these emerging impacts, rigorous climate change impact studies in Bhutan are constrained by the country's relatively short meteorological record, which only began in 1996 (NCHM, 2018). Although investigating the impacts of climate change is imperative, such studies can only be done using spatially consistent and reliably long-term climate datasets that can accurately represent the unique climatic conditions of Bhutan. While spatially consistent and long-term gridded climate products and satellite-derived datasets offer promising alternatives, their accuracy in representing Bhutan's intricate microclimates remains unverified. This study provides the first comprehensive evaluation of these datasets across Bhutan's elevation gradients, quantifying their biases and assessing their suitability for climate impact research. Our findings not only address a critical knowledge gap for Bhutan but also establish a methodological framework applicable to other data-scarce mountainous regions, ultimately supporting more reliable climate change adaptation strategies.

* Corresponding author
E-mail address: a.zerihun@curtin.edu.au (A. Zerihun).

https://doi.org/10.1016/j.srs.2025.100275
Received 24 March 2025; Received in revised form 18 August 2025; Accepted 20 August 2025
Available online 23 August 2025

Some of the valuable gridded data products used for climate research include; remotely sensed MODerate Resolution Imaging Spectroradiometer Land Surface Temperature (MODIS LST, Wan et al., 2015a), reanalysis datasets such as the European Center for Medium-range Weather Forecasts (ECMWF) Re-analysis version 5 (ERA5, Hersbach et al., 2020; ERA5-Land, Muñoz-Sabater et al., 2021), Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA2, Gelaro et al., 2017) by NASA's Global Modeling and Assimilation Office (GMAO) and the University of East Anglia's Climate Research Unit observation-based gridded Time Series (CRU TS, Harris et al., 2020). However, the precision of these datasets depends on the effectiveness of data assimilation (ERA5, ER5-Land and MERRA2), interpolation methods (CRU TS), spatial resolution, and the representation of land surface processes (Forsythe et al., 2015). Given the diverse techniques used to create gridded datasets, some degree of variability is unavoidable (Sun et al., 2018). Thus, it is necessary to assess the final products of the datasets. The spatial resolution is highly relevant in complex topographical regions, but the local effects are largely removed in the interpolation of surface station anomalies in the CRU TS (Rapaiˇc et al., 2015), which can introduce uncertainties as the efficiency of these techniques may vary depending on the type of terrain (Rana et al., 2015). The reanalyses are also affected by inhomogeneities in observational data streams and biases linked to incomplete model physics and inadequate spatial representation of topography (Screen and Simmonds, 2011). The limitations of accurately capturing the mesoscale climatic conditions are especially evident in the mountainous regions such as Bhutan, characterized by complex topography, sparse monitoring networks, and diverse microclimates (Buytaert et al., 2010; Böhner and Lehmkuhl, 2005). While in recent times, MODIS satellite products gained popularity in climate studies (Phan and Kappas, 2018), it is also not applicable directly, as MODIS LST may differ from actual station readings due to varied atmospheric conditions and land cover (Benali et al., 2012; Duan et al., 2019; Li et al., 2014; Wan et al., 2004; Yang et al., 2013). Consequently, evaluating the performance of these datasets is a crucial prerequisite for their application in impact studies.

The BhutanClim 1 km gridded dataset, developed by Lehner and Formayer (2023) using station data and ERA5, improves spatial detail but remains temporally limited (1996–2019) and potentially biased due to unvalidated reliance on ERA5. Extensive evaluations of reanalysis datasets—including ERA5, ERA5-Land and MERRA2—reveal systematic cold biases in mountainous regions such as the Tibetan Plateau and northeast India (Liu et al., 2024; Yang et al., 2022; Wang and Zeng, 2012; Huang et al., 2022; Li et al., 2022; Ghodichore et al., 2018). These biases, often linked to terrain complexity and elevation discrepancies, are corroborated across diverse geographies including Canada, Brazil, East Asia, and Africa (Goswami et al., 2024; Arau'jo et al., 2022; Kim and Lee, 2022; Tesfaye et al., 2017). Nonetheless, improved accuracy is observed in low-relief regions like South India (Valappil et al., 2023) and Siberia (Clelland et al., 2024). In fact, several other studies, for example Liu et al. (2024), Yang et al. (2022) and McNicholl et al. (2021) highlight reasonable accuracy in temperate regions, although challenges remain in tropical and topographically complex areas. The CRU TS dataset, though applied in dendroclimatic and glacier studies in Bhutan (Khandu et al., 2022; Krusic et al., 2015), has demonstrated poor performance over the Tibetan Plateau and Southeast Asia (Peng et al., 2019b; Kim and Lee, 2022; Salehie et al., 2022; Xu et al., 2009), calling into question its use as a reference in regional evaluations (Khandu et al., 2017).

Besides the elevation difference, the discrepancies can also stem from erroneous *in-situ* temperature observations due to their sparseness, location and technical faults in the instruments. Unlike reanalyses, MODIS LST dataset is derived from standalone, remote sensing observations. Previous efforts (e.g. Wan et al., 2004; Coll et al., 2005; Duan et al., 2019) have validated MODIS LST at multiple sites and land surfaces with accuracy <1 °C. However, their assessments are based on a homogeneous surface with known emissivity. Over the heterogeneous surfaces, its performance differs spatially and seasonally (Duan et al., 2019). To correct for such issues, air temperature have been derived from LST using empirical models such as linear regression (e.g. Benali et al., 2012; Hereher, 2019; Recondo et al., 2022; Shen and Leptoukh, 2011) with reference to station observations. The inclusion of MODIS products in the assessment would help ascertain whether the discrepancies in representing surface air temperatures in Bhutan are exclusively attributable to the limitations of reanalyses and CRU datasets or if other factors, such as observational gaps, sensor inaccuracies, or complex terrain effects, also contribute to the observed biases.

Although previous studies provide a robust foundation for the underlying issues of the gridded climate products in many parts of the world, they do not specifically address the issue over Bhutan. The climatic conditions in Bhutan may differ significantly from those of the Tibetan Plateau, Northeast India and other regions with changing natural conditions and socio-economic developments. Therefore, conducting similar research tailored to Bhutan is crucial to gaining a deeper understanding of how well the gridded datasets represent the unique local climate characteristics of the region. Unlike the CRU dataset, the reanalysis datasets have only been used in a limited number of applications in Bhutan; furthermore, no comprehensive evaluations have been carried out to assess their suitability for local climate studies. The utility of such long-term datasets in offering insights and their capacity to represent regional climatic conditions accurately remains questionable unless subjected to rigorous validation. To address the data gap that constrains climate change impact studies in Bhutan, this study aims to comprehensively evaluate the CRU, ERA5, ERA5-Land and MERRA2 temperature datasets against *in-situ* data of Bhutan. The application of various standard accuracy evaluation metrics, such as root mean squared error, mean absolute error, mean bias error and correlation coefficients, quantifies the systematic bias, which will ultimately determine the performance and reliability of satellite and gridded climate products in Bhutan. This critical study and the findings will enhance understanding of the applicability of these datasets in Bhutan and other regions with similar topography, providing a robust foundation for reliable climate studies and impact assessments.

2. Data and methods

2.1. Study area

The study area covers whole of Bhutan. Situated on the southern slopes of the eastern Himalayas, Bhutan exemplifies a mountainous environment where climate studies are lacking due to the scarcity of temperature datasets, further exacerbated by challenging topography. Bhutan's rugged terrain, spanning from lowland subtropical areas to high alpine regions, results in diverse climatic conditions over short distances, influenced by both topography and atmospheric circulation patterns (Hoy et al., 2016; NCHM, 2019). Moreover, the spatial coverage of existing weather stations is sparse, primarily in urban areas, limiting comprehensive climate monitoring and analysis (Khandu et al., 2017; Lehner and Formayer, 2023).

Bhutan spans approximately 38,394 km^2, with a geographical extent ranging from about.

88.7°E to 92.2°E longitude and 26.7°N to 28.5°N latitude, bordered by China to the north and India to the south, east, and west. Bhutan's altitude ranges dramatically from about 100 meters above sea level (m.a.s.l.) in the southern plains to over 7500 m.a.s.l. in the northern Himalayan mountains, creating diverse ecological zones and varied climates. Fig. 1 illustrates Bhutan, its geography and the location of weather stations.

Bhutan experiences four seasons: Spring, summer, autumn, and winter. Additionally, the country's diverse climate and the impact of microclimates across its steep terrain create six unique agro-ecological zones, from wet and humid regions in the south to cold alpine in the north, which are vulnerable to the impact of climate change. According to Bhutan's state of climate report NCHM (2023), the country's annual

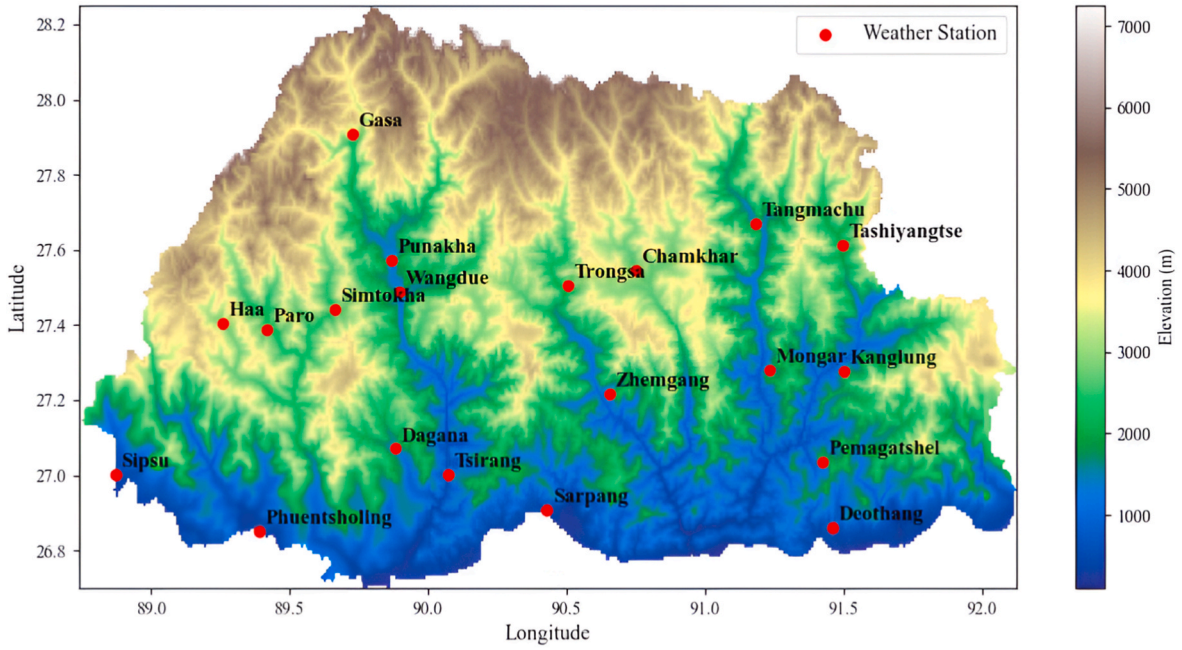

Fig. 1. The study area showing the map of Bhutan, topography and location of Class 'A' weather stations. The text associated with red dots is the names of the weather stations.

average maximum and minimum temperatures are 23.18 °C and 12.18 °C across the country, respectively. The highest maximum and minimum recorded temperatures are 38.5 °C and 10.5 °C, respectively during the period of 1996–2023.

2.2. Datasets

In this study, we use satellite (MODIS LST), reanalyses (ERA5, ERA5-Land and MERRA2) and gauge-based (CRU) temperature products, to evaluate their reliability in representing surface air temperature over Bhutan's complex topography. The performance of these datasets is specifically tested against *in-situ* surface air temperature data. The details of the datasets are briefly explained in the following sections.

2.2.1. In-situ data

A temporally consistent weather dataset, commencing in 1996, is derived from 20 Class 'A' stations—one each in Bhutan's 20 districts. While a few stations existed prior to 1996 (Lehner and Formayer, 2023), their records are largely absent. As noted by Lehner and Formayer (2023), the number of operational stations increased post-1996 but declined sharply after 2012. Class 'A' stations represent the longest-operating network in Bhutan, maintained continuously by the National Center for Hydrology and Meteorology (NCHM). Daily temperatures are measured twice—minimum at 9:00 a.m. and maximum at 3:00 p.m.—using mercury thermometers installed 2 m above ground level (personal communication, NCHM). Although Lehner and Formayer (2023) utilized data from 67 stations, we restrict our analysis to monthly mean temperatures from Class 'A' stations to prevent artefacts associated with imputed data post-2012. These monthly records, unavailable online, are acquired through formal requests in accordance with NCHM's data protocols. Station details are presented in Table A1.

2.2.2. Reanalysis data

ERA5 and ERA5-Land are the fifth-generation reanalysis datasets offering a broad range of temporal resolution of global climate and atmospheric data from 1950 onward (Hersbach et al., 2020; Bell et al., 2021; Muñoz-Sabater et al., 2021) and monthly ERA5 reanalysis available from 1940 (Soci et al., 2024). They provide a comprehensive record of past weather and climate by blending multiple observations with model outputs through advanced 4D-Var (Four-Dimensional Variational Data Assimilation) techniques, ensuring temporal consistency and improved accuracy by accounting for the evolution of observations within an assimilation window. ERA5 and ERA5-Land offer high temporal resolutions, ranging from hourly to monthly. ERA5 2 m screen level temperature is produced with a ~31 km horizontal resolution via 2-dimensional optimal interpolation (Hersbach et al., 2020) whereas ERA5-Land 2 m temperature is produced using linear interpolation based on triangular mesh and adjusted for elevational differences using a lapse rate derived from ERA5, resulting in a finer grid resolution of ~9 km (Muñoz-Sabater et al., 2021). The final product of ERA5-Land is the replay of ERA5's land variables and atmospheric forcings, with the key refinement being improved spatial resolution (Muñoz Sabater, 2019; Muñoz-Sabater et al., 2021), crucial for complex topography. These datasets include other key variables such as precipitation, wind speed, and radiative fluxes, making it suitable for applications like climate trend analysis, hydrological modeling, and agricultural research. While their global coverage and rigorous quality control ensure reliability, their performance may differ at local scale with complex topography (Wang and Zeng, 2012; Buytaert et al., 2010; Huang et al., 2021) such as Bhutan. ERA5 and ERA5-Land datasets are freely accessible through the Copernicus Climate Data Store, supporting diverse research and operational applications. This study uses monthly 2 m temperature for the accuracy assessment.

MERRA2 atmospheric reanalysis dataset is designed to provide a comprehensive record of weather and climate from 1980 to the present, using the Goddard Earth Observing System (GEOS) atmospheric model and data assimilation system. MERRA2 assimilates satellite observations, including aerosols, land surface data, and precipitation, to generate consistent and high-quality meteorological datasets (Gelaro et al., 2017). It employs a 3D-Var (Three-Dimensional Variational Data Assimilation) system to accommodate satellite-era data effectively. It is specifically tailored to represent the climate of the modern satellite era, incorporating aerosols' direct radiative effects and their interaction with meteorology, a unique feature compared to the earlier version, MERRA. The dataset is produced at a spatial resolutions of 0.5° × 0.625° and provides data at hourly and sub-hourly intervals (Rienecker et al.,

2011). It is one of the latest reanalysis products which focuses on capturing land, atmosphere, and ocean interactions with high observational accuracy, integrating satellite data to mitigate shifts in observational quality over time (Gelaro et al., 2017; Rienecker et al., 2011). This emphasis on observational reliability makes MERRA2 highly effective for studying climate processes, especially in regions with complex topographies or pronounced seasonal variability (Randles et al., 2017). MERRA2 2 m temperature is downloaded from https://daac.gsfc.nasa.gov.

2.2.3. Gauge-based gridded data

CRU TS dataset provides monthly climate data from observational records dating back to 1901. With a spatial resolution of 0.5° × 0.5°, CRU TS incorporates extensive station data, offering a valuable observational basis for analyzing century-long climate trends (Harris et al., 2020). Unlike reanalysis datasets, which incorporate observational data into model-generated fields, CRU TS relies heavily on direct station data, making it especially useful for validating trends over long periods (Jones et al., 2013). It is gridded using a climate anomaly method, where observed station data are interpolated using angular distance weighting by combining long-term climatology, which ensures that gridded values represent local climatic conditions adjusted for spatial variability (Harris et al., 2020). The CRU TS version 4.08 was downloaded from https://crudata.uea.ac.uk/cru/data/hrg/cru ts 4.08/.

2.2.4. Satellite data

The MODIS sensors, launched aboard NASA's Earth Observing System (EOS) Terra and Aqua satellites in 1999 and 2002, respectively (Salomonson et al., 2006; Wan et al., 2015a). The Terra and Aqua satellites overpass at 10:30 a.m./p.m. and 1:30 a.m./p.m., solar local time, respectively (Wang et al., 2006). Each satellites provide daytime and nighttime LST products, MOD11A1 (Wan et al., 2021) and MYD11A1 (Wan et al., 2015b), respectively. Both products offer daily LST at a 1 km spatial resolution under clear-sky conditions. These products are derived from thermal infrared (TIR) observations in bands 31 (10.78–11.28 μm) and 32 (11.77–12.27 μm) using a generalized split-window algorithm that corrects for atmospheric water vapour and surface emissivity effects (Wan, 2014). Together, there are four daily products, each sensor providing separate daytime and nighttime LST values. In this study, basic preconditions, such as the percentage of cloud cover and the closeness of overpass time to *in-situ* temperature reading time, were considered for the adoption of LST products. Thus, we use both the Terra and Aqua satellite products, MOD11A1 V061 and MYD11A1 V061. Although they are limited in representing the complete diurnal temperature cycle (Jin et al., 2011; Zhang et al., 2014), their overpass times are very close to station reading times, MOD11A1 for morning and MYD11A1 for afternoon reading. From our investigation on clear-sky pixels using quality flags embedded in the metadata, both the Terra and Aqua nighttime LST products contain 25 %–60 % cloud coverage, shown in Figure A1. The daytime LST from Aqua satellite overpassing in the afternoon has maximum cloud coverage of 30 %–70 %, while Terra, which overpasses in the morning, has the least cloud coverage of 10 %–50 %. Wan (2013) sets two levels of confidence for clear sky pixels, 95 % and 66 % on land below and above 2000 m.a.s.l., respectively. The former is a very strict criterion that in mountainous regions, such as Bhutan, which are often covered by clouds, cannot achieve representative spatial coverage; and the latter is flexible, which allows researchers to obtain better spatial coverage. Thus, in this study, we retained quality pixels at a confidence of ≥66 %, ensuring good spatial coverage. Example spatial maps are shown in Figure A2. Based on the recording time of *in-situ* stations, MOD11A1 and MYD11A1 are integrated to derive monthly composites of LST for our study. All good quality pixels with common dates and coordinates in MOD11A1 and MYD11A1, are used to calculate daily mean LST using Equation (2) of Xing et al. (2021) and Li et al. (2024). For the calculation of monthly LST, only pixels with valid LST observations for more than 60 % of the days in a given month were included. Using the *in-situ* station coordinates, we cross-checked the available pixels/grid corresponding to ground stations. We found that the MODIS monthly LST contains valid pixels at station locations, but has missing values in some months. These LST products are downloaded from the Application for Extracting and Exploring Analysis Ready Samples (AppEEARS), https://appeears.earthdatacloud.nasa.gov.

2.2.5. Digital elevation model (DEM)

The DEM of 30 arcseconds (~1 km) provided by NASA's Shuttle Radar Topography Mission (SRTM, NASA, 2013) is used in this study. SRTM DEM is freely available with several data partners' websites, and for our study we downloaded it from https://earthexplorer.usgs.gov/. The summary of datasets is provided in Table 1.

2.3. Methods

To evaluate the performance of MODIS LST, ERA5, ERA5-Land, MERRA2, and CRU TS datasets in accurately representing surface air temperature, we followed well-established methods for climate data evaluation, including monthly and seasonal mean temperature time series for grid-to-station and sub-regional subsets over Bhutan.

2.3.1. Harmonization of datasets

The datasets exhibit varying temporal extents. Specifically, CRU TS provides monthly mean temperatures from 1901, ERA5 from 1940, ERA5-Land from 1950, MERRA2 from 1980, and MODIS LST from the Terra and Aqua satellites from 2001 and 2003, respectively. Despite the pursuit of long-term, reliable climate records, our evaluation period is standardized to 1996–2023, aligning with the temporal span of the reference station dataset. MODIS LST products (MOD11A1 and MYD11A1) are temporally harmonized over their overlapping period, restricting their use in this study to 2003–2023. While satellite, reanalysis, and CRU datasets are available in standardized gridded formats, the station data remain spatially discrete, representing point-based observations at fixed locations. To enable meaningful comparisons and to accurately evaluate model performance, comprehensive quality inspections and interpolating the point-based station data into gridded formats is crucial. Harmonization of the gridded datasets to a common grid by resampling or interpolation is also essential. These steps facilitate spatiotemporally explicit performance assessments of ERA5, ERA5-Land, MERRA2 and CRU datasets against *in-situ* observational data. Before conducting spatial interpolation, the station dataset underwent rigorous quality control measures, including temporal coverage and data continuity assessments, to address missing values. Two ground stations, Tangmachu and Gasa, are excluded from all forms of analyses and comparisons due to the limited temporal extent of their records, which commenced in 2013 and 2003, respectively. This exclusion ensured a consistent observation period across all stations. A thorough evaluation of the station dataset revealed minimal instances of missing data. For the few months with missing records, values were imputed by calculating the mean of the preceding and subsequent months, thereby preserving the temporal integrity of the dataset.

2.3.2. Interpolation of datasets

As summarized in Table 1, the spatial resolutions of reanalysis datasets vary markedly leading to spatial misalignment where grid cells may either omit station data or aggregate multiple stations. To standardize spatial comparability, all datasets were interpolated to a common 0.025° × 0.025° grid. Spatial interpolation methods vary in complexity and applicability (Mitas and Mitasova, 1999; Li and Heap, 2008, 2011). Inverse distance weighting (Shepard, 1968) is straightforward but assumes uniform spatial influence, making it less reliable in complex terrains (Li and Heap, 2008). Spline interpolation offers smooth surfaces but may overshoot in sparse data regions (Wahba, 1990). Linear, bilinear, and cubic methods are fast but lack spatial awareness

Table 1
Summary information of the datasets used in the present study.

Datasets	Temporal Resolution	Spatial Resolution	Temporal Coverage	Study	Reference	Source Period
Observed	Monthly	point data	1996–2023	1996–2023	NCHM	NCHM
ERA5	Monthly	0.25° × 0.25°	1940–2023	1996–2023	Hersbach et al. (2020) Soci et al. (2024)	https://cds.climate.copernicus.eu
ERA5-Land	Monthly	0.1° × 0.1°	1940–2023	1996–2023	Hersbach et al. (2020) Soci et al. (2024)	https://cds.climate copernicus.eu
MERRA2	Monthly	0.5° × 0.625°	1980–2023	1996–2023	Gelaro et al. (2017)	https://disc.gsfc.nasa.gov
CRU TS	Monthly	0.5° × 0.5°	1901–2023	1996–2023	Harris et al. (2020)	https://crudata.uea.ac.uk/
MODIS LST	Daily	1 km × 1 km	2001–2023	2003–2023	Wan et al. (2015a)	https://appeears.earth datacloud.nasa.gov
SRTM DEM	–	30 arc-second	–	–	NASA (2013)	https://earthexplorer.usgs.gov

and are unsuitable for irregularly spaced data (Li and Heap, 2011). Parameter-elevation Regressions on Independent Slopes Model (PRISM, Daly et al., 2002), is yet another interpolation method incorporating various physiographic features in the simulation, but varies its performance in complex topography and sparse observation network (Daly et al., 2008). Park and Jang (2016) state that there is no significant difference in terms of efficiency between kriging and PRISM. Moreover, kriging has been widely used in surrounding regions (e.g. Zhao et al., 2008; Song et al., 2016; Huang et al., 2021; Ali et al., 2024) and several other studies. Universal kriging, a geostatistical interpolation method, is widely regarded as one of the most robust techniques for spatial interpolation (Zimmerman et al., 1999) and better than other methods (Ibrahim and Nasser, 2017). It incorporates trend estimation errors and minimizes spatial interpolation uncertainties (Brus and Heuvelink, 2007; Koike et al., 2001; Goovaerts, 1997), making it ideal for this application. Following Nguyen et al. (2015); Holdaway (1996) and Shen et al. (2023), we interpolate all datasets using universal kriging, incorporating spatial dependence via a spherical variogram model.

In spatial climate analyses, computing experimental variograms is sometimes impractical due to quadratic scaling of pairwise semivariance calculation with sample size, which is especially prohibitive for multi-temporal datasets (Cressie, 2015; Webster and Oliver, 2007; Atkinson and Lloyd, 2007). The challenge is further compounded by non-convergent data at some spatial locations due to incomplete spatial coverage, depending on the grid size of modeled datasets, hindering accurate empirical variogram estimation, and if untreated, can bias spatial correlation modeling (Moyeed and Papritz, 2002; Sun et al., 2011). To address these constraints, theoretical variogram models (linear, exponential and spherical) are directly parameterized using established heuristics from climatology and geostatistics, and best variogram models are selected based on Akaike's information criterion (AIC, Akaike, 2011; Portet, 2020) and leave-one-out cross validation (LOOCV) to quantify predictive accuracy and avoid overfitting (Isaaks and Srivastava, 1989; Hoeting et al., 2006). The AIC and LOOCV model parameterization and selection identified the spherical model yielding the lowest AIC value and RMSE <1 as the good-fit variogram for kriging interpolation (Figure A3). Kriging is also employed in spatial interpolation as an optimal linear unbiased predictor that minimizes estimation variance while accounting for spatial correlation (Oliver and Webster, 1990), returning the exact value when coincided with sample locations (Burrough and McDonnell, 1998; Sajid et al., 2013; Carter et al., 2018). The coarser temperature datasets are treated as point samples located at grid cell centers which are then interpolated to finer grids, an approach also used in Shen et al. (2022) and Hofstra et al. (2008). For each time step, valid temperature data points were systematically extracted and evaluated for variability. Where variability was negligible (standard deviation < 0.01), the grid was filled with the mean of available values. In rare cases where kriging encountered computational limitations, nearest-neighbour interpolation was employed as a fallback to ensure consistent grid generation. ERA5-Land and in-situ temperature were selected for validation with respect to their respective pre- and post-interpolation datasets in terms of monthly temperature time series. The comparison reveals RMSE< 1 °C as shown in Fig. A4(a) and (b) which aligns with the findings of comparative studies such as Aalto et al. (2013); Declercq (1996) and Hofstra et al. (2008) all of whom acknowledging kriging with least deviations both in temporal and spatial scales. The temporal adjustments are not required, as all datasets, including station data, are inherently provided at monthly temporal resolutions.

2.3.3. Gridding in-situ data using lapse rate

Although station-based surface air temperature observations are accurate at their specific sites and immediate surroundings, they are frequently influenced by local topographic and urban-induced warm air circulations, given their typical placement in valleys and urbanized zones. Consequently, their spatial representativeness diminishes, particularly in complex terrains like Bhutan. As described in Section 2.3.2, in this study, station temperatures are interpolated via universal kriging. However, White and Saunders (2000) and Phillips and Marks (1996) highlight that interpolation inherently introduces uncertainty, which is exacerbated in data-scarce regions. Bhutan's sparse observational network amplifies this challenge, where even advanced interpolation techniques may yield uncertain estimates in remote and high-altitude areas lacking proximal stations. Moreover, the concentration of stations in the interior leads to extrapolation over peripheral zones, further compromising the fidelity of the gridded reference surface air temperature. To derive regionally representative surface air temperature, we employ a dynamically estimated lapse rate (DELR) and generate gridded fields using all 18 Class 'A' stations. The DELR is estimated using linear regression following Dorji et al. (2016) which varies at each time step. Our DELR ranges from 4 °C/km to 8.4 °C/km (Fig. 2). Using DELR, we propagate using Equation (1) which expresses that temperature drops as we move towards higher elevation (Fairbridge and Oliver, 1987; Daidzic, 2019).

$$T_{grid,t} = T_{station,t} - \Gamma_t \cdot (z_{grid} - z_{station}) \quad (1)$$

Where, $T_{grid,t}$ is the estimated temperature of a grid at time step t, (in °C) $T_{station,t}$ is the known station temperature (in °C), Γ_t (°C/km) is DELR at each time step and $z_{grid} - z_{station}$ is the elevation difference of the grid with respect to the known station.

2.3.4. Estimation of air temperature (T_{air}) from MODIS LST

The land surface temperatures do not represent air temperature due to the heterogeneity of land surface (Benali et al., 2012; Wan, 2008; Duan et al., 2019; Shen and Leptoukh, 2011). Therefore, MODIS LST cannot be compared with station observations directly unless the necessary LST correction is taken into consideration. The atmospherically corrected MODIS LST yields RMSE of 8.88 °C and MBE of 8.68 °C despite its high correlation (R = 0.92) with monthly mean time series. Given such high discrepancy, we estimated MODIS air temperature (T_{air}) from MODIS LST using linear regression (Benali et al., 2012; Shen and Leptoukh, 2011; Hereher, 2019; Yan et al., 2009) given by Equation (2).

$$T_{air} = a + b \cdot T_{LST} \quad (2)$$

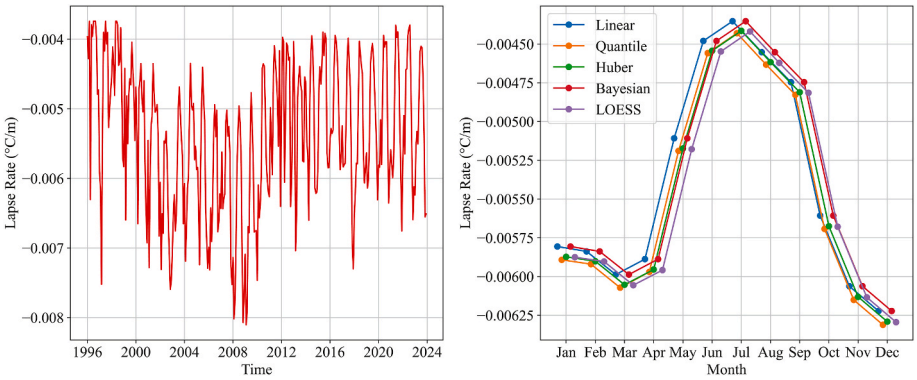

Fig. 2. Dynamically estimated temperature lapse rates (DELR): Five regression models (left): Linear, Quantile, Huber (Robust), Bayesian, and LOESS regression were first evaluated based on monthly climatology from 1996 to 2023. There are minor differences in lapse rates among the models. The monthly temperature data (station) consists of 336 time steps. The linear regression coincides with the most common lapse rate values (133 time steps) among the models; hence, the linear model is chosen to calculate the DELR (right). Lapse rates tend to be steeper in winter months and weaker during the monsoon season.

Where, T_{air} is the estimated temperature (in °C), T_{LST} is the LST from MODIS (in °C), a and b are model parameters for tuning which may differ from region to region. Although the calibration of model parameters, produced (T_{air}) series with reduced RMSE and MBE, it was less effective for correcting the seasonal cycles. To incorporate cyclic patterns of time series, we use harmonic components following Recondo et al. (2022) given by Equation (3).

$$T_{air} = a + b \cdot T_{LST} + c.\sin\left(\frac{2\pi t}{12}\right) + d.\cos\left(\frac{2\pi t}{12}\right) \quad (3)$$

Where the *a, b, c* and *d* are model parameters. Sine and cosine functions account for the correction of seasonal variations and *t* is time. The utilization of Equation (3) yields RMSE = 2.56 °C and MBE = 0.85 °C. The comparison of MODIS LST (raw), T_{air} with linear regression and harmonic regression using Fourier terms are given in Fig. A5. In our detailed evaluation, we use T_{air} obtained from harmonic regression with elevation-corrected temperature time series of reanalyses, CRU and station datasets.

2.3.5. *Elevation correction of reanalyses datasets*

The elevation correction was done using several methods by following Gao et al. (2012, 2017); Zhao and Qian (2025) and Qin et al. (2020). In these studies, station elevations are used as reference elevations to correct the corresponding grid elevations of reanalyses data. In Method I, we corrected the elevation of reanalyses data using vertical profile lapse rates (VPLR) calculated at different pressure levels given by Equations (4)–(8). In Method II, the elevation correction and corresponding temperature correction is done by using a single lapse rate derived from the temperature difference of two extreme pressure/elevation levels as shown in Fig. 3. Method III uses a universal lapse rate of 6.5 °C/km. According to our station network and their respective locations from Table S1, the stations' elevation ranges from 392 m to 2764 m above mean sea level. Based on Bhutan's topography, we use five pressure levels from reanalyses to determine the temperature lapse rate. The representative pressure levels in the context of Bhutan's topography are 975 hPa, 850 hPa, 750 hPa, 600 hPa, 450 hPa, and 350 hPa and their corresponding heights derived from geopotential heights over acceleration due to gravity (9.80655 m/s^2) as determined by Gao et al. (2012).

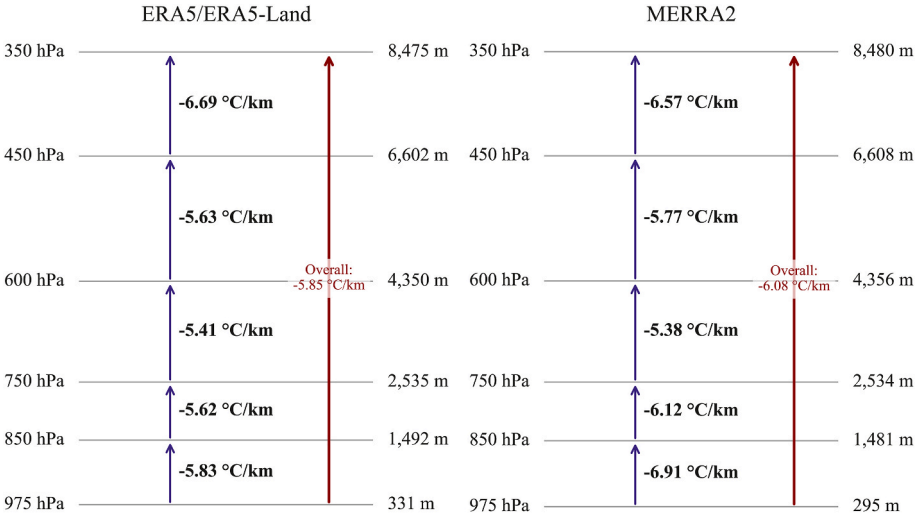

Fig. 3. Temperature lapse rates estimated from pressure level data of ERA5/ERA5-Land and MERRA2 datasets across six standard atmospheric pressure levels (975, 850, 750, 600, 450, and 350 hPa). Corresponding elevations in (m) are shown along with calculated lapse rates (in °C/km) between each pressure level and elevation. For ERA5/ERA5-Land, the lapse rate is steepest in the upper atmosphere (e.g., between 350 and 450 hPa) and shallower at lower levels (e.g., 850–975 hPa). For MERRA2, lapse rates were highest at the foothills and highest elevations, with the intermediate elevations having the lowest lapse rate. The overall lapse rate is calculated as −5.85 °C/km for ERA5/ERA5-Land and −6.08 °C/km for MERRA2, reflecting the average vertical temperature gradient across all levels. These values are crucial for correcting surface temperatures in complex terrain.

$$\Gamma_{350-450} = \frac{T_{350} - T_{450}}{z_{350} - z_{450}} \quad (4)$$

$$\Gamma_{450-600} = \frac{T_{450} - T_{600}}{z_{450} - z_{600}} \quad (5)$$

$$\Gamma_{600-750} = \frac{T_{600} - T_{750}}{z_{600} - z_{750}} \quad (6)$$

$$\Gamma_{750-850} = \frac{T_{750} - T_{850}}{z_{750} - z_{850}} \quad (7)$$

$$\Gamma_{850-975} = \frac{T_{850} - T_{975}}{z_{850} - z_{975}} \quad (8)$$

Where Γ and T are the LR in °C/km and °C, respectively, at various pressure levels and corresponding heights z in meters. The pressure levels for calculating the LR are set at 350–450, 450–600, 600–750, 750–850, and 850–975 in hPa.

In the elevation correction and temperature adjustment at the station grids, we apply Equation 9.

$$T_{corrected,t} = T_{model,t} + \Gamma \cdot \Delta h \quad (9)$$

Where, $T_{corrected,t}$ is the corrected temperature of grid corresponding to station, $T_{model,t}$ is 2-m temperature from reanalyses, Γ is LR in °C/km at various pressure level and heights calculated using Equations (4)–(8) and Δh is the elevation difference in m, between the nearest grid and the corresponding station.

2.3.6. Statisitcal Aanalysis

The monthly mean and seasonal time series of the satellite, reanalyses and the CRU datasets are statistically compared with *in-situ* temperatures following the methods from Liu et al. (2024); Chen et al. (2021); Ghodichore et al. (2018); Huang et al. (2022); Sun et al. (2023); Wang et al. (2024); Yan et al. (2020); Yang et al. (2022) and Hu et al. (2014). The statistical error metrics such as (a) mean absolute error (MAE, Equation (10)), (b) root mean squared error (RMSE, Equation (11)), (c) mean bias error (MBE, Equation (12)), and (d) Pearson correlation (R, Equation (13)), are applied to evaluate the reliability and robustness for quantifying the accuracy and bias in temperature datasets.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2} \quad (11)$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i) \quad (12)$$

$$R = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}\cdot\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2}} \quad (13)$$

Where N is the number of observations, x_i represents the observed values, y_i represents the estimated values, and $\overline{x}$ and $\overline{y}$ denote the mean of observed and estimated values, respectively.

2.3.7. Evaluation approaches

A comprehensive evaluation of satellite-derived, reanalysis, and observation-based gridded datasets is conducted using raw, spatially interpolated, and elevation-adjusted temperature datasets. The raw datasets comprise point-based *in-situ* observations, and quality-screened MODIS LST. The *in-situ,* the CRU and reanalysis datasets that are interpolated using universal kriging are referred to as interpolated, whereas station data adjusted using DELR, and estimated MODIS T_{air}, elevation-adjusted temperature datasets are referred to as corrected datasets. Monthly and seasonal mean temperature time series comparisons are conducted among the interpolated datasets, including raw MODIS LST, with reference to interpolated station data, to understand discrepancies related to statistical errors and correlation. Building upon these results, a second approach was adopted, wherein well-aligned grid points from the reanalysis datasets were extracted, and their time series were directly compared against the corresponding stations (e.g. Scherrer, 2020; Wang and Zeng, 2012; Huang et al., 2021, 2023; Chen et al., 2021; Ghodichore et al., 2018; Frauenfeld et al., 2005; Nelli et al., 2024). The second approach was conducted explicitly for all 18 stations of Bhutan, providing a station-wise validation framework. Following this detailed station-level evaluation, a third phase of analysis was undertaken. Here, the study area was delineated into four distinct regions to assess elevation-dependent variations in model performance and observational alignment, as outlined in Table 2. This regional analysis adds an altitudinal dimension to the evaluation process, enhancing the understanding of model accuracy across different geographic contexts. This approach is similar to Huang et al. (2022, 2021); Scherrer (2020) and Begert and Frei (2018).

Statistical error metrics were employed across all three approaches to assess their comparative performance. To provide insights into the elevation-dependent performance of temperature datasets, we examined the relationship between elevation and MBE in satellite, reanalysis and the CRU datasets relative to observed data (1996–2023) which determines the change of bias with elevation using linear regression.

3. Results

3.1. Evaluation of raw and interpolated datasets

The temporal evolution of mean monthly and seasonal temperatures from the uncorrected MODIS LST, reanalysis datasets, and CRU TS was assessed against *in-situ* observations from 1996 to 2023 (Fig. 4a–e). Despite exhibiting broadly similar temporal patterns, systematic biases are apparent, particularly a persistent cold bias across all datasets. The observed station temperature spans 8–24 °C, whereas reanalysis and CRU counterparts consistently remain below 16 °C (Fig. 4a), signifying pronounced underestimation. In contrast, MODIS LST overestimates station temperatures, ranging up to nearly 30 °C. Although CRU is an observation-based gridded product, its temporal dynamics closely resemble those of the reanalysis datasets.

Error metrics substantiate these discrepancies: RMSE (Fig. 5a) and MAE (Fig. 5b) vary from 2.5 °C (MODIS LST) to over 8 °C (ERA5-Land), while MBE (Fig. 5c) ranges from −8 °C to 2 °C (MODIS LST). Correlation coefficients (Fig. 5d) show strong monthly agreement overall, though seasonal variability is evident, with R-values from 0.2 (summer, MODIS LST) to 0.8 (winter, MERRA2).

Spatial diagnostics of these error metrics (Fig. 6) reveal increasing RMSE, MAE, and.

MBE with elevation for CRU and the reanalyses. Both RMSE and MAE are lowest (~0.4 °C) in the southern lowlands and highest (~18 °C) in

Table 2
Delineation of sub-regions based on altitudes and number of representative stations within each region. NIL** indicates that, for this study period, there is no station in the delineated zone. Further information is provided in Section 2.2.1. Featured number of stations in this table is 20, and the exclusion of some stations in this study is explained in Section 2.3.1.

Region	Bounding Altitude (m)	No. of stations
Southern Foothills (SF)	≤ 700	4
Sub Himalaya (SH)	$700 < altitude \leq 1600$	7
Inner Himalaya (IH)	$1600 < altitude \leq 3000$	9
Greater Himalaya (GH)	> 3000	NIL**

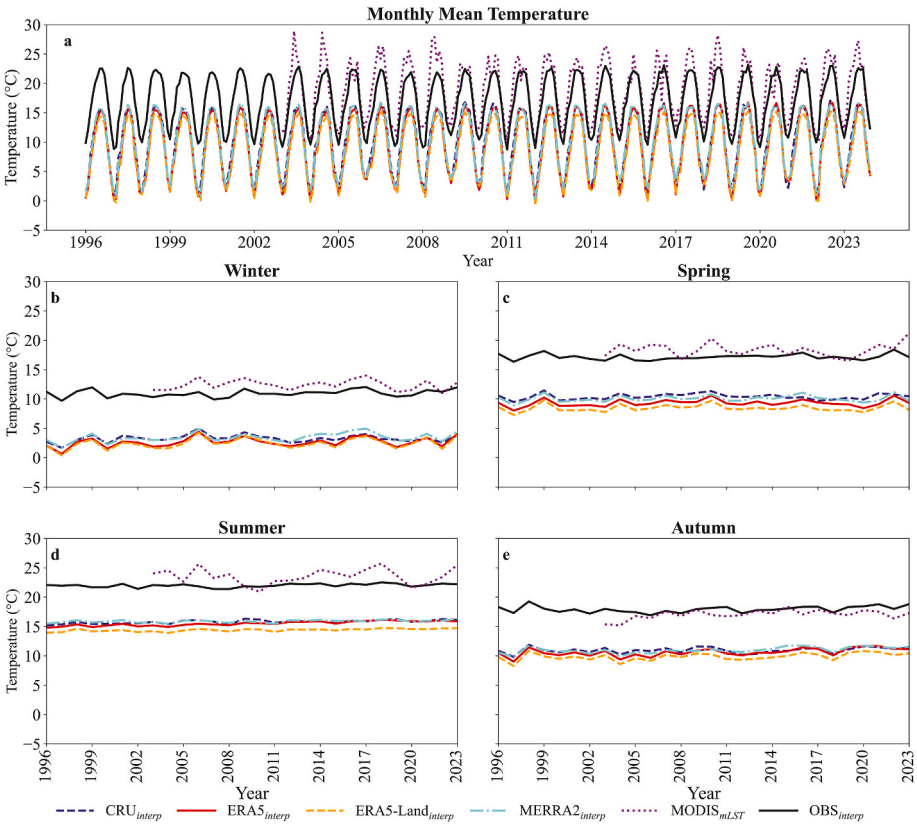

Fig. 4. Mean monthly (a) and seasonal (b–e) temperature time series from 1996 to 2023 for interpolated CRU, ERA5, ERA5-Land, MERRA2 and station datasets, and from 2003 to 2023 for MODIS LST. The time series drawn in this figure are determined from the spatially interpolated datasets, but elevation differences of reanalyses are not corrected.

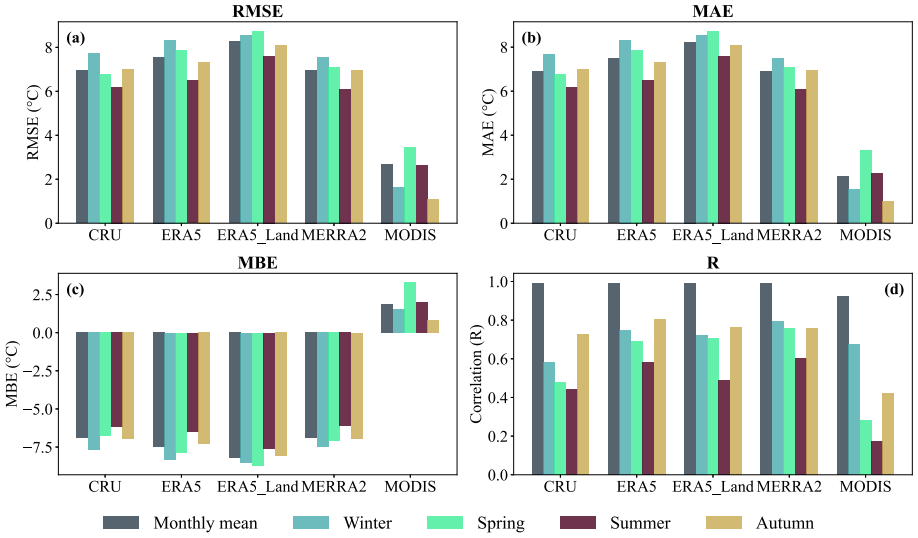

Fig. 5. Comparison of statistically computed errors and correlations (R) of the datasets with respect to interpolated station observation from 1996 to 2023 for raw CRU, ERA5, ERA5-Land, MERRA2, and from 2003 to 2023 for MODIS LST (raw). The values of RMSE (a), MAE (b), MBE (c) and R (d) are derived from the time series given in Fig. 4. All datasets except MODIS LST shows cold bias across all seasons. R reveals a uniformly strong association in the case of mean monthly temperature, but it varies across the seasons.

the northern high mountains, particularly for ERA5 and ERA5-Land. MODIS LST displays heterogeneous spatial patterns but similarly elevated errors in mountainous terrain. The datasets exhibit a gradient from warm bias in the south (e.g., MBE = 3.4 °C, ERA5) to severe cold bias in the north (>17 °C, ERA5-Land). MODIS LST shows both extreme overestimation (~14 °C) and underestimation (~–19 °C) in complex topography.

Grid-wise comparison with station observations (Fig. 7) confirms elevated errors at higher-altitude stations or those enclosed by mountainous barriers. The heatmap in Fig. 8 further illustrates station-level disparities, with the lowest MBE (0.03 °C, MERRA2) recorded at 1567 m.a.s.l., and the highest (−11.94 °C, ERA5-Land) at 2764 m.a.s.l. All

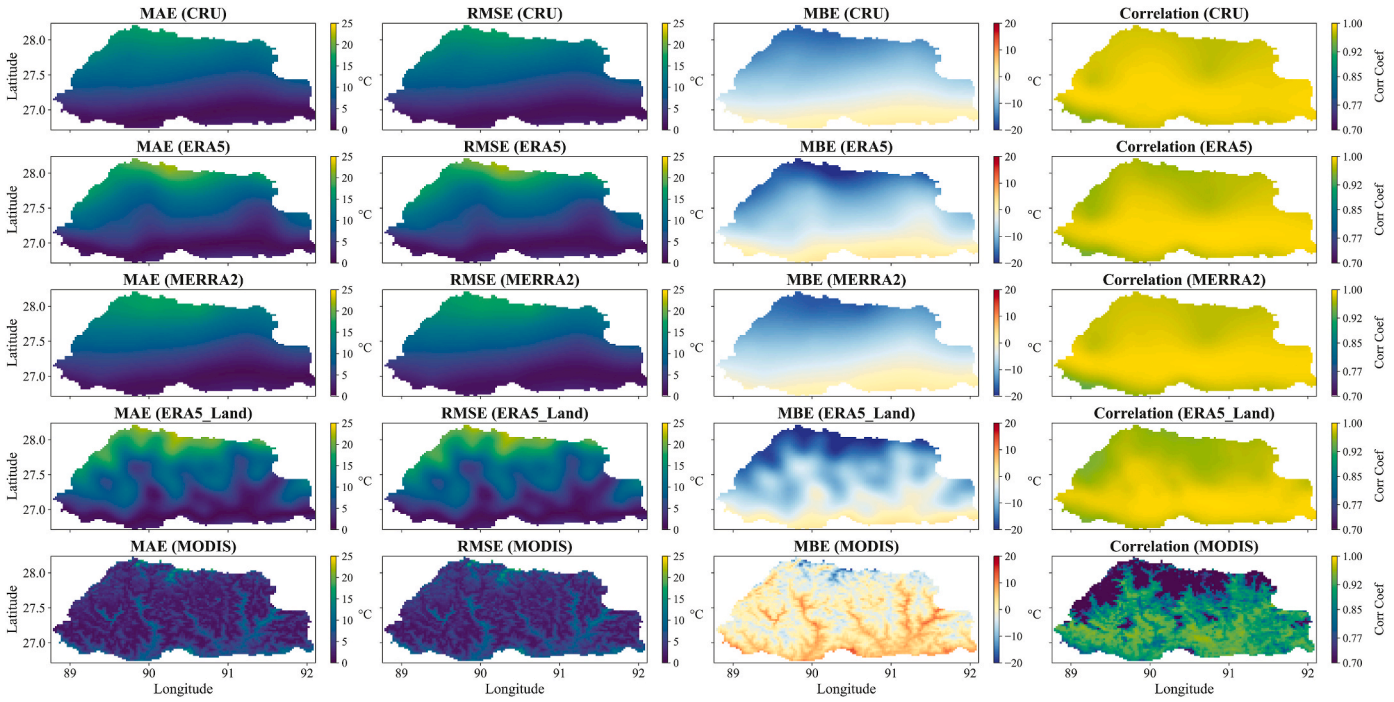

Fig. 6. Spatial patterns of statistical errors (RMSE, MAE and MBE) and correlation coefficient (R) across Bhutan for all products. The spatial characteristics are determined from interpolated CRU, ERA5, ERA5-Land, MERRA2 and raw MODIS LST. Temporally, all datasets represent from 1996 to 2023 except MODIS LST from 2003 to 2023.

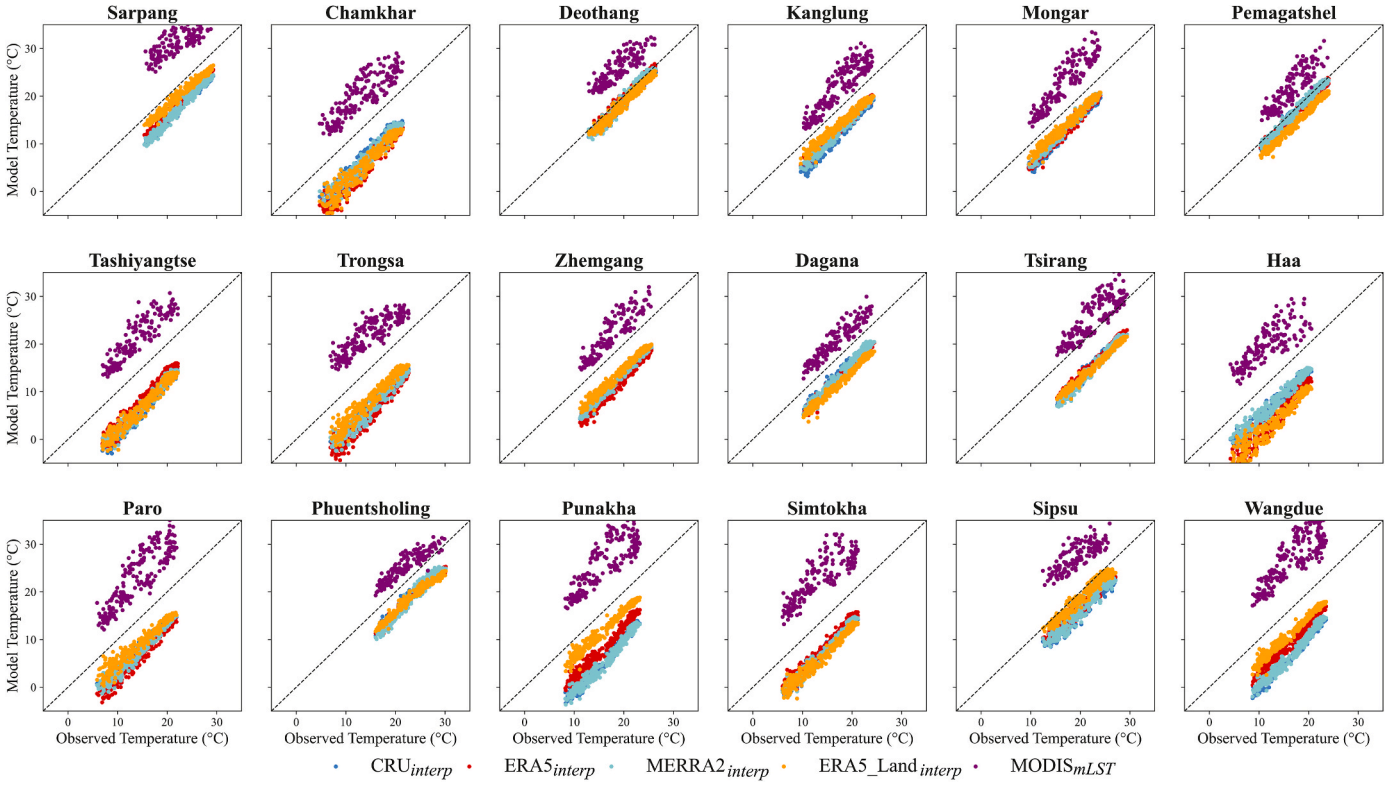

Fig. 7. Evaluation of interpolated CRU, ERA5, ERA5-land and MERRA2 temperature (1996–2023) and raw MODIS LST (2003–2023) against raw station temperature. Bias varies with station locations; for most stations, the CRU and reanalyses datasets exhibit a consistently cold bias, while MODIS LST shows a consistently warm bias across all stations.

datasets exhibit altitude-dependent error escalation, with stronger agreement at lower elevations (see Table 3 for elevation differences). Notably, some lowland stations also show warm bias in CRU and reanalyses. MODIS LST consistently overestimates across all stations (MBE: 4.5–13.98 °C), particularly in low-elevation zones. Trends in RMSE and MAE mirror those of MBE. The correlation remains robust for

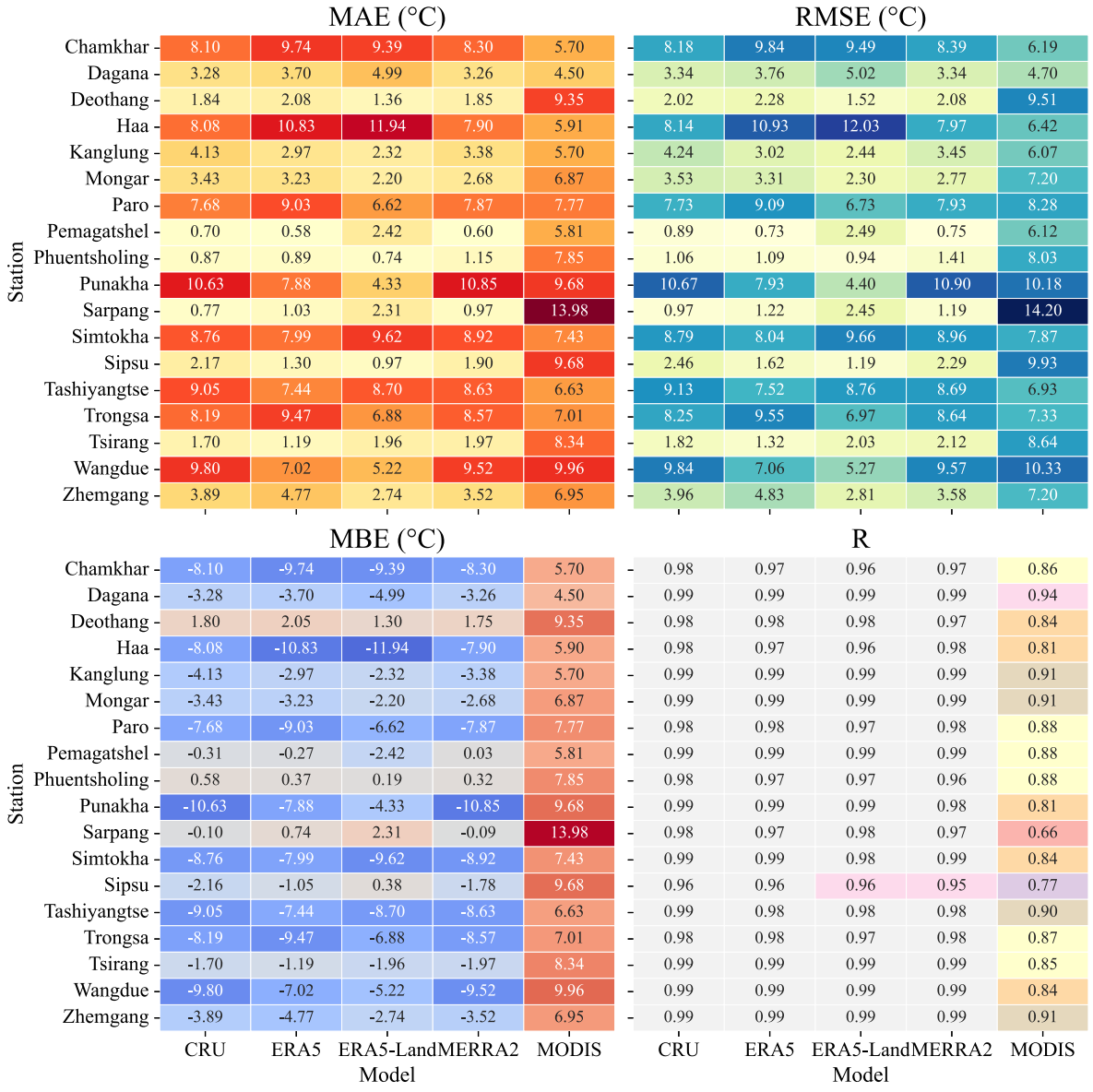

Fig. 8. Heatmap of RMSE, MAE, MBE and R for interpolated CRU and reanalysis datasets, including raw MODIS LST, with reference to point-based in-situ data. In the colour patterns, the darker the red colour, the higher the MAE is; the darker the blue, the higher the RMSE; in the heatmap of MBE, the darker the blue, the more severe the cold bias is, and the darker the red, the warmer the bias is.

CRU and reanalyses ($R \geq 0.99$), while MODIS LST demonstrates comparatively variable agreement ($R = 0.77–0.94$).

Sub-regional stratification by altitude corroborates systematic underestimation by CRU and reanalyses in higher-elevation zones and overestimation below 700 m.a.s.l. (Fig. 9). The temperature gap widens progressively from SF to GH subregions, both in monthly and seasonal means. MODIS LST exhibits an inverse trend, with narrowing gaps toward GH and eventual underestimation. Quantified differences (Fig. 10) show that MAE and RMSE for CRU and reanalyses increase from ~1 °C (SF) to ~14 °C (GH), whereas MODIS LST errors peak in SF (~9 °C) and minimize in IH (~3 °C). The elevation–bias relationship (Fig. 11) indicates a clear lapse-rate-like gradient, with underestimation intensifying at a rate of 3.2 °C/km (CRU) to 4.2 °C/km (ERA5-Land). MODIS LST exhibits a shallower lapse-like trend (3.1 °C/km) compared to the reanalyses and CRU.

3.2. Evaluation of elevation corrected datasets

As outlined in Section 2.3.5, three methodological frameworks were implemented for elevation correction. The CRU TS dataset does not incorporate elevation explicitly in its interpolation, relying instead on angular-distance weighting (ADW) of station anomalies, which considers spatial proximity but omits elevation differentials (Harris et al., 2020; Jones et al., 2013). Consequently, elevation-corrected versions of CRU TS are unavailable. Nevertheless, CRU TS is retained in subsequent comparisons to assess its performance relative to elevation-corrected reanalyses and the MODIS-derived T_{air}.

Observed temperatures corrected using DELR serve as the reference. Elevation-adjusted reanalysis and MODIS datasets were generated using all three methods, and the corresponding statistical metrics are summarized in the heatmap in Fig. 12. Among them, MERRA2 demonstrates the most pronounced reduction in MBE under Method I, as also shown in Table 4. Therefore, subsequent evaluations adopt Method I as the benchmark. Table 4 further quantifies the impacts of elevation correction across three comparative groups: the original comparison (refer to Section 3.1), a direct comparison between raw uncorrected *in-situ* and Metho I-elevation-corrected counterparts (yielding 23–31 % reductions in RMSE and MBE), and a third comparison between Method I-corrected

Table 3
Elevation (z) and elevation difference (z_{diff}) between station and corresponding grids of reanalysis datasets. The elevation difference presented in the table is obtained by subtracting the reanalysis elevation from station elevation.

Station	$Station_z$	$ERA5_z$	$ERA5_{zdiff}$	$ERA5\text{-}Land_z$	$ERA5\text{-}Land_{zdiff}$	$MERRA2_z$	$MERRA2_{zdiff}$
Sarpang	392	882	−490	413	−21	1288	−896
Chamkhar	2573	3395	−822	3199	−626	3069	−496
Deothang	658	713	−55	711	−53	1097	−439
Kanglung	1987	1906	81	1732	255	2160	−173
Mongar	1564	1878	−314	1574	−10	2138	−574
Pemagatshel	1567	1224	343	1445	122	1488	79
Tashiyangtse	1839	2840	−1001	2786	−947	3134	−1295
Trongsa	2167	3300	−1133	2670	−503	2996	−829
Zhemgang	1905	2233	−328	1733	172	2137	−232
Dagana	1531	1949	−418	2060	−529	1881	−350
Tsirang	1568	1388	180	1292	276	1639	−71
Haa	2764	3539	−775	3670	−906	2869	−105
Paro	2393	3235	−842	2798	−405	2803	−410
Phuentsholing	415	970	−555	784	−369	1140	−725
Punakha	1276	2908	−1632	1856	−580	3297	−2021
Simtokha	2310	3011	−701	3127	−817	2981	−671
Sipsu	424	1172	−748	608	−184	1672	−1248
Wangdue	1204	2744	−1540	2022	−818	3085	−1881

datasets and the DELR-based observed gridded data, which shows substantial MBE reductions—residuals ranging from 0.64 °C (MERRA2) to 1.647 °C (ERA5-Land)—though the datasets still slightly underestimate observations.

The post-correction grid-station comparison depicted in Fig. 13 support the conclusion that elevation correction substantially improves dataset fidelity. The raw station temperature remains as the reference in post-correction grid-station time series comparison. Notably, Punakha and Wangdue stations, which initially exhibited MBEs of 10.85 °C and 9.52 °C respectively, see reductions to under 1 °C post-correction. While elevation correction improved accuracy for most stations, elevated locations (>2000 m.a.s.l.) still exhibit residual biases of up to ~6 °C. MODIS-derived T_{air} similarly shows significant MBE reduction (−0.04 to −6.8 °C), with largest post-correction errors concentrated in the valley and urban areas.

Though elevation correction is not feasible for CRU, a long-term monthly climatology comparison (1996–2023) reveals minimal overall bias (MBE = −0.13 °C; Fig. 14). However, spatially disaggregated analyses (Fig. 12) indicate that CRU still underperforms at individual stations, despite closely matching observed climatologies at broader spatial scales. Finally, the elevation–bias relationships for the corrected datasets (Fig. 15) reveal substantial attenuation of the altitudinal bias gradients seen in raw data (Fig. 11). For instance, the MERRA2 bias gradient is now nearly flat (0.1 °C/km), whereas MODIS T_{air} shows a comparatively steeper lapse-like gradient. In general, the direction of bias with elevation in corrected datasets has reversed—shifting from cold (negative) to warm (positive) slopes—highlighting the efficacy of elevation correction in mitigating systematic underestimation in high-elevation zones.

4. Discussion

This study demonstrates that near-surface temperature datasets—including reanalyses, observation-based gridded products (CRU), and satellite-derived LST—exhibit notable elevation-dependent biases in Bhutan's complex terrain. Uncorrected reanalysis datasets such as ERA5, ERA5-Land, and MERRA2 consistently underestimate temperatures at higher elevations. This cold bias is well-documented in global assessments of reanalyses over mountainous regions and arises from limited topographic representation, coarse spatial resolution, and sparse assimilated observations in complex orography (McNicholl et al., 2021; Thorne et al., 2016; Li et al., 2014; Gao et al., 2021; Liu et al., 2024; Yan et al., 2024; Scherrer, 2020; Arau'jo et al., 2022; Goswami et al., 2024; Wang and Zeng, 2012; Huang et al., 2023).

Although the CRU TS dataset is derived from station anomalies, it also exhibits systematic underestimation at high altitudes. This stems from its angular-distance weighting (ADW) approach, which does not incorporate elevation explicitly during interpolation (Harris et al., 2020; Jones et al., 2013). Despite relatively good agreement in spatially averaged climatology, CRU demonstrates poor station-level fidelity—a limitation highlighted in other mountainous regions as well (Liu et al., 2024; Peng et al., 2019b; Kim and Lee, 2022; Salehie et al., 2022; Xu et al., 2009).

Raw MODIS LST shows overestimation, especially in lowland areas. These discrepancies are driven by the physical difference between land surface temperature and near-surface air temperature, and are further amplified by variations in emissivity, land cover, and atmospheric effects (Wan et al., 2004; Coll et al., 2005; Duan et al., 2019). While MODIS LST maintains high accuracy over homogeneous surfaces (Wan et al., 2004; Duan et al., 2019; Coll et al., 2005; Pepin et al., 2019), its performance deteriorates significantly in regions with steep terrain or heterogeneous land surfaces (Jin et al., 2011; He et al., 2019; Firozjaei et al., 2022).

To mitigate these systematic errors, we applied three elevation correction methodologies to the reanalyses using VPLR, while a DELR approach was used to generate a station-based gridded reference. The application of VPLR corrections notably improved the performance of reanalysis datasets, especially MERRA2 under Method I. These findings corroborate other studies that emphasize the importance of incorporating lapse rate-based adjustments to reanalyses for enhancing temperature representation in data-sparse high-altitude environments (Zhao and Qian, 2025; Rozante et al., 2022; Cannon et al., 2012).

The corrected reanalyses exhibited substantial reductions in RMSE and MBE, confirming the efficacy of VPLR-based adjustments. However, residual biases remained, particularly at elevations exceeding 2000 m, where atmospheric processes, land surface interactions, and unresolved topographic effects may still induce persistent discrepancies (Zhao and Li, 2015). Although ERA5-Land has enhanced spatial resolutions, there is no significant difference from ERA5 (Clelland et al., 2024; Vanella et al., 2022), and our study found that ERA5 slightly better representing Bhutan's surface air temperature. However, MERRA2 outperformed all other datasets, both in regional and point-based evaluation demonstrating most consistent improvement post-correction, suggesting it may have comparatively better vertical structure or assimilation fidelity in this region (Gao et al., 2021; Valappil et al., 2023).

In parallel, we generated MODIS-derived surface air temperature (T_{air}) by adjusting LST using harmonic regression. This transformation yielded a product that more closely approximated *in-situ* air temperatures, particularly in mid- and high-elevation zones. MODIS T_{air} demonstrated markedly reduced biases compared to raw LST, aligning

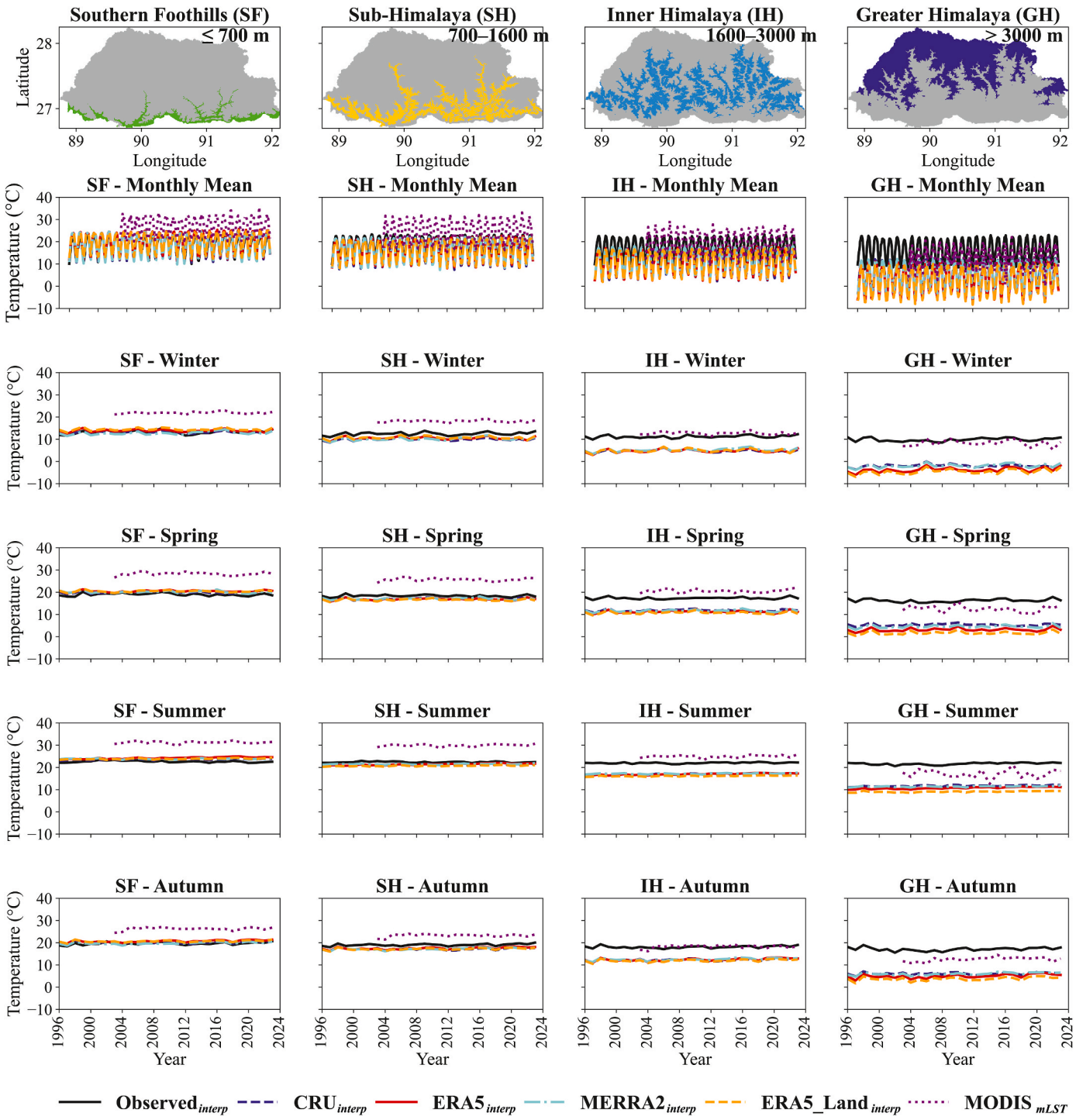

Fig. 9. The top panel represents the altitudinally stratified regional subsets, while the subsequent columns display the monthly and seasonal mean temperature time series for each region from 1996 to 2023 for all interpolated datasets, except raw MODIS LST, which spans 2003 to 2023. The results indicate that CRU and reanalyses datasets generally overestimate temperatures in low-altitude areas, whereas a pronounced cold bias is evident in high-altitude regions. In contrast, MODIS LST exhibits a moderate to severe warm bias across the SF, SH, and IH regions, and a substantial cold bias in the GH region.

with recent efforts to retrieve T_{air} from MODIS through physically consistent models and lapse-rate adjustments (Venkatraman et al., 2021; Peng et al., 2019a; Zhao et al., 2022). Although MODIS T_{air} displayed improved fidelity overall, residual errors persisted in certain valley and urbanized locations, likely due to unresolved microclimates, urban heat effects, and land cover heterogeneity (Pepin et al., 2019; Oyler et al., 2016). Nevertheless, MODIS T_{air} outperformed several reanalysis datasets in some regions, underscoring its value as a supplementary temperature product in topographically complex areas.

The DELR-based gridded station temperature dataset served as a reliable reference, exhibiting low RMSE and minimal bias across Bhutan. This method leverages DELR derived directly from station observations and has been proven effective in capturing fine-scale thermal gradients in other mountainous contexts (Nepal et al., 2015; Cannon et al., 2012; Rozante et al., 2022; Gao et al., 2012, 2017). Its superior agreement across the study domain further justifies its use for benchmarking other

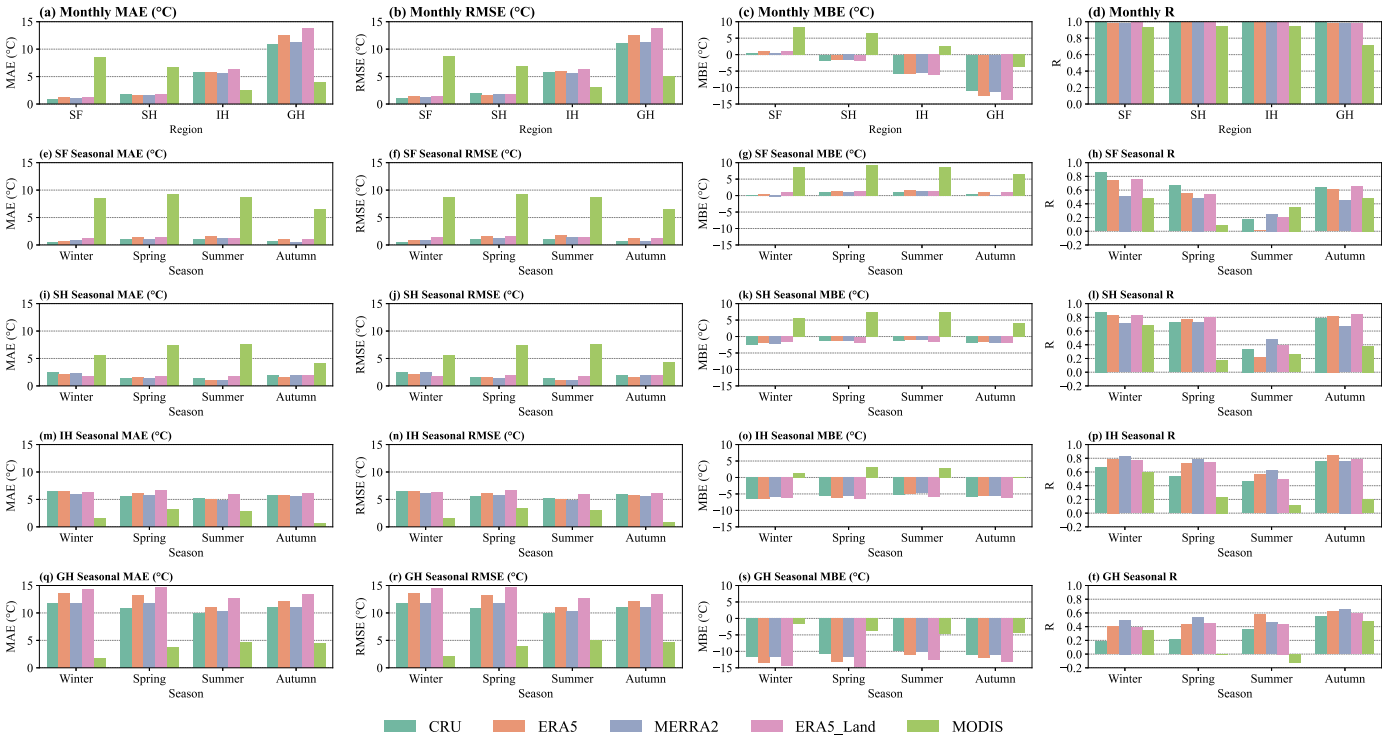

Fig. 10. The magnitude of statistical metrics, with reference to interpolated observed data, from altitudinally delineated regional subset analysis illustrated in Fig. 9. Subplots (a)–(d) represent the magnitude of MAE, RMSE, MBE and R for the monthly mean temperature across all subset regions. Subplots (e)–(h), (i)–(l), (m)–(p) and (q)–(t) show the systematic errors of all datasets in SF, SH, IH and GH regions, respectively. The elevation differences are not yet corrected for this result.

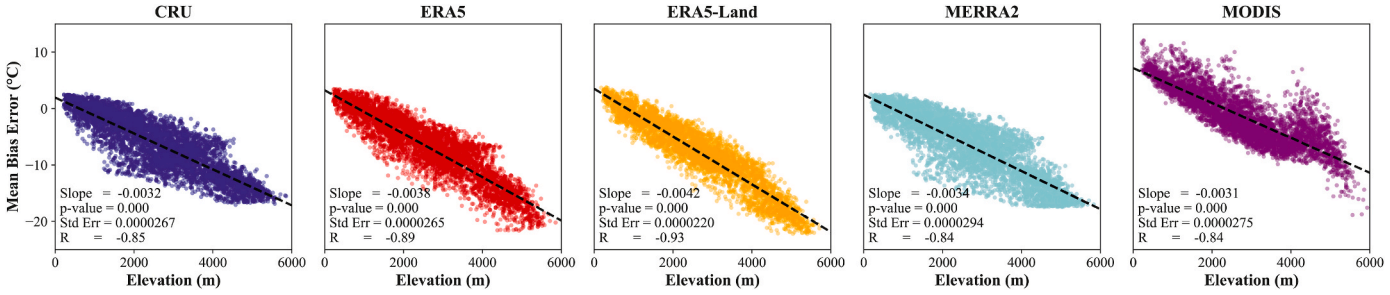

Fig. 11. The illustration of the altitude-dependent bias for interpolated CRU and reanalysis datasets using MBE as an example. It summarizes that all the datasets have systematic warm bias in the low altitude areas and cold bias in high altitude areas. The slope from linear regression in each subplot represents the overall increase or decrease of MBE with elevation. This linear relationship between bias and elevation is derived from uncorrected reanalyses and MODIS LST.

datasets.

Notably, the elevation–bias relationship across all datasets shifted markedly following correction. In uncorrected reanalyses, biases increased steeply with elevation (up to 3.1–3.9 °C/km), a pattern reported globally in reanalyses without elevation correction (Huang et al., 2021; Zhao et al., 2008). Post-correction, this lapse-rate-like bias was substantially reduced or even reversed in slope. MERRA2 showed particularly flat elevation–bias relationships, indicating enhanced spatial consistency. These results are consistent with high-resolution downscaling studies emphasizing the importance of elevation-sensitive models for improving thermal representation in complex terrain (Karger et al., 2017; Cannon et al., 2012).

In summary, while all uncorrected datasets exhibit systematic biases driven by terrain complexity, their performance improves significantly when elevation-aware corrections are applied. VPLR-corrected reanalyses, MODIS-derived T_{air}, and DELR-interpolated station data each demonstrate strengths, with MERRA2 emerging as particularly promising options post-correction. These findings reinforce the growing consensus that lapse-rate-based temperature correction is essential for climate monitoring and impact studies in mountainous regions with sparse observation networks (Nepal et al., 2015).

While this study presents the first comprehensive evaluation of reanalysis datasets, the CRU product, and satellite-based temperature estimates in the context of Bhutan's climate, Lehner and Formayer (2023) introduced BhutanClim—an innovative 1 km gridded temperature dataset for the region. In their approach, station temperatures were adjusted for elevation using a linearly regressed lapse rate, and for elevations exceeding 3500 m, vertical profile lapse rates (VPLR) from ERA5 were employed. However, our post-correction analysis indicates that ERA5 continues to exhibit a cold bias, particularly above 2000 m.a.s.l. Given that BhutanClim uses ERA5 as a reference source, this raises concerns regarding its reliability for climate impact assessments. Notably, Lehner and Formayer (2023) themselves acknowledge the presence of uncertainties in areas distant from representative observational networks, underscoring the need for more robust and spatially comprehensive validation.

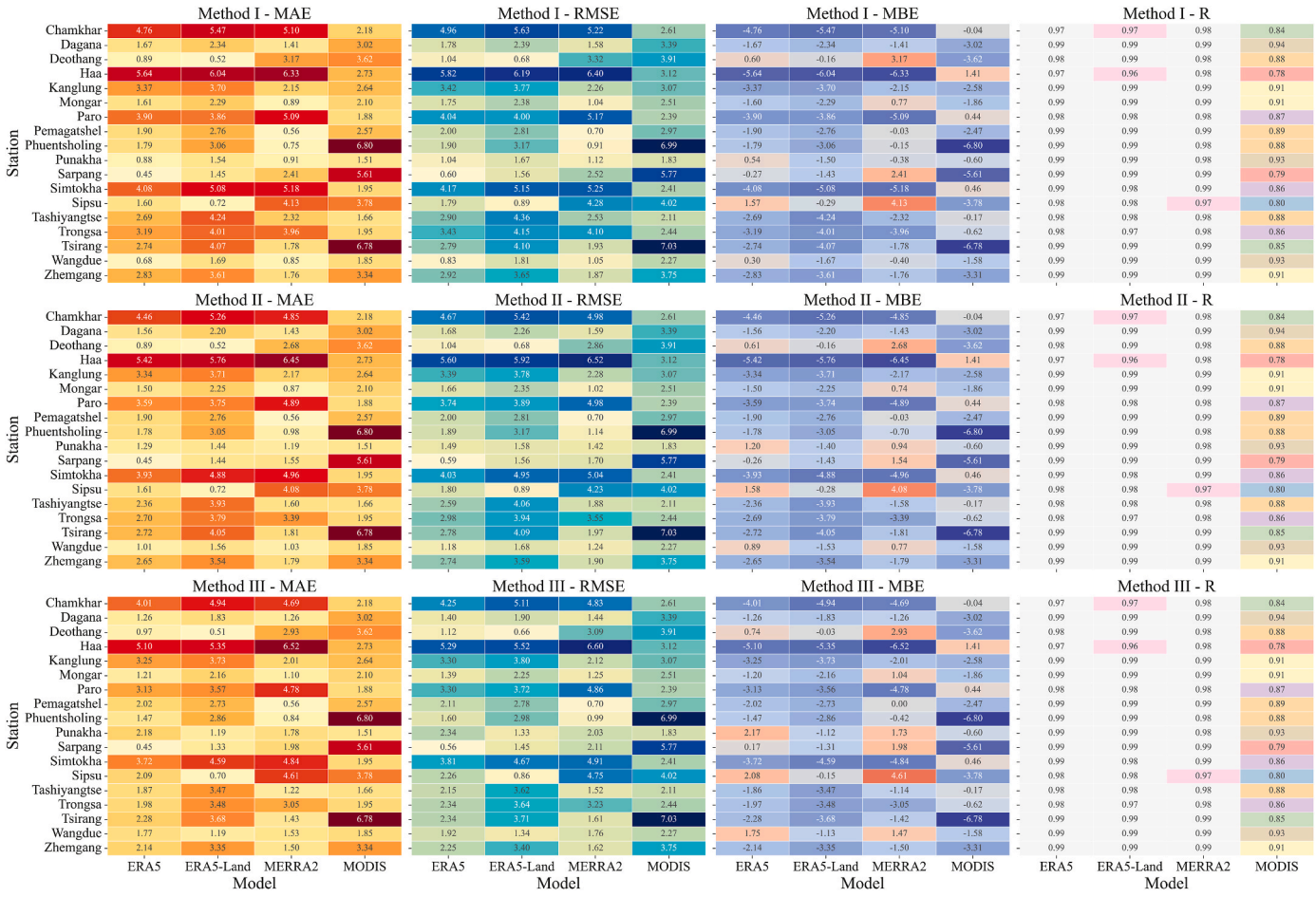

Fig. 12. Heatmap illustrating the statistical errors of the model grids with reference to the corresponding station observation. The heatmap also compares the ability to reduce the bias between station and reanalyses temperatures by three different methods, to determine specific methods for further evaluation. MODIS $_{Tair}$ remains unchanged as we use the same data across all methods.

Table 4
RMSE and MBE for different methods and reanalyses datasets. The statistical metrics are calculated from mean monthly temperature time series with reference to *in-situ* data. Due to rounding of the numbers, the values of RMSE and MBE for ERA5 and ERA5-Land appear identical but differ slightly in actual values. The RMSE and MBE derived from pre-corrected, interpolated datasets are given under Comparison 1. The results from elevation-corrected reanalyses datasets and raw station temperature is shown under Comparison 2. The results in Comparison 3 are quantified through comparison of DELR corrected station and elevation corrected reanalyses datasets.

Comparison	Method	RMSE			MBE		
		ERA5	ERA5-Land	MERRA2	ERA5	ERA5-Land	MERRA2
1	Interpolated[a]	7.554	8.262	6.962	−7.493	−8.225	−6.917
2	Method I	5.763	6.431	5.434	−5.679	−6.381	−5.375
	Method II	5.763	6.431	5.505	−5.680	−6.381	−5.446
	Method III	5.766	6.428	5.531	−5.683	−6.378	−5.473
3	Method I	1.396	1.843	1.060	−0.953	−1.652	−0.640
	Method II	1.381	1.840	1.110	−0.950	−1.651	−0.720
	Method III	1.374	1.837	1.130	−0.946	−1.647	−0.740

[a] Comparison refers to values estimated using the interpolated datasets, but prior elevation and lapse rate correction.

5. Conclusion

This study comprehensively evaluated the performance of reanalysis datasets (ERA5, ERA5-Land, and MERRA2), the CRU TS product, and MODIS-derived temperatures against *in situ* station observations across Bhutan's complex terrain. The results revealed that uncorrected reanalyses and CRU TS data systematically underestimate temperature, particularly in high-elevation regions, while MODIS LST shows consistent overestimation, especially in the lowlands. These biases, driven by coarse resolution, simplified interpolation schemes, and limited topographic representation, undermine the reliability of these datasets for regional climate applications in mountainous regions.

Applying elevation-sensitive corrections significantly enhanced dataset performance. Vertical profile lapse rate-based adjustments for reanalyses yielded notable improvements, especially for MERRA2, and reduced the magnitude and altitudinal trend of biases. Similarly, the transformation of MODIS LST to surface air temperature (T_{air}) substantially improved agreement with station data. The dynamically estimated lapse rate applied to *in-situ* station data further yielded a high-fidelity gridded observational reference for validation.

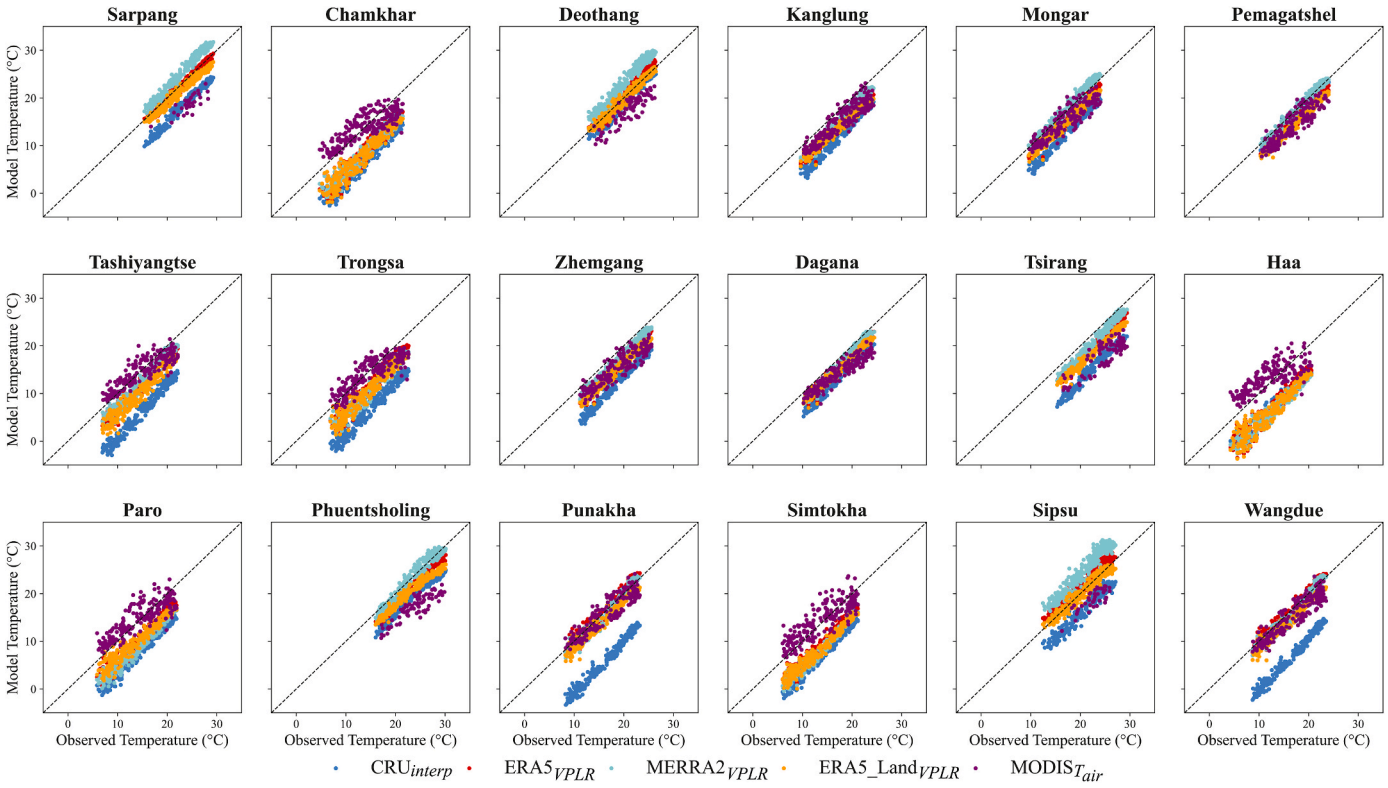

Fig. 13. The time series of representative model grids from Method I (VPLR)-based corrected reanalyses, interpolated CRU and MODIS $_{Tair}$ compared with reference to the corresponding point-based raw station temperature. The CRU's time series remains unchanged from Fig. 7, as the dataset has not been processed for any form of correction. However, CRU is brought into this comparison to visualize that elevation-aware correction in reanalyses is an essential step to reduce the bias.

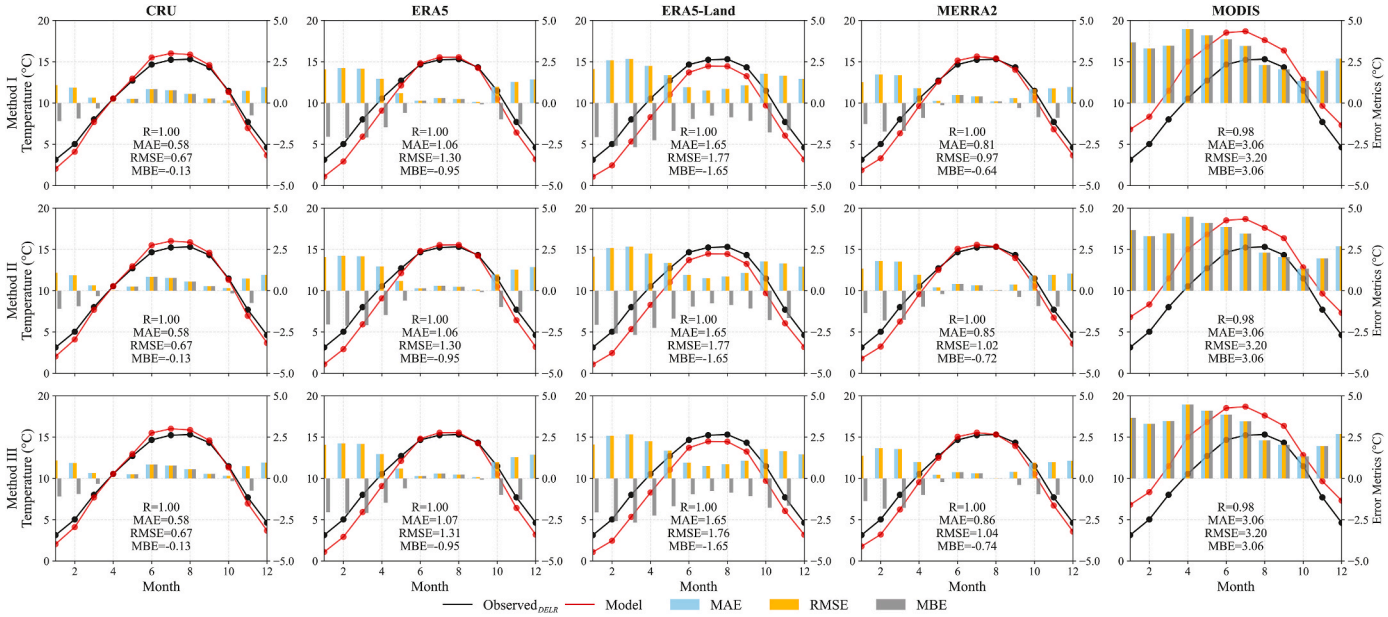

Fig. 14. Monthly climatology of interpolated CRU, Method I-corrected reanalyses (1996–2023) and MODIS Tair (2003–2023) evaluated with DELR-corrected in-situ as reference data. The comparison is with respect to RMSE, MAE, MBE and R. CRU, although not corrected for any discrepancy, closely matches the monthly climatology of observed temperature. CRU, ERA5 and MERRA2 tend to overestimate summer temperature, whereas underestimate other seasons. However, ERA5-Land and MODIS Tair consistently underestimate and overestimate the station temperature, respectively.

Post-correction analyses showed that elevation-dependent biases were either flattened or reversed, emphasizing the critical importance of topography-aware methodologies. Among the datasets, corrected MERRA2 showed the most promising results for high-resolution climate monitoring. While residual errors persist in high mountain areas, the integration of elevation corrections demonstrably enhances the spatial and temporal consistency of these datasets.

These findings underscore that in topographically complex and data-scarce regions like Bhutan, applying physically and statistically consistent elevation correction techniques is essential for accurate climate

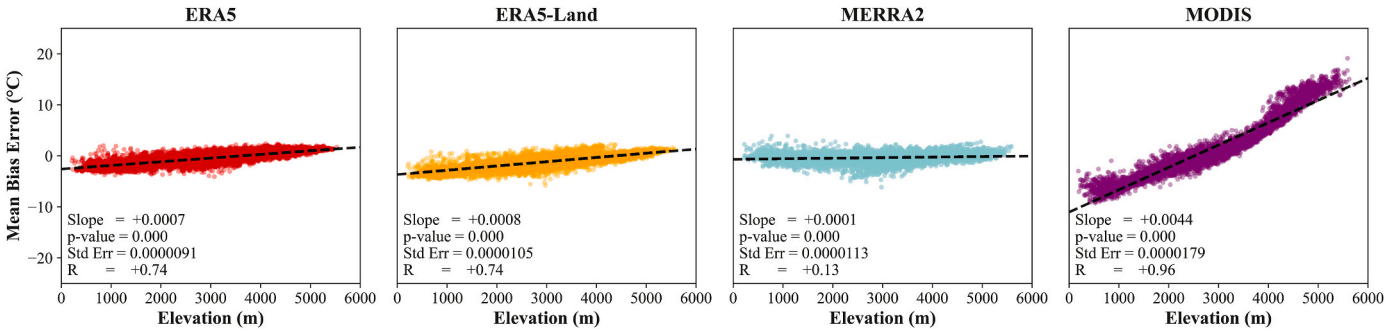

Fig. 15. MBE-elevation relationship after elevation correction of reanalyses and MODIS $_{Tair}$. The slope of linear regression signifies that ERA5-Land has higher rate of change of bias comparing within reanalyses, but MODIS Tair has poorer performance in different altitude regions in complex topography.

assessment. Such corrections improve the utility of global datasets for agricultural planning and climate impact studies, ultimately supporting more reliable policy and adaptation strategies under changing climate conditions.

CRediT authorship contribution statement

Nima Dorji: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joseph L. Awange:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Ayalsew Zerihun:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

6. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Grammarly in order to check and correct the grammatical errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

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

Acknowledgments

Nima expresses heartfelt gratitude to Curtin University for the opportunity to pursue a Doctorate of Philosophy (Ph.D.). He appreciates the Discipline of Spatial Sciences, School of Earth and Planetary Sciences, for their unwavering technical and professional support. The authors of this article also acknowledge the invaluable contributions of data providers, including the NCHM, the CRU, ECMWF, and GMAO. Special thanks are extended to AppEEARS for providing ready-to-use MODIS LST datasets in a timely manner.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.srs.2025.100275.

Data availability

Data will be made available on request.

References

Aalto, J., Pirinen, P., Heikkinen, J., Venäläinen, A., 2013. Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. Theor. Appl. Climatol. 112, 99–111. https://doi.org/10.1007/s00704-012-0716-9.

Akaike, H., 2011. Akaike's information criterion. In: International Encyclopedia of Statistical Science. Springer. https://doi.org/10.1007/978-3-642-04898-2110, 25–25.

Ali, S., Bhutta, Z.A., Reboita, M.S., Goheer, M.A., Ebrahimi, S., Rozante, J.R., Kiani, R.S., Muhammad, S., Khan, F., Rahman, M.M., et al., 2024. A 5-km gridded product development of daily temperature and precipitation for Bangladesh, Nepal, and Pakistan from 1981 to 2016. Geosci. Data J. 11, 292–302. https://doi.org/10.1002/gdj3.217.

Araújo, C.S.P.d., Silva, I.A.C.e., Ippolito, M., Almeida, C.D.G.C.d., 2022. Evaluation of air tem- perature estimated by ERA5-Land reanalysis using surface data in Pernambuco, Brazil. Environ. Monit. Assess. 194, 381. https://doi.org/10.1007/s10661-022-10047-2.

Atkinson, P.M., Lloyd, C.D., 2007. Non-stationary variogram models for geostatistical sampling optimisation: an empirical investigation using elevation data. Comput. Geosci. 33, 1285–1300. https://doi.org/10.1016/j.cageo.2007.05.011.

Begert, M., Frei, C., 2018. Long-term area-mean temperature series for Switzerland—combining homogenized station data and high resolution grid data. Int. J. Climatol. 38, 2792–2807. https://doi.org/10.1002/joc.5460.

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., et al., 2021. The ERA5 global reanalysis: preliminary extension to 1950. Q. J. R. Meteorol. Soc. 147, 4186–4227. https://doi.org/10.1002/qj.4174.

Benali, A., Carvalho, A., Nunes, J., Carvalhais, N., Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ. 124, 108–121. https://doi.org/10.1016/j.rse.2012.04.024.

Böhner, J., Lehmkuhl, F., 2005. Environmental change modelling for central and high Asia: pleistocene, present and future scenarios. Boreas 34, 220–231. https://doi.org/10.1111/j.1502-3885.2005.tb01017.x.

Brus, D.J., Heuvelink, G.B., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138, 86–95. https://doi.org/10.1016/j.geoderma.2006.10.016.

Burrough, P.A., McDonnell, R.A., 1998. Principles of Geographical Information Systems. Clarendon Press, Oxford. https://books.google.com.au/books?hl=en&lr=&id=kvoJCAAAQBAJ&oi=fnd&pg=PP1&dq=+Principals+of+Geographical+Information+Systems.&ots=aMF54lt2CD&sig=tb0972rLf9Jdd8Cl45NzyomUN4&redirescy#v=onepage&q=Principals%20of%20Geographical%20Inform.a.tion%20Systems.&f=false. (Accessed 11 August 2025).

Buytaert, W., Vuille, M., Dewulf, A., Urrutia, R., Karmalkar, A., Célleri, R., 2010. Uncertainties in climate change projections and regional downscaling in the tropical andes: implications for water resources management. Hydrol. Earth Syst. Sci. 14, 1247–1258. https://doi.org/10.5194/hess-14-1247-2010.

Cannon, A.J., Neilsen, D., Taylor, B., 2012. Lapse rate adjustments of gridded surface temperature normals in an area of complex terrain: atmospheric reanalysis versus statistical up-sampling. Atmos.-Ocean 50, 9–16. https://doi.org/10.1080/07055900.2011.649035.

Carter, A.L., Kearney, M.R., Hartley, S., Porter, W.P., Nelson, N.J., 2018. Geostatistical interpolation can reliably extend coverage of a very high-resolution model of temperature-dependent sex determination. J. Biogeogr. 45, 652–663. https://doi.org/10.1111/jbi.13152.

Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X., Hu, X., Khadka, N., 2021. Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central himalaya. Atmos. Res. 250, 105365. https://doi.org/10.1016/j.atmosres.2020.105365.

Clelland, A.A., Marshall, G.J., Baxter, R., 2024. Evaluating the performance of key ERA-Interim, ERA5 and ERA5-Land climate variables across siberia. Int. J. Climatol. https://doi.org/10.1002/joc.8456.

Coll, C., Caselles, V., Galve, J.M., Valor, E., Niclos, R., Sánchez, J.M., Rivas, R., 2005. Ground measurements for the validation of land surface temperatures derived from

AATSR and MODIS data. Remote Sens. Environ. 97, 288–300. https://doi.org/10.1016/j.rse.2005.05.007.
Cressie, N., 2015. Statistics for Spatial Data. John Wiley & Sons. https://books.google.com.au/books?hl=en&lr=&id=MzNBwAAQBAJ&oi=fnd&pg=PP1&dq=Cressie,+1993%3B&ots=NN.jC0.43N&sig=bl-CMoSmruu1G6ZGnu5RBqbF-2U&redir.esc=y#v=onepage&q=Cressie%2C%201993%3B&f=false. (Accessed 11 August 2025).
Daidzic, N.E., 2019. On atmospheric lapse rates. Intl. J. Aviat. Aeronautics Aerospace 6. https://doi.org/10.15394/ijaaa.2019.1374.
Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P., 2002. A knowledge-based approach to the statistical mapping of climate. Clim. Res. 22, 99–113. https://www.int-res.com/articles/cr2002/22/c022p099.pdf. (Accessed 5 June 2025).
Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol.: J. Royal Meteorol. Soc. 28, 2031–2064. https://doi.org/10.1002/joc.1688.
Declercq, F.A.N., 1996. Interpolation methods for scattered sample data: accuracy, spatial patterns, processing time. Cartogr. Geogr. Inf. Syst. 23, 128–144. https://doi.org/10.1559/152304096782438882.
Dorji, U., Olesen, J.E., Bøcher, P.K., Seidenkrantz, M.S., 2016. Spatial variation of temperature and precipitation in Bhutan and links to vegetation and land cover. Mt. Res. Dev. 36, 66–79. https://doi.org/10.1659/MRD-JOURNAL-D-15-00020.1.
Duan, S.B., Li, Z.L., Li, H., Göttsche, F.M., Wu, H., Zhao, W., Leng, P., Zhang, X., Coll, C., 2019. Validation of collection 6 MODIS land surface temperature product using in situ measurements. Remote Sens. Environ. 225, 16–29. https://doi.org/10.1016/j.rse.2019.02.020.
Fairbridge, R.W., Oliver, J.E., 1987. LAPSE RATE Lapse Rate. Springer, US, Boston, MA. https://doi.org/10.1007/0-387-30749-4105, 540–543.
Firozjaei, M.K., Kiavarz, M., Alavipanah, S.K., 2022. Satellite-derived land surface temperature spatial sharpening: a comprehensive review on current status and perspectives. Eur. J. Remote Sens. 55, 644–664. https://doi.org/10.1080/22797254.2022.2144764.
Forsythe, N., Blenkinsop, S., Fowler, H., 2015. Exploring objective climate classification for the Himalayan arc and adjacent regions using gridded data sources. Earth Syst. Dynam. 6, 311–326. https://doi.org/10.5194/esd-6-311-2015.
Frauenfeld, O.W., Zhang, T., Serreze, M.C., 2005. Climate change and variability using European centre for medium-range weather forecasts reanalysis (ERA-40) temperatures on the Tibetan Plateau. J. Geophys. Res. Atmos. 110. https://doi.org/10.1029/2004JD005230.
Gao, L., Bernhardt, M., Schulz, K., 2012. Elevation correction of ERA-interim temperature data in complex terrain. Hydrol. Earth Syst. Sci. 16, 4661–4673. https://doi.org/10.5194/hess-16-4661-2012.
Gao, L., Bernhardt, M., Schulz, K., Chen, X., 2017. Elevation correction of ERA-interim temperature data in the Tibetan Plateau. Int. J. Climatol. 37, 3540–3552. https://doi.org/10.1002/joc.4935.
Gao, M., Ding, Y., Zhang, S., 2021. Evaluation of ERA5, ERA-Interim, and NCEP-NCAR reanalyses with observations over complex terrain in southwestern China. Int. J. Climatol. 41, E246–E262. https://doi.org/10.1002/joc.6671.
Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., et al., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.
Ghodichore, N., Vinnarasi, R., Dhanya, C., Roy, S.B., 2018. Reliability of reanalyses products in simulating precipitation and temperature characteristics over India. J. Earth Syst. Sci. 127, 1–21. https://doi.org/10.1007/s12040-018-1024-2.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press. https://books.google.com/books?hl=en&lr=&id=CW7tHAaVR0C&oi=fnd&pg=PA3&dq=Geostatistics+for+natural+resources+evaluation&ots=zltcCU6MmD&sig=IR93hRiLuD7dZrvddxNgzqsJ-HE. (Accessed 1 June 2025).
Goswami, U.P., Déry, S.J., Fortin, V., 2024. Performance evaluation of high-resolution reanalysis datasets over north-central British Columbia. Atmos. Ocean 62, 222–242. https://doi.org/10.1080/07055900.2024.2308878.
Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci. Data 7, 109. https://doi.org/10.1038/s41597-020-0453-3.
He, J., Zhao, W., Li, A., Wen, F., Yu, D., 2019. The impact of the terrain effect on land surface temperature variation based on Landsat-8 observations in mountainous areas. Int. J. Rem. Sens. 40, 1808–1827. https://doi.org/10.1080/01431161.2018.1466082.
Hereher, M.E., 2019. Estimation of monthly surface air temperatures from MODIS LST time series data: application to the deserts in the Sultanate of Oman. Environ. Monit. Assess. 191, 592. https://doi.org/10.1007/s10661-019-7771-y.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.
Hoeting, J.A., Davis, R.A., Merton, A.A., Thompson, S.E., 2006. Model selection for geostatistical models. Ecol. Appl. 16, 87–98. https://doi.org/10.1890/04-0576.
Hofstra, N., Haylock, M., New, M., Jones, P., Frei, C., 2008. Comparison of six methods for the interpolation of daily, european climate data. J. Geophys. Res. Atmos. 113. https://doi.org/10.1029/2008JD010100.
Holdaway, M.R., 1996. Spatial modeling and interpolation of monthly temperature using kriging. Clim. Res. 6, 215–225. https://doi.org/10.3354/cr006215.
Hoy, A., Katel, O., Thapa, P., Dendup, N., Matschullat, J., 2016. Climatic changes and their impact on socio-economic sectors in the Bhutan Himalayas: an implementation strategy. Reg. Environ. Change 16, 1401–1415. https://doi.org/10.1007/s10113-015-0868-0.
Hu, Z., Zhang, C., Hu, Q., Tian, H., 2014. Temperature changes in central Asia from 1979 to 2011 based on multiple datasets. J. Clim. 27, 1143–1167. https://doi.org/10.1175/JCLI-D-13-00064.1.
Huang, L., Fang, X., Zhang, T., Wang, H., Cui, L., Liu, L., 2023. Evaluation of surface temperature and pressure derived from MERRA-2 and ERA5 reanalysis datasets and their applications in hourly GNSS precipitable water vapor retrieval over China. Geodesy Geodynamics 14, 111–120. https://doi.org/10.1016/j.geog.2022.08.006.
Huang, X., Han, S., Shi, C., 2021. Multiscale assessments of three reanalysis temperature data systems over China. Agriculture 11, 1292. https://doi.org/10.3390/rs14184447.
Huang, X., Han, S., Shi, C., 2022. Evaluation of three air temperature reanalysis datasets in the alpine region of the Qinghai–Tibet Plateau. Remote Sens. 14. https://doi.org/10.3390/rs14184447.
Ibrahim, A.M., Nasser, R.H.A., 2017. Comparison between inverse distance weighted (IDW) and kriging. Intl. Sci. Res. 6, 249–254. https://doi.org/10.21275/ART20177562.
IPCC, 2018. Summary for Policymakers. Cambridge University Press, pp. 1–24. https://doi.org/10.1017/9781009157940.001.
IPCC, 2023. IPCC, 2023: climate change 2023: synthesis report. In: Lee, H., Romero, J. (Eds.), Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team. IPCC, Geneva, Switzerland. https://doi.org/10.59327/IPCC/AR6-9789291691647.
Isaaks, E., Srivastava, R., 1989. An Introduction to Applied Geostatistics, vol. 561. Oxford University Press. https://doi.org/10.1016/0098-3004(91)90055-I.
Jin, M.S., Kessomkiat, W., Pereira, G., 2011. Satellite-observed urbanization characters in Shanghai, China: aerosols, urban heat island effect, and land–atmosphere interactions. Remote Sens. 3, 83–99. https://doi.org/10.3390/rs3010083.
Jones, P., Harris, I., 2013. Cru TS3. 21: Climatic research unit (CRU) time-series (TS) version 3.21 of high resolution gridded data of month-by-month variation in climate (Jan. 1901– Dec. 2012). NCAS British Atmospheric Data Centre 34, 623–642. https://doi.org/10.5285/4c7fdfa6-f176-4c58-acee-683d5e9d2ed5.
Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the Earth's land surface areas. Sci. Data 4, 1–20. https://www.nature.com/articles/sdata2017122. (Accessed 12 June 2025).
Khandu, J., Awange, L., Kuhn, M., Anyah, R., Forootan, E., 2017. Changes and variability of precipitation and temperature in the ganges–brahmaputra–meghna river Basin based on global high-resolution reanalyses. Int. J. Climatol. 37, 2141–2159. https://doi.org/10.1002/joc.4842.
Khandu, Y., Polthanee, A., Isarangkool Na Ayutthaya, S., 2022. Dendroclimatic reconstruction of mean annual temperatures over treeline regions of northern Bhutan Himalayas. Forests 13, 1794. https://doi.org/10.3390/f13111794.
Kim, M., Lee, E., 2022. Validation and comparison of climate reanalysis data in the East Asian monsoon region. Atmosphere 13, 1589. https://doi.org/10.3390/atmos13101589.
Koike, K., Matsuda, S., Gu, B., 2001. Evaluation of interpolation accuracy of neural kriging with application to temperature-distribution analysis. Math. Geol. 33, 421–448. https://doi.org/10.1023/A:1011084812324.
Krusic, P.J., Cook, E., Dukpa, D., Putnam, A., Rupper, S., Schaefer, J., 2015. Six hundred thirty-eight years of summer temperature variability over the Bhutanese Himalaya. Geophys. Res. Lett. 42, 2988–2994. https://doi.org/10.1002/2015GL063566.
Kuensel, 2021. Climate change impacts make cardamom farming a less predictable livelihood. Kuensel Online. https://kuenselonline.com/climate-change-impacts-make-cardamom-farming-a-less-predictable-livelihood. (Accessed 12 August 2024).
Lehner, F., Formayer, H., 2023. Insights on the climate of Bhutan from a new daily 1 km gridded data set for temperature and precipitation. Int. J. Climatol. 43, 4927–4943. https://doi.org/10.1002/joc.8125.
Li, B., Liang, S., Ma, H., Liu, X., He, T., Zhang, Y., 2024. Generation of global 1 km all-weather instantaneous and daily mean land surface temperature from MODIS data. Earth Syst. Sci. Data Discuss. 2024, 1–45. https://doi.org/10.5194/essd-16-3795-2024.
Li, H., Sun, D., Yu, Y., Wang, H., Liu, Y., Liu, Q., Du, Y., Wang, H., Cao, B., 2014. Evaluation of the VIIRS and MODIS LST products in an arid area of Northwest China. Remote Sens. Environ. 142, 111–121. https://doi.org/10.1016/j.rse.2013.11.014.
Li, J., Heap, A.D., 2008. A review of spatial interpolation methods for environmental scientists. https://www.researchgate.net/profile/Jin-Li-74/publication/246546630A.Review.of.Spatial.Interpolation.Methods.for.Environmental.Scientists/links/56f9ccb408ae95e8b6d40461/A-Review-of-Spatial-Interpolation-Methods-for-Environmental-Scientists.pdf. (Accessed 15 June 2025).
Li, J., Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecol. Inform. 6, 228–241. https://doi.org/10.1016/j.ecoinf.2010.12.003.
Li, Y., Qin, X., Liu, Y., Jin, Z., Liu, J., Wang, L., Chen, J., 2022. Evaluation of long-term and high- resolution gridded precipitation and temperature products in the Qilian Mountains, Qinghai– Tibet Plateau. Front. Environ. Sci. 10, 906821. https://doi.org/10.3389/fenvs.2022.906821.
Liu, R., Zhang, X., Wang, W., Wang, Y., Liu, H., Ma, M., Tang, G., 2024. Global-scale ERA5 product precipitation and temperature evaluation. Ecol. Indic. 166, 112481. https://doi.org/10.1016/j.ecolind.2024.112481.
McNicholl, B., Lee, Y.H., Campbell, A.G., Dev, S., 2021. Evaluating the reliability of air temperature from ERA5 reanalysis data. IEEE Geosci. Remote Sensing Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3137643.

Mitas, L., Mitasova, H., 1999. Spatial interpolation. Geographical information systems: principles, techniques. Manag. Appl. 1, 481–492. http://fatra.cnr.ncsu.edu/~hmitaso/gmslab/papers/mitas.mitasova.1999-2005.pdf. (Accessed 28 May 2025).
Moyeed, R.A., Papritz, A., 2002. An empirical comparison of kriging methods for nonlinear spatial point prediction. Math. Geol. 34, 365–386. https://doi.org/10.1023/A:1015085810154.
Muñoz-Sabater, J., Dutra, E., Agusti-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Bous- setta, S., Choulga, M., Harrigan, S., Hersbach, H., et al., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.
Muñoz Sabater, J., 2019. ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.68d2bb30.
NASA, 2013. NASA Shuttle Radar Topography Mission: Global 1 Arc Second. https://doi.org/10.5067/measures/srtm/srtmgl1.003.
NCHM, 2018. Royal Government of Bhutan Climate Data Book of Bhutan. (Accessed 11 November 2024).
NCHM, 2019. Analysis of Historical Climate and Climate Projection for Bhutan. NCHM. https://www.nchm.gov.bt/attachment/ckfinder/userfiles/files/Analysis%20of%20Historical%20Climate%20and%20Climate%20Change%20Projection.pdf. (Accessed 13 December 2024).
NCHM, 2023. State of the climate 2022 national centre for hydrology and meteorology royal government of Bhutan 2023. https://www.nchm.gov.bt/attachment/ckfinder/userfiles/files/State%20of%20Climate(1).pdf. (Accessed 12 December 2024).
Nelli, N., Francis, D., Alkatheeri, A., Fonseca, R., 2024. Evaluation of reanalysis and satellite products against ground-based observations in a desert environment. Remote Sens. 16. https://doi.org/10.3390/rs16193593.
Nepal, S., Shrestha, A.B., Krakauer, N.Y., Wahid, S.M., Pradhananga, D., Bajracharya, S. R., 2015. Evaluation of uncertainties in gridded precipitation and temperature in the Hindukush-Karakoram-Himalaya region. Environ. Res. Lett. 10, 065006. https://doi.org/10.1088/17489326/10/6/065006.
Nguyen, X.T., Nguyen, B.T., Do, K.P., Bui, Q.H., Nguyen, T.N.T., Vuong, V.Q., Le, T.H., 2015. Spatial interpolation of meteorological variables in Vietnam using the Kriging method. J. Inform. Proc. Syst. 11, 134–147. https://doi.org/10.3745/JIPS.02.0016.
Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. Int. J. Geogr. Inf. Syst. 4 (3), 313–332. https://doi.org/10.1080/02693799008941549.
Oyler, J.W., Dobrowski, S.Z., Holden, Z.A., Running, S.W., 2016. Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. J. Appl. Meteorol. Climatol. 55, 1441–1457. https://doi.org/10.1175/JAMC-D-15-0276.1.
Park, J., Jang, D.H., 2016. Application of MK-PRISM for interpolation of wind speed and comparison with co-kriging in South Korea. GIScience Remote Sens. 53, 421–443. https://doi.org/10.1080/15481603.2016.1192373.
Peng, J., Dong, J., Wang, J., 2019a. Modelling near-surface air temperature from MODIS LST in a mountainous region of northwest China. Theor. Appl. Climatol. 138, 1717–1730. https://doi.org/10.1007/s00704-019-02900-5.
Peng, S., Ding, Y., Liu, W., Li, Z., 2019b. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. Earth Syst. Sci. Data 11, 1931–1946. https://doi.org/10.5194/essd-11-1931-2019.
Pepin, N., Bradley, R., Diaz, H., Baraer, M., Caceres, E., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M., Liu, X., et al., 2015. Elevation-dependent warming in mountain regions of the world. Nat. Clim. Change 5 (5), 424–430. https://doi.org/10.1038/nclimate2563.
Pepin, N., Deng, H., Zhang, H., Zhang, F., Kang, S., Yao, T., 2019. An examination of temperature trends at high elevations across the Tibetan plateau: the use of MODIS LST to understand patterns of elevation-dependent warming. J. Geophys. Res. Atmos. 124, 5738–5756. https://doi.org/10.1029/2018JD029798.
Phan, T.N., Kappas, M., 2018. Application of MODIS land surface temperature data: a systematic literature review and analysis. J. Appl. Remote Sens. 12. https://doi.org/10.5555/20193441854, 041501–041501.
Phillips, D.L., Marks, D.G., 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. Ecol. Model. 91, 213–229. https://doi.org/10.1016/0304-3800(95)00191-3.
Portet, S., 2020. A primer on model selection using the Akaike information criterion. Infect. Dis. Model. 5, 111–128. https://doi.org/10.1016/j.idm.2019.12.010.
Qin, Y., Zhang, P., Liu, W., Guo, Z., Xue, S., 2020. The application of elevation corrected MERRA2 reanalysis ground surface temperature in a permafrost model on the Qinghai-Tibet Plateau. Cold Reg. Sci. Technol. 175, 103067. https://doi.org/10.1016/j.coldregions.2020.103067.
Rana, S., McGregor, J., Renwick, J., 2015. Precipitation seasonality over the Indian subcontinent: an evaluation of gauge, reanalyses, and satellite retrievals. J. Hydrometeorol. 16, 631–651. https://doi.org/10.1175/JHM-D-14-0106.1.
Randles, C., Da Silva, A., Buchard, V., Colarco, P., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., et al., 2017. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: system description and data assimilation evaluation. J. Clim. 30, 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1.
Rapaić, M., Brown, R., Markovic, M., Chaumont, D., 2015. An evaluation of temperature and precipitation surface-based and reanalysis datasets for the Canadian Arctic, 1950–2010. Atmos.-Ocean 53, 283–303. https://doi.org/10.1080/07055900.2015.1045825.
Recondo, C., Corbea-Pérez, A., Peón, J., Pendás, E., Ramos, M., Calleja, J.F., de Pablo, M.Á., Fernández, S., Corrales, J.A., 2022. Empirical models for estimating air temperature using MODIS land surface temperature (and spatiotemporal variables) in the hurd peninsula of livingston Island. Antarctica, between 2000 and 2016. Remote Sensing 14, 3206. https://doi.org/10.3390/rs14133206.
Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K., et al., 2011. MERRA: nasa's modern-era retrospective analysis for research and applications. J. Clim. 24, 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1.
Rozante, J.R., Ramirez, E., Fernandes, A.d.A., 2022. A newly developed South American mapping of temperature with estimated lapse rate corrections. Int. J. Climatol. 42, 2135–2152. https://doi.org/10.1002/joc.7356.
Sajid, A., Rudra, R., Parkin, G., 2013. Systematic evaluation of kriging and inverse distance weighting methods for spatial analysis of soil bulk density. Canad. Biosyst. Eng. J. 55. https://doi.org/10.7451/CBE.2013.55.1.1.
Salehie, O., Ismail, T.b., Shahid, S., Sammen, S.S., Malik, A., Wang, X., 2022. Selection of the gridded temperature dataset for assessment of thermal bioclimatic environmental changes in the Amu Darya river basin. Stoch. Environ. Res. Risk Assess. 36, 2919–2939. https://doi.org/10.1007/s00477-022-02172-8.
Salomonson, V.V., Barnes, W., Masuoka, E.J., 2006. Introduction to MODIS and an overview of associated activities. Earth Sci. Satel. Remote Sens. 1, 12–32. https://doi.org/10.1007/978-3-540-37293-62. Science and Instruments.
Scherrer, S.C., 2020. Temperature monitoring in mountain regions using reanalyses: lessons from the Alps. Environ. Res. Lett. 15, 044005. https://doi.org/10.1088/1748-9326/ab702d.
Screen, J.A., Simmonds, I., 2011. Erroneous Arctic temperature trends in the ERA-40 reanalysis: a closer look. J. Clim. 24, 2620–2627. https://doi.org/10.1175/2010JCLI4054.1.
Shen, B., Song, S., Zhang, L., Wang, Z., Ren, C., Li, Y., 2022. Temperature trends in some major countries from the 1980s to 2019. J. Geogr. Sci. 32, 79–100. https://doi.org/10.1007/s11442-022-1937-1.
Shen, F., Xu, C., Hu, M., 2023. Comparison of approaches to spatiotemporally interpolate land surface air temperature for the Qinghai–Tibet Plateau. Environ. Earth Sci. 82, 452. https://doi.org/10.1007/s12665-023-11151-3.
Shen, S., Leptoukh, G.G., 2011. Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. Environ. Res. Lett. 6, 045206. https://doi.org/10.1088/1748-9326/6/4/045206.
Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM National Conference, pp. 517–524. https://doi.org/10.1145/800186.810616.
Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., et al., 2024. The ERA5 global reanalysis from 1940 to 2022. Q. J. R. Meteorol. Soc. https://doi.org/10.1002/qj.4803.
Song, C., Ke, L., Richards, K.S., Cui, Y., 2016. Homogenization of surface temperature data in High Mountain Asia through comparison of reanalysis data and station observations. Int. J. Climatol. 36. https://doi.org/10.1002/joc.4403.
Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.L., 2018. A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev. Geophys. 56, 79–107. https://doi.org/10.1002/2017RG000574.
Sun, Y., Li, B., Genton, M.G., 2011. Geostatistics for large datasets. In: Advances and Challenges in space-time Modelling of Natural Events. Springer, pp. 55–77. https://doi.org/10.1007/978-3-642-17086-73.
Sun, Y., Yang, F., Liu, M., Li, Z., Gong, X., Wang, Y., 2023. Evaluation of the weighted mean temperature over China using multiple reanalysis data and radiosonde. Atmos. Res. 285, 106664. https://doi.org/10.1016/j.atmosres.2023.106664.
Tesfaye, T.W., Dhanya, C., Gosain, A., 2017. Spatiotemporal evaluation of reanalysis and *in-situ* surface air temperature over Ethiopia. In: 2017 Fall Meeting. https://www.ewra.net/ew/pdf/EW.2017.59.29.pdf. (Accessed 24 December 2024).
The Bhutanese, 2018. Orange trees climb higher in Bhutan with climate change – the Bhutanese. https://thebhutanese.bt/orange-trees-climb-higher-in-bhutan-with-climate-change. (Accessed 26 August 2024).
Thorne, P., Donat, M., Dunn, R., Williams, C., Alexander, L., Caesar, J., Durre, I., Harris, I., Hausfather, Z., Jones, P., et al., 2016. Reassessing changes in diurnal temperature range: Intercomparison and evaluation of existing global data set estimates. J. Geophys. Res. Atmos. 121, 5138–5158. https://doi.org/10.1002/2015JD024584.
Valappil, N.K.M., Hamza, V., de Oliveira Júnior, J.F., 2023. Evaluation of daily average temperature trends in Kerala, India, using MERRA-2 reanalysis data: a climate change perspective. Environ. Sci. Pollut. Control Ser. 30, 26663–26686. https://doi.org/10.1007/s11356-022-23895-9.
Vanella, D., Longo-Minnolo, G., Belfiore, O.R., Ramírez-Cuesta, J.M., Pappalardo, S., Consoli, S., D'Urso, G., Chirico, G.B., Coppola, A., Comegna, A., et al., 2022. Comparing the use of ERA5 reanalysis dataset and ground-based agrometeorological data under different climates and topography in Italy. J. Hydrol.: Reg. Stud. 42, 101182. https://doi.org/10.1016/j.ejrh.2022.101182.
Venkatraman, V., Lal, R., Yadav, P., 2021. Estimating near-surface air temperature from MODIS land surface temperature and reanalysis data in data-scarce regions. Remote Sens. Appl.: Soc. Environ. 22, 100471. https://doi.org/10.1016/j.rsase.2021.100471.
Wahba, G., 1990. Spline Models for Observational Data. SIAM. https://doi.org/10.1137/1.9781611970128.bm.
Wan, Z., 2008. New refinements and validation of the MODIS land-surface temperature/emissivity products. Remote Sens. Environ. 112, 59–74. https://doi.org/10.1016/j.rse.2006.06.026.
Wan, Z., 2013. Collection-6 MODIS Land Surface Temperature Products, Users' Guide, ERI. University of California, Santa Barbara. (Accessed 7 August 2025).
Wan, Z., 2014. New refinements and validation of the MODIS land-surface temperature/emissivity products. Remote Sens. Environ. 140, 36–45. https://doi.org/10.1016/j.rse.2006.06.026.

Wan, Z., Hook, S., Hulley, G., 2015a. MOD11C1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 0.05 Deg CMG V006. https://doi.org/10.5067/MODIS/MOD11A1.006.

Wan, Z., Hook, S., Hulley, G., 2015b. MYD11A1 MODIS/aqua land surface temperature/emissivity daily L3 global 1km SIN grid V006. NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC) Data Set, MYD11A1–006doi. https://doi.org/10.5067/MODIS/MYD11A1.061.

Wan, Z., Hook, S., Hulley, G., 2021. MODIS/terra land surface temperature/emissivity daily L3 global 1km SIN grid V061. NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC) Data Set , MOD11A1–061doi. https://doi.org/10.5067/MODIS/MOD11A1.061.

Wan, Z., Zhang, Y., Zhang, Q., Li, Z.L., 2004. Quality assessment and validation of the MODIS global land surface temperature. Int. J. Rem. Sens. 25, 261–274. https://doi.org/10.1080/0143116031000116417.

Wang, A., Zeng, X., 2012. Evaluation of multireanalysis products with *in-situ* observations over the Tibetan Plateau. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2011JD016553.

Wang, K., Li, Z., Cribb, M., 2006. Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: a new method to determine the priestley–Taylor parameter. Remote Sens. Environ. 102, 293–305. https://doi.org/10.1016/j.rse.2006.02.007.

Wang, Y., Sun, W., Huai, B., Wang, Y., Ji, K., Yang, X., Du, W., Qin, X., Wang, L., 2024. Comparison and evaluation of the performance of reanalysis datasets for compound extreme temperature and precipitation events in the qilian Mountains. Atmos. Res. 304, 107375. https://doi.org/10.1016/j.atmosres.2024.107375.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons. https://books.google.com.au/books?hl=en&lr=&id=WBwSyvIvNY8C&oi=fnd&pg=PR5&dq=Webster+%26+Oliver,+2007&ots=CEMqSSkR26&sig=JCaWOBrvPIIt3TIKuzOmkBBf.pls&redir.esc=y. (Accessed 9 August 2025).

White, D., Saunders, P., 2000. The propagation of uncertainty on interpolated scales, with examples from thermometry. Metrologia 37, 285. https://doi.org/10.1088/0026-1394/37/4/4.

Xing, Z., Li, Z.L., Duan, S.B., Liu, X., Zheng, X., Leng, P., Gao, M., Zhang, X., Shang, G., 2021. Estimation of daily mean land surface temperature at global scale using pairs of daytime and nighttime MODIS instantaneous observations. ISPRS J. Photogrammetry Remote Sens. 178, 51–67. https://doi.org/10.1016/j.isprsjprs.2021.05.017.

Xu, Y., Gao, X., Shen, Y., Xu, C., Shi, Y., Giorgi, a., 2009. A daily temperature dataset over China and its application in validating an RCM simulation. Adv. Atmos. Sci. 26, 763–772. https://doi.org/10.1007/s00376-009-9029-z.

Yan, H., Zhang, J., Hou, Y., He, Y., 2009. Estimation of air temperature from MODIS data in east China. Int. J. Rem. Sens. 30, 6261–6275. https://doi.org/10.1080/01431160902842375.

Yan, X., Zhang, M., Yin, F., You, J., Chen, Y., Gao, L., 2024. Multi-Scale evaluation of ERA5 air temperature and precipitation data over the Poyang Lake Basin of China. Water 16, 3123. https://doi.org/10.3390/w16213123.

Yan, Y., You, Q., Wu, F., Pepin, N., Kang, S., 2020. Surface mean temperature from the observational stations and multiple reanalyses over the Tibetan Plateau. Clim. Dyn. 55, 2405–2419. https://doi.org/10.1007/s00382-020-05386-0.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nat. Clim. Change 3, 875–883. https://doi.org/10.1038/nclimate1908.

Yang, Y., Li, Q., Song, Z., Sun, W., Dong, W., 2022. A comparison of global surface temperature variability, extremes and warming trend using reanalysis datasets and CMST-interim. Int. J. Climatol. 42, 5609–5628. https://doi.org/10.1002/joc.7551.

Zhang, P., Bounoua, L., Imhoff, M.L., Wolfe, R.E., Thome, K., 2014. Comparison of MODIS land surface temperature and air temperature over the continental USA meteorological stations. Can. J. Rem. Sens. 40, 110–122. https://doi.org/10.1080/07038992.2014.935934.

Zhao, F., Zhu, X., Wang, T., 2022. Improving air temperature estimation in complex terrain by integrating MODIS LST and a machine learning approach. Remote Sens. 14, 303. https://doi.org/10.3390/rs14020303.

Zhao, P., Qian, L., 2025. Elevation correction of ERA5 reanalysis temperature over the Qilian Mountains of China. Atmosphere 16, 324. https://doi.org/10.3390/atmos16030324.

Zhao, T., Guo, W., Fu, C., 2008. Calibrating and evaluating reanalysis surface temperature error by topographic correction. J. Clim. 21, 1440–1446. https://doi.org/10.1175/2007JCLI1463.1.

Zhao, W., Li, A., 2015. A review on land surface processes modelling over complex terrain. Adv. Meteorol. 2015, 607181. https://doi.org/10.1155/2015/607181.

Zimmerman, D., Pavlik, C., Ruggles, A., Armstrong, M.P., 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. Math. Geol. 31, 375–390. https://doi.org/10.1023/A:1007586507433.