Atmospheric Research 336 (2026) 108779

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

The June 2024 middle east compound heatwave: Dynamical drivers and AI-weather forecast models' evaluation

Narendra Nelli [a], Diana Francis [a,*], Ricardo Fonseca [a], Pedram Hassanzadeh [b,c], Charfeddine Cherif [a]

[a] Environmental and Geophysical Sciences (ENGEOS) Lab, Earth Sciences Department, Khalifa University, Abu Dhabi 127788, United Arab Emirates
[b] University of Chicago, Department of the Geophysical Sciences, Chicago, IL 60637, USA
[c] University of Chicago, Committee on Computational and Applied Mathematics, Chicago, IL 60637, USA

ARTICLE INFO

Keywords:
Middle East heatwaves
Compound day–night extremes
Arabian heat low
GraphCast
PanguWeather
FourCastNetv2
Aurora

ABSTRACT

Heatwaves are intensifying across the Middle East with distinct daytime and nighttime impacts. We quantify the climatology and trends of daytime, nighttime, and compound heatwaves using station observations for 2005 to 2025, and evaluate four AI-weather models for near-surface air temperature during a recent high-impact episode.

Daytime heatwaves recur from the eastern Mediterranean through northern Iraq and Iran into Anatolia during warm seasons, whereas nighttime and compound events cluster along the southern Red Sea and Arabian Gulf coasts. Trends show significant summer and autumn increases across all classes (daytime: + 4.09 and + 6.09; nighttime: + 4.13 and + 6.22; compound: + 3.42 and + 2.40 per year), a winter increase led by nighttime events, and a spring increase in both nighttime and compound events, consistent with asymmetric diurnal warming, strong land–atmosphere coupling over arid interiors, and coastal humidity that limits nocturnal cooling.

A case study of 10 to 27 June 2024 documents record-breaking daytime anomalies over the northern Middle East and persistent nocturnal warmth along maritime margins, with peak impacts around 17 June in the Mecca region. ERA-5 diagnostics indicate a quasi-stationary Rossby wavetrain, an anomalous ridge over southwest Asia, and an intensified Arabian Heat Low that weakened low-level ventilation and sustained humid coastal nights.

We assess GraphCast, PanguWeather, FourCastNetv2, and Aurora using six-hourly forecasts initialized at 00 UTC on 12 June 2024 and verified against ERA-5 from 12 to 22 June. All capture synoptic timing and multi-day persistence but underestimate daytime peaks and show lead-dependent cold biases. PanguWeather provides the strongest deterministic temperature guidance; GraphCast corroborates synoptic evolution. Operationally, a bias-corrected PanguWeather and GraphCast blend is recommended.

1. Introduction

Heatwaves are among the most rapidly intensifying weather-related hazards in a warming climate, with well-documented increases in frequency, intensity, duration and escalating risks to health, energy, water, and infrastructure (Perkins-Kirkpatrick and Lewis, 2020). The Middle East is a recognized heatwave hotspot, warming faster than the global means and exhibiting pronounced summertime amplification of extremes (Lelieveld et al., 2016; Zittis et al., 2022; Syed et al., 2024). Regionally, subsidence associated with the subtropical anticyclones, pervasive aridity, and strong land–sea thermal contrasts, amplified by rapid urbanization, create chronically high background heat stress, which in turn compounds the risks during multi-day heatwave events. Beyond hot daytime interiors, the southern Arabian Gulf and Red Sea coastal regions frequently experience evening and nighttime humid-heat maxima that curtail physiological recovery and elevate the risk of heatstroke (Raymond et al., 2024). These attributes underscore the need to quantify heatwave occurrence and to resolve the processes that generate heat extremes. Projections further indicate outsized future increases in the intensity, duration, and persistence of Middle East heatwaves, reinforcing the urgency of robust climatology and mechanism-aware prediction for the region (Lelieveld et al., 2016; Zittis et al., 2021; Zittis et al., 2022). Recent reviews and attribution studies across the Northern Hemisphere also point to the combined roles of persistent anticyclones, quasi-stationary Rossby waves, and land-surface feedbacks in setting the severity and persistence of heatwaves, with a growing

* Corresponding author.
E-mail address: diana.francis@ku.ac.ae (D. Francis).

https://doi.org/10.1016/j.atmosres.2026.108779
Received 25 September 2025; Received in revised form 8 January 2026; Accepted 15 January 2026
Available online 22 January 2026

anthropogenic contribution (Xu et al., 2021; Dong et al., 2021; Perkins-Kirkpatrick et al., 2024; Gong et al., 2024; Ma et al., 2024).

Beyond broad trends, a process-based regional framing is essential. Over the Arabian Peninsula, the summertime thermal low, Arabian Heat Low (AHL), is a first-order control on near-surface temperatures and interacts with monsoon-related circulations (Fonseca et al., 2022b). The AHL's position and intensity modulate pressure gradients and therefore the northwesterly Shamal winds (Yu et al., 2016), sea- and land-breeze interactions, and moisture transport (Steinhoff et al., 2018; Fonseca et al., 2022b; Fonseca et al., 2025a, 2025b). Heatwaves in Iraq and the northern Arabian Peninsula frequently coincide with quasi-stationary anticyclonic ridging embedded in Rossby wave trains, a configuration that weakens the Shamal and promotes adiabatic warming and foehn-like downslope flows in the lee of the Zagros mountains (Al-Shamarti et al., 2025). In addition, strong land–atmosphere coupling and dust–radiation interactions can amplify surface heating and boundary-layer stratification, sustaining extremes across the Arabian Peninsula and adjacent basins (Parajuli et al., 2023; Francis et al., 2022). This picture is consistent with broader hemispheric evidence that soil-moisture depletion reduces evaporative cooling and, together with blocking highs, enhances near-surface heating and limits nighttime relief (Xu et al., 2021; Perkins-Kirkpatrick et al., 2024). These circulation regimes and surface feedback, interacting with sharp coast–interior gradients and rapidly urbanizing corridors, shape the diurnal structure of heat extremes across the region.

Recent global analyses distinguish independent daytime, independent nighttime, and compound day–night heatwaves, each with distinct physical controls and impacts (Wu et al., 2023). Daytime events are typically linked to clear skies, large-scale subsidence, and reduced surface evaporative cooling; nighttime events are favored by elevated humidity, enhanced downwelling longwave radiation, and weak nocturnal ventilation. Compound events inherit both mechanisms and produce the largest anomalies in 2-m air temperature and cumulative heat stress (Wu et al., 2023). In arid coastal settings along the southern Arabian Gulf and the Red Sea, late-day moisture advection by sea breezes and limited nighttime ventilation can lock in high minimum temperatures, creating particularly hazardous compound heat (Raymond et al., 2024). These distinctions motivate a climatology that explicitly separates daytime, nighttime, and compound heatwaves across the Middle East and interprets their dynamics in an Arabian Peninsula's context. Motivated by this, we first develop a region-wide occurrence climatology that distinguishes daytime, nighttime, and compound heatwaves across the Middle East, resolved seasonally. We then assess trends in heatwave occurrence in the past few decades, exploring whether compound events show the strongest increases relative to daytime- or nighttime-only episodes. To situate these patterns physically, we interpret spatial and temporal heterogeneity in relation to canonical regional drivers, AHL position and intensity, Shamal variability, and moisture advection.

According to the forecast near-surface temperature in the Middle East remains challenging. Physics-based reanalyses and numerical weather prediction (NWP) models provide essential guidance yet can under-represent daily maxima and nocturnal minima in arid coastal settings where sea breezes, urban heat islands, and aerosol–radiation interactions modulate the diurnal cycle (Hersbach et al., 2020; Ullah et al., 2024). In parallel, global Artificial Intelligence (AI)-weather models, such as GraphCast, Pangu-Weather, FourCastNetv2, and Aurora, have demonstrated competitive medium-range skill at far lower computational cost (Lam et al., 2023; Bi et al., 2023; Pathak et al., 2022; Lang et al., 2024). Recent AI-weather model evaluations show mixed performance on heat extremes. Pasche et al. (2025) compare GraphCast, Pangu-Weather, and FourCastNet against the European Centre for Medium Range Weather Forecasts - High Resolution Integrated Forecast System (ECMWF-HRES) across three high-impact cases: the 2021 Pacific Northwest heatwave, the 2023 South Asian humid-heat event, and the 2021 North American winter storm. They report that machine learning (ML) models can match HRES locally for the Pacific Northwest heatwave but tend to underperform when errors are aggregated; for the humid-heat case, impact-relevant variables are limited and danger hotspots are underestimated. Liu et al. (2024) conduct a regional benchmark over East Asia and the Western Pacific for June–November 2023, evaluating five global AI-weather models—Pangu-Weather, FourCastNet v2, GraphCast, FuXi, and FengWu—initialized and verified with ERA5. Their analysis is general-purpose rather than heatwave-targeted; FengWu ranks highest on root-mean-square-error (RMSE) and anomaly correlation coefficient (ACC), and a simple multi-model mean performs comparably to the best single model. Beyond temperature extremes, event papers on unprecedented precipitation also highlight strengths and gaps for extremes: for the April 2024 Dubai storm, GraphCast predicted the event eight days ahead while FuXi underestimated totals at longer leads (Qiang et al., 2025). Taken together, these studies indicate that AI models generally reproduce synoptic evolution of hot spells but often mute near-surface amplitude, especially for the most intense phases, and that model rankings can be region- and metric-dependent. However, systematic verification for humid coastal and nighttime regimes in the Middle East, as well as stability across multi-day leads during compound heat, remains largely untested. This research gap motivates our focused evaluation.

A severe early-summer heat episode affected the Middle East from roughly 14–19 June 2024, peaking on 17 June in western Saudi Arabia and parts of Iraq and Kuwait. Reported air temperatures along the Hajj route reached the upper 40s °C during 16–18 June, with 51.8 °C observed at Mecca's Grand Mosque. The authorities recorded >2700 heat-related illnesses on 16 June and later reported ~1300 fatalities among the pilgrims, underscoring the compound risks of hot days and humid nights (Nashwan et al., 2024; Reuters, 2024). In this study, we (i) diagnose the multiscale circulation that triggered the 17 June peak—including the AHL's characteristics, local- and regional-scale circulations, and inland moisture transport—and its expression in the diurnal cycle; (ii) quantify where daytime, nighttime, and compound heat anomalies were largest (with emphasis on southern Arabian Gulf and Red Sea coastal regions); and (iii) evaluate four AI weather models for 2-m temperature across 1–10-day leads during the event, benchmarking the model performance also to the initialization robustness. The June 2024 heatwave episode, peaking on 17 June, offers a stringent, policy-relevant stress test that links mechanistic understanding to forecast performance in precisely those regimes where impacts are greatest. Salient results previewed here are: (1) an anomalously strong/westward-shifted lower-tropospheric thermal trough with ridging to the north and a weakened Shamal favouring hot, humid nights along the coasts; (2) compound (day–night) temperature anomalies dominated in coastal areas, while interior maxima were primarily daytime-driven; and (3) AI models retained useful land-average skill through ~D + 5 but degraded more rapidly for nocturnal minima and in coastal humidity regimes, where the underestimation of nighttime heat was common.

This paper is structured as follows. Section 2 describes datasets and AI-weather models. Section 3 details the heatwave taxonomy, verification diagnostics and metric considered in the study. Section 4 first establishes a seasonally resolved climatology of daytime, nighttime, and compound heatwaves over 2005 to 2025 and documents their trends (Section 4.1). Guided by these regimes, Section 4.2 examines the drivers and structure of the 10 to 27 June 2024 compound event, characterized by extreme daytime maxima over inland deserts and persistently elevated nighttime minima along the Red Sea and Arabian Gulf coasts. Section 4.3 then evaluates four AI weather models during this event to assess their operational readiness for heat-risk early warning in the Middle East. Section 5 synthesizes the implications for early-warning heat and for improving data-driven models in arid coastal environments.

2. Data and AI-Weather Models

We used the United States National Oceanic and Atmospheric Administration (NOAA) Global Surface Summary of the Day (GSOD)

station data over the Middle East to identify daytime, nighttime, and compound heatwaves, and ERA-5 (Hersbach et al., 2020) reanalysis data to (i) diagnose the drivers of the June 2024 compound heatwave, and (ii) initialize the AI weather models and evaluate their forecasts.

2.1. Observations and reanalysis

NOAA GSOD is a global, quality-controlled daily product derived from the Integrated Surface Hourly/Integrated Surface Database (ISH/ISD) archive by aggregating hourly and synoptic reports to daily summaries (data access: https://www.ncei.noaa.gov/data/global-summary-of-the-day/access). Our analysis domain (30°-60°E, 10°-40°N) comprises 491 GSOD stations across the Middle East; station locations and record lengths are shown in Fig. S1a. The number of active stations varies by year over 1970–2025 (Fig. S1b). To ensure adequate spatial coverage, we restrict climatology and trend analyses of heatwaves to December 2005–February 2025, during which the annual network availability exceeds ~58% of the 491-station inventory and station counts vary little (Fig. S1b). Additionally, for each station and season we apply a ≥ 90% data-availability threshold; only stations meeting this criterion are included in the seasonal frequency aggregation and trend estimation. The resulting seasonal samples are nearly equal in size across seasons (DJF n = 194, MAM n = 195, JJA n = 194, SON n = 196), providing balanced spatial coverage for the analyses (Fig. 1). We use daily 2-m air temperature, dew-point temperature, and relative humidity fields. In what follows, we denote daily maximum and minimum 2-m air temperatures from the GSOD stations as T_{max} and T_{min}, respectively, and hourly 2-m air temperature from ERA-5 and the AI-weather models as T2m. Prior to analysis, we (i) apply gross-range checks to remove physically implausible values (for example, cases with $T_{min} > T_{max}$). We treat GSOD missing-value codes for T_{max} and T_{min} (9999.9) as NaN and require seasonal completeness: for any station and season, at least 50% of the calendar days must be present for inclusion in frequency aggregation and trend analysis.

ERA-5 reanalysis data is used to investigate the drivers of the mid-June 2024 heatwave in the Middle East. This dataset extends from 1940 to present and is at 0.25° × 0.25° (~27 km) spatial and hourly temporal resolution. Its performance over the region has been thoroughly assessed and is found to be in close agreement with ground-based observations. For example, Nelli et al. (2024), and at a site in the Rub'Al Khali Desert and for the summer season targeted here, reported air temperature biases generally within 2 °C, with the observed and ERA-5 wind speed typically within 2 m s^{-1} of each other. Owing to deficiencies in the simulation of the surface radiation fluxes and properties, the surface temperature in the reanalysis can be as much as 5 °C higher than that observed. Ullah et al. (2024) notes that ERA-5 has a tendency to underestimate cold and hot extremes in the Arabian Peninsula, even though the inter-annual variability seen in station data is typically replicated in the reanalysis product.

2.2. Description of AI-weather models

We conduct an independent evaluation of four prominent AI-based global forecasting systems, namely,FourCastNetv2 (FCN2), PanguWeather, GraphCast, and Aurora. Together, these models span the major architectural families used for medium-range guidance: Spherical-harmonic neural operator, 3D transformers, and graph neural networks, alongside a recent Earth-system foundation model. Prior studies have shown competitive skill relative to leading NWP baselines at dramatically lower computational cost (e.g., Pathak et al., 2022; Bi et al., 2023; Lam et al., 2023; Lang et al., 2024; Bodnar et al., 2025), motivating their evaluation during a compound heat event in the Middle East. To ensure consistency, we used similarly configured models at 0.25° horizontal resolution with a 6-h fundamental time step and 13 vertical levels (Table 1).

The main features of the four AI-weather models are summarized below:

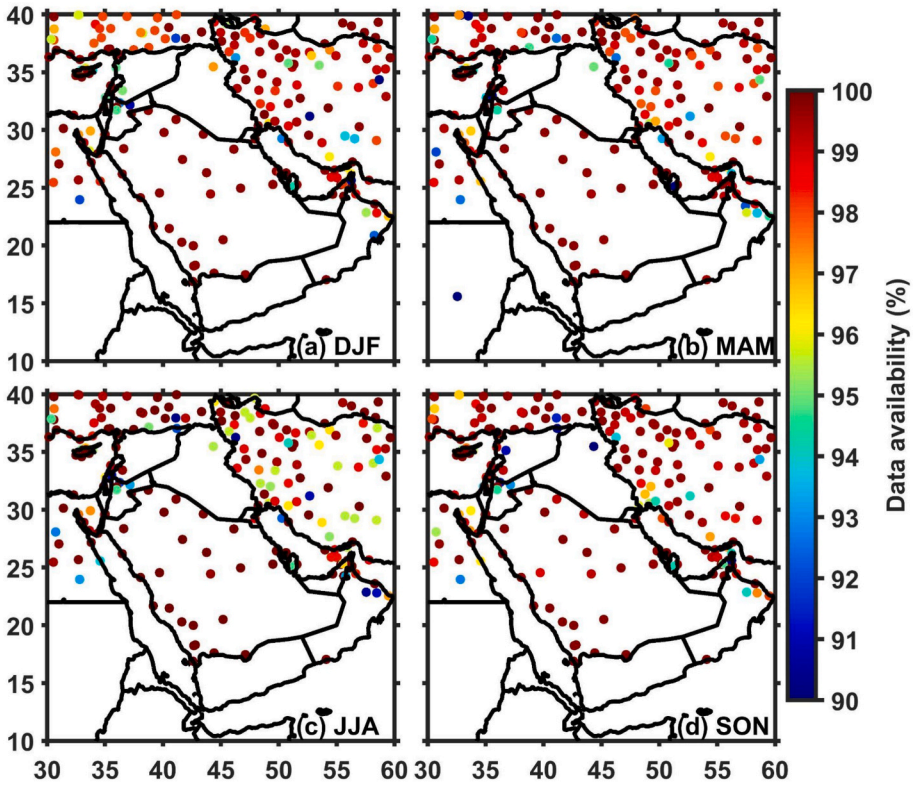

Fig. 1. NOAA GSOD stations across the Middle East with ≥90% seasonal data availability during December 2005–February 2025. (a) DJF (n = 194), (b) MAM (n = 195), (c) JJA (n = 194), (d) SON (n = 196). Marker colour indicates seasonal availability (%) relative to the total possible days in that season.

Table 1
Summary of key features of recent AI-weather forecasting models.

Feature	FourCastNetv2	PanguWeather	GraphCast (operational)	Aurora Fine-tune
Architecture	Spherical-harmonic neural operator	Transformer	Graph neural network	Multiscale 3D Swin-Transformer U-Net
Resolution	0.25° × 0.25°	0.25° × 0.25°	0.25° × 0.25°	0.25° × 0.25°
No. of surface variables	7 T2m, u10m, v10m, u100m, v100m, mean sea level pressure, Total column integrated water vapour	4 T2m, u10m, v10m, mean sea level pressure	5 T2m, u10m, v10m, mean sea level pressure, geopotential height	4 T2m, u10m, v10m, mean sea level pressure
Vertical levels (13) and atmospheric variables	T, u, v,RH, geopotential height	T, u, v,q, geopotential height	T, u, v, w, q, geopotential height	T, u, v,q, geopotential height
	50 hPa, 100 hPa, 150 hPa, 200 hPa, 250 hPa, 300 hPa, 400 hPa, 500 hPa, 600 hPa, 700 hPa, 850 hPa, 925 hPa and 1000 hPa			
Fundamental time step (h)	6	6	6	6
Reference	Pathak et al. (2022); Lam et al. (2023)	Bi et al. (2023)	Lam et al. (2023)	Bodnar et al. (2025)

- FourCastNetv2 (FCN2): A successor to FourCastNet's Adaptive Fourier Neural Operator (AFNO) design, FCN2 replaces planar Fourier operators with spherical-harmonic neural operators, improving stability for global autoregressive rollouts while retaining fast inference. In our setup, FCN2 ingests ERA-5 initial states and produces 6-hly forecasts on 13 levels to a 10-day horizon (Pathak et al., 2022; Lam et al., 2023).
- PanguWeather: Uses a 3D Earth-specific Transformer with an encoder–decoder stack derived from Swin-Transformer to represent multiscale atmospheric structure; surface and upper-air predictors are embedded jointly. The model achieves strong medium-range performance with multi-timescale rollout strategies (Bi et al., 2023).
- GraphCast: Formulates global weather as a message passing on an icosahedral, multiresolution mesh, learning prognostic updates via a graph neural network. Autoregressive fine-tuning stabilizes longer leads; published results show broad skill advantages over a leading deterministic baseline at substantially lower cost (Lam et al., 2023).
- Aurora: A large-scale foundation model for the Earth system that couples a multiscale 3D Transformer–U-Net–style backbone with pretraining on diverse geophysical data, then task-specific fine-tuning for global weather, air quality, waves, and TC tracking. Recent reports document performance competitive with operational systems (Bodnar et al., 2025).

Forecasts for all four systems are produced using the ECMWF ai-models framework with model-specific backends (https://github.com/ecmwf-lab/ai-models); Aurora inference uses Microsoft's released weights on Hugging Face for the fine-tuned aurora model. All forecasts are initialized from ERA-5 analyses and integrated to a 10-day horizon. We extract T2m from each model and evaluate their performance against ERA-5 over the intense heatwave period 12–22 June 2024, with additional station-based validation where applicable.

3. Methodology

3.1. Heatwave classification

We adopt a diurnal classification that distinguishes daytime, nighttime, and compound (day–night) heatwaves, as the mechanisms and impacts differ between daily maxima and minima extremes. Let T_{max} denote daily maximum temperature and T_{min} denote daily minimum temperature, derived from NOAA GSOD station observations (Section 2.1). This taxonomy follows Wu et al. (2023) and is consistent with established percentile-based extremes practice. A "hot" day (night) is defined when T_{max} (T_{min}) exceeds a sliding 90th percentile threshold, T_{max90} (T_{min90}), computed from a 15-day moving window (±7 day) centered on each calendar day, using a station-specific reference climatology (here, 1970–2025 where available). A heatwave requires ≥3 consecutive hot days (for daytime events) or hot nights (for nighttime events), as summarized below:

- Daytime heatwave: ≥3 consecutive days with $T_{max} \geq T_{max90}$ while the corresponding nights remain $T_{min} < T_{min90}$ (i.e., no concurrent hot nights);
- Nighttime heatwave: ≥3 consecutive nights with $T_{min} \geq T_{min90}$ while the corresponding days remain $T_{max} < T_{max90}$ (i.e., no concurrent hot days);
- Compound heatwave: ≥3 consecutive days during which both $T_{max} \geq T_{max90}$ and $T_{min} \geq T_{min90}$ on each day.

The sliding-percentile method preserves seasonality and local heterogeneity better than absolute thresholds and is widely used in regional and global heatwave studies (Barriopedro et al., 2023; Wu et al., 2023). This choice of a 90th percentile threshold is consistent with recent climatological analyses that employ percentile-based definitions near the 90th percentile (e.g., Barriopedro et al., 2023; Wu et al., 2023; Korhonen et al., 2025) and aligns with common practice in epidemiological assessments, where exceedance of local 90th percentile thresholds over multiple days is associated with elevated risks of mortality and morbidity (e.g., Arsad et al., 2022; Guo et al., 2017). It is also appropriate for the study region, where T2m are systematically high in summer. A 95th percentile criterion isolates only the most extreme cases and substantially reduces event counts, whereas an 85th percentile criterion identifies too many events for the objectives of this study.

Seasonal trends in heatwave for three classifications are estimated with the Theil–Sen robust slope (Sen, 1968; Theil, 1992) and tested for significance using the Mann–Kendall test (Mann, 1945) at the 90% confidence level; slopes are reported in events per year.

3.2. Arabian heat low and thermal indices

An important feature of the summertime circulation in the Arabian Peninsula is the AHL (Fonseca et al., 2022b). The AHL, which develops in response to the strong surface heating by the Sun, modulates the occurrence of convection around the Al Hajar mountains in northern Oman (e.g., Fonseca et al., 2025) and the Hijaz mountains in Saudi Arabia and Yemen (e.g., Parajuli et al., 2023) in addition to the tracks of tropical disturbances approaching the Arabian Peninsula (Francis et al., 2022). It is typically diagnosed based on the low-level atmospheric thickness (LLAT), defined as the geopotential height difference between the 700 and 925 hPa pressure surfaces. In particular, the AHL corresponds to the region with the top 10% LLAT values in the domain 10°-35°N and 40°-60°E at 03 UTC, excluding water bodies and areas where the 925 hPa pressure level is below the surface.

Three thermal indices are considered in this work: wet-bulb temperature (T_w), defined in eq. (1), extracted from Stull (2011); a simplified version of the wet-bulb globe temperature (WBGT), taken from Leroyer et al. (2018) and defined in eq. (2); the heat index (HI), extracted from Anderson et al. (2011). In eqs. (1)–(3),T is the air temperature in °C and *RH* is the relative humidity in %. Heat stress can become life-threatening at $T_W > 31$ °C (Wilson et al., 2024), WBGT>30 °C (Leroyer et al., 2018), and HI > 41 °C (Anderson et al., 2011).

$$T_w = T\ atan\left[0.151977\sqrt{RH + 8.313659}\right] + atan[T + RH] - atan[RH - 1.676331] + 0.00391838[RH]^{\frac{3}{2}}\ atan[0.023101\ RH] - 4.686035 \quad (1)$$

$$WBGT = 0.567\ T + 0.393\left(\frac{RH}{100}\right) 6.105^{\frac{17.27\ T}{237.7+T}} + 3.94 \quad (2)$$

$$\mathrm{HI} = -8.78469475556 + 1.61139411\ \mathrm{T} + 2.33854883889\ \mathrm{RH} - 0.14611605\ \mathrm{T\ RH} - 0.012308094\ T^2 - 0.0164248277778\ RH^2 + 2.211732 \times 10^{-3}\ T^2\ RH + 7.2546 \times 10^{-4}\ T\ RH^2 - 3.582 \times 10^{-6}\ T^2\ RH^2 \quad (3)$$

3.3. AI-weather models evaluation metrics

We use ERA-5 as the common reference for cross-model comparison because the AI-weather models evaluated here are trained to predict the next ERA-5 state and are typically initialized from ERA-5 analyses (Pasche et al., 2025; Qiang et al., 2025). This choice ensures internal consistency across models. At the same time, verification against ERA-5 primarily quantifies fidelity to the ERA-5 target and may understate biases relative to independent observations, particularly in the distribution tails (Qiang et al., 2025). Over the Arabian Peninsula, ERA-5 performs comparatively well for T2m among available reanalysis products (Ullah et al., 2024). We therefore adopt ERA-5 for the primary comparison while noting this limitation explicitly. For each NOAA GSOD site, we collocated T2m from the four AI-weather models by averaging all grid points whose centers fall within a 25 km radius of the station. The 25 km radius approximates the effective 0.25° grid spacing of the AI models and ERA-5 in this domain and reduces station-to-grid representativeness error relative to using a single nearest grid point. To ensure symmetry, we applied the same 25 km radius-mean procedure to ERA-5 at the station locations. As all model outputs and ERA-5 are provided on the same 0.25° grid, no horizontal regridding was required; the radius-mean is used to reduce representativeness error between point observations and gridded fields. The models' skill is evaluated over the 12–22 June 2024 heatwave.

We compute the Pearson correlation coefficient (r), mean bias (MB), and root-mean-square error (RMSE), widely used in regional evaluation studies (e.g., Fonseca et al., 2022a; Temimi et al., 2020; Nelli et al., 2024; Yarragunta et al., 2025). Let O_i denote the observation and M_i the corresponding model value at time i, with N paired samples and $\underline{O}$, $\underline{M}$ their means after removing missing pairs. Then.

$$r = \frac{\sum_{i=1}^{N}[(O_i - \overline{O})(M_i - \overline{M})]}{\sum_{i=1}^{N}(O_i - \overline{O})^2 \sum_{i=1}^{N}(M_i - \overline{M})^2} \quad (4)$$

$$\mathrm{RMSE} = \left(\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2\right)^{\frac{1}{2}} \quad (5)$$

$$\mathrm{MB} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) \quad (6)$$

By this sign convention, MB > 0 indicates a positive bias (model > observation).

4. Results

4.1. Climatology and trends of heatwaves over the middle east

We first quantify the spatial occurrence and seasonal trends of daytime, nighttime, and compound heatwaves during 2005–2025 to establish the regimes and mechanisms that motivate the subsequent case study and model evaluation. Heatwaves—segregated into daytime, nighttime, and compound events—are compiled season-wise (spring:

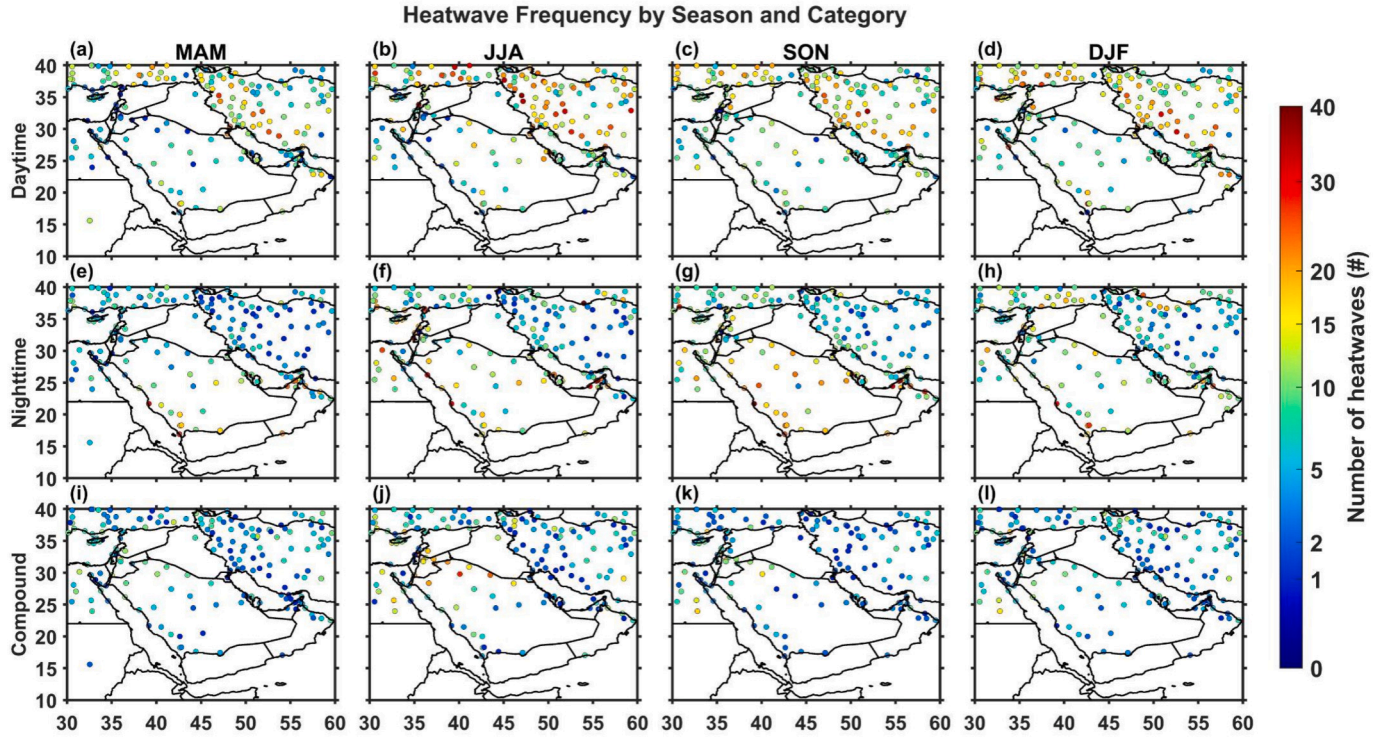

Fig. 2. Seasonal distribution of heatwave occurrences at weather stations across the Middle East for December 2005–February 2025. Rows denote event type—(a–d) daytime, (e–h) nighttime, and (i–l) compound—while columns show seasons in the order MAM, JJA, SON, and DJF. Colors indicate the accumulated number of events per station over the period.

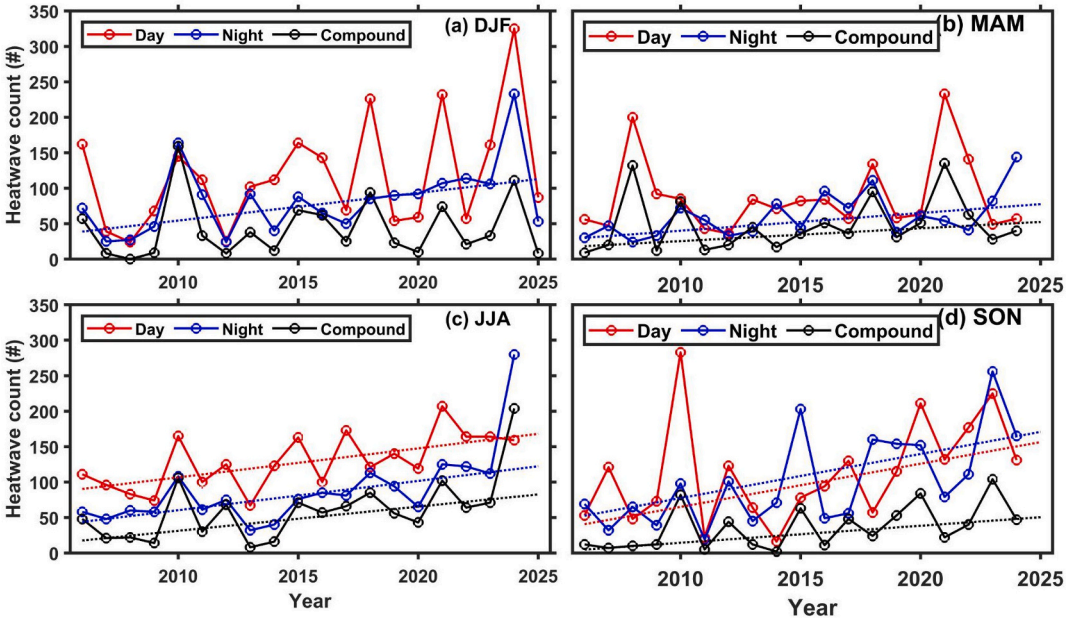

Fig. 3. Regional annual counts of heatwave occurrences and least-squares trends over the Middle East domain shown in Fig. 1 for December 2005–February 2025. Solid lines denote daytime (red), nighttime (blue), and compound (black) events summed across the Middle East stations; dotted lines indicate the corresponding linear fits, plotted only when significant at the 90% confidence level ($p \leq 0.10$). Panels show seasons: (a) DJF, (b) MAM, (c) JJA, (d) SON. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

March–April–May (MAM); summer: June–July–August (JJA); autumn: September–October–November (SON); winter: December–January–February (DJF)), and the accumulated counts across 491 stations for December 2005–February 2025 are shown in Fig. 2 (daytime: a–d; nighttime: e–h; compound: i–l), with regional totals and linear trends in Fig. 3. A prominent belt of high seasonal counts extends from the eastern Mediterranean through northern Iraq–Iran into Anatolia (Asian Türkiye), especially in the warm seasons, as evident in the daytime maps for MAM–SON (Fig. 2a–c) and, to a lesser extent, the compound maps for JJA–SON (Fig. 2j–k). This pattern is consistent with frequent placement beneath the subsiding branch of the subtropical anticyclone and the monsoon–desert circulation that favor clear skies and strong insolation (Rodwell and Hoskins, 1996; Rodwell and Hoskins, 2001). In contrast, nighttime and compound events are relatively common along the southern Arabian Gulf shore and adjacent inland areas, most clearly in summer and autumn (nighttime: Fig. 2f–g; compound: Fig. 2j–k). Warm sea-surface temperatures (SSTs), which in these seasons can exceed 30 °C (Nesterov et al., 2021), and the evening sea-breeze transport the moist marine air inland after sunset, elevating near-surface humidity and downward longwave radiation and thereby suppressing nocturnal cooling—processes that coincide with observed peaks in wet-bulb temperature and heat index along this coast (Im et al., 2017; Raymond et al., 2020). Over continental interiors in late spring and summer, soil-moisture depletion and reduced evaporative cooling further amplify daytime extremes, consistent with elevated counts in MAM–JJA (Fig. 2a–b) (Seneviratne et al., 2010; Miralles et al., 2014). These spatial patterns and diurnal contrasts are consistent with process-based assessments that distinguish shortwave-driven daytime heatwaves from humidity/longwave-dominated nighttime events (Wu et al., 2023).

Linear trends presented in Fig. 3 (counts yr^{-1}) show a statistically significant rise in nighttime events during winter (+3.90; Fig. 3a), in both nighttime and compound events during spring (+2.50 and + 1.80; Fig. 3b), and in all classes increase in summer (daytime +4.09; nighttime +4.13; compound +3.42; Fig. 3c) and again in autumn (daytime +6.09; nighttime +6.22; compound +2.40; Fig. 3d). The seasonal contrast—strong summertime growth dominated by interior daytime heating, and sustained autumn increases led by coastal nocturnal/compound events—is consistent with diurnally asymmetric warming over land, whereby nights warm faster than days and the likelihood of nighttime and compound heatwaves increases (Vose et al., 2005; IPCC, 2021; Francis and Fonseca, 2024a, 2024b). It also reflects the growing influence of Gulf humidity, which is in line with the up to 0.08 °C yr^{-1} increase in SSTs during 2003–2019 (Nesterov et al., 2021): warm marginal seas and the sea-breeze deliver moisture inland after sunset, enhancing downward longwave flux and wet-bulb temperatures and sustaining nocturnal heat stress along the coast (Im et al., 2017; Raymond et al., 2020). Basin-scale warming of the Red Sea and Arabian Gulf further supports this interpretation (Chaidez et al., 2017; Hereher, 2020). Collectively, these patterns are consistent with the Eastern Mediterranean–Middle East "hot-spot" behavior documented in regional assessments, in which hot extremes have already intensified and are projected to continue increasing (Lelieveld et al., 2016; Zittis et al., 2022; IPCC, 2021). The significant winter increase in nighttime events (+3.90 yr^{-1})is consistent with two mechanisms: (i) diurnally asymmetric warming, with minimum temperatures rising faster than daytime maxima and more warm nights across many land regions, including the Arabian Peninsula (Donat et al., 2013; Lewis and Karoly, 2013; IPCC, 2021), and (ii) warmer marginal seas (Arabian Gulf, Red Sea) that sustain a more humid boundary layer even in winter. Enhanced humidity/cloudiness increases downwelling longwave radiation and suppresses nocturnal cooling, a known driver of nighttime heatwaves (Chaidez et al., 2017; Hereher, 2020; Wu et al., 2023). In spring, the significant rise in both nighttime and compound events (+2.80 yr^{-1} and + 1.80 yr^{-1}) likely reflects an earlier and lengthening warm season that pushes both daytime and nighttime temperatures above thresholds during April–May (Zittis et al., 2022; Founda et al., 2019; Francis and Fonseca, 2024a, 2024b), with soil-moisture depletion and clear-sky, Shamal-favourable conditions intensifying daytime heating and concurrent humid–longwave processes elevating nocturnal minima, increasing the odds that day-and-night exceedances co-occur (Mueller and Seneviratne, 2012; Miralles et al., 2014; Wu et al., 2023).

4.2. 10–27 June 2024 compound Heatwave over the Middle East

4.2.1. Observational characterization from station records

Figs. 4 and 5 show daily anomalies of GSOD-station T_{max} and T_{min}

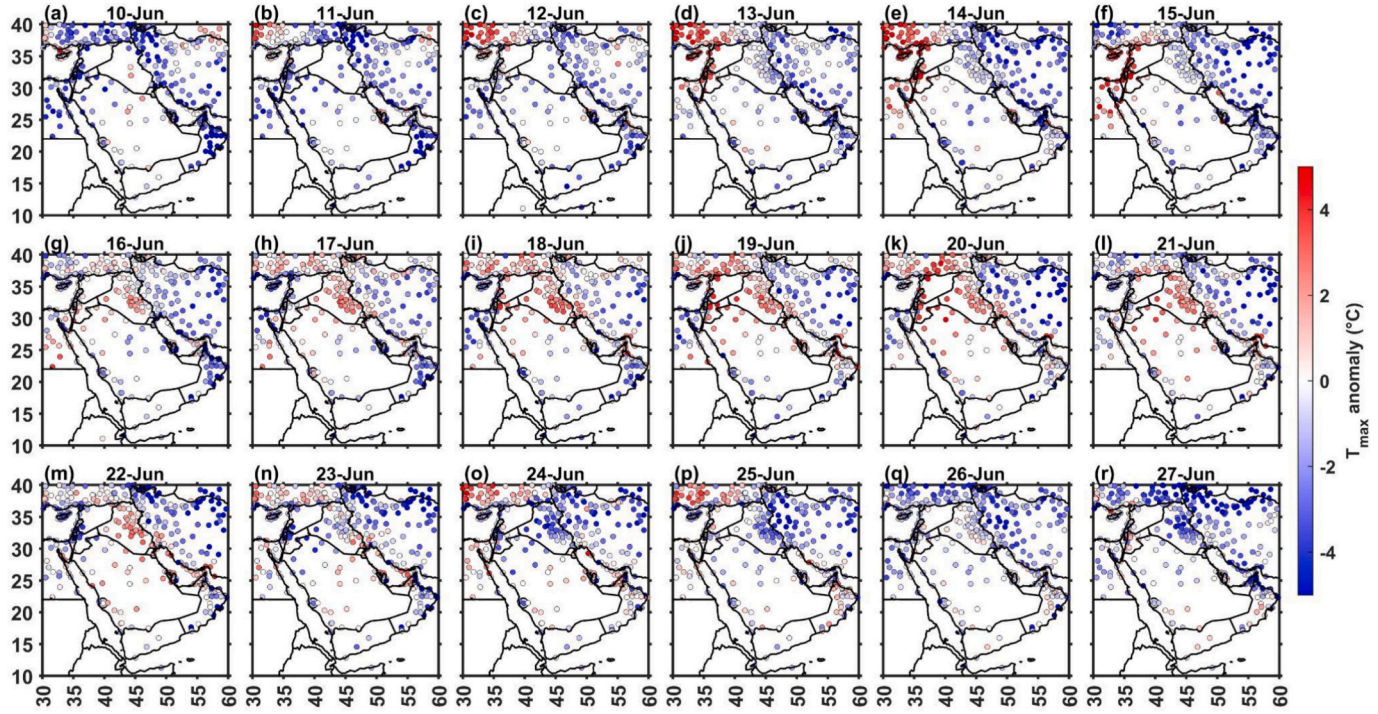

Fig. 4. Daily anomalies of the daily maximum 2-m air temperature (T2m, °C) at GSOD stations across the Middle East during 10–27 June 2024. The anomalies are computed relative to each station's day-of-year climatology using the maximum available record within 1970–2025. The circle in coastal western Saudi Arabia indicates the location of Jeddah.

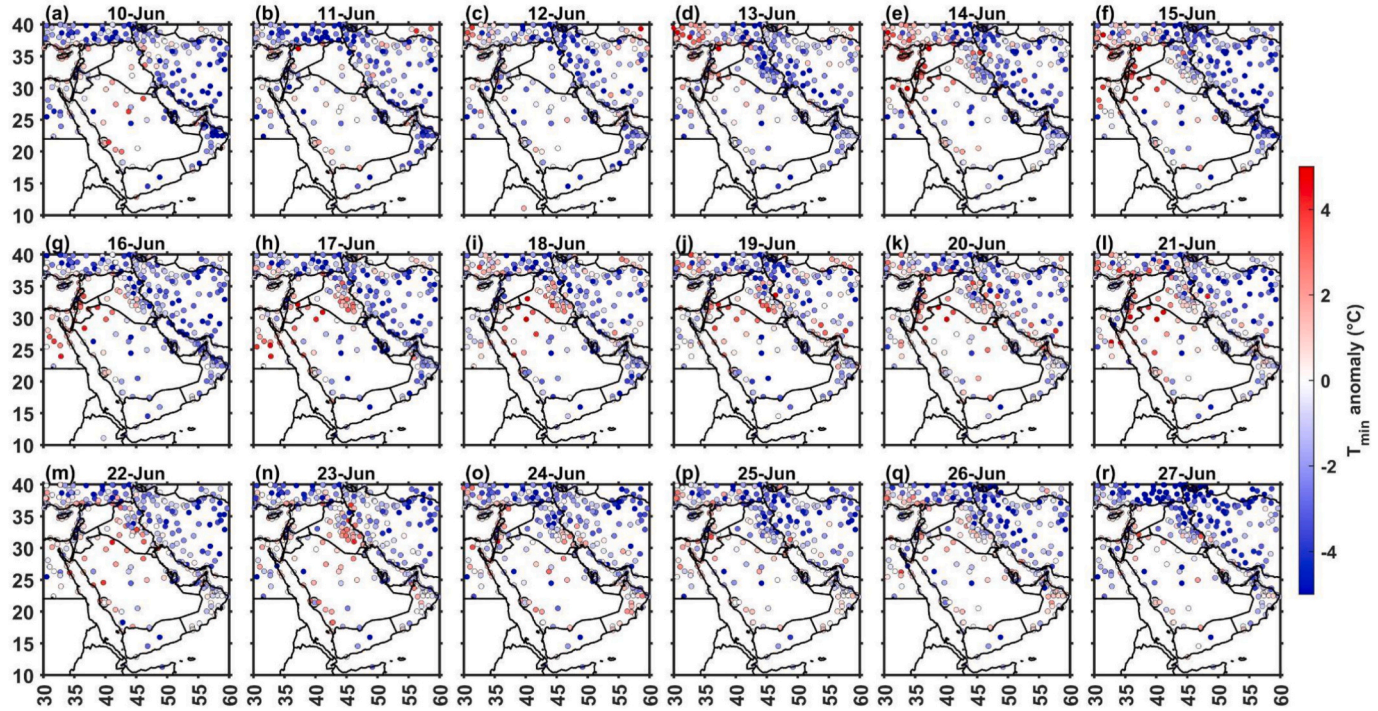

Fig. 5. As in Fig. 4, but for daily-minimum T2m anomalies (°C).

across the Middle East during 10–27 June 2024, computed relative to each station's day-of-year climatology. Figs. S2–S4 show spatial maps of ERA-5 T2m and anomalies relative to the corresponding monthly climatologies over the Middle East at 00 and 12 UTC on 16–18 June 2024. The warm spell first emerges over the northwestern sector (Anatolia–Levant; Fig. 4a–f) and then intensifies after ~15 June, expanding eastward and southward (Fig. 4g–l). Concurrently, elevated T_{min} appears along the southern Red Sea and Arabian Gulf coasts (Fig. 5a–f) and strengthens during 16–23 June (Fig. 5g–n), underscoring the compound character of the event. The latter can be at least partially explained by the positive SST anomalies in these two water bodies, which during 10–28 June 2024 are typically of 1°-3 °C, but exceeding 4 °C in parts of

the northern Arabian Gulf (not shown). In the daytime (T_{max}) field, the strongest and most spatially coherent positive anomalies are evident by 16–21 June across the northern Middle East, including the eastern Mediterranean coast (Lebanon, Syria, Israel/Palestine, and Jordan), northern Iraq, western and northern Iran, and Anatolia, with many stations registering +3–5 °C departures from their day-of-year climatologies (Fig. 4g–l).

In the nighttime temperature (T_{min}) field, large positive anomalies persist along maritime margins—especially the southern Red Sea and Arabian Gulf coasts—where warmer than average seas and the evening onshore flow maintain moist boundary layers and suppress nocturnal cooling through 16–23 June (Fig. 5g–n). By late month (22–27 June), anomalies weaken heterogeneously: interior plateaus cool more rapidly, while positive departures linger near the southern Arabian Gulf, parts of the Red Sea, and sectors of the northern Middle East possibly because of the warmer SSTs, leaving many coastal and northern stations still above normal as the event decays (Figs. 4m–r, 5m–r). This inland–coastal contrast, strong daytime anomalies over interiors and pronounced nighttime anomalies along warm coasts, is a classic feature of compound heat in arid and semi-arid regions (Wu et al., 2023). Daytime extremes are favored by clear-sky shortwave heating and strong land–atmosphere coupling over dry soils, whereas coastal nighttime warmth is sustained by humidity-enhanced downward longwave radiation that limits nocturnal cooling (Seneviratne et al., 2010; Miralles et al., 2014; Wu et al., 2023).

Daily percentile ranks place these departures in statistical context (Figs. 6–7). During 16–21 June, many northern Middle East stations reached the 90th–100th percentiles of their long-term day-of-year T_{max} distributions (Fig. 6g–l), indicating near-record to record daytime heat across the region. Simultaneously, extensive stretches of the southern Red Sea and Arabian Gulf coasts registered T_{min} in the upper decile—i.e., ≥90th percentile relative to each station's day-of-year distribution (Fig. 7g–l)—evidencing widespread nocturnal heat. In the Mecca corridor, stations near Mina and Jeddah repeatedly fell in the upper decile for both T_{max} and T_{min} around 16–18 June (Figs. 6h–j, 7h–j), implying compound conditions in which daytime and nighttime exceedances co-occur within the same multi-day window. Such compound exceedances are of particular concern because reduced nocturnal relief increases cumulative heat strain and risk (Perkins-Kirkpatrick and Lewis, 2020; Raymond et al., 2020; Zittis et al., 2022).

To document persistence and geographic reach, we estimate for 10 to 27 June the daily fraction of GSOD stations, relative to all available stations, with percentile rank at or above the 90th, 95th, and 99th percentiles for T_{max} and for T_{min} (Fig. S5). Fractions peak on 16 to 20 June, when approximately 36% to 49% of stations exceed the 90th percentile for T_{max} and 27% to 40% exceed the 90th percentile for T_{min}, underscoring the compound character of the episode (Fig. S5).

To further characterize the temporal features and local drivers of the compound heatwave, we plotted time series of air temperature and thermal indices (wet-bulb temperature, eq. (1); WBGT, eq. (2); and heat index, eq. (3)), daily anomalies of T_{max} and T_{min}, and wind speed and direction for Jeddah and Mina during 10–28 June 2024 (Fig. 8), which are the nearest GSOD stations to Mecca's Grand Mosque and represent contrasting coastal (Jeddah) and slightly elevated inland valley (Mina) environments. At Jeddah, T_{min} anomalies remain positive for much of 16–24 June while T_{max} anomalies are smaller (Fig. 8c), and the thermal indices stay elevated over the same interval (Fig. 8a). The wind analysis shows frequent onshore-flow episodes during mid-June, consistent with moist evening sea-breeze intrusions that suppress nocturnal cooling (Fig. 8e). At Mina (farther inland and slightly elevated), T_{max} spikes on 17 June while T_{min} increases more intermittently (Fig. 8d); the thermal indices mirror this sharp peak (Fig. 8b), and the wind panel indicates weaker, more variable low-level flow relative to the coast (Fig. 8f). Taken together, these site records show concurrent day–night exceedances over several days, a persistent, multi-day compound heatwave, with interior/northern stations experiencing exceptional daytime heat and maritime margins (notably the Mecca vicinity) maintaining nocturnal heat that keeps heat-stress indices high across successive days. This spatial–diurnal partitioning and its persistence are consistent with the established physics of hot spells in the Middle East, strong land-surface coupling under clear skies inland, and humidity-driven nighttime warming along warm seas (Rodwell and Hoskins, 1996, 2001;

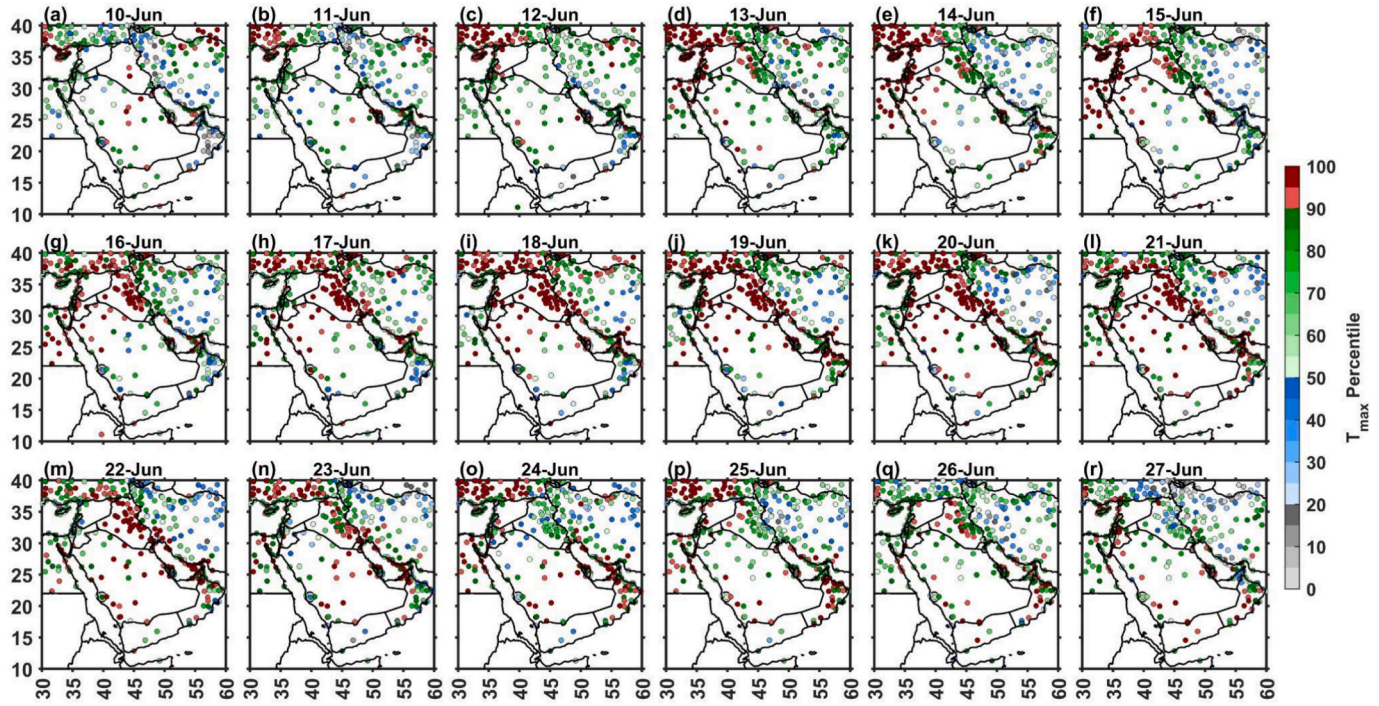

Fig. 6. Percentile rank of daily maximum temperature at GSOD stations during 10–27 June 2024, evaluated against each station's day-of-year distribution from its maximum available record within 1970–2025. Values near 100 indicate record highs. The circle in coastal western Saudi Arabia indicates the location of Jeddah.

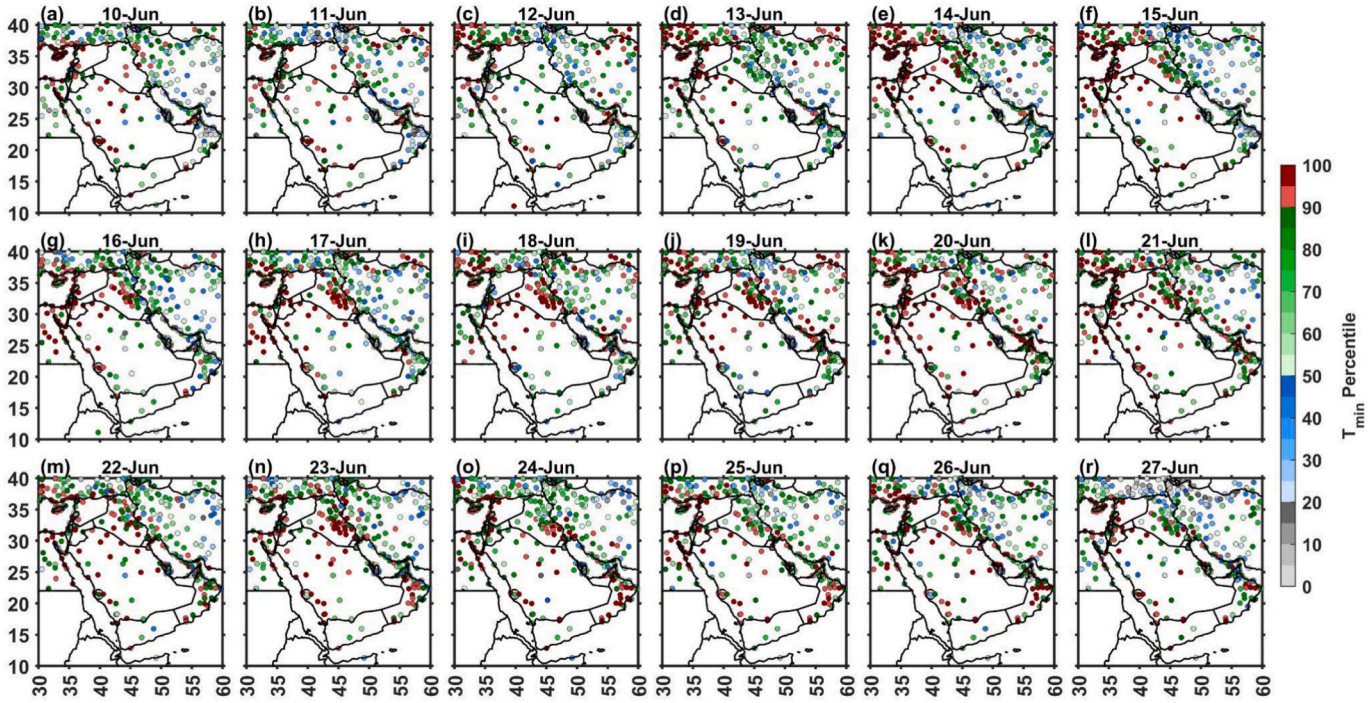

Fig. 7. As in Fig. 6, but for daily minimum temperature.

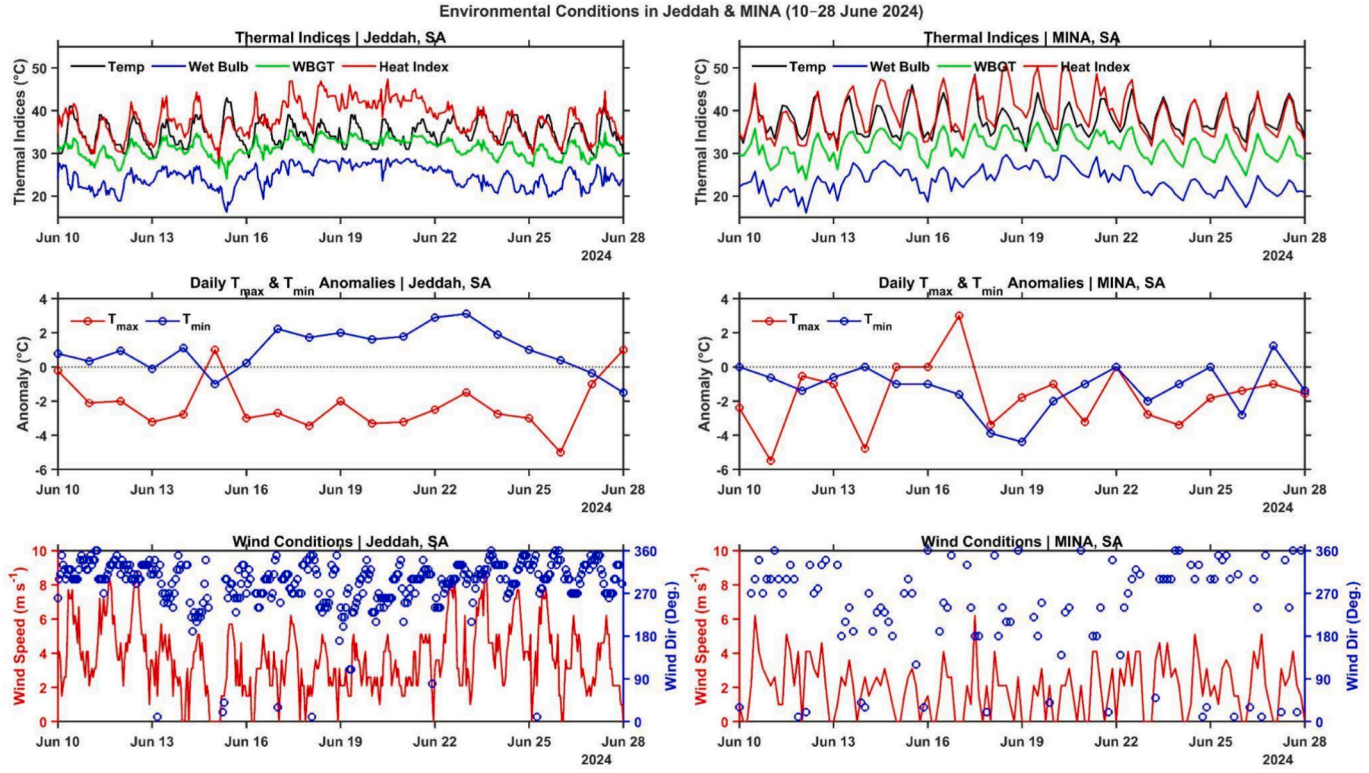

Fig. 8. Time series of meteorological variables at Jeddah (21.680° N, 39.157° E; circle in Figs. 4–7) and Mina (21.417° N, 39.867° E)—the nearest available stations to Mecca (21.433° N, 39.767° E), Saudi Arabia—during 10–28 June 2024. (a) Hourly air temperature (black; °C), wet bulb temperature (blue; °C), wet-bulb globe temperature (green; °C), and heat index (red; °C), (b) anomalies in the daily maximum (red; °C) and minimum (blue; °C) temperatures, and (c) hourly wind speed (red; m s^{-1}) and direction (blue; °) at Jeddah from 10 to 28 June 2024. (d)-(f) are as (a)-(c) but at Mina. These two stations span the coastal–inland gradient of the Mecca region and were selected to illustrate the distinct local drivers of coastal nocturnal heat and inland daytime extremes during the June 2024 compound heatwave. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

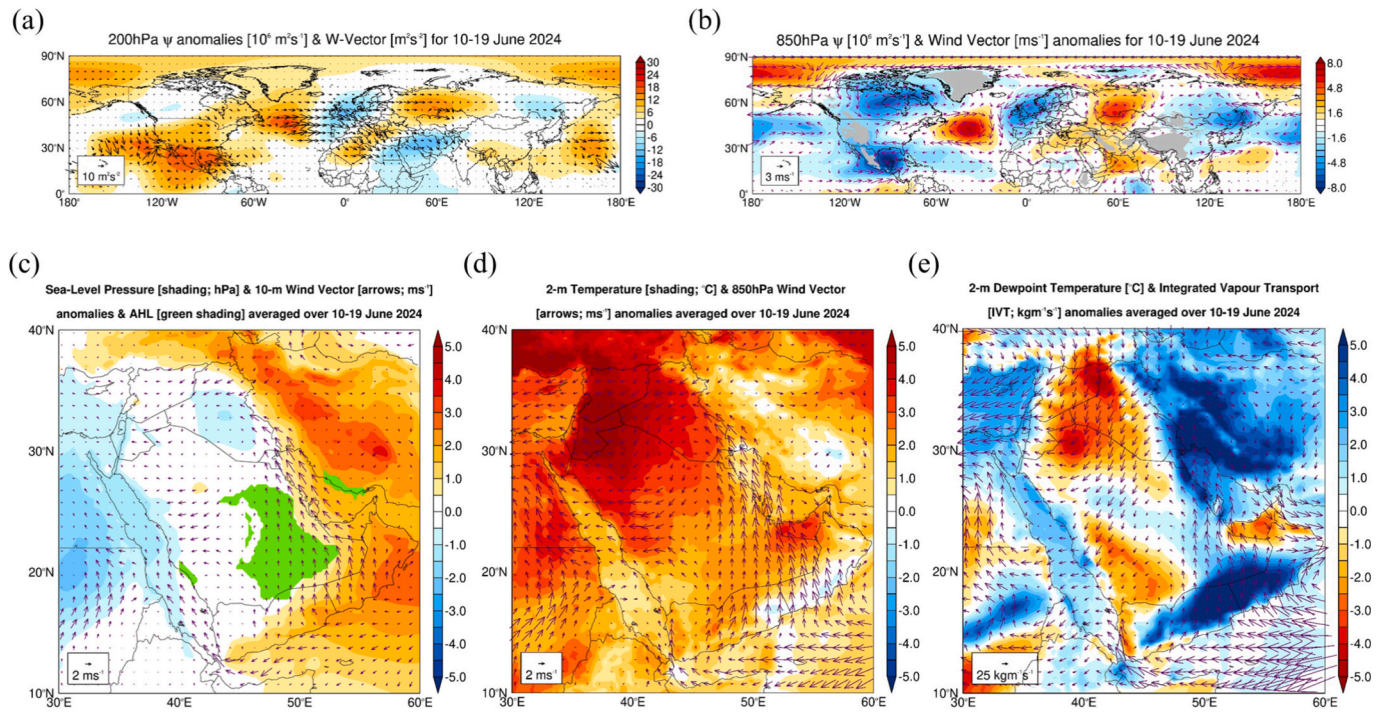

Fig. 9. Drivers of Middle East mid-June 2024 Heatwave: (a) 200 hPa streamfunction anomalies (shading; 10^6 m^2 s^{-2}), with respect to 1979–2023 climatology, and horizontal component of the Rossby wave activity flux (arrows; m^2 s^{-2}), as defined by Takaya and Nakamura's (2001) W-vector, averaged over 10–19 June 2024 for the Northern Hemisphere. (b) is as (a) but with the shading giving the 850 hPa streamfunction (10^6 m^2 s^{-2}) and the arrows the 850 hPa wind vector (m s^{-1}) anomalies. (c)-(e) are zoomed in plots over the Arabian Peninsula (10–40°N; 30–60°E) showing the anomalies of (c) sea-level pressure (shading; hPa) and 10-m wind vector (arrows; m s^{-1}), (d) 2-m temperature (shading; °C) and 850 hPa wind vector (arrows; m s^{-1}), and (e) 2-m dewpoint temperature (shading; °C) and integrated vapour transport (IVT; arrows; kg m^{-1} s^{-1}), also averaged over 10–19 June 2024. The green shading in (c) gives the Arabian Heat Low (AHL). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

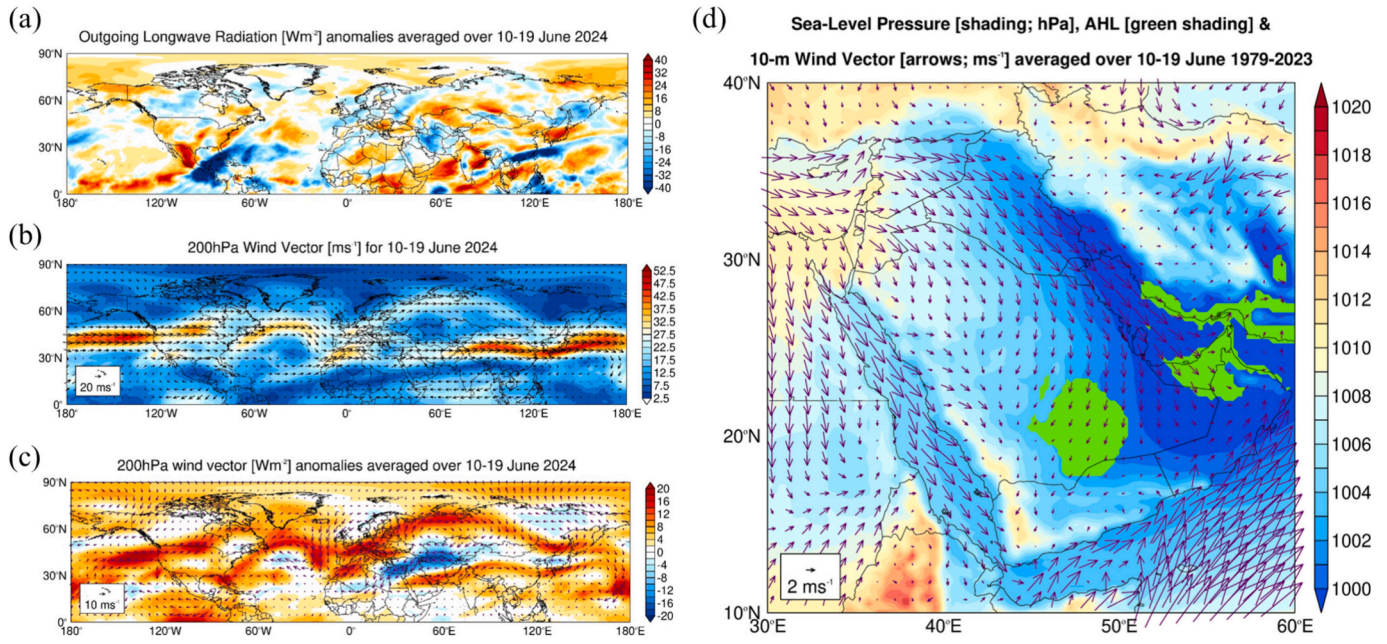

Fig. 10. Large-Scale Circulation during 10–19 June 2024: (a) Outgoing longwave radiation (OLR; W m^{-2}) anomalies with respect to the 1979–2023 climatology averaged over 10–19 June 2024 for the Northern Hemisphere. (b) and (c) are as (a) but showing the 200 hPa wind vector (m s^{-1}) actual values and anomalies, respectively. (d) Sea-level pressure (shading; hPa) and 10-m wind vector (arrows; m s^{-1}) 1979–2023 climatological values averaged over 10–19 June. The green shading gives the Arabian Heat Low (AHL), and the fields are shown for the Arabian Peninsula (10–40°N; 30–60°E). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Miralles et al., 2014; Im et al., 2017; Wu et al., 2023; Zittis et al., 2022).

4.2.2. Synoptic-to-planetary drivers from ERA-5 reanalysis

The large-scale circulation during 10–19 June 2024, which comprises the build-up and onset of the heatwave (Figs. 4–5), features an arc-shaped Rossby wavetrain extending from the Gulf of Mexico and Caribbean Sea to southwest Asia (Fig. 9a). This wavetrain is triggered by anomalous convection over Central America and Caribbean Sea extending to the southwest tropical North Atlantic Ocean (Fig. 10a). The enhanced convection arises mostly from the interaction of an active Central American Gyre, a broad lower-tropospheric cyclonic system that occasionally develops over Central America in particular in May–June and September–November (Papin et al., 2017), with a decaying cold front coming off the eastern United States. The wavetrain leads to a southward shift in the position of the subtropical jet over the Middle East (Figs. 10b-c), with an anomalous ridge over southwestern Asia extending southwards into the northern Arabian Sea and southwestwards into southeastern Europe (Figs. 9b-c). The anomalous ridge over the eastern Arabian Peninsula and anomalous low to its west lead to a weakening of the Shamal winds, the primary mechanism whereby summertime heatwaves develop in the northern and central Arabian Peninsula (Al-Shamarti et al., 2025). There are also noteworthy changes to the AHL: with respect to climatology (cf. Figs. 9c and 10d), its western core over Saudi Arabia is more meridionally and zonally elongated, which favours the weakening of the Shamal winds, while the eastern core that typically sits over the United Arab Emirates (UAE) and southeastern Iran in mid-June is restricted to a small patch over southern coastal Iran, owing to the anomalous ridge (+2–3 hPa pressure anomalies; Fig. 9c) over the region.

The anomalous offshore winds resulted in hotter (air temperature anomalies exceeding +5 °C in parts of Jordan, Syria, and Iraq; Fig. 9d) and drier (dewpoint temperature anomalies below −5 °C; Fig. 9e) conditions over the northwestern Arabian Peninsula. On the other hand, the increased onshore flow over coastal Oman and northern and central parts of the Arabian Gulf leads to a more moist environment here (dewpoint temperature anomalies above +5 °C; Fig. 9e). Fig. 8 shows a persistent positive anomaly in the daily minimum air temperature around Jeddah from 16 to 25 June, exceeding +3 °C on 22–23 June. Figs. 9c-e highlight this is a reflection of the weakening of the background northwesterly flow over the Red Sea, which in the summer typically extends throughout the whole basin (Fig. 10d), with moisture transport from the Arabian Sea into the central Red Sea through the Gulf

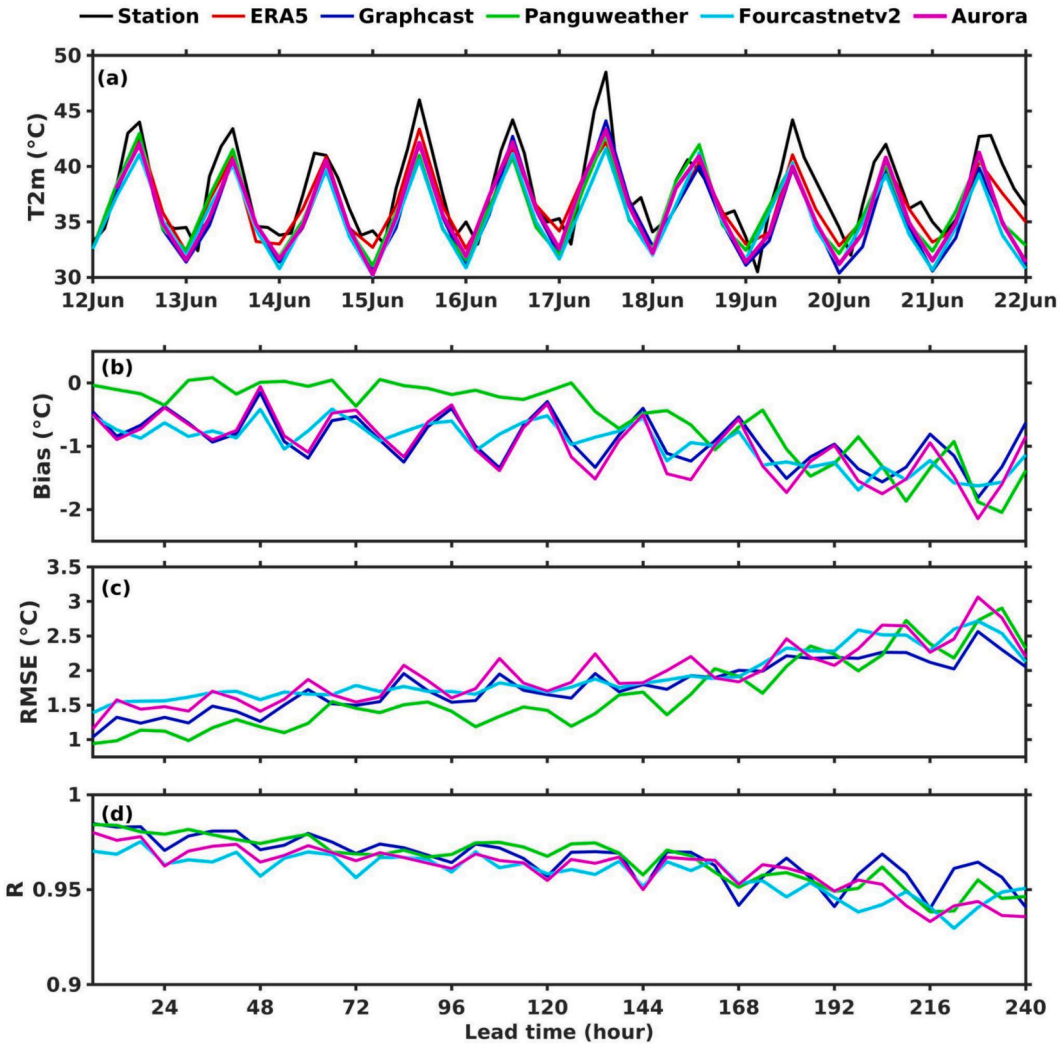

Fig. 11. (a) Time-series of T2m (°C) averaged over a 25 km radius around the Mina site: station observations (black), six hourly ERA-5 reanalysis (red), and six-hourly AI-weather models forecasts initialized at 00 UTC on 12 June 2024 (blue for Graphcast; green for Panguweather; cyan for Fourcastnetv2; pink for Aurora). (b-d) Mean-bias, defined in eq. (6), RMSE, defined in eq. (5), and correlation coefficient (R), defined in eq. (4) of AI-weather models with respect to ERA-5 over the Middle East land region (10–40°N; 30–60°E) for the same period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of Aden. Positive SST anomalies of 1–3 °C (not shown) also contribute to the higher moisture levels in the region (Fig. 9e).

4.2.3. AI-model performance

We evaluate four AI weather models(GraphCast, PanguWeather, FourCastNetv2, and Aurora) using six-hourly forecasts initialized at 00 UTC on 12 June 2024 and verified against ERA-5. For operational use, these models are initialized from an analysis such as ERA-5; as ERA-5 is released with ~five-day latency, a same-day analysis is not available in real time. Our target verification day is 17 June (peak T_{max} anomaly, ≥99th percentile at the Mina station, Fig. 8d), so we use forecasts initialized at 00 UTC on 12 June, the latest ERA-5 state available before the event, which places the 16 to 18 June within a five-day lead. This 12 to 22 June verification window lies within the broader 10 to 27 June episode documented in Figs. 6, 7, and S5. To assess robustness to initialization time, we also run staggered start dates that enter the peak at different leads; results are summarized in Fig. 12 and extended in Fig. S6. All AI-models and ERA-5 T2m grid points within a 25 km radius of the Mina site are averaged at each time step and plotted as a time series alongside the Mina station observations in Fig. 11a. All models reproduce the diurnal cycle, and the multi-day warming that culminates on 15–18 June. They underestimate the hottest daytime peaks relative to the station record (black, Fig. 11a) and show very good agreement with ERA-5 (red, Fig. 11a). The daytime shortfall is mostly apparent on 16–17 June, and is consistent with known difficulties in capturing near-surface extremes under strong land–atmosphere coupling (reduced evaporative cooling over dry soils) and with representativeness differences between point stations and grid-cell means (Seneviratne et al., 2010; Hofstra et al., 2009). Daily minima are also slightly low: all four models exhibit a modest cold bias in T_{min} relative to ERA-5 (Fig. 11a). Among them, PanguWeather shows the smallest nighttime offset, followed by GraphCast, whereas Aurora and FourCastNetv2 tend to be cooler at night, consistent with recent reports that AI models achieve strong synoptic skill but still under-amplify near-surface extremes (Lam et al., 2023; Bi et al., 2023; Pathak et al., 2022).

We next verified the 00 UTC 12 June initialization over the Middle East land mask (Fig. 1) by extracting all land grid points from the AI-weather models and ERA-5 and computing mean bias, RMSE, and correlation (R) (Figs. 11b–d). Region-wide diagnostics indicate small cool biases overall (typical median ≈ −0.5 to −1.5 °C, Fig. 11b) and monotonic error growth with lead time (the RMSE rises from ~1–1.5 °C at 1–2 days to ~2–2.5 °C by 8–10 days; Fig. 11c), while the correlations remain high throughout (R ≳ 0.94, Fig. 11d). The model ranking is consistent across leads: PanguWeather exhibits the smallest bias, lowest RMSE, and highest R, followed closely by GraphCast; Aurora is intermediate with modest, fairly steady cool bias; FourCastNetv2 tends toward a stronger cold bias and a slightly larger RMSE, especially around the peak-heat window. The spread increases with lead time for all systems, but none shows instability or abrupt skill loss. These patterns mirror the station-based comparison (Fig. 11a) and align with recent evaluations of AI-weather models, which report strong synoptic skill but under-amplification of near-surface extremes over dry, clear-sky regimes.

To test robustness beyond ERA-5, we verified the same 00 UTC 12 Jun 2024 initializations against an independent 9-km Weather Research and Forecasting Model (WRF, see supplementary text S1 for model configuration, Nelli et al., 2020; Abida et al., 2022; Fonseca et al., 2022a, Fonseca et al., 2025a, 2025b; Yarragunta et al., 2025) and the operational Global Forecast System (GFS, supplementary Figs. S7–S10). Across both baselines, all AI-weather models preserved the event's phasing and multi-day persistence with high temporal coherence (r ≈ 0.93–0.98), while errors increased gradually with lead and exhibited the expected diurnal modulation. Relative to WRF, the AI models showed systematic cool biases of roughly −1 to −3 °C and RMSE of ~2–4 °C, consistent with WRF's larger diurnal amplitude at 9 km (stronger land–sea-breeze and boundary-layer contrasts) making the AI guidance appear too cool at daytime peaks. Against GFS, biases were smaller (typically −0.5 to −1.5 °C) and RMSEs ~1.5–3 °C; GraphCast and Aurora yielded the most stable lead-wise statistics, FourCastNet v2 remained the coolest with the largest spread, and PanguWeather, though still competitive, was less uniformly superior to in the ERA5-based verification, reflecting its closer calibration to reference data.

In general, ERA5 exhibits a modest but systematic daytime near-surface cold tendency over arid land, reflecting limitations in surface-energy partitioning, turbulent mixing, and daytime boundary-layer growth (e.g., −0.3 to −2.5 °C T2m at the inland Madinat Zayed site in the UAE; Nelli et al., 2024). In addition, AI-weather models trained on historical ERA5 show a systematic tendency to regress toward a cooler training climatology during anomalously warm periods, yielding ~0.5 to 1.5 °C cold bias in T2m (Landsberg and Barnes, 2025). Taken together, these baseline comparisons and physical considerations indicate that the muted daytime maxima and lead-dependent cool biases are intrinsic features of current AI guidance under dry-soil, clear-sky conditions rather than artifacts of a single verification dataset.

To evaluate forecast robustness to initialization time, we conducted experiments with multiple forecasts starting at different leads relative to the peak heatwave day. Fig. 12 shows the distribution of daily-maximum T2m bias at 12 UTC on 17 June 2024 relative to ERA-5 for the four AI-weather models (boxplots across the various initialization times). An analogous analysis for 00 UTC on 17 June is provided in Supplementary Fig. S5. PanguWeather has a near-zero median for most starts (|median| ≲ 0.3 °C, Fig. 12b), with a tight IQR of about 1–1.3 °C; the whiskers generally lie within −6 to +5 °C, with one cold-outlier episode near T-9 reaching roughly −10 °C. GraphCast and Aurora have comparable medians across most starts, with GraphCast slightly cooler (≈ −0.8 to −1.2 °C, Fig. 12a) and Aurora closer to zero (≈ −0.3 to −0.5 °C, Fig. 12d). Aurora's cores are slightly narrower (IQR ~1.0–1.3 °C) but it shows more extreme outliers, whereas GraphCast's cores are wider (IQR ~1.2–1.8 °C) with fewer outliers. FourCastNetv2 exhibits the largest cold shift (medians ≈ −2 to −3 °C, Fig. 12c), broader spread (IQR ~1.5–2.5 °C), and frequent cold tails to −8 to −10 °C, especially at longer lead times. The nighttime verification at 00 UTC (Fig. S5) shows the same ranking but with smaller cold shifts: PanguWeather again clusters near zero (|median| ≲ 0.2–0.3 °C; IQR ~1 °C), GraphCast remains modestly cool (≈ −0.6 to −1.0 °C), Aurora is close to zero but with more outliers than GraphCast, and FourCastNetv2 has the largest cold bias of all (≈ −1.5 to −2.5 °C) also with the broadest spread. Across both hours, the spread increases with lead time for all models, yet PanguWeather maintains the narrowest, near-zero–centred boxes over most starts, indicating the highest robustness.

PanguWeather's relative advantage is consistent with its design and training. Its 3-D Earth-Specific Transformer with an Earth-aware positional bias improves vertical–horizontal coupling and location-dependent dynamics, which benefits near-surface variables such as T2m; its hierarchical multi-step scheme (1/3/6/24 h) reduces error accumulation compared with strictly autoregressive 6-hour stepping; and its deterministic training on ERA-5 at 0.25° emphasizes medium-range skill and calibration (Bi et al., 2023). By contrast, GraphCast advances forecasts in repeated 6-h increments (Lam et al., 2023), which can lead to drift under strong land–atmosphere coupling, while FourCastNetv2's spectral/attention architecture shows larger shifts over dry, clear-sky regimes (Pathak et al., 2022). Aurora's broader, foundation-model objective trades some deterministic sharpness for generality, consistent with its slightly greater number of outliers in the boxplots (Liu et al., 2024). Taken together, Figs. 11 and 12 indicate that all four AI-weather models capture the timing and persistence of the event with high temporal coherence yet tend to under-amplify daytime extremes and display lead-dependent cold biases, a pattern expected in dry-soil, clear-sky heatwaves where land–atmosphere coupling, boundary-layer processes, and station–grid representativeness differences are most influential. Part of this lead-dependent cool bias is consistent with spectral bias, an optimization-driven preference for large spatial scales

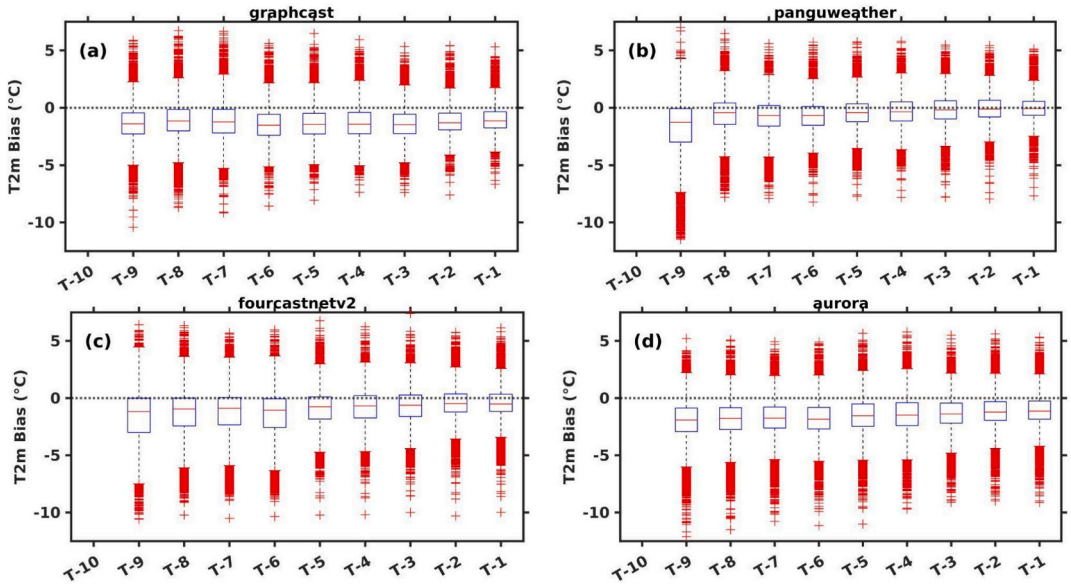

Fig. 12. Box-and-whisker diagrams of T2m bias (°C) at 12 UTC on the peak heatwave day, 17 June 2024, evaluated over the Middle East land region (10–40°N; 30–60°E). (a-d) corresponds to the four AI models. Each panel shows ten box plots representing forecasts initialized 1 to 10 days before the verification time (T-1 to T-10). The central line, box limits, and whiskers denote the median, inter-quartile range, and the 1st – 99th percentiles, respectively. Positive values indicate warm bias; negative values indicate a cold bias; narrower boxes centred near zero denote greater robustness to initialization time.

that suppresses high-wavenumber variance and smooths extremes. This manifests as a loss of small-scale spectral power in AI-weather models forecasts relative to ERA-5 (eg., Qiang et al., 2025).

For heatwave forecasting over the Middle East, we recommend PanguWeather as the primary deterministic guidance at medium range, with GraphCast used to corroborate the synoptic timing. Aurora can add useful secondary context, but its occasional outliers argue for caution in threshold-based decisions, while FourCastNetv2 should be bias-corrected before use in amplitude-sensitive heat-risk triggers.

Practical workflow and applications:A bias-corrected multi-model blend (e.g., PanguWeather with GraphCast) is preferable to any single system for operational decision-making, as reported in other studies for other applications (e.g., Liu et al., 2024). For medium-range guidance with lead times of 3 to 10 days, PanguWeather should serve as the primary deterministic source of near-surface air temperature forecasts. GraphCast should be used to independently verify synoptic timing and persistence by confirming that large-scale features, including mid-tropospheric ridges and troughs, heat-dome development, and low-level flow, arrive at the correct time, persist for the expected duration, and align with regional circulation patterns that promote heat stress. For short-range decisions with lead times of one to three days, interpolate the forecasts to station points or to small urban buffers and apply a rolling, site-specific bias correction to T2m by hour of day using the most recent 30 to 60 days of forecast–observation pairs. Use a simple bias-corrected blend, for example starting with weights of 0.6 for PanguWeather and 0.4 for GraphCast, and update these weights on a weekly or monthly basis based on verification statistics such as mean absolute error. From the bias-corrected blend, compute Heat Index using the corrected T2m and relative humidity, and evaluate heat-risk triggers against local climatological percentiles and required persistence.

This configuration supports operational products tailored to risk contexts in the Middle East. For inland areas, it enables daytime-focused alerts that inform occupational heat-exposure management for outdoor labor and short-term electricity demand forecasting for grid operators. For coastal cities, it enables nighttime and compound-heat alerts that support health services and municipal planning where humid boundary layers sustain elevated nocturnal minima and impede physiological recovery. The coastal emphasis is critical for nocturnal heat-health risk and for crowd management and mass-gathering situations, whereas the inland emphasis supports midday exposure mitigation, work and rest scheduling, and demand-response actions in the energy sector.

5. Summary

We assessed Middle East heatwaves using weather station records from December 2005 to February 2025, classifying the events as daytime, nighttime, and compound. The analysis reveals a persistent spatial structure: daytime heatwaves recur from the eastern Mediterranean through northern Iraq–Iran into Anatolia during the warm seasons, while nighttime and compound events cluster along the southern Red Sea and Arabian Gulf coasts. This pattern reflects dry, high-insolation interiors that favor daytime extremes, and humid coastal boundary layers that retain heat at night. Heatwave trends are robust: summer and autumn exhibit increases across all heatwave classes, winter increases are led by nighttime events, and spring shows a rise in compound events. These shifts are consistent with faster nighttime warming over land, strong land–atmosphere coupling in arid interiors, and coastal humidity that curtails nocturnal cooling and has increased in recent decades in light of warmer SSTs (Nesterov et al., 2021).

The high-impact compound event of 10–28 June 2024 exemplifies these behaviours. Daytime anomalies were record-breaking across the northern Middle East, while nocturnal warmth persisted along maritime margins, with peak impacts in the Mecca region around 17 June. ERA-5 fields indicate a quasi-stationary Rossby wavetrain, a reinforced ridge over southwest Asia, and an intensified, elongated Arabian Heat Low. During the peak of the event, low-level ventilation weakened as pressure gradients and northerlies relaxed, limiting advective cooling; inland skies remained largely cloud-free, enabling strong daytime radiative heating; and coastal boundary layers retained moisture and warmth overnight because warmer than average seas, with SST anomalies exceeding 4 °C in parts of the northern Arabian Gulf, and the recurrent evening onshore flow sustained humid conditions.

AI-based weather models provide actionable guidance for operational forecasting. We assessed forecasts of T2m from four state-of-the-art AI-weather models, naemely,GraphCast, PanguWeather, FourCastNetv2, and Aurora,against the ERA-5 reanalysis during the record-breaking Middle East heatwave of June 2024. All AI models reproduced the synoptic timing and multi-day persistence of the event but

consistently underestimated the daytime peaks and exhibited cold, lead-dependent near-surface biases as large as 10 °C. Among them, PanguWeather emerged as the most skillful; GraphCast offered a stable synoptic evolution; Aurora achieved a near-zero median bias but showed frequent outliers; and FourCastNetv2 produced the coldest temperatures and the widest spread. From an operational perspective, we recommend using PanguWeather as the primary medium-range temperature guide, corroborating event timing and persistence with GraphCast. Site-specific bias correction of near-surface temperatures is essential before triggering heat-risk thresholds (e.g., for heat advisories). A bias-corrected multi-model blend outperforms any single system and should be preferred for routine decision support.

The key limitation of this study is that assessment centers on a single, extreme event. Future work should (i) extend the evaluation across multiple heatwaves spanning seasons and synoptic regimes, (ii) implement dynamic, site-specific bias correction for T2m and derived heat-stress indices, and (iii) build a regional early-warning pipeline that blends bias-corrected AI guidance with continuous calibration from local observations. Together, these steps will translate recent advances in AI weather prediction into impact-ready heat-risk services for the Middle East.

CRediT authorship contribution statement

Narendra Nelli: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Diana Francis:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Ricardo Fonseca:** Writing – review & editing, Validation, Methodology, Formal analysis. **Pedram Hassanzadeh:** Writing – review & editing, Formal analysis. **Charfeddine Cherif:** Formal analysis, Data curation.

Declaration of competing interest

The authors declare they have no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank Khalifa University's high-performance computing and research computing facilities for their support of this research work. The authors gratefully acknowledge the support of Aim for Scale.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2026.108779.

References

Abida, R., Addad, Y., Francis, D., Temimi, M., Nelli, N., Fonseca, R., Nesterov, O., Bosc, E., 2022. Evaluation of the performance of the WRF model in a hyper-arid environment: a sensitivity study. Atmosphere 13, 985. https://doi.org/10.3390/atmos13060985.

Al-Shamarti, H., Rupp, P., Birner, T., 2025. Circulation Aspects Associated with Heatwave events over Iraq. Int. J. Climatol. e70017. https://doi.org/10.1002/joc.70017.

Anderson, G.B., Bell, M.L., Peng, R.D., 2011. Methods to calculate the heat index as an exposure metric in environmental health research. Environ. Health Persp. 121, 1111–1119. https://doi.org/10.1289/ehp.1206273.

Arsad, F.S., Hod, R., Ahmad, N., Ismail, R., Mohamed, N., Baharom, M., Osman, Y., Radi, M.F.M., Tangang, F., 2022. The impact of heatwaves on mortality and morbidity and the associated vulnerability factors: a systematic review. Int. J. Environ. Res. Pu. 19, 16356. https://doi.org/10.3390/ijerph192316356.

Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D.G., Salcedo-Sanz, S., 2023. Heat waves: physical understanding and scientific challenges. Rev. Geophys. 61, e2022RG000780. https://doi.org/10.1029/2022RG000780.

Bi, K., Zhang, Y., Guo, Y., Fan, L., Chen, J., et al., 2023. Accurate medium-range global weather forecasting with Pangu-Weather. Nature 619, 533–538. https://doi.org/10.1038/s41586-023-06185-3.

Bodnar, C., Bruinsma, W.P., Lucic, A., et al., 2025. A foundation model for the Earth system. Nature 641, 1180–1187. https://doi.org/10.1038/s41586-025-09005-y.

Chaidez, V., Dreano, D., Agustí, S., Duarte, C.M., Hoteit, I., 2017. Decadal trends in Red Sea maximum surface temperature. Sci. Rep. 7, 8144. https://doi.org/10.1038/s41598-017-08146-z.

Donat, M.G., et al., 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. J. Geophys. Res. Atmos. 118 (5), 2098–2118. https://doi.org/10.1002/jgrd.50150.

Dong, Z., Wang, L., Sun, Y., Hu, T., Limsakul, A., Singhruck, P., Pimonsree, S., 2021. Heatwaves in Southeast Asia and their changes in a warmer world. Earth's. Future 9, e2021EF001992. https://doi.org/10.1029/2021EF001992.

Fonseca, R., Francis, D., Nelli, N., Farrah, S., Wehbe, Y., Al Hosari, T., Al Mazroui, A., 2022a. Assessment of the WRF model as a guidance tool into cloud seeding operations in the United Arab Emirates. Earth Space Sci. 9, e2022EA002269. https://doi.org/10.1029/2022EA002269.

Fonseca, R., Francis, D., Nelli, N., Thota, M., 2022b. Climatology of the heat low and the intertropical discontinuity in the Arabian Peninsula. Int. J. Climatol. 42, 1092–1117. https://doi.org/10.1002/joc.7291.

Fonseca, R., Francis, D., Nelli, N., Yarragunta, Y., Paparella, F., Pauluis, O.M., 2025a. Summertime secondary convection and interaction with sea-breeze circulations. Q. J. Roy. Meteorol. Soc. 151, e4907. https://doi.org/10.1002/qj.4907.

Fonseca, R., Francis, D., Nelli, N., Yarragunta, Y., Paparella, F., Pauluis, O.M., 2025b. Summertime secondary convection and interaction with sea-breeze circulations. Q. J. Roy. Meteorol. Soc. 151, e4907. https://doi.org/10.1002/qj.4907.

Founda, D., Varotsos, K.V., Pierros, F., Giannakopoulos, C., 2019. Observed and projected shifts in hot extremes' season in the Eastern Mediterranean. Global Planet. Change 175, 190–200. https://doi.org/10.1016/j.gloplacha.2019.02.012.

Francis, D., Fonseca, R., 2024a. Recent and projected changes in climate patterns in the Middle East and North Africa (MENA) region. Sci. Rep. 10279. https://doi.org/10.1038/s41598-024-60976-w.

Francis, D., Fonseca, R., 2024b. Recent and projected changes in climate patterns in the Middle East and North Africa. Sci. Rep. 14, 16166. https://doi.org/10.1038/s41598-024-60976-w.

Francis, D., Fonseca, R., Nelli, N., 2022. Key factors modulating the threat of the Arabian Sea's tropical cyclones to the Gulf countries. J. Geophys. Res. Atmos. 127, e2022JD036528. https://doi.org/10.1029/2022JD036528.

Gong, H., Ma, K., Hu, Z., Dong, Z., Ma, Y., Chen, W., Wu, R., Wang, L., 2024. Attribution of the August 2022 extreme heatwave in southern China: role of dynamical and thermodynamical processes. Bull. Amer. Meteor. Soc. 105, E193–E199. https://doi.org/10.1175/BAMS-D-23-0175.1.

Guo, Y., Gasparrini, A., Armstrong, B.G., Tawatsupa, B., Tobias, A., Lavigne, E., de Sousa Zanotti Stagliorio Coelho, M, Pan, X., Kim, H., Hashizume, M., Honda, Y., Guo, Y.L., Wu, C., Zanobetti, A., Schwartz, J.D., Bell, M.L., Scortichini, M., Michelozzi, P., Punnasiri, K., Li, S., Tian, L., Osorio Garcia, S.D., Seposo, X., Overcenco, A., Zeka, A., Goodman, P., Dang, T.N., Van, D.D., Mayvaneh, F., Saldiva, P.H.N., Williams, G., Tong, S., 2017. Heat wave and mortality: A multicountry, multicommunity study. Environ. Health Persp. 125, 087006. https://doi.org/10.1289/EHP1026.

Hereher, M.E., 2020. Assessment of climate-change impacts on sea surface temperatures and sea level rise—the Arabian Gulf. Climate 8 (4), 50. https://doi.org/10.3390/cli8040050.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bachtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thepaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. Roy. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Hofstra, N., Haylock, M., New, M., Jones, P., 2009. Testing E-OBS European high-resolution gridded dataset of daily precipitation and surface temperature. J. Geophys. Res. Atmos. 114, D21101 (often cited as D11103 in some sources). https://doi.org/10.1029/2009JD011799.

Im, E.-S., Pal, J.S., Eltahir, E.A.B., 2017. Deadly heat waves projected in the densely populated regions of South Asia. Sci. Adv. 3 (8), e1603322. https://doi.org/10.1126/sciadv.1603322.

IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC. Cambridge University Press. https://doi.org/10.1017/9781009157896.

Korhonen, N., Hyvärinen, O., Kollanus, V., Lanki, T., Jokisalo, J., Kosonen, R., Richardson, D.S., Jylhä, K., 2025. The probabilistic skill of extended-range heat wave forecasts over Europe. Nat. Hazards Earth Syst. Sci. 25, 1865–1879. https://doi.org/10.5194/nhess-25-1865-2025.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., et al., 2023. Learning skillful medium-range global weather forecasting (GraphCast). Science 382, 1416–1421. https://doi.org/10.1126/science.adi2336.

Landsberg, J.B., Barnes, E.A., 2025. Forecasting the future with yesterday's climate: Temperature bias in AI weather and climate models. arXiv preprint 1–20 arXiv: 2509.22359. https://arxiv.org/abs/2509.22359.

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., et al., 2024. AIFS — ECMWF's Data-driven Forecasting System. arXiv:2406.01465. https://doi.org/10.48550/arXiv.2406.01465.

Lelieveld, J., Proestos, Y., Hadjinicolaou, P., Tanarhte, M., Tyrlis, E., Zittis, G., 2016. Strongly increasing heat extremes in the Middle East and North Africa (MENA) in the 21st century. Clim. Change 137, 245–260. https://doi.org/10.1007/s10584-016-1665-6.

Leroyer, S., Belair, S., Spacek, L., Gultepe, I., 2018. Modeling of radiation-based thermal stress indicators for urban numerical weather prediction. Urban Clim. 25, 64–81. https://doi.org/10.1016/j.uclim.2018.05.003.

Lewis, S.C., Karoly, D.J., 2013. Evaluation of historical diurnal temperature range trends in CMIP5 models. J. Climate 26 (22), 9077–9089. https://doi.org/10.1175/JCLI-D-13-00032.1.

Liu, C.-C., Hsu, K., Peng, M.S., Chen, D.-S., Chang, P.-L., Hsiao, L.-F., Fong, C.-T., Hong, J.-S., Cheng, C.-P., Lu, K.-C., Chen, C.-R., Kuo, H.-C., 2024. Evaluation of five global AI models for predicting weather in Eastern Asia and Western Pacific. NPJ Clim. Atmos. Sci. 7, 2024. https://doi.org/10.1038/s41612-024-00769-0.

Ma, K., Gong, H., Wang, L., et al., 2024. Anthropogenic forcing intensified internally driven concurrent heatwaves in August 2022 across the Northern Hemisphere. NPJ Clim Atmos Sci 7, 290. https://doi.org/10.1038/s41612-024-00828-6.

Mann, H.B., 1945. Nonparameteric tests against trend. Econ. Soc. 13, 245–259. https://doi.org/10.2307/1907187.

Miralles, D.G., Teuling, A.J., van Heerwaarden, C.C., de Arellano, J.V.-G., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7 (5), 345–349. https://doi.org/10.1038/ngeo2141.

Mueller, B., Seneviratne, S.I., 2012. Hot days induced by precipitation deficits at the global scale. Proc. Natl. Acad. Sci. 109 (31), 12398–12403. https://doi.org/10.1073/pnas.1204330109.

Nashwan, A.J., Aldosari, N., Hendy, A., 2024. Hajj 2024 heatwave: addressing health risks and safety. Lancet 404 (10451), 427–428. https://doi.org/10.1016/S0140-6736(24)01440-5.

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, 3593. https://doi.org/10.3390/rs16193593.

Nelli, N.R., Temimi, M., Fonseca, R.M., Weston, M.J., Thota, M.S., Valappil, V.K., Branch, O., Wulfmeyer, V., Wehbe, Y., Al Hosary, T., Shalaby, A., Al Shamsi, N., Al Naqbi, H., 2020. Impact of roughness length on WRF simulated land-atmospheres interactions over a hyper-arid region. Earth Space Sci. 7, e2020EA001165. https://doi.org/10.1029/2020EA001165.

Nesterov, O., Temimi, M., Fonseca, R., Nelli, N.R., Addad, Y., Bosc, E., Abida, R., 2021. Validation and statistical analysis of the Group for High Resolution Sea Surface Temperature data in the Arabian Gulf. Oceanologia 63, 497–515. https://doi.org/10.1016/j.oceano.2021.07.001.

Parajuli, S.P., Stenchikov, G.L., Ukhov, A., Morrison, H., Shevchenko, I., Mostamandi, S., 2023. Simulation of a dust-and-rain event across the Red Sea using WRF-Chem. J. Geophys. Res. Atm. 128, e2022JD038384. https://doi.org/10.1029/2022JD038384.

Pasche, O.C., Wider, J., Zhang, Z., Zscheischler, J., Engelke, S., 2025. Validating deep learning weather forecast models on recent high-impact extreme events. Artif. Intell. Earth Syst. 4, e240033. https://doi.org/10.1175/AIES-D-24-0033.1.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., et al., 2022. FourCastNet: A global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. arXiv:2202.11214. https://doi.org/10.48550/arXiv.2202.11214.

Perkins-Kirkpatrick, S., Barriopedro, D., Jha, R., et al., 2024. Extreme terrestrial heat in 2023. Nat. Rev. Earth Environ. 5, 244–246. https://doi.org/10.1038/s43017-024-00536-y.

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

Qiang, S.Y., Hassanzadeh, P., Shaw, T., Pahlavan, H.A., 2025. Predicting beyond Training Data via Extrapolation versus Translocation: AI Weather Models and Dubai's Unprecedented 2024 Rainfall. https://doi.org/10.48550/arXiv.2505.10241.

Raymond, C., Matthews, T., Horton, R.M., 2020. The emergence of heat and humidity too severe for human tolerance. Sci. Adv. 6 (19), eaaw1838. https://doi.org/10.1126/sciadv.aaw1838.

Raymond, C., Matthews, T., Tuholske, C., 2024. Evening humid-heat maxima near the southern Persian/Arabian Gulf. Commun. Earth Environ. 5, 1763. https://doi.org/10.1038/s43247-024-01763-3.

Rodwell, M.J., Hoskins, B.J., 1996. Monsoons and the dynamics of deserts. Q. J. Roy. Meteorol. Soc. 122 (534), 1385–1404. https://doi.org/10.1002/qj.49712253408.

Rodwell, M.J., Hoskins, B.J., 2001. Subtropical anticyclones and summer monsoons. J. Climate 14 (15), 3192–3211. https://doi.org/10.1175/1520-0442(2001)014<3192:SAASM>2.0.CO;2.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's Tau. J. Amer. Stat. Assoc. 63, 1379–1389. https://doi.org/10.1080/01621459.1968.10480934.

Seneviratne, S.I., et al., 2010. Investigating soil moisture–climate interactions in a changing climate: a review. Earth Sci. Rev. 99 (3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.07.004.

Steinhoff, D.F., Bruintjes, R., Hacker, J., Keller, T., Williams, C., Jensen, T., Al Mandous, A., Al Yazeedi, O.A., 2018. Influences of the monsoon trough and Arabian heat low on summer rainfall over the United Arab Emirates. Mon. Wea. Rev. 146, 1383–1403. https://doi.org/10.1175/MWR-D-17-0296.1.

Stull, R., 2011. Wet-bulb temperature from relative humidity and air temperature. J. Appl. Meteorol. Climatol. 50, 2267–2269. https://doi.org/10.1175/JAMC-D-11-0143.1.

Syed, F.S., Al-Azemi, M.A., Zamreeq, A., Islam, M.N., Ghulam, A., 2024. Observed heatwaves characteristics and variability over Saudi Arabia. Meteorol. Atmos. Phys. 136, 13. https://doi.org/10.1007/s00703-024-01010-6.

Takaya, K., Nakamura, H., 2001. A formulation of a phase-independent wave-activity flux for stationary and migratory quasigeostrophic eddies on a zonally varying basic flow. J. Atmos. Sci. 6, 608–627. https://doi.org/10.1175/1520-0469(2001)058<0608:AFOAPI>2.0.CO;2.

Temimi, M., Fonseca, R., Nelli, N., Weston, M., Thota, M., Valappil, V., Branch, O., Wizemann, H.-D., Kondapalli, N.K., Wehbe, Y., Al Hosary, T., Shalaby, A., Al Shamsi, N., Al Naqbi, H., 2020. Assessing the impact of changes in land surface conditions on WRF predictions in arid regions. J. Hydrometeorol. 21, 2829–2853. https://doi.org/10.1175/JHM-D-20-0083.1.

Theil, H., 1992. A rank-invariant method of linear and polynomial regression analysis. In: Raj, B., Koerts, J. (Eds.), Henri Theil's Contributions to Economics and Econometrics Advanced Studies in Theoretical and Applied Econometrics, 23. Spring, Dordrecht, pp. 345–351. https://doi.org/10.1007/978-94-011-2546-8_20.

Ullah, W., Alabduoli, K., Ullah, S., Al-Ghamdi, S.G., Alhebsi, K., Almazroui, M., Assiri, M. E., Azeem, W., Abuelgasim, A., Hagan, D.F.T., 2024. Comparison of 2-m surface temperature data between reanalysis and observations over the Arabian Peninsula. Atmos. Res. 311, 107725. https://doi.org/10.1016/j.atmosres.2024.107725.

Vose, R.S., Easterling, D.R., Gleason, B., 2005. Maximum and minimum temperature trends for the globe: an update through 2004. Geophys. Res. Lett. 32 (23), L23822. https://doi.org/10.1029/2005GL024379.

Wilson, A.J., Bressler, R.D., Ivanovich, C., Tuholske, C., Raymond, C., Horton, R.M., Sobel, A., Kinney, P., Cavazos, T., Shrader, J., 2024. Heat disproportionately kills young people: evidence from wet-bulb temperature in Mexico. Sci. Adv. 10, 49. https://doi.org/10.1126/sciadv.adq3367.

Wu, S., Luo, M., Zhao, R., Li, J., Sun, P., Liu, Z., Wang, X., Wang, P., Zhang, H., 2023. Local mechanisms for global daytime, nighttime, and compound heatwaves. NPJ Clim. Atmos Sci. 6, 36. https://doi.org/10.1038/s41612-023-00365-8.

Xu, P., Wang, L., Huang, P., Chen, W., 2021. Disentangling dynamical and thermodynamical contributions to the record-breaking heatwave over Central Europe in June 2019. Atmos. Res. 252, 105446. https://doi.org/10.1016/j.atmosres.2020.105446.

Yarragunta, Y., Francis, D., Fonseca, R., Nelli, N., 2025. Evaluation of the WRF-Chem performance for the air pollutants over the United Arab Emirates. Atmos. Chem. Phys. 25, 1685–1709. https://doi.org/10.5194/acp-25-1685-2025.

Yu, Y., Notaro, M., Kalashnikova, O., Garay, M., 2016. Climatology of summer Shamal wind in the Middle East. J. Geophys. Res. Atmos. 121, 289–305. https://doi.org/10.1002/2015JD024063.

Zittis, G., Hadjinicolaou, P., Almazroui, M., Bucchignani, E., Driouech, F., et al., 2021. Business-as-usual will lead to super and ultra-extreme heatwaves in the Middle East and North Africa. NPJ Clim. Atmos. Sci. 4, 20. https://doi.org/10.1038/s41612-021-00178-7.

Zittis, G., Hadjinicolaou, P., Fnais, M., Lelieveld, J., 2022. Climate change and weather extremes in the Eastern Mediterranean and Middle East. Rev. Geophys. 60, e2021RG000762. https://doi.org/10.1029/2021RG000762.