Article

A data-to-forecast machine learning system for global weather

Xiuyu Sun [1,12], Xiaohui Zhong [2,12], Xiaoze Xu [1,3,4], Yuanqing Huang[1], Hao Li [1,2] ✉, J. David Neelin [5] ✉, Deliang Chen [6,7], Jie Feng[1,8], Wei Han [4] ✉, Libo Wu [1,9,10,11] ✉ & Yuan Qi [1,2] ✉

Weather forecasting traditionally relies on numerical weather prediction (NWP) systems that integrate global observations, data assimilation (DA), and physics-based models. However, further advances are increasingly constrained by high computational costs, the underutilization of vast observational datasets, and challenges in obtaining finer resolution. Recent advances in machine learning present a promising alternative, but still depend on the initial conditions generated by NWP systems. Here, we introduce FuXi Weather, a machine learning-based global forecasting system that assimilates multi-satellite data and is capable of cycling DA and forecasting. FuXi Weather generates reliable 10-day forecasts at 0.25° resolution using fewer observations than conventional NWP systems. It demonstrates the value of background forecasts in constraining the analysis during DA. FuXi Weather outperforms the European Centre for Medium-Range Weather Forecasts high-resolution forecasts beyond day one in observation-sparse regions such as central Africa, highlighting its potential to improve forecasts where observational infrastructure is limited.

Accurate weather forecasting is essential for informed decision-making and serves as the foundation of early warning systems[1,2] that help to mitigate the impacts of extreme weather events and save lives. Since the first successful numerical weather prediction (NWP)[3] using the ENIAC computer in 1950[4], forecast accuracy has steadily improved[5], driven by advances in data assimilation (DA), spatial resolution, computational power, observational infrastructure, and physical parameterizations. However, substantial global disparities remain, with wealthier nations benefiting from better resources and more accurate forecasting[6], while many low-income countries, particularly in Africa, continue to struggle with forecasts only marginally better than climatology[7]. These disparities are especially concerning as many low-income countries are particularly vulnerable to the impacts of climate change and extreme weather[8].

Expanding observational infrastructure could help to alleviate this issue, but the financial investment required is prohibitive for many poorer nations. Additionally, the further enhancement of traditional NWP systems is increasingly challenging owing to high computational costs and the complexities of parallelizing models on modern supercomputers[9]. Meanwhile, recent advances in machine learning present a promising alternative, offering more efficient and accurate forecasts using the same initial conditions as traditional NWP[9,10]. State-

[1]Shanghai Academy of Artificial Intelligence for Science, Shanghai, China. [2]Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, China. [3]School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, China. [4]Earth System Modeling and Prediction Centre, China Meteorological Administration, Beijing, China. [5]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA. [6]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China. [7]University of Gothenburg, Gothenburg, Sweden. [8]Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai, China. [9]School of Data Science, Fudan University, Shanghai, China. [10]Institute for Big Data, Fudan University, Shanghai, China. [11]MOE Laboratory for National Development and Intelligent Governance, Fudan University, Shanghai, China. [12]These authors contributed equally: Xiuyu Sun, Xiaohui Zhong. ✉e-mail: lihao_lh@fudan.edu.cn; neelin@atmos.ucla.edu; hanwei@cma.gov.cn; wulibo@fudan.edu.cn; qiyuan@fudan.edu.cn

of-the-art machine learning models, such as Pangu-Weather, Graph-Cast, FuXi, and AIFS[11–15], have demonstrated forecasting skills that rival or even surpass traditional high-resolution forecasts (HRES) from the European Centre for Medium-Range Weather Forecasts (ECMWF)[16]. While early machine learning applications focused primarily on deterministic forecasts, recent developments have shown their potential for ensemble forecasting as well[17–19]. Nevertheless, NWP models and DA systems remain indispensable because they provide the initial conditions necessary for both traditional and machine learning forecasting models[20]: this raises the question of whether machine learning-based DA could further improve forecast accuracy.

DA is a complex, nonlinear process that incorporates vast, multi-source, and multi-resolution observational data, often plagued by noise and missing values[21], involving challenges such as distinguishing the effects of clouds on satellite radiance from those of temperature and moisture, while ensuring consistency with dynamic models to minimize error growth. Leading weather centers employ sophisticated DA methods[22,23], such as hybrid four-dimensional ensemble-variational approaches[24–26], which leverage ensembles of short-range forecasts to incorporate flow-dependent background error covariances and enhance forecast accuracy[27–29]. These methods, though effective, are computationally expensive and typically use only 5%–10%[5] of available observational data to deliver timely analyses. This limited usage is partly due to constraints related to observation error correlations. Although progress has been made in all-sky radiance assimilation for microwave sounders, challenges remain in fully leveraging satellite data across all grids, surfaces, and channels. With the volume of observational data projected to exceed 100 terabytes per day in the coming decade[30] and higher model resolutions further exacerbating computational demands[31], more efficient DA systems are urgently required[32].

The mathematical similarities between machine learning and DA, particularly in variational methods, have inspired efforts to improve DA efficiency through machine learning[33]. Early attempts focused on simplified dynamical systems, such as the Lorenz63[34,35] and Lorenz96[36,37] models, which are far less complex than NWP models. However, extending these approaches to operational NWP models is challenging owing to the markedly higher dimensionality of such models (on the order of 10^9)[38]. Recent studies have demonstrated the potential of machine learning for specific tasks within the DA work-flow, such as developing linear and adjoint models for parameterizations through automatic differentiation[39]. The rise of machine learning forecasting models[40] has reignited interest in developing fully integrated machine learning-based DA frameworks for end-to-end weather prediction.

One such attempt is FengWu-4DVar[41], which uses a simplified FengWu[42] model to assimilate ERA5 data[43]. However, its reliance on simulated observations and lower dimensionality limits its effectiveness in real-world scenarios. Aardvark Weather[44] processes raw observations for forecasts but falls short of the accuracy achieved by ECMWF HRES. These cases highlight the difficulties in developing machine learning-based DA systems for real-world forecasts using actual observational data. FuXi-DA[45], a machine learning-based DA framework, has shown promise by assimilating raw Fengyun-4B satellite data alongside background forecasts, but its limited spatial coverage constrains its global and cycling DA capabilities.

To address these challenges, we present FuXi Weather, an end-to-end machine learning-based weather forecasting system capable of running cycling DA and forecasting every 6 h using raw observations. FuXi Weather integrates a substantially enhanced version of FuXi-DA[45] with fine-tuned FuXi. Both FuXi-DA and FuXi are trained using ERA5 reanalysis data at a spatial resolution of 0.25° as the reference. Key updates to FuXi-DA include variable- and instrument-specific encoders for diverse satellite data and a modified PointPillars[46] approach for processing sparse observations. The FuXi-Short model is fine-tuned using FuXi-DA analysis for initial conditions, while a replay-based incremental learning strategy updates FuXi-DA monthly, ensuring the system's stability as satellite data quality and availability evolve.

FuXi Weather assimilates raw brightness temperature data from three polar-orbiting meteorological satellites (FengYun-3E (FY-3E), Meteorological Operational Polar Satellite-C (Metop-C), and National Oceanic and Atmospheric Administration (NOAA)-20), along with the radio occultation (RO) data from the Global Navigation Satellite System (GNSS), across all grids, surfaces, and channels under all weather conditions. This represents the first realization of all-grid, all-surface, all-channel, and all-sky DA capability. FuXi Weather demonstrates comparable 10-day forecast performance to that of ECMWF HRES, extending the skillful lead time for key variables while using considerably less observational data compared with that used by ECMWF HRES. Furthermore, FuXi Weather consistently outperforms ECMWF HRES in regions with sparse land-based observations, such as Africa, demonstrating its potential to provide more accurate forecasts and enhance climate resilience. To the best of our knowledge, FuXi Weather is the first system to successfully perform cycling DA and weather forecasting over a continuous 1-year testing period[47]. This achievement challenges the prevailing view that standalone machine learning-based weather forecasting systems are not viable for operational use.

Results

FuXi Weather operates in a cycling analysis and forecasting mode, utilizing the full range of available satellite data. Because DA is inherently an ill-posed problem[38,48] requiring background forecasts to improve analysis accuracy, we developed a variant of FuXi-DA without these forecasts to evaluate their contribution to the DA process. This variant, which relies exclusively on observations, represents a direct-from-observation prediction approach. Performance was assessed by comparing the accuracy of analysis fields and forecasts globally and in specific regions such as central Africa and northern South America, using ERA5 as the reference. The performance of FuXi Weather was compared with that of ECMWF HRES, which was evaluated using the time series of its 0-h lead time analysis, HRES-fc0 (see Section "Evaluation method"). This comparison inherently favors HRES at early lead times, since by definition it starts with a low root mean square error (RMSE) and a high anomaly correlation coefficient (ACC). Consistent with the common practices in the NWP community, FuXi Weather was also evaluated against its analyses. Statistical significance testing was conducted following the methodology outlined by Geer[49]. Single observation tests validated DA responses against theoretical expectations while data denial experiments (see Supplementary Information Section 5) evaluated the impact of excluding certain observations.

Global analysis fields

This subsection evaluates the performance of FuXi Weather analyses and 42-h FuXi forecasts (initialized with ERA5), against ERA5 as the reference. Figure 1 presents the globally-averaged and latitude-weighted RMSE for two FuXi Weather configurations: one incorporating background forecasts and one without. Performance varied markedly across different variables and pressure levels. The RMSE of analysis fields relative to forecasts is higher at 850 hPa than 300 and 500 hPa, likely owing to the lower information content from satellite observations at lower altitudes.

For relative humidity (R), the analyses of FuXi Weather outperform forecasts at 300 and 500 hPa, but have slightly higher RMSE values at 850 hPa. For temperature (T), geopotential (Z), and wind components (U and V), the RMSE values are comparable to those of forecasts at higher altitudes but were consistently higher at 850 hPa. Although satellite data primarily capture temperature and moisture information, their assimilation also improves wind fields through the
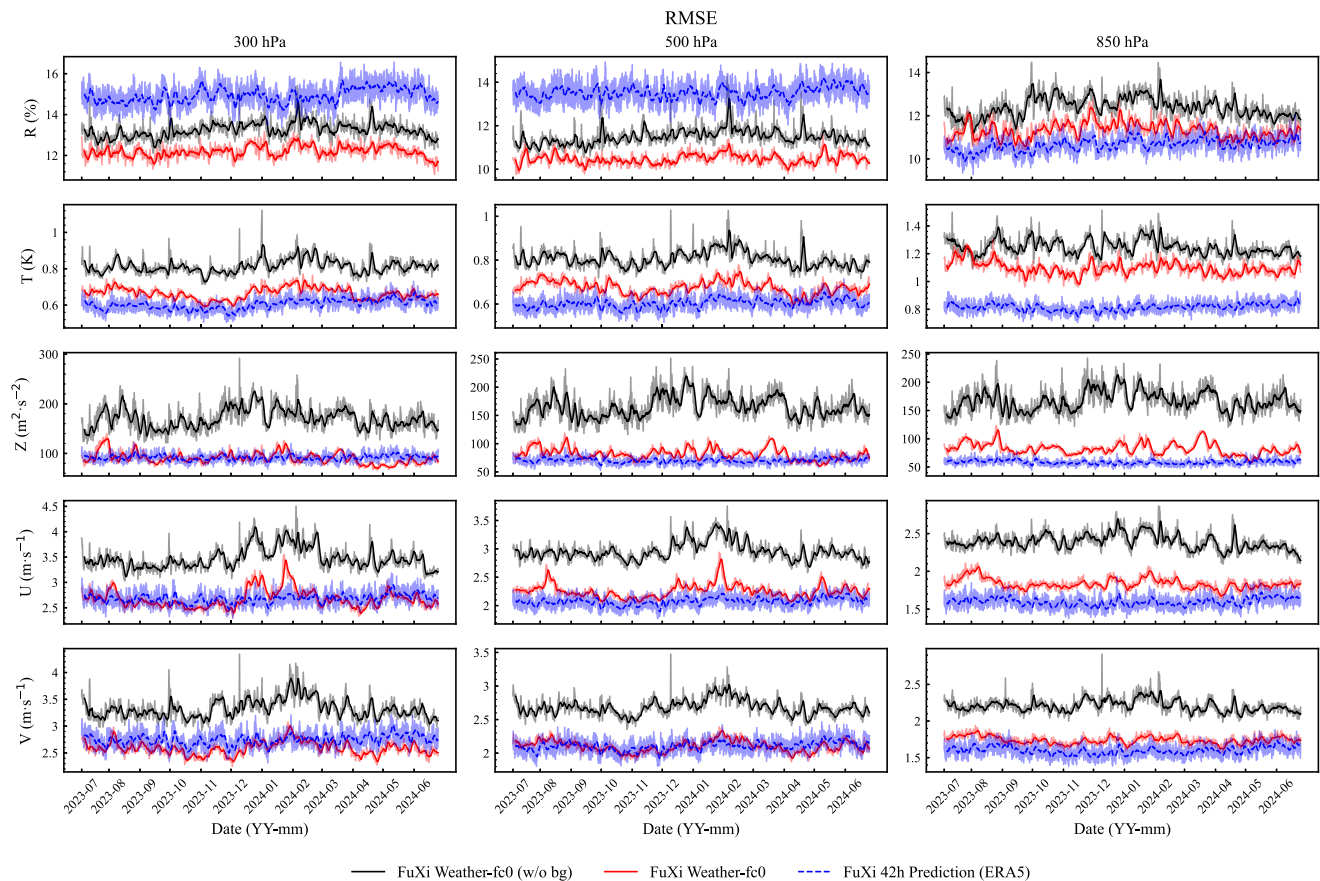
Fig. 1 | Comparison of analysis fields produced by FuXi Weather and 42-h FuXi forecasts over a 1-year testing period from July 03, 2023 to June 30, 2024. The time series shows the globally-averaged and latitude-weighted root mean square error (RMSE) relative to ERA5 for: the analysis fields of FuXi Weather with (solid red lines) and without (solid black lines) background (bg) forecasts, along with 42-h FuXi forecasts initialized using ERA5 (dashed blue lines). The comparison includes five variables: relative humidity (R), temperature (T), geopotential (Z), u component of wind (U), and v component of wind (V), at three pressure levels (300, 500, and 850 hPa). The five rows and three columns correspond to five variables and three pressure levels, respectively. To improve clarity, the original data are shown with reduced opacity, while solid lines represent smoothed values using a 12-point moving average. Both FuXi Weather analyses (black and red) and 42-h FuXi forecasts (blue) are evaluated against ERA5.

dynamic relationship between wind, temperature, and moisture. Wind can be inferred from temperature gradients (geostrophic balance) and the movement of atmospheric constituents, such as humidity, known as the "generalized tracer effect"[23].

Incorporating background forecasts yields statistically significant improvements in the accuracy of FuXi Weather analysis fields, as demonstrated by systematically lower RMSE values. This highlights the crucial role of background forecasts in DA, which is ill-posed without prior information (as detailed in Supplementary Information Section 9). Both configurations of FuXi Weather show similar trends over time, but the analyses without background forecasts exhibit more pronounced error peaks, especially when some satellite data were missing (see Supplementary Figs. 1 and 2), underscoring the stabilizing effect of background forecasts.

The shaded area in Fig. 1 represents variations across initialization times; this is more pronounced in forecasts. Forecasts initialized at 00/12 UTC consistently outperform those at 06/18 UTC, likely because the 12-h observation windows of ERA5 (09-21 UTC and 21-09 UTC)[43] provide 9 h of look-ahead time for 00/12 UTC but only 3 h for 06/18 UTC[13]. In contrast, the analysis fields of FuXi Weather demonstrate more consistent accuracy across initialization times, likely due to its fixed 8-h assimilation window, and its use of cycled background fields initialized from previous analyses. Additional evaluations, including the analysis activity and mean bias error (MBE), are provided in Supplementary Information Section 6.

Global weather forecasts

The primary criterion for evaluating an end-to-end weather forecasting system is its ability to provide reliable and accurate forecasts in a cycling analysis and forecasting mode. This subsection evaluates the performance of 6-h cycle forecasts generated by FuXi Weather, initialized using two types of FuXi-DA analysis fields: one incorporating background forecasts and one without. The forecasts are compared with those from ECMWF HRES.

Figure 2 shows the globally-averaged and latitude-weighted RMSE as a function of forecast lead times over 10 days. FuXi Weather forecasts are initialized using FuXi-DA analysis fields either with (red solid and green dashed lines) or without (black lines) background forecasts. Forecasts depicted by red and black lines are evaluated against ERA5, while the green dashed lines represent forecasts assessed against the FuXi-DA analyses. Statistically significant improvements in FuXi Weather forecasts (red lines) over ECMWF HRES are indicated by red dots, based on t-test at the 95% confidence level. When validated against ERA5, FuXi Weather forecasts initialized with background-inclusive analyses (red lines) consistently demonstrate lower RMSE values than those without, aligning with results in Fig. 1. Regardless of the evaluation reference (ERA5 or FuXi-DA analyses), the performance gap between forecasts (red and green dashed lines) diminishes over lead time and becomes negligible by day 10.

When evaluated against their respective analyses, both FuXi Weather and ECMWF HRES show small initial errors. Against ERA5,
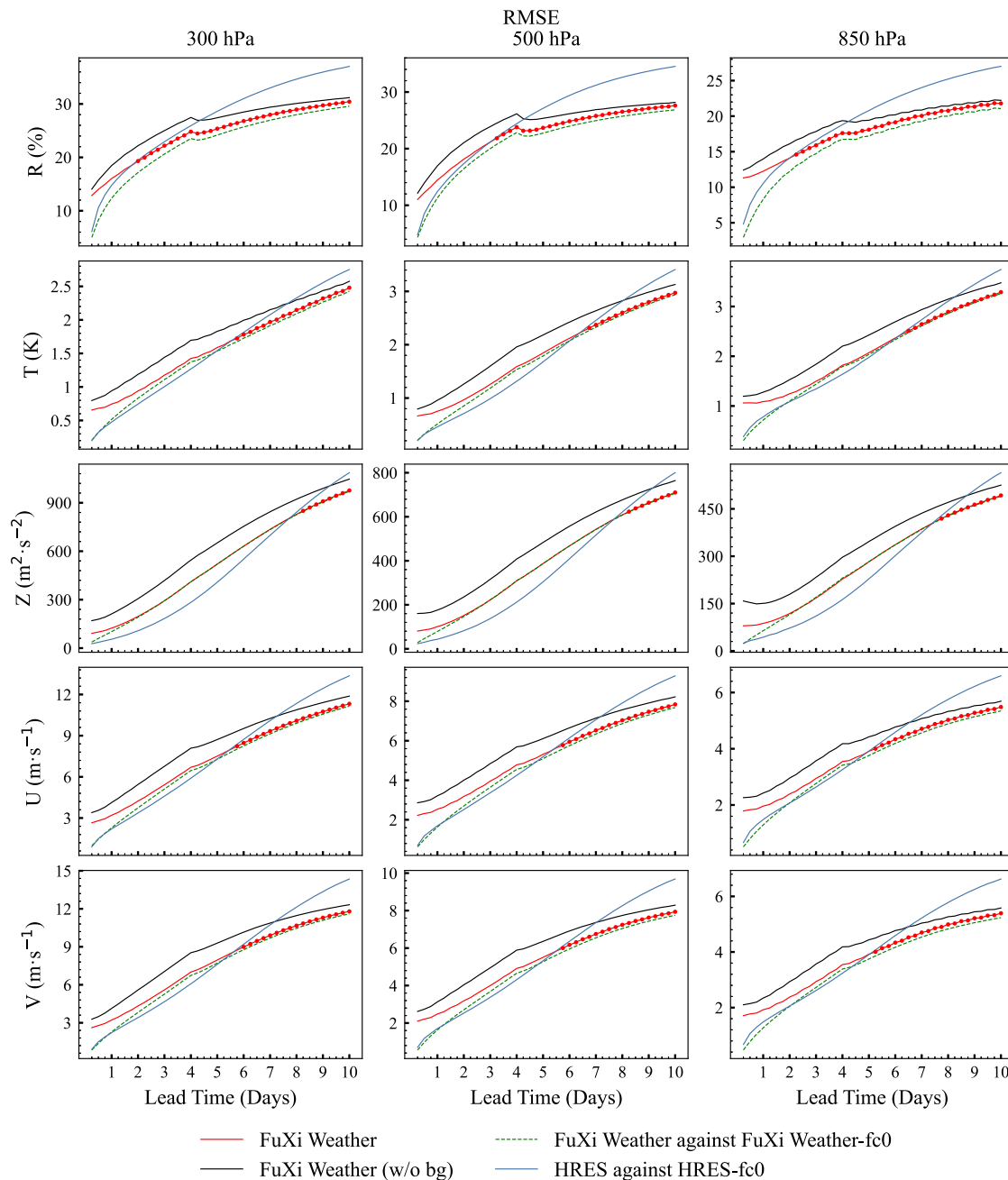
Fig. 2 | Comparison of 10-day forecast performance across models over a 1-year testing period from July 03, 2023, to June 30, 2024. The figure presents the globally-averaged and latitude-weighted root mean square error (RMSE) for forecasts generated by the FuXi model and ECMWF HRES (blue) in 10-day forecasts. FuXi forecasts are initialized using analysis fields produced by FuXi-DA with (red solid and green dashed lines) and without (black) background forecasts. The evaluation includes 5 variables: relative humidity (R), temperature (T), geopotential (Z), u component of wind (U), and v component of wind (V), at three pressure levels (300, 500, and 850 hPa). The five rows and three columns correspond to five variables and three pressure levels, respectively. FuXi forecasts (red and black lines) are verified against ERA5, and also against FuXi-DA analyses (green dashed lines). When FuXi (green dashed lines) and ECMWF HRES (blue) forecasts are evaluated against their respective initialization time series, they inherently exhibit lower RMSE at early lead times. Red dots indicate time steps where FuXi Weather significantly outperforms ECMWF HRES, based on the *t*-test at the 95% confidence level. The performance change on day 4 arises from the model transition from FuXi-Short to FuXi-Medium.

FuXi Weather initially shows higher RMSE values than ECMWF HRES, but outperforms ECMWF HRES after a lead time of 2–8 days, depending on the variable and pressure level. For R, FuXi Weather outperforms ECMWF HRES at lead times of 2.00, 3.25, and 2.25 days for 300, 500, and 850 hPa, respectively. For T, Z, U, and V, the critical lead times are later owing to the lower accuracy of their corresponding analysis fields. For Z, these times are 8.00, 7.75, and 7.50 days at 300, 500, and 850 hPa, respectively. The performance discontinuity on day 4 reflects the transition between FuXi-Short and FuXi-Medium forecast components.

Figure 3 shows similar trends for the globally-averaged and latitude-weighted ACC. FuXi Weather forecasts initialized without background forecasts perform worse, as expected. However, FuXi Weather forecasts initialized with analyses incorporating background forecasts, though initially less accurate than ECMWF HRES, improve over time and eventually achieve higher ACC values across all
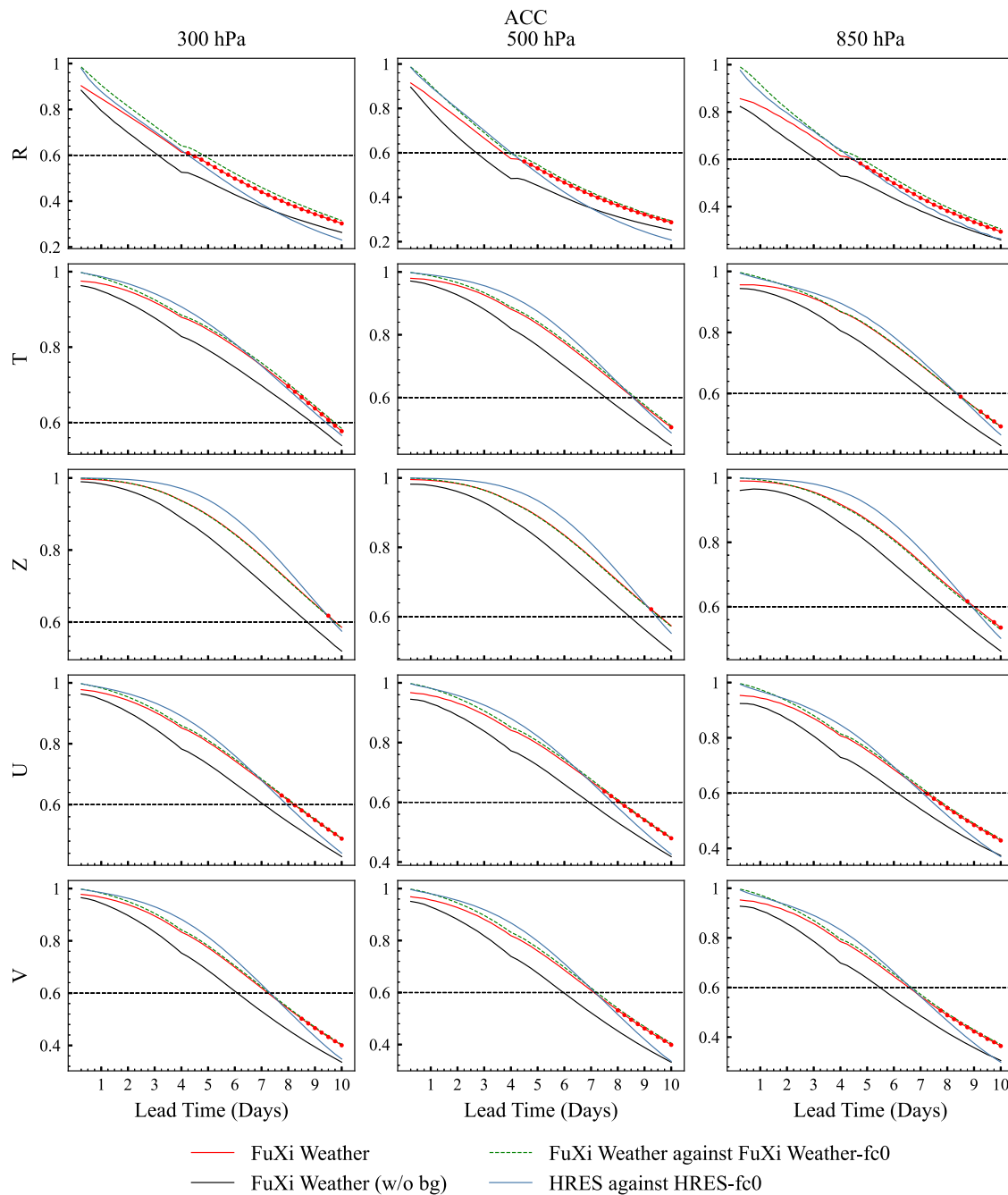
Fig. 3 | Comparison of 10-day forecast performance across models over a 1-year testing period, spanning July 03, 2023–June 30, 2024. The figure presents the globally-averaged and latitude-weighted anomaly correlation coefficient (ACC) for forecasts generated by the FuXi model and ECMWF HRES in 10-day forecasts. FuXi forecasts are initialized using analysis fields produced by FuXi-DA with (red solid and green dashed lines) and without (black) background (bg) forecasts. The analysis includes five variables: relative humidity (RH), temperature (T), geopotential (Z), u component of wind (U), and v component of wind (V), at three pressure levels (300, 500, and 850 hPa). FuXi forecasts (red and black lines) are verified against ERA5, and also against FuXi-DA analyses. The five rows and three columns correspond to five variables and three pressure levels, respectively. When FuXi (green dashed lines) and ECMWF HRES (blue) forecasts are evaluated against their respective initialization time series, they inherently exhibit higher ACC in early lead times. Red dots indicate time steps where FuXi Weather significantly outperforms ECMWF HRES, based on the t-test at the 95% confidence level. The performance change on day 4 arises from the model transition from FuXi-Short to FuXi-Medium.

examined variables. Using an ACC threshold of 0.6 to define a skillful forecast, Fig. 4 compares skillful lead times. FuXi Weather extends skillful lead times for 7 out of 15 variables, matching ECMWF HRES for 6 others. For example, for Z500, FuXi Weather extends the skillful lead time from the ECMWF HRES value of 9.25 days to 9.50 days for forecasts initialized with background forecasts (forecasts initialized without background forecasts show a skillful lead time of only 8.25 days). Additional forecast comparisons, including spatial RMSE distributions, are provided in the Supplementary Information Section 7.

Forecast performance in central Africa

Operational evaluations of NWP systems routinely assess both global and regional performance metrics[16], covering geographical areas such as Europe, North America, East Asia, and Australia. However, forecast accuracy tends to be lower in low-income countries, largely due to limited investment in weather observation infrastructure. This issue is especially concerning for many low-income countries, where agriculture is a major economic sector that relies heavily on accurate weather forecasts. Climate change further exacerbates weather-
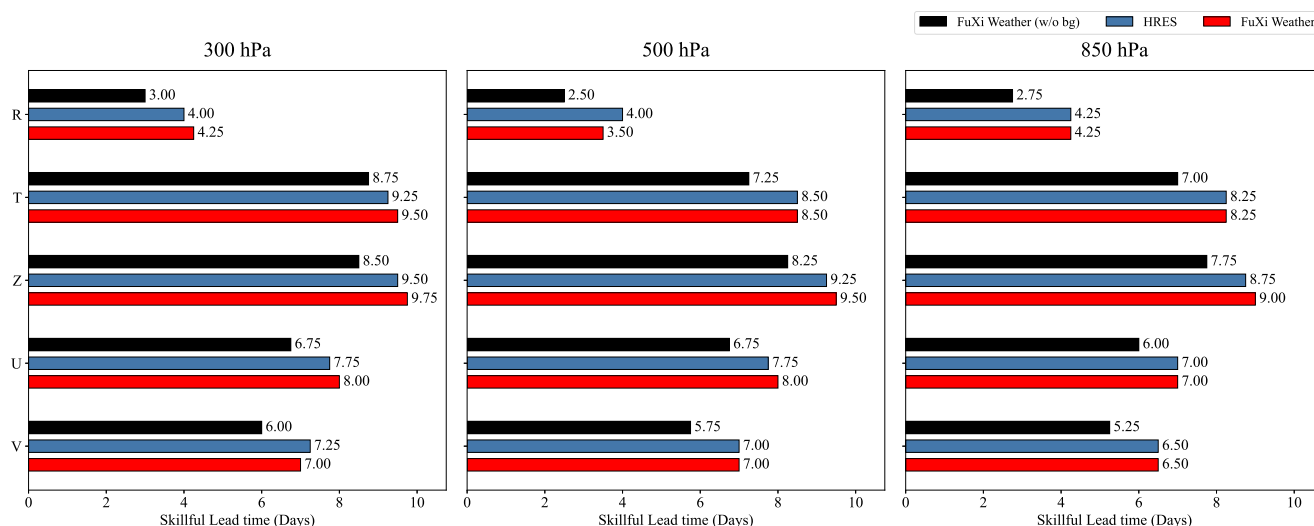
Fig. 4 | Skillful forecast lead time comparisons with an anomaly correlation coefficient (ACC) value of 0.6 as the threshold. Skillful forecast lead times of ECMWF HRES and FuXi Weather for five variables: relative humidity (R), temperature (T), geopotential (Z), u component of wind (U), and v component of wind (V), at three pressure levels (300, 500, and 850 hPa), using all testing data over a 1-year testing period, spanning July 03, 2023–June 30, 2024. The five rows and three columns correspond to five variables and three pressure levels, respectively.
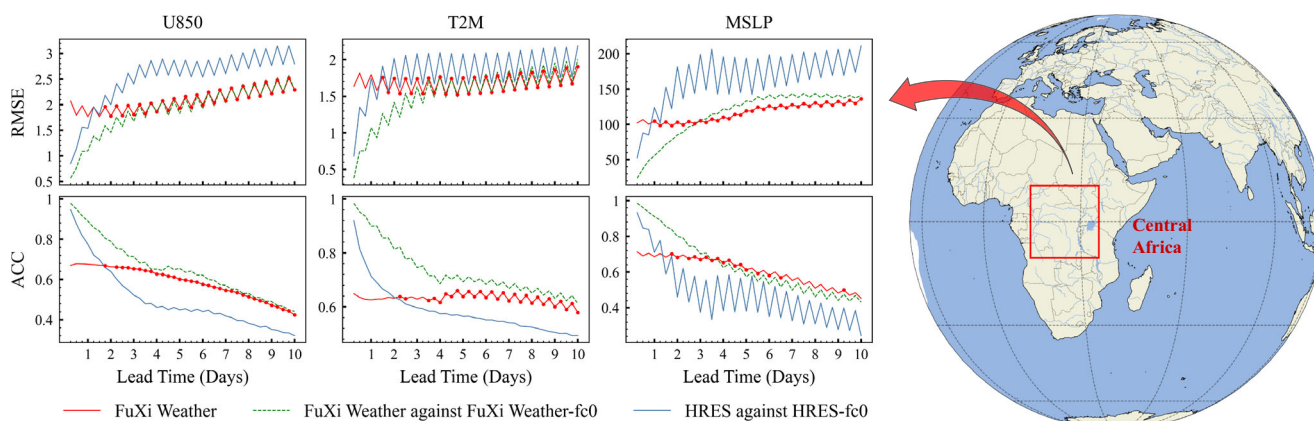

Fig. 5 | Comparison of forecast performance over central Africa during the 1-year testing period from July 03, 2023 to June 30, 2024. Central Africa is defined as the region spanning 15° E to 35° E in longitude and 10° N to 10° S in latitude. Rows 1 and 2 show the root mean square error (RMSE), and anomaly correlation coefficient (ACC) for forecasts generated by FuXi Weather (red solid and green dashed lines) and ECMWF HRES (blue). FuXi Weather is initialized using analysis fields produced by FuXi-DA incorporating background forecasts. This figure includes three variables: 850 hPa u wind component (U850), 2-meter temperature (T2M), and mean sea level pressure (MSLP). FuXi forecasts (red) are verified against ERA5, and also against FuXi-DA analyses (green dashed lines). When FuXi (green dased lines) and ECMWF HRES (blue) forecasts are evaluated against their respective initialization time series, they inherently exhibit lower RMSE and higher ACC in early lead times. Red dots indicate time steps where FuXi Weather significantly outperforms ECMWF HRES, as paired difference passed the 95%-confidence-level the *t*-test of significance.

related risks, disproportionately affecting vulnerable populations with low adaptive capacities in these countries. Therefore, improving forecast accuracy in underserved regions, especially Africa, is crucial for enhancing climate resilience[50,51].

This subsection compares the performance of FuXi Weather and ECMWF HRES in underserved regions, with a particular focus on central Africa. Similar to Fig. 2, FuXi Weather forecasts are evaluated against both ERA5 (red lines) and its analyses (green dashed lines). Figure 5 illustrates that, when verified against their respective analyses, FuXi Weather (green dashed lines) consistently outperforms ECMWF HRES (blue lines) in forecasting the 850 hPa u wind component (U850), 2-meter temperature (T2M), and mean sea level pressure (MSLP), throughout the 10-day forecast period. When evaluated against ERA5, FuXi Weather (red lines) has nontrivial initial error, but the magnitude and growth of this error are sufficiently modest that ECMWF HRES−

even compared to its own analyses, so with inherently zero initial error −exhibits larger error after two days. In particular, FuXi Weather (red lines) achieves lower RMSE and higher ACC, with ACC values for T2M consistently exceeding 0.6 across the 10-day forecasts, indicating meaningful predictive skill. In contrast, ECMWF HRES maintains skillful T2M forecasts for approximately two days.

Forecast errors are further decomposed into systematic and random components by calculating the MBE and the standard deviation (std) of errors (STD$_{ERROR}$). Supplementary Fig. 22 reveals that FuXi Weather (red lines) exhibits both lower MBE and smaller STD$_{ERROR}$ across all five evaluated variables: U850, 850 hPa temperature (T850), T2M, MSLP, and total precipitation (TP). These results suggest that FuXi Weather more effectively reduces both systematic bias and random errors compared to ECMWF HRES, contributing to its overall superior forecast performance. Improvements relative to HRES in TP

forecasts are of note due to precipitation's socioeconomic importance in central Africa, although with the caveat that HRES performance is relatively poor for TP in this region. Forecast behavior is further characterized using forecast activity[40], defined as the std of forecast anomalies relative to climatological means and normalized by ECMWF HRES forecast activity. As shown in Supplementary Fig. 22, FuXi Weather normalized forecast activity values indeed drop below 1, suggesting smoother predictions relative to ECMWF HRES. This reduction in forecast activity may partially account for FuXi Weather's improved performance. However, FuXi Weather's superior forecast skill (red lines) over ECMWF HRES becomes evident as early as day 1, prior to any considerable reduction in forecast activity. The forecast activity of FuXi Weather decreases gradually until around day 2 and then stabilizes, indicating that FuXi Weather's enhanced accuracy arises earlier than the substantial reduction in forecast activity and cannot be fully attributed to it.

Notably, FuXi Weather achieves superior forecasts for surface variables without assimilating surface-based observations, pointing to its strength in utilizing satellite data in regions with limited in-situ observational infrastructure. Further analysis (see Supplementary Information Section 7) reveals that FuXi Weather also outperformed ECMWF HRES in other data-sparse regions, such as tropical oceans and South America, although it is less competitive in areas with dense surface observations. In central Africa, where observational networks are sparse, the efficient use of satellite data by FuXi Weather closes the performance gap with ECMWF HRES, resulting in superior forecasts.

Supplementary Fig. 23 illustrates two 10-day forecast time series for two randomly selected initialization times, while Supplementary Fig. 24 presents forecasts at a fixed 3-day lead time. Both figures confirm that FuXi Weather more closely aligns with its benchmark than ECMWF HRES, reinforcing the results in Fig. 5. Additionally, Supplementary Fig. 25 shows FuXi Weather's superior performance, particularly for T2M, MSLP, and TP over northern South America, where observational coverage is also sparse relative to Europe or North America. However, the reduction in forecast activity may partially contribute to these improvements. A detailed discussion on the trade-offs between forecast accuracy and activity is provided in Supplementary Information Section 12. While incorporating generative models or differentiable solvers for atmospheric dynamics could potentially enhance forecast activity without compromising accuracy[18,52,53], an in-depth investigation of these approaches is beyond the scope of this study.

Due to substantial biases in TP data from ERA5[54], the Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) (IMERG)[55,56] is used to evaluate TP forecasts over central Africa and northern South America, respectively. As shown in Supplementary Fig. 26, FuXi Weather achieves lower RMSE than ECMWF HRES, relative to IMERG. However, both FuXi Weather and ECMWF HRES exhibit undesirably low ACC and substantial MBE when evaluated against IMERG. In FuXi Weather, this deficiency is likely inherited from its training with ERA5, underscoring the potential advantages of training with more accurate observational datasets, such as IMERG, to further improve FuXi Weather's precipitation forecasts.

Overall, these preliminary results suggest that FuXi Weather can produce forecasts of comparable or potentially improved accuracy relative to traditional NWP systems, despite relying on substantially fewer observations. The superior performance of FuXi Weather relative to ECMWF HRES may be attributed to two primary factors: (1) enhanced ability to mitigate both systematic biases and random errors, and (2) reduced forecast activity. While further advancements, such as improving forecast activity, are necessary, FuXi Weather represents a promising and cost-effective alternative for regions with limited observational infrastructure. Future work will include further validation against independent observational datasets to better evaluate its performance advantages.

Physical consistency of analysis changes

FuXi Weather, as a data-driven machine learning system, does not inherently encode prior physical knowledge of atmospheric processes. This subsection examines the impact of assimilating a single observation on background fields and assesses whether the resulting changes align with theoretical expectations.

Two FuXi-DA runs were conducted: the first using a 6-h forecast with original observations, and the second with a perturbation introduced to raw satellite data from individual channels at a specific observation location. The differences between these two runs reflected the changes in analysis fields caused by the perturbation (details in Supplementary Information Section 4.1). The first run, initialized at 06 UTC on July 24, 2023, assimilated all available data to generate the analysis. In the second run, a +5 K perturbation was introduced into the NOAA-20 ATMS raw observation at 19.9° N, 125.5° E (marked as a purple dot in Supplementary Fig. 9), near Typhoon Doksuri over the ocean. The impact of this perturbation was evaluated by comparing outputs from both runs. The satellite observations were independently perturbed for each channel.

Figure 6 shows the horizontal and vertical distributions of changes in the analysis fields resulting from three separate perturbations, each applied to a different humidity channel. The spatial patterns of these changes in analysis fields aligned with the radiative transfer theory: an increase in brightness temperature corresponds to a decrease in humidity, resulting in less radiation absorption[57]. The vertical distribution showed progressive increases in the peak heights of the Jacobian functions for channels 18, 19, and 20, matched by corresponding increases in the peak heights of the humidity increments. This pattern suggests that the DA system effectively captures the varying detection altitudes of these channels. Additionally, flow-dependent characteristics were observed in the humidity field. The perturbation introduced at 05 UTC, 1 h before the analysis, generated changes in analysis fields mainly localized near the perturbation location, with a moderate eastward extension along the prevailing flow, consistent with downwind propagation. Supplementary Fig. 10 illustrates the changes in wind vector analysis fields, overlaid with relative humidity analysis fields. The perturbation results in increased northerly flow near the perturbed location. This change enhances the advection of drier air, characterized by lower relative humidity, into a more humid region. Consequently, the perturbation leads to a localized reduction in relative humidity, consistent with the results shown in Fig. 6.

In summary, FuXi Weather effectively captures the horizontal and vertical dependencies of analysis changes on satellite observations without explicitly incorporating prior knowledge. Data denial experiments (Supplementary Information Section 5) further confirm FuXi Weather's physical consistency with satellite observations, while additional tests demonstrate the robustness of its performance.

Discussion

In this paper, we introduce FuXi Weather, an end-to-end machine learning-based weather forecasting system that performs global-scale DA and forecasting on a 6-h cycle through processing raw satellite observations across all grids, surfaces, channels, and sky conditions. The system matches the global forecasting performance of state-of-the-art ECMWF HRES and outperforms it in observation-sparse regions such as central Africa and northern South America. Moreover, FuXi Weather extends the skillful forecast lead time achieved by ECMWF HRES in many regions, despite using considerably fewer observations. Single observation tests confirm that DA responses align with theoretical expectations, and data denial experiments demonstrate the system's robustness, with only moderate error growth when specific observations are excluded. Notably, FuXi Weather performs continuous cycling DA and weather forecasting over a full one-year testing period. Due to its computational efficiency and reduced complexity
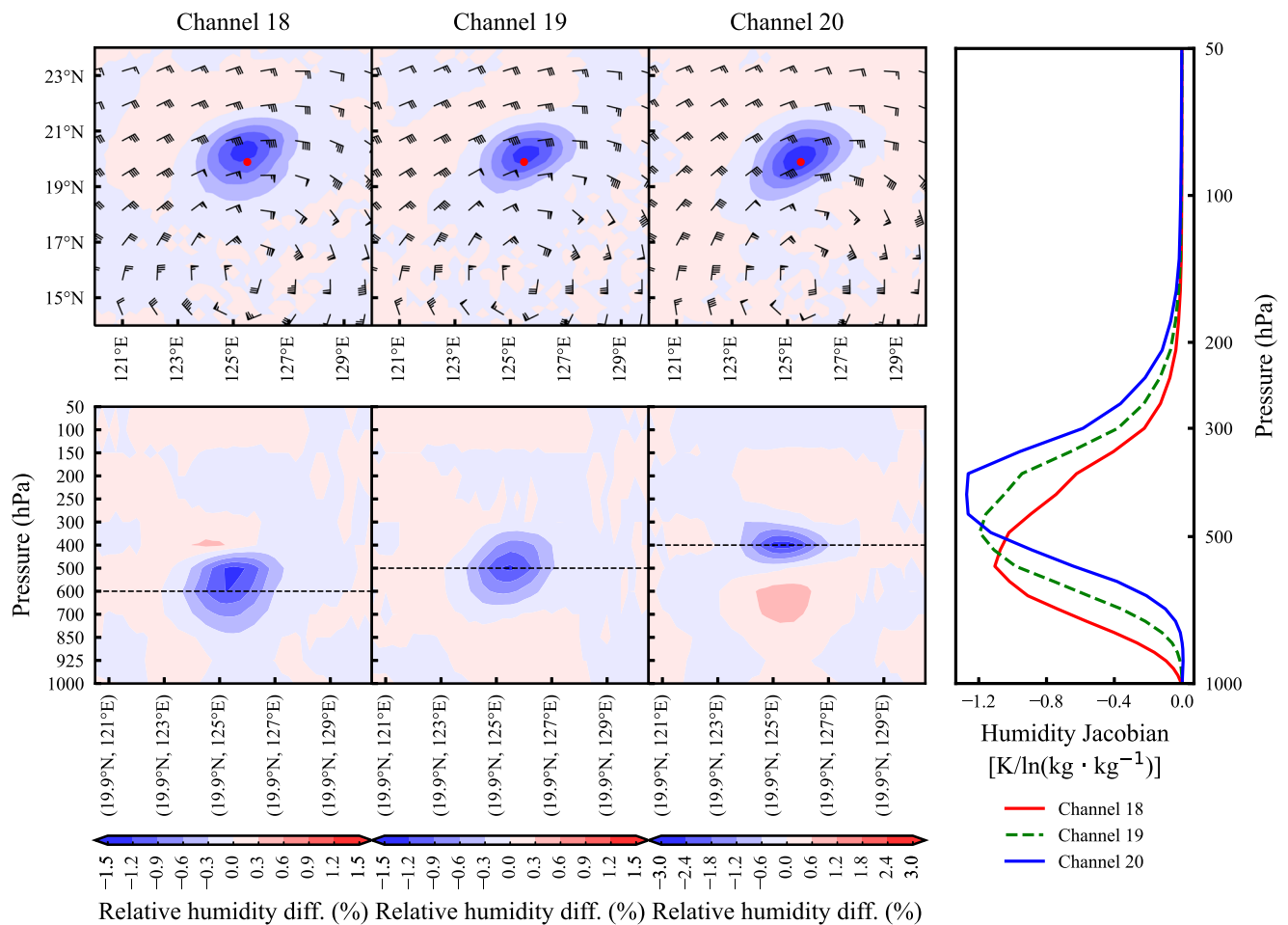
Fig. 6 | Changes in analysis fields resulting from a 5 K perturbation to the NOAA-20 ATMS observation at a selected location, based on the background field for 06 UTC on July 24, 2023. The perturbation, located over the ocean near Typhoon Doksuri at 19.9° N, and 125.5° E (red dot), is introduced at 05 UTC, 1 h before the analysis time. The two rows show, in the left panel, the horizontal spatial distribution of the analysis changes for channels 18–20 at 600, 500, and 400 hPa, with wind fields overlaid, as well as the corresponding vertical distribution along the same west-east cross-section. The dashed lines on the second row indicate the pressure levels for the horizontal spatial distribution. The right panel shows the Jacobian functions for three humidity channels derived from ATMS aboard NOAA-20. The atmospheric profile is based on the US Standard Atmosphere, and radiative transfer calculations are performed using RTTOV version 13.2. In the wind vector plots, a long barb represents 4 m/s, a short barb 2 m/s, and a pennant indicates 20 m/s.

compared to traditional NWP systems, FuXi Weather offers a cost-effective alternative for improving operational forecasts in regions with limited land-based observations, thus enhancing climate resilience.

Despite these promising results, several challenges remain. While FuXi Weather extends lead times for multiple variables, its short-term forecast accuracy requires further improvement. This limitation is likely due to its reliance on a limited subset of satellite observations, whereas the ECMWF system assimilates observations from approximately 90 satellite instruments operationally[58]. Furthermore, FuXi Weather learns the relationship between satellite observations and background forecasts entirely in latent space, without relying on traditional DA components such as observation operators, adjoint models, or explicit estimation of observation and background error covariance matrices. This design dramatically simplifies model development and reduces computational demands and domain-specific expertise requirements. However, extending FuXi Weather to integrate conventional observations, such as radiosonde soundings, and surface, marine, and radar measurements, remains challenging due to their spatial and temporal sparsity, inhomogeneity, and varying quality. To address these challenges, tailored preprocessing pipelines and observation-specific quality control algorithms[59] must be developed to identify and remove outliers. In addition, like many machine learning

weather forecasting models, FuXi Weather exhibits reduced forecast activity, which partially accounts for its improved forecast skill. The system outperforms ECMWF HRES at longer forecast lead times, particularly where its forecasts become smoother. Potential solutions include integrating generative models or enforcing physical constraints to better capture atmospheric variability. Incorporating ensemble-based[19,60] DA methods offers further potential to enhance model performance. By lowering technical barriers, such as eliminating reliance on legacy Fortran-based NWP infrastructures[61,62], systems like FuXi Weather, could pave the way for closer interdisciplinary collaboration between meteorologists and machine learning scientists.

FuXi Weather, built upon the foundation of traditional NWP systems and ERA5 reanalysis, developed over several decades, inherently inherits both their strengths and limitations[63]. Although ERA5 provides a consistent, high-quality dataset, this dependency may cap the ultimate performance gains achievable by machine learning approaches. For instance, documented discrepancies between ERA5 precipitation data and observations[54] suggest that ERA5 precipitation may not be the most appropriate target for training precipitation forecasts in FuXi Weather. Instead, more accurate observational datasets, such as IMERG precipitation, could be used as reference data to enhance the model's predictive skill. Currently, FuXi Weather retains an explicit DA step and forecasts meteorological variables rather than raw

observations (e.g., brightness temperature), enabling rigorous evaluation against reanalysis and direct comparisons with ECMWF HRES. Recent advances in Artificial Intelligence Direct Observation Prediction (AI-DOP) frameworks have demonstrated the feasibility of bypassing explicit DA entirely[59,64]. Unlike NWP systems, which require initial conditions to solve partial differential equations, machine learning models can generate forecasts directly from observations. However, the success of AI-DOP demands two prerequisites: (1) sufficient spatiotemporal observational coverage and (2) long-term and high-quality historical records. For instance, ECMWF's AI-DOP model is trained on 18 years of observational data (2004–2021) encompassing primary observation categories used in NWP systems[59], whereas FuXi Weather has thus far leveraged only one year. Explicit DA approaches benefit from pretrained forecasting models (often trained using decades of ERA5), enhancing temporal consistency when observational data are limited. In contrast, implicit DA requires substantially more data to learn these relationships from scratch and resolve inconsistencies in historical observational datasets. With sufficiently extensive and high-quality observational datasets, we expect the performance gap between explicit and implicit approaches to close. Future iterations of FuXi Weather may eliminate the explicit DA step by learning to forecast directly from sequences of past and present observations, reducing dependence on reanalysis data and advancing toward a fully independent and robust forecasting system.

As the volume of assimilated observations grows, scaling FuXi Weather to accommodate larger models and datasets will be essential. Optimal hybrid parallelization strategies[65–67] that combine pipeline parallelism and data parallelism, could enable efficient training with increased observations. The flexible, multi-branch architecture of FuXi Weather supports scalable implementation for additional observational data. Data denial experiments also suggest that selectively excluding less informative satellite data could improve efficiency without compromising accuracy.

Methods
FuXi Weather
Figure 7 illustrates FuXi Weather, which generates global weather forecasts every 6 h. It has three main components: satellite data preprocessing (detailed in Supplementary Information Section 2.1), DA via FuXi-DA, and forecasting using the FuXi model. A complete list of variables and abbreviations is provided in Table 1.

The preprocessing step addresses the heterogeneity in satellite data across space and time (see Fig. 8). While FuXi Weather can directly process raw observational data, the data are interpolated to a regular 0.25° grid using nearest-neighbor interpolation for simplicity. This approach enhances the system's scalability and ensures consistent integration across diverse observation types. This study utilized brightness temperature from five microwave instruments aboard three polar-orbiting satellites (FY-3E, Metop-C, and NOAA-20) and GNSS-RO data[68] (see Supplementary Table 1), processed using a modified PointPillars[46] approach initially designed for three-dimensional point clouds[69]. Missing data are handled using a masking technique, assigning a value of 1 where data are available and 0 otherwise. Further details are provided in Supplementary Information Sections 1.2 and 2.1.

FuXi-DA assimilates the preprocessed data with background forecasts within 8 h to produce analysis fields. Key improvements include separate processing of different upper-air and surface variables, and a refinement module for improved accuracy (see Supplementary Information Section 2.2). The multi-branch architecture handles satellite data and meteorological variables in background forecasts separately, allowing for flexible integration of additional observations. DA is performed four times per day (at 00, 06, 12, and 18 UTC), using observations from 3 h before to 4 h after forecast initialization, generating global analysis fields at 0.25° resolution.

The FuXi-Short produces 0–4 day forecasts, which serve as initial conditions for the FuXi-Medium model to generate 4–10 day predictions.

FuXi Weather is trained through joint optimization of analyses and forecasts, using ECMWF ERA5 reanalysis data at 0.25° resolution as the reference. While both the DA and forecasting components rely on ERA5 during training, the operational system operates independently of ERA5 during inference. To mimic varying operational conditions, FuXi forecasts (initialized with ERA5 data) are randomly sampled across lead times of 6 h to 5 days and used as background forecasts to train FuXi-DA. Owing to the limited amount of satellite data, FuXi-DA is trained on a 1-year dataset (June 1, 2022–June 30, 2023); this contrasts with the 37-year dataset used to train FuXi models[14]. A replay-based incremental learning strategy adapts the system to changes in satellite data quality and availability[70,71] (see Supplementary Figs. 1 and 2), retraining FuXi-DA monthly with data from the previous year. Further details are in the Supplementary Information Section 3.2.

The FuXi-Short model is fine-tuned with FuXi-DA analysis fields to reconcile accuracy differences with ERA5 (Supplementary Information Section 3.3). During testing, FuXi Weather is initialized with zero values for cycling DA and forecasting, using one year of data spanning from July 1, 2023 to June 30, 2024.

Evaluation method
Forecasts are evaluated against benchmark datasets at corresponding forecast times. For FuXi model forecasts, whether initialized with ERA5 or analysis fields generated by FuXi-DA, ERA5 is used as the benchmark. Consistent with standard practices in NWP, FuXi Weather is also evaluated against its analyses. In evaluating the performance of ECMWF high-resolution (HRES) forecasts, the time series of HRES-fc0 data used to initialize these forecasts at time t_0 is also used as the benchmark at the evaluation time $t_0 + \tau$. When FuXi and ECMWF HRES forecasts are evaluated against their respective initialization time series, both systems inherently exhibit higher accuracy at shorter lead times.

Deterministic forecasts are evaluated using established metrics, including the RMSE and ACC, defined as follows:

$$\text{RMSE}(c, \tau) = \frac{1}{|D|} \sum_{t_0 \in D} \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{X}_{c,i,j}^{t_0+\tau} \right)^2} \quad (1)$$

$$\text{ACC}(c, \tau) = \frac{1}{|D|} \sum_{t_0 \in D} \frac{\sum_{i,j} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right) \left(\mathbf{X}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right)}{\sqrt{\sum_{i,j} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right)^2 \sum_{i,j} a_i \left(\mathbf{X}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right)^2}} \quad (2)$$

where, t_0 denotes the forecast initialization time within the testing dataset (D), and τ is the forecast lead time. The climatological mean (**M**), calculated from ERA5 over the period 1993–2016, reflects the average conditions over these years. To better distinguish forecast performance between models with minor differences, the normalized RMSE difference between model A and baseline model B is calculated as $(\text{RMSE}_A - \text{RMSE}_B)/\text{RMSE}_B$. Similarly, the normalized ACC difference is calculated as $(\text{ACC}_A - \text{ACC}_B)/(1 - \text{ACC}_B)$. A negative RMSE difference and positive ACC difference indicate that model A outperforms model B.

Furthermore, RMSE can be decomposed into systematic and random error components through calculation of the MBE and standard deviations of errors ($\text{STD}_{\text{ERROR}}$). These metrics distinguish whether forecast errors originate from consistent bias or random variations around observed values. The MBE and $\text{STD}_{\text{ERROR}}$ are
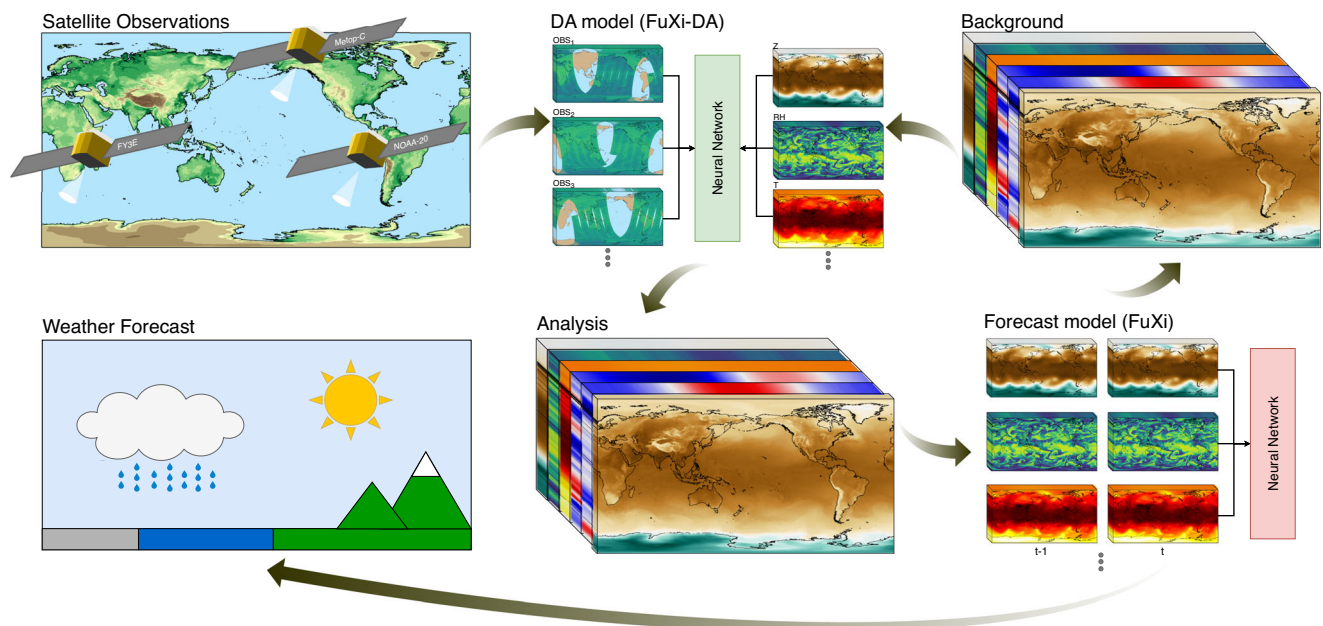
Fig. 7 | Schematic of the FuXi Weather system. Satellite radiance observations are brought in through machine learning data assimilation (DA) coordinated with the FuXi forecast model.

Table 1 | Summary of all input and output variables of the FuXi model

Type	Full name	Abbreviation	Role
Upper-air variables	Geopotential	Z	Input and Output
	Temperature	T	Input and Output
	u component of wind	U	Input and Output
	v component of wind	V	Input and Output
	Relative humidity	R	Input and Output
Single-level variables	2-meter temperature	T2M	Input and Output
	Mean sea-level pressure	MSLP	Input and Output
	10-meter u wind component	U10	Input and Output
	10-meter v wind component	V10	Input and Output
	Total precipitation	TP	Input and Output
Geographical	Orography	OR	Input
	Latitude	LAT	Input
	Longitude	LON	Input
Temporal	Hour of day	HOUR	Input
	Day of year	DOY	Input
	Step	STEP	Input

The "Type" indicates whether the variable is a time-varying variable, including upper-air, single-level, geographical variable, or temporal variable. The "Full name" and "Abbreviation" columns refer to the complete name of each variable and their corresponding abbreviations in this paper. The "Role" column clarifies whether each variable serves as both an input and an output, or is solely utilized as input by our model.

calculated as follows:

$$\text{MBE}(c, \tau) = \frac{1}{|D|} \sum_{t_0 \in D} \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{X}_{c,i,j}^{t_0+\tau} \right) \quad (3)$$

$$\text{STD}_{\text{ERROR}}(c, \tau) = \frac{1}{|D|} \sum_{t_0 \in D} \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left[\left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{X}_{c,i,j}^{t_0+\tau} \right) - \text{MBE}(c, \tau) \right]^2} \quad (4)$$

Machine learning-based weather forecasting models often produce excessively smooth predictions as lead time increases. We quantify this forecast smoothness using two complementary activity

metrics: (1) the standard deviation (std) of forecast anomalies relative to climatological means[40], and (2) the RMSE between forecasts and climatological means[44]. For both metrics, lower activity values indicate smoother fields. The std-based activity metric measures spatial variability in forecast anomalies with respect to the climatological mean M:

$$\text{Act}_{\text{STD}}(c, \tau) = \frac{1}{|D|} \sum_{t_0 \in D}$$

$$\sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left[\left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right) - \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right) \right]^2} \quad (5)$$
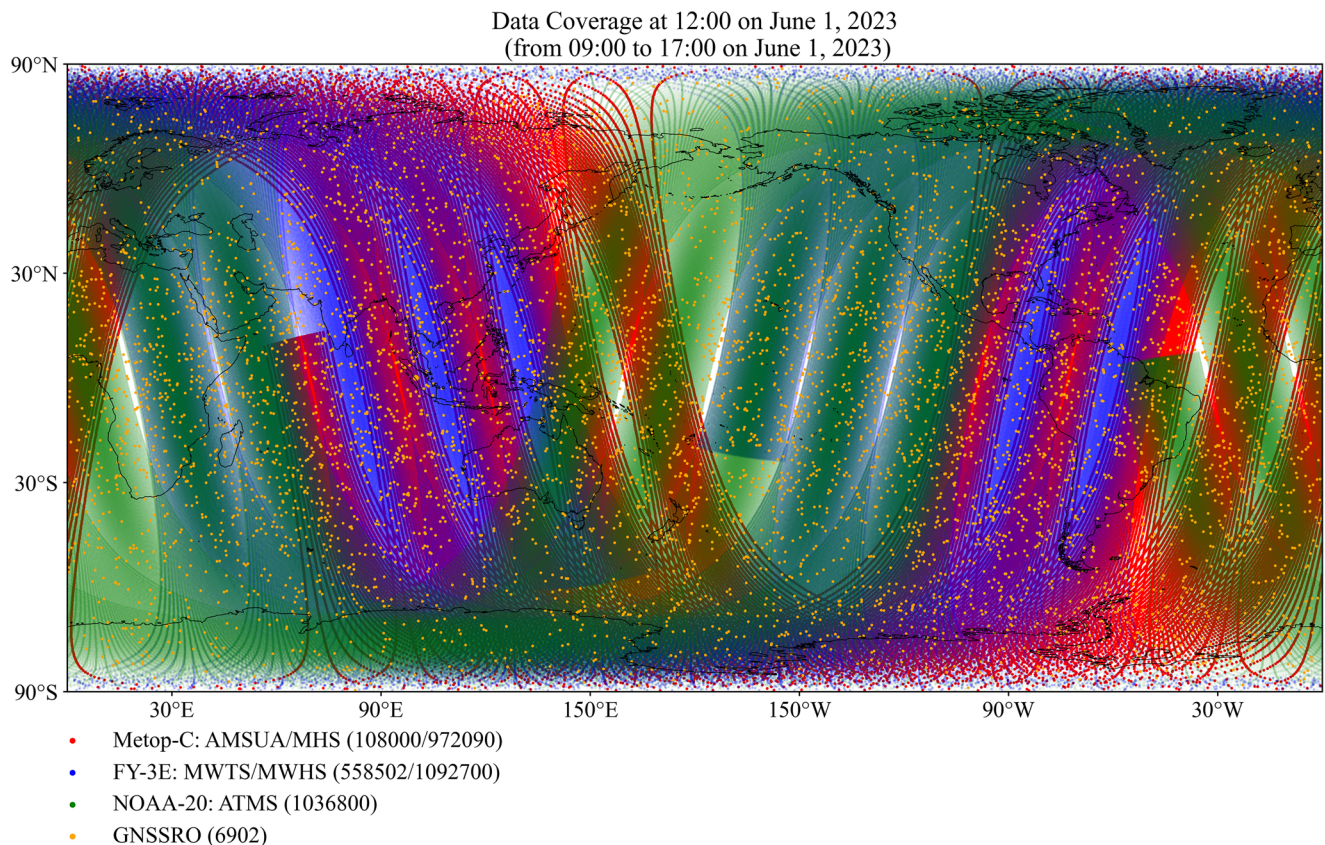
Fig. 8 | Typical data coverage from observations collected by three polar-orbiting meteorological satellites and the Global Navigation Satellite System (GNSS). The satellite observations are FengYun-3E (blue), Meteorological Operational Polar Satellite-C (Metop-C) (red), National Oceanic and Atmospheric Administration-20 (NOAA-20) (green), and GNSS radio occultation (RO) (yellow). This represents data from 3 h before to 4 h after 12 UTC on June 1, 2023. These data are utilized to generate analysis fields for 12 UTC on the same date.

The RMSE-based activity metric directly measures forecast deviations from climatological means:

$$\text{ACT}_{\text{RMSE}}(c,\tau) = \frac{1}{|\mathbf{D}|} \sum_{t_0 \in \mathbf{D}} \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right)^2} \tag{6}$$

To assess the quality of analysis fields, we calculate the RMSE and MBE using the same formulations as for forecast evaluation. Furthermore, we introduce analysis activity, which is defined as the ratio of the std to the climatological mean. This metric quantifies the degree to which analyses deviate from the climatological average state. The analysis activity is calculated as follows:

$$\text{Act}_{\text{STD}}(c,\tau) = \frac{1}{|\mathbf{D}|} \sum_{t_0 \in \mathbf{D}}$$

$$\sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left[\left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right) - \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} a_i \left(\hat{\mathbf{X}}_{c,i,j}^{t_0+\tau} - \mathbf{M}_{c,i,j}^{t_0+\tau} \right) \right]^2} \tag{7}$$

Data availability

The ERA5 reanalysis data are accessible through the Copernicus Climate Data Store at https://cds.climate.copernicus.eu/. ECMWF HRES forecasts can be retrieved from https://apps.ecmwf.int/archive-catalogue/?type=fc&class=od&stream=oper&expver=1. Satellite data can be obtained from the portal of the National Satellite Meteorological Center at http://satellite.nsmc.org.cn/DataPortal/en/home/index.html. The IMERG precipitation data are available at https://gpm.nasa.gov/data/imerg.

Code availability

The FuXi model is available at https://zenodo.org/records/10401602[72]. The model for FuXi Weather used in this study is available at https://zenodo.org/records/15762985[73].

References
1. Grasso, V. F. & Singh, A. Early warning systems: state-of-art analysis and future directions. *Draft Rep. UNEP* **1**, 1–70 (2011).
2. Rogers, D. & Tsirkunov, V. Costs and benefits of early warning systems. *Global Assessment Rep.* **1**, 1–17 (2011).
3. Pu, Z. & Kalnay, E. in *Handbook of Hydrometeorological Ensemble Forecasting* (eds Duan, Q. et al.) 67–97 (Springer, 2019).
4. Charney, J. G., Fjörtoft, R. & Neumann, J. V. Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237–254 (1950).
5. Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015).
6. Linsenmeier, M. & Shrader, J. G. *Global Inequalities in Weather Forecasts* (SocArXiv, 2023). https://doi.org/10.31235/osf.io/7e2jf.
7. Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A. & Gneiting, T. Skill of global raw and postprocessed ensemble predictions of rainfall in the tropics. *Weather Forecast.* **35**, 2367–2385 (2020).
8. Carleton, T. et al. Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *Q. J. Econ.* **137**, 2037–2105 (2022).
9. Bauer, P. What if? numerical weather prediction at the crossroads. *J. Eur. Meteorol. Soc.* **1**, 1–12 (2024).
10. de Burgh-Day, C. O. & Leeuwenburg, T. Machine learning for numerical weather and climate modelling: a review. *Geosci. Model Dev.* **16**, 6433–6477 (2023).

11. Pathak, J. et al. Fourcastnet: a global data-driven high-resolution weather model using adaptive fourier neural operators. Preprint at https://arxiv.org/abs/2202.11214 (2022).

12. Bi, K. et al. Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **619**, 533–538 (2023).

13. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).

14. Chen, L. et al. Fuxi_ A cascade machine learning forecasting system for 15-day global weather forecast. *NPJ Clim. Atmos. Sci.* **6**, 1–11 (2023).

15. Lang, S. et al. AIFS - ECMWF's data-driven forecasting system. Preprint at https://arxiv.org/abs/2406.01465 (2024).

16. Haiden, T. et al. *Evaluation of ECMWF Forecasts, Including the 2021 Upgrade* (ECMWF, 2021).

17. Li, L., Carver, R., Lopez-Gomez, I., Sha, F. & Anderson, J. Generative emulation of weather forecast ensembles with diffusion models. *Sci. Adv.* **10**, 1–11 (2024).

18. Price, I. et al. Probabilistic weather forecasting with machine learning. *Nature* **637**, 84–90 (2025).

19. Zhong, X. et al. Fuxi-ens: a machine learning model for medium-range ensemble weather forecasting. Preprint at https://arxiv.org/abs/2405.05925 (2024).

20. Bonavita, M. On some limitations of current machine learning weather prediction models. *Geophys. Res. Lett.* **51**, 1–10 (2024).

21. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A. & Kumar, V. Machine learning for the geosciences: challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* **31**, 1544–1554 (2019).

22. Bannister, R. N. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143**, 607–633 (2017).

23. Geer, A. J. et al. All-sky satellite data assimilation at operational weather forecasting centres. *Q. J. R. Meteorol. Soc.* **144**, 1191–1217 (2018).

24. Hamill, T. M. & Snyder, C. A hybrid ensemble Kalman filter–3d variational analysis scheme. *Monthly Weather Rev.* **128**, 2905–2919 (2000).

25. Buehner, M. Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**, 1013–1043 (2005).

26. Wang, X. Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: a mathematical framework. *Monthly Weather Rev.* **138**, 2990–2995 (2010).

27. Clayton, A. M., Lorenc, A. C. & Barker, D. M. Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **139**, 1445–1461 (2013).

28. Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R. & Fairbairn, D. Comparison of hybrid-4denvar and hybrid-4dvar data assimilation methods for global nwp. *Monthly Weather Rev.* **143**, 212–229 (2015).

29. Buehner, M. et al. Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part i: the Global System. *Monthly Weather Rev.* **143**, 2532–2559 (2015).

30. Brotzge, J. A. et al. Challenges and opportunities in numerical weather prediction. *Bull. Am. Meteorol. Soc.* **104**, E698–E705 (2023).

31. Carrassi, A., Bocquet, M., Bertino, L. & Evensen, G. Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev.: Clim. Change* **9**, 1–50 (2018).

32. Gettelman, A. et al. The future of earth system prediction: advances in model-data fusion. *Sci. Adv.* **8**, 1–12 (2022).

33. Cheng, S. et al. Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA J. Autom. Sin.* **10**, 1361–1387 (2023).

34. Arcucci, R., Zhu, J., Hu, S. & Guo, Y.-K. Deep data assimilation: integrating deep learning with data assimilation. *Appl. Sci.* **11**, 1–21 (2021).

35. Fablet, R. et al. Learning variational data assimilation models and solvers. *J. Adv. Model. Earth Syst.* **13**, e2021MS002572 (2021).

36. Brajard, J., Carrassi, A., Bocquet, M. & Bertino, L. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model. *J. Comput. Sci.* **44**, 1–11 (2020).

37. Wang, W. et al. A four-dimensional variational constrained neural network-based data assimilation method. *J. Adv. Model. Earth Syst.* **16**, 1–21 (2024).

38. Nichols, N. Overview: data assimilation and model reduction. *PAMM* **7**, 1026501–1026502 (2007).

39. Hatfield, S. et al. Building tangent-linear and adjoint models for data assimilation with neural networks. *J. Adv. Model. Earth Syst.* **13**, 1–16 (2021).

40. Bouallègue, Z. B. et al. The rise of data-driven weather forecasting: a first statistical assessment of machine learning-based weather forecasts in an operational-like context. *Bull. Am. Meteorol. Soc.* **105**, 864–883 (2024).

41. Xiao, Y. et al. Towards a self-contained data-driven global weather forecasting framework. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24, 1–21 https://openreview.net/forum?id=Y2WorV5ag6 (2024).

42. Chen, K. et al. The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time. *Commun. Earth Environ.* **6**, 518 (2025).

43. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

44. Allen, A. et al. End-to-end data-driven weather prediction. *Nature* 1–30 https://www.nature.com/articles/s41586-025-08897-0 (2025).

45. Xu, X. et al. Fuxi-da: a generalized deep learning data assimilation framework for assimilating satellite observations. *NPJ Clim. Atmos. Sci.* **8**, 1–15 (2025).

46. Lang, A. H. et al. PointPillars: fast encoders for object detection from point clouds. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12689–12697 https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.01298 (IEEE Computer Society, Los Alamitos, CA, USA, 2019).

47. Slivinski, L. C., Whitaker, J. S., Frolov, S., Smith, T. A. & Agarwal, N. Assimilating observed surface pressure into ML weather prediction models. *Geophys. Res. Lett.* **52**, 1–8 (2025).

48. Blum, J., Dimet, F.-X. L. & Navon, I. M. Data assimilation for geophysical fluids. *Handb. Numer. Anal.* **14**, 385–441 (2009).

49. Geer, A. J. Significance of changes in medium-range forecast scores. *Tellus A: Dyn. Meteorol. Oceanogr.* **68**, 1–23 (2016).

50. Pachauri, R. K. et al. Climate change 2014: synthesis report. In *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf (2014).

51. Dinku, T. Chapter 7 - challenges with availability and quality of climate data in Africa. *Extreme Hydrology and Climate Variability* (eds. Melesse, A. M., Abtew, W. & Senay, G.) 71–80 (Elsevier, 2019).

52. Zhong, X. et al. FuXi-Extreme: improving extreme rainfall and wind forecasts with diffusion model. *Sci. China Earth Sci.* **67**, 3696–3708 (2024).

53. Kochkov, D. et al. Neural general circulation models for weather and climate. *Nature* **632**, 1060–1066 (2024).

54. Lavers, D. A., Simmons, A., Vamborg, F. & Rodwell, M. J. An evaluation of era5 precipitation for climate monitoring. *Q. J. R. Meteorol. Soc.* **148**, 3152–3165 (2022).

55. Huffman, G. J. et al. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). *Algor. Theor. Basis Doc. Vers.* **4**, 1–39 (2015).

56. Kirschbaum, D. B. et al. Nasa's remotely sensed precipitation: a reservoir for applications users. *Bull. Am. Meteorol. Soc.* **98**, 1169–1184 (2017).

57. Sieglaff, J. M., Schmit, T. J., Menzel, W. P. & Ackerman, S. A. Inferring convective weather characteristics with geostationary high spectral resolution ir window measurements: a look into the future. *J. Atmos. Ocean. Technol.* **26**, 1527–1541 (2009).

58. ECMWF. Observations used in ECMWF (accessed 07 July 2024). https://www.ecmwf.int/en/research/data-assimilation/observations

59. Alexe, M. et al. Graphdop: towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations. Preprint at https://arxiv.org/abs/2412.15687 (2024).

60. Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y. & Toth, Z. Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Rev.* **136**, 463–482 (2008).

61. Ott, J. et al. A Fortran-keras deep learning bridge for scientific computing. *Sci. Program.* **2020**, 1–13 (2020).

62. Zhong, X., Yu, X. & Li, H. Machine learning parameterization of the multi-scale Kain–Fritsch (MSKF) convection scheme and stable simulation coupled in the Weather Research and Forecasting (WRF) model using WRF–ML v1.0. *Geosci. Model Dev.* **17**, 3667–3685 (2024).

63. Maddy, E. S., Boukabara, S. A. & Iturbide-Sanchez, F. Assessing the feasibility of an NWP satellite data assimilation system entirely based on AI techniques. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **17**, 9828–9845 (2024).

64. McNally, A. et al. Data driven weather forecasts trained and initialised directly from observations. Preprint at https://arxiv.org/abs/2407.15586 (2024).

65. Song, L. et al. HyPar: towards hybrid parallelism for deep learning accelerator array. In *2019 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 56–68 https://doi.ieeecomputersociety.org/10.1109/HPCA.2019.00027 (IEEE Computer Society, Los Alamitos, CA, USA, 2019).

66. Rasley, J., Rajbhandari, S., Ruwase, O. & He, Y. DeepSpeed: system optimizations enable training deep learning models with over 100 billion parameters. In *Proc. 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD'20* 3505–3506 (Association for Computing Machinery, 2020).

67. Fan, S. et al. Dapple: a pipelined data parallel approach for training large models. In *Proc. 26th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming* 431–445 (Association for Computing Machinery, 2021).

68. Foelsche, U. et al. An observing system simulation experiment for climate monitoring with GNSS radio occultation data: setup and test bed study. *J. Geophys. Res. (Atmospheres)* **113**, 1–14 (2008).

69. Guo, Y. et al. Deep learning for 3d point clouds: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 4338–4364 (2020).

70. Harris, B. & Kelly, G. A satellite radiance-bias correction scheme for data assimilation. *Q. J. R. Meteorol. Soc.* **127**, 1453–1468 (2001).

71. Dee, D. P. Variational bias correction of radiance data in the ECMWF system. In *Proc. ECMWF Workshop on Assimilation of High Spectral Resolution Sounders in NWP, Reading, UK* 97–112 (ECMWF, 2004).

72. Chen, L. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast (Version 1.0) [Dataset] [Software]. Zenodo. https://zenodo.org/records/10401602 (2023).

73. Sun, X. et al. A data-to-forecast machine learning system for global weather (Version 1.0) [Figure Dataset]. Zenodo. https://zenodo.org/records/15762985 (2025).

Author contributions
H.L, X.S., and X.Z. designed the project. X.S. designed and performed the model training. X.Z., X.X., Y.H., and X.S. performed the analysis under supervision of H.L., D.N., W.H., D.C., L.W., and Y.Q. X.Z., D.N., and X.S. wrote and revised the manuscript. X.X., W.H., D.C., and J.F. contributed to interpreting results.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-62024-1.

Correspondence and requests for materials should be addressed to Hao Li, J. David Neelin, Wei Han, Libo Wu or Yuan Qi.

Peer review information *Nature Communications* thanks Massimo Bonavita and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

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

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