npj Clean Water

https://doi.org/10.1038/s41545-026-00571-w

Article in Press

Machine learning reveals disruptive nutrient pollution shifts in Chinese rivers to 2100

Received: 20 January 2026

Accepted: 11 March 2026

Cite this article as: Zhang, X., Zhang, H., Yin, D. *et al.* Machine learning reveals disruptive nutrient pollution shifts in Chinese rivers to 2100. *npj Clean Water* (2026). https://doi.org/10.1038/s41545-026-00571-w

Xiaoyue Zhang, Hong Zhang, Dingkun Yin, Baojing Gu & Lei Chen

Machine learning reveals disruptive nutrient pollution shifts in Chinese rivers to 2100

Xiaoyue Zhang[1,2], Hong Zhang[1,2]*, Dingkun Yin[3], Baojing Gu[4], Lei Chen[5]*

[1]Key Laboratory of Environmental Aquatic Chemistry, State Key Laboratory of Regional Environment and Sustainability, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Laboratory of Transport Pollution Control and Monitoring Technology, Transport Planning and Research Institute, Ministry of Transport, Beijing, China

[4]State Key Laboratory of Soil Pollution Control and Safety, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China

[5]State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, China

*Corresponding Authors: Hong Zhang (hongzhang@rcees.ac.cn) and Lei Chen (chenlei1982bnu@bnu.edu.cn)

Abstract: Anticipating nutrient pollution under changing conditions is urgent for water security. Nonetheless, high-resolution predictive frameworks capturing nonlinear driver responses remain limited. Here, we present a nationwide assessment of China's water quality evolution from 2023 to 2100, integrating over 3 million daily records with 41 climatic, landscape, and socioeconomic drivers via regionally tailored Random Forest models (R^2 of 0.88–0.92). Our results reveal a disruptive spatiotemporal shift, with projected NPI ranges from –50.9% to +218.1% under SSP5-8.5. Seasonal patterns restructure toward unimodal peaks, with pollution increases in spring/autumn (up to 28.3%) but decreases in summer (up to 27.0%). Spatial homogenization emerges via a westward/southward shift of pollution centers, with localized increases exceeding 200% from coldspots with low baselines (<0.4 vs >1.0 in hotspots). Landscape configuration dominates (64.5% feature importance) over climatic forcing (7.2%–35.5%), reinforced by minimal climatic projection uncertainty. Strategic land-use planning could be a cornerstone of future water security.

Keywords: Nutrient pollution; Machine learning; Spatiotemporal shifts; Critical thresholds; River basins; China

Introduction

Water security is a cornerstone of the United Nations Sustainable Development Goals (SDGs), yet it faces escalating global threats. Nitrogen pollution alone already exposes approximately 22% of global basins to water scarcity [1], a figure projected to double by 2050, endangering the drinking water safety of over 3 billion people [1, 2]. Despite decades of mitigation efforts, progress remains inconsistent, and many regions continue to experience worsening trends [3, 4, 5]. A core scientific challenge persists in that the nonlinear mechanisms and spatiotemporal evolution of nutrient pollution remain insufficiently understood [6], particularly at high resolution. While extensive research has revealed large-scale patterns of nutrient pollution, critical temporal and spatial gaps persist. Existing global and regional assessments have primarily focused on source identification, flux quantification, and ecological impacts, emphasizing hotspots such as the Mississippi, Rhine, and major Asian rivers [7, 8]. However, the integration of high-frequency monitoring data across large spatial scales remains limited, leaving it unresolved whether transient nutrient pulses driven by rainfall, runoff, or human activities are adequately captured by conventional sampling regimes [9, 10, 11]. This limitation constrains our ability to quantify short-term nutrient dynamics and their responses to hydrological and anthropogenic variability. Moreover, most large-scale assessments still rely on monthly or seasonal sampling and/or long-term averages, which can smooth episodic export events that contribute disproportionately to annual nutrient loads, especially under intensifying hydro-climatic extremes. A shift toward fine-scale spatiotemporal quantification is therefore crucial for accurately assessing nutrient dynamics in large river systems.

Identifying and quantifying the drivers of nutrient pollution are vital for effective management. Numerous studies have examined the influence of land use, agricultural practices, industrial emissions, and socioeconomic factors on nutrient loading in river basins [12, 13]. Yet, the interactions among these drivers are complex and frequently nonlinear [10, 14]. Quantifying such nonlinearity remains limited, and the role of landscape configuration, an important process-related factor, is often overlooked [15, 16]. While multivariate statistical and process-based watershed models such as SWAT and SPARROW have been widely used to attribute nutrient loads to sources or land uses [17, 18], they often struggle to disentangle the synergistic, antagonistic, or threshold-dependent interactions among climatic, anthropogenic, and biogeochemical processes. Although prior research has recognized the importance of multi-factor interactions in shaping nutrient dynamics [19], the thresholds governing these relationships remain largely unexplored at basin or national scales. Understanding these thresholds is essential for predicting abrupt shifts in nutrient behavior and for developing adaptive mitigation strategies [20, 21]. Furthermore, while significant progress has been made in identifying drivers of nutrient pollution, research remains disproportionately focused on its future spatiotemporal trajectories under climate change alone [5, 19, 22], often overlooking the combined effects of climate and human activities. This knowledge gap undermines the formulation of forward-looking management and risks misallocating resources amid evolving environmental pressures. In parallel, recent studies have begun to apply machine-learning methods (e.g., Random Forests, gradient boosting, and deep learning) to predict nutrient concentrations and loads, often achieving high accuracy [23, 24]; however, many applications provide limited insight into nonlinear interactions, threshold behavior, and the management-relevant conditions under which pollution regimes may reorganize.

Despite substantial progress in monitoring and modeling riverine nutrient pollution, three gaps limit the ability to anticipate future water-quality risks under non-stationary climate and

socioeconomic change. First, most projections are based on long-term trends or climate-only drivers and therefore lack the resolution and integration necessary to capture nationwide and cross-basin heterogeneity at management-relevant scales. Second, although machine learning has improved prediction performance for complex environmental systems, existing applications seldom provide interpretable, driver-specific diagnostics of nonlinear effects and threshold transitions that can reshape pollution regimes. Third, the extent to which future nutrient pollution will reorganize in both space and time remains poorly quantified at high resolution. These gaps are especially consequential for rapidly transitioning regions where land-use intensification, infrastructure expansion, and climate-driven hydrological shifts co-occur, potentially producing nonlinear responses that are not captured by linear trend extrapolation or climate-only scenarios [10, 14].To bridge these critical gaps, we developed a novel analytical framework that integrates nationwide daily monitoring data with explainable machine learning to quantify future dynamics of nutrient pollution at an unprecedented spatiotemporal resolution. Applying this framework to China, a macrocosm of complex environmental transitions[17, 25], we aimed to achieve three key advances: (i) to project basin-scale trajectories of nutrient pollution in China under combined climate and socioeconomic change; (ii) to quantify the nonlinear effects and critical thresholds of interacting climatic, landscape, and anthropogenic drivers governing pollution patterns; and (iii) to elucidate the mechanistic basis for any projected shifts in pollution geography and seasonality. Specifically, we integrated >3 million daily water-quality records from a nationwide monitoring network (forming a nationwide Nutrient Pollution Index, Fig. 1) with gridded climatic, landscape, and socioeconomic predictors, and coupled Random Forest prediction with SHAP-based interpretation to diagnose driver importance, nonlinear responses, and critical transition thresholds. We anticipate that the quantification of nonlinear driver contributions and critical thresholds could redefine the priorities for future water resources management. This forward-looking analysis aims to provide a scalable template for anticipating and managing nonlinear water quality shifts in large river systems globally.

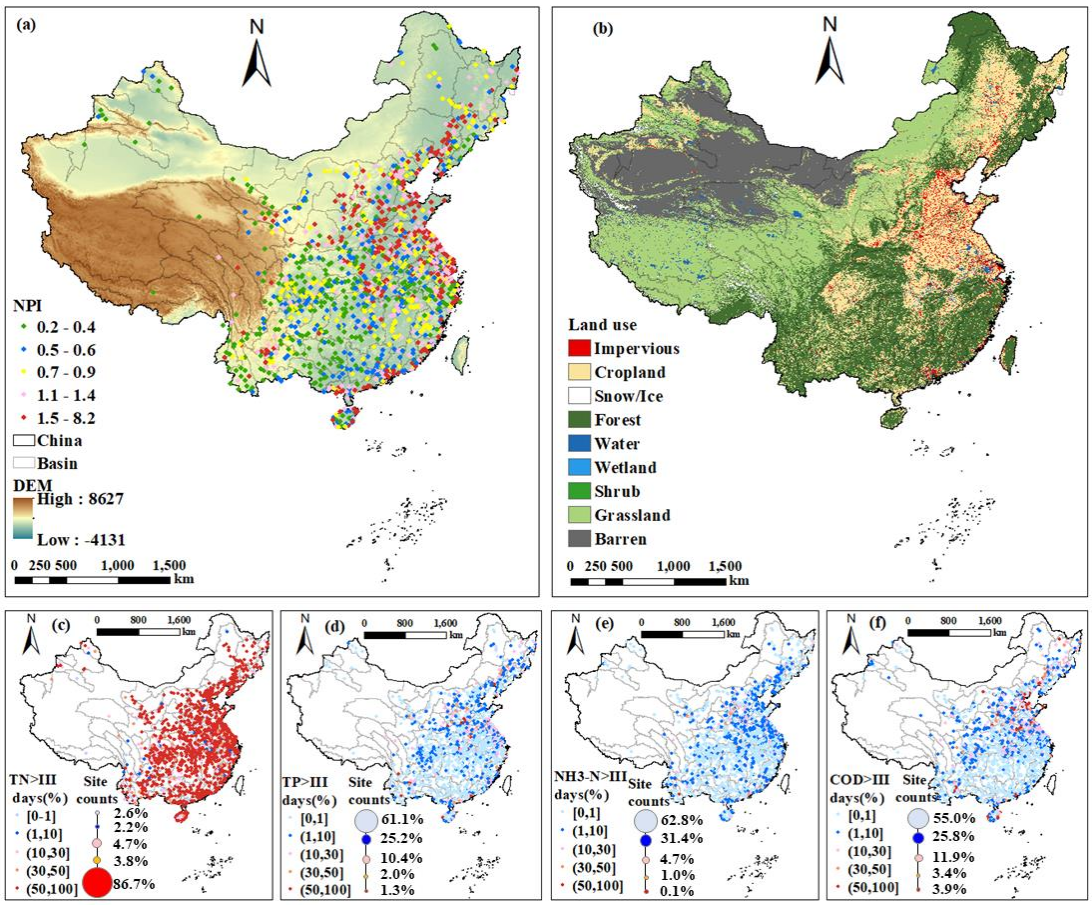

Fig. 1 Quantification of nutrient pollution in China. a: Site-based nutrient pollution index (NPI). c-f: Exceedance rates for each site against the Class III water quality standards for surface water [26] and site counts in each level.

Results

Temporal restructuring in nutrient pollution

Firstly, we quantified the annual averages of TN, TP, NH_3–N, and COD across basins as 0.6–8.0 $mg \cdot L^{-1}$, 0.0–0.2 $mg \cdot L^{-1}$, 0.0–0.6 $mg \cdot L^{-1}$, and 0.6–5.8 $mg \cdot L^{-1}$, respectively (Figs. S1), and over 40% of the monitoring stations exceeded national Class III standards [26] (Fig. 1) according to the water quality records. It indicated marked spatial heterogeneity and severe nutrient pollution [27]. The NPI demonstrated high consistency across all tested configurations, with Pearson correlations reaching up to 0.91-0.95 across seasons and exceeding 0.86 for spatial fields across river basins (Fig. S2), confirming its robustness for assessing nutrient pollution dynamics. Temporally, the coefficient of variation (CV) for daily NPI was 0.2, while the monthly CV was 0.1, signifying substantial short-term fluctuations. Seasonally, approximately 30% of sites exhibited seasonal fluctuations exceeding a twofold difference between maximum and minimum values (Fig. 2e), likely reflecting the influence of non-point source pollution, hydrological regulation, and seasonal human activities [28, 29]. These patterns underscore inherent vulnerabilities in aquatic ecosystems with limited capacity to buffer seasonal nutrient shocks [21, 25]. Specifically, summer and winter were identified as critical pollution periods, accounting for 42.6% and 30.3% of peak sites, respectively (Fig. S3). Winter exhibited a distinctive bimodal pattern, with concurrent peaks at both low-pollution (L1, 22.5%) and high-pollution (L5, 24.8%) sites (Fig. 2a). This duality likely represents a thermally

mediated bifurcation. In high-altitude western regions such as the Tibetan Plateau, subfreezing conditions suppress biogeochemical activity, producing an L1-dominant low NPI regime. In contrast, in temperate urban basins like the North China Plain, hydraulic stagnation concentrates pollutants, generating L5-dominant high NPI conditions [30]. However, it is noteworthy that high seasonal fluctuation occurred nationwide (Fig.2e), implying that regional climatic rhythms drive pollution variability across watersheds, while local characteristics such as basin morphology and land use intensity regulate the magnitude of these responses. During summer, pollution polarization intensified, with L5 sites increasing to 25.94% and L1 sites declining to 15.80% (Fig. 2c), confirming that warm-season hydrology acts as a major amplifier of extreme nutrient fluxes [12, 31]. Spring and autumn showed smoother transitions toward moderate pollution levels (L2–L3: 39.5–44.8%, Figs. 2b and 2d), suggesting the combined effects of dilution and loading balance.

Then, we identified four temporal variation types by daily NPI values and specific pollutants using STL decomposition (Figs. 2f–i and S4). For type I, the NPI displayed a pronounced bimodal cycle, characterized by an initial decline, a mid-year (August) peak, and a rebound toward year-end. This dual peak arose from contrasting seasonal mechanisms. Winter peaks corresponded to elevated TN (7.1 ± 0.6 mg L^{-1}) and NH_3–N (0.3 ± 0.1 mg L^{-1}) concentrations, whereas summer peaks were associated with concurrent increases in TP (0.10 ± 0.02 mg L^{-1}) and COD (4.0 ± 0.2 mg L^{-1}). The winter enhancement of TN and NH_3–N likely resulted from inhibited microbial denitrification and delayed mobilization of accumulated nitrogen during early rainfall or snowmelt events, consistent with previous studies [25]. Elevated groundwater nitrogen during the dry season may also contribute to this winter accumulation [32]. Although type II and III exhibited a similar intra-annual trajectory, characterized by relative stability in spring, abrupt intensification in June–July, and a sharp decline after October; their biogeochemical contexts differed. Type II exhibited simultaneously high TP (0.05–0.25 mg L^{-1}) and COD (2.9–6.8 mg L^{-1}) levels, with peak concentrations nearly double those in other types. Type III was characterized by elevated TP, COD, and NH_3–N concentrations, low DO, and extreme WT polarization (Fig. S1), indicating complex pollutant interactions. The reduced winter nitrogen levels in these types (TN: 2.9 ± 0.1 mg L^{-1} and 3.1 ± 0.2 mg L^{-1}; NH_3–N: 0.23 ± 0.04 mg L^{-1} and 0.19 ± 0.04 mg L^{-1}, respectively; Figs. 2f–h) dampened winter NPI amplification, potentially suggesting that climatic and hydrological forcing overrides nitrogen seasonality. Type IV followed a unimodal trajectory with a gradual rise toward an August peak, accompanied by lower pollution levels for all pollutants throughout the year. These findings demonstrate that climate- and region-sensitive dynamic frameworks are essential to replace static seasonal management approaches, enabling more adaptive and predictive strategies for nutrient pollution control.

During 2040–2100, nutrient pollution dynamics are projected to undergo a temporal restructuring. Nationwide, the annual NPI was predicted to rise by an average of 8.82% by 2100. The monthly CV of NPI was expected to decline from 0.1 to 0.01–0.02 (Fig. S5), reflecting the collapse of seasonal contrasts. For sites in type I-III, the NPI will rise sharply in spring (+12.6% to +20.1%) and autumn (+19.7% to +28.3%), but decline markedly in summer (–14.8% to –27.0%) (Figs. 2j–m). This finding aligns with recent global-scale projections of earlier runoff timing [33] and increased eutrophication risk under climate change [23]. Type I, in particular, will shift from a dual-peak regime (winter–summer) to a single winter-dominated peak. These transitions are tightly linked to climate-induced modifications in precipitation and temperature patterns. On the one hand, decreasing summer rainfall (Fig.S6–S10) will constrain hydrological transport, preventing the flushing of

agricultural nutrients into river systems despite continued fertilizer use and irrigation. On the other hand, stronger winter and spring warming (Fig.S11–S15) [34] will alter soil processes, crop cycles, and water resource management, decoupling pollution peaks from historical hydrological maxima. This decoupling reflects the disruption of the seasonally synchronized interplay among fertilizer application, monsoon rainfall, and temperature-mediated nutrient cycling [21, 35]. In contrast, type IV (low pollution type) exhibited consistent year-round NPI increases after 2040, with monthly mean growth ranging from +1.8% to +35.9% (Fig. 2n). The absence of seasonal suppression indicates that even low-intensity human activities, when coupled with climate change, can erode natural buffering capacities. Processes such as agricultural expansion, earlier ice-off, and enhanced soil nutrient mobility collectively promote continuous nutrient export [36, 37]. This temporal restructuring signals that future environmental management must transcend traditional seasonal frameworks and adopt more forward-looking and dynamically adaptive governance strategies to meet the novel challenges posed by climate change.

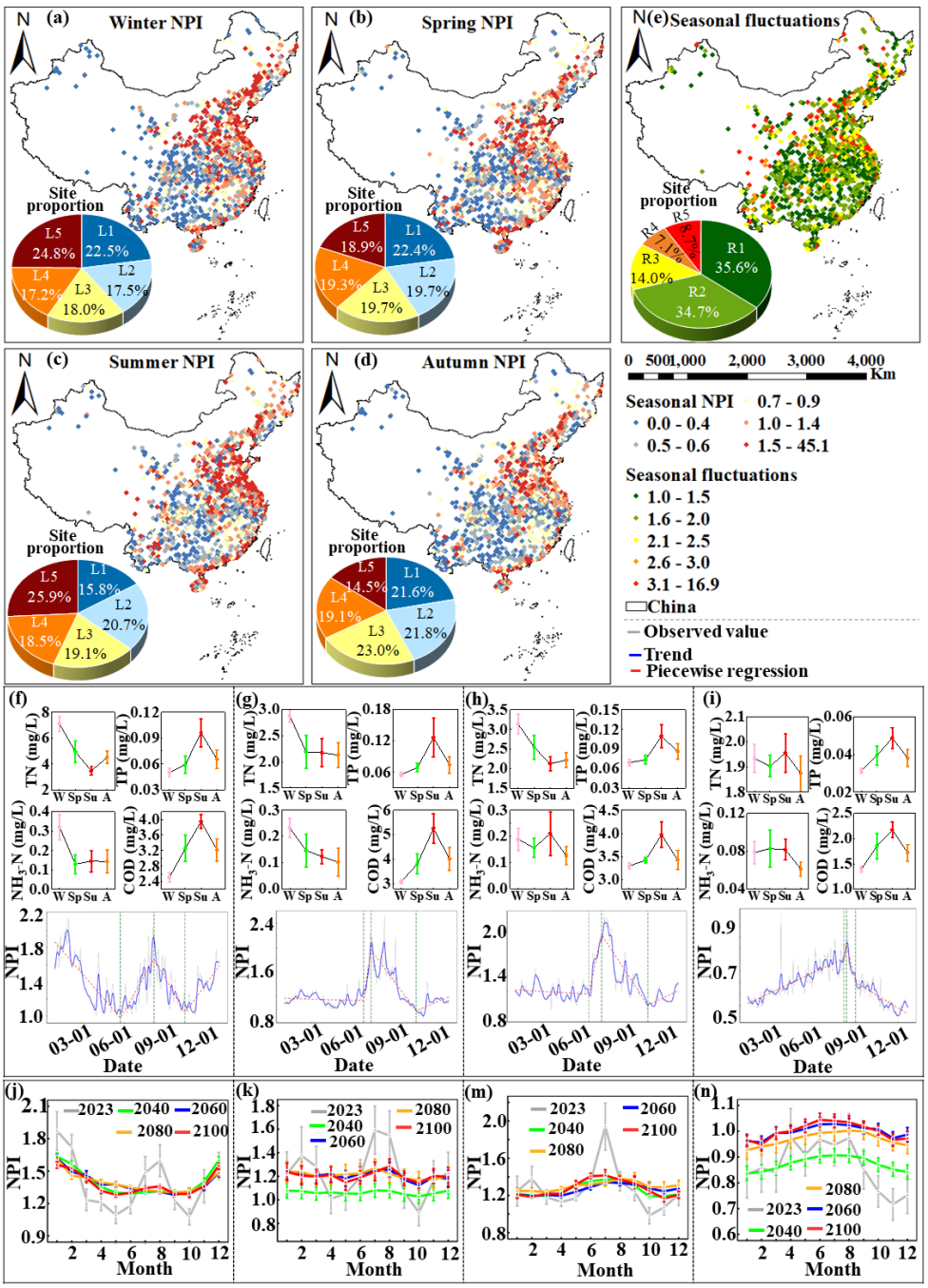

Fig. 2 Intra-annual temporal variations and future evolution of nutrient pollution. a-e: Seasonal NPI across sites and related fluctuations. f-i: Four identified variation types by daily NPI values and specific pollutants. j-n: Predicted monthly NPI for 2040–2100 under four variation types. L1-L5 in Figs. a-d indicate 5 levels integrated four-season data with quantile-defined boundaries. R1-R5 in Fig. e represent seasonal fluctuation rate calculated from the seasonal maximum and minimum values.

Spatial shift in nutrient pollution

We identified three spatially coherent process-based functional zones (PFZs) (Fig. 3a), integrating hotspot analysis (Figs. 3a and S1) and temporal variation types (Fig. 2). Hotspot analysis revealed

strong spatial clustering of nutrient pollution. Statistically significant hotspots were concentrated in densely populated and industrialized regions of eastern China, particularly the North China Plain and the Yangtze River Delta, where intensive agriculture and urbanization coincide with high summer precipitation. In contrast, coldspots dominated the western high-altitude regions, such as the Tibetan Plateau, and the southwestern mountainous areas, which experience low anthropogenic pressure and rapid hydrological flushing. This spatial contrast aligns with China's "Hu Line" socioecological divide, emphasizing the coupled influence of anthropogenic and hydrological factors [38]; and is consistent with the findings of Liu et al. [25]. Such a stark dichotomy highlights profound spatial heterogeneity in nutrient pollution, necessitating finer-scale partitioning to uncover the governing mechanisms. Beyond intensity gradients, we determined the PFZs by the hotspots and types I-III of site temporal variation, characterized by distinct biogeochemical signatures. These zones correspond broadly to China's major basin boundaries (Fig. S16). PFZ-1, defined by pronounced TN dominance (4.9 ± 1.4 mg L^{-1}, approximately twice that of PFZs-2 and -3), encompasses the water-stressed North China Plain and the Northeast Industrial Corridor, including the downstream Yellow, Hai, and Liao River basins. The elevated TN concentrations are largely attributable to intensive fertilizer application [4, 20, 32] and coal-based industrial emissions [39]. PFZ-2, dominated by TP and COD pollution, was concentrated within the Songhua River Basin, representing cold-climate urban-industrial watersheds where legacy pollution interacts with freeze-thaw cycles to generate combined phosphorus and organic pollution [40]. PFZ-3 was a combined multi-pollutant contaminated area. This zone, distributed across the Huai River Basin, the lower Yangtze River, and the upper Pearl River Basin, encompasses densely urbanized regions and erosion-prone subtropical foothills. It reflects the strong coupling between urbanization intensity and monsoon hydrology [41, 42, 43], leading to rapid and nearly linear increases in multipollutant loads. This functional zoning transcends administrative boundaries, revealing that similar pollution syndromes (type II and type III) arise from distinct underlying mechanisms (PFZ-2 and PFZ-3). Such mechanistic heterogeneity challenges uniform pollution control standards and underscores the need for syndrome-specific strategies tailored to the dominant drivers within each functional zone.

We identified a disruptive spatial shift in future nutrient pollution across Chinese basins (Fig. 3). The NPI was projected to increase in an average of 56.7–64.2%, of which coldspot basins accounted for 62.2–71.1%. The projected median NPI increases of individual basins displayed wide variation, ranging from –50.9% to +218.1% under the SSP5-8.5 scenario (Figs. 3b and S17), highlighting the heterogeneous sensitivity of watershed systems to future environmental stressors. Although PFZs-1–3 will continue to record higher absolute NPI values (median 1.3, compared to 0.9 in coldspots in 2100), their relative increases (–26.5% to 85.9%) will remain moderate compared with those in coldspots (–50.9% to 218.1%). Especially, basins along the Tibetan Plateau fringe and in the Yunnan-Guizhou highlands, currently characterized by low human activity and rapid hydraulic flushing, showed median NPI increases exceeding 70%, with localized growth surpassing 200%. These dramatic relative changes, while partly reflecting the low baseline NPI (<0.4 in 2023, compared to >1.0 in hotspots) in these pristine basins (NPI<0.9 even after the projected increase), nonetheless challenge the long-held assumption that such ecosystems possess inherent resilience. Instead, it reveals that even minimally disturbed systems could experience accelerating nutrient pressures as climate change and anthropogenic encroachment reshape long-standing biogeochemical rhythms governing nutrient export.

Spatial homogenization of nutrient pollution thus occurred, confirmed by a decrease in CV from 0.5 in 2023 to 0.2 in 2100 (Figs. 3c-g), alongside the gradual southwestward shift of pollution centers into historically pristine basins (Fig. 3b). It may signal a profound shift in the geography of environmental equity [44]. This flattening is primarily driven by climate-induced smoothing of seasonal cycles and the extension of anthropogenic influences into previously resilient regions [7, 19]. Quantitative socio-economic analysis reveals that these coldspot regions have substantially lower population density (288.4 vs. 429.6 persons/km²), impervious surface coverage (2.6% vs. 10.6%), and GDP (3,200.9 vs. 4,937.8 10^8CNY) compared to hotspots (Fig. S18), indicating limited adaptive capacity to cope with rising pollution pressures. While stringent mitigation efforts in eastern hotspots have curbed further deterioration [45], the disproportionate escalation in western coldspots possibly exposes a form of spatial injustice. Regions such as the Tibetan Plateau margin and Yunnan-Guizhou highlands, historically minor contributors to China's nutrient emissions due to limited socioeconomic development, are expected to bear a growing share of the future burden, even though their absolute NPI levels will remain comparatively low. This distinction between relative and absolute burdens underscores the need for equity-oriented governance that addresses both the erosion of ecological integrity and the imposition of new risks on vulnerable communities.

This inversion of risk reflects the combined influence of expanding agricultural frontiers and climate-driven perturbations, including permafrost thaw and altered hydrological regimes, in ecologically fragile and socioeconomically vulnerable landscapes [46, 47]. The apparent flattening of spatial variability thus masks a fundamental realignment of environmental burdens. Although absolute NPI levels in western coldspots will remain comparatively low, the disproportionate escalation of relative pollution pressure signals a critical loss of natural buffering capacity in regions least equipped to adapt. Such trends demand a reorientation of basin management toward principles of distributive equity, aligning resource allocation and mitigation responsibilities with both current emissions and projected vulnerabilities. Critically, these divergent seasonal and spatial trajectories identified here underscore the urgent need for regionally tailored monitoring and intervention frameworks that incorporate both the magnitude and timing of future nutrient risks. Adaptive management must therefore move beyond static thresholds to embrace dynamic, equity-oriented governance that anticipates where and when pressures will intensify in a rapidly changing climate.

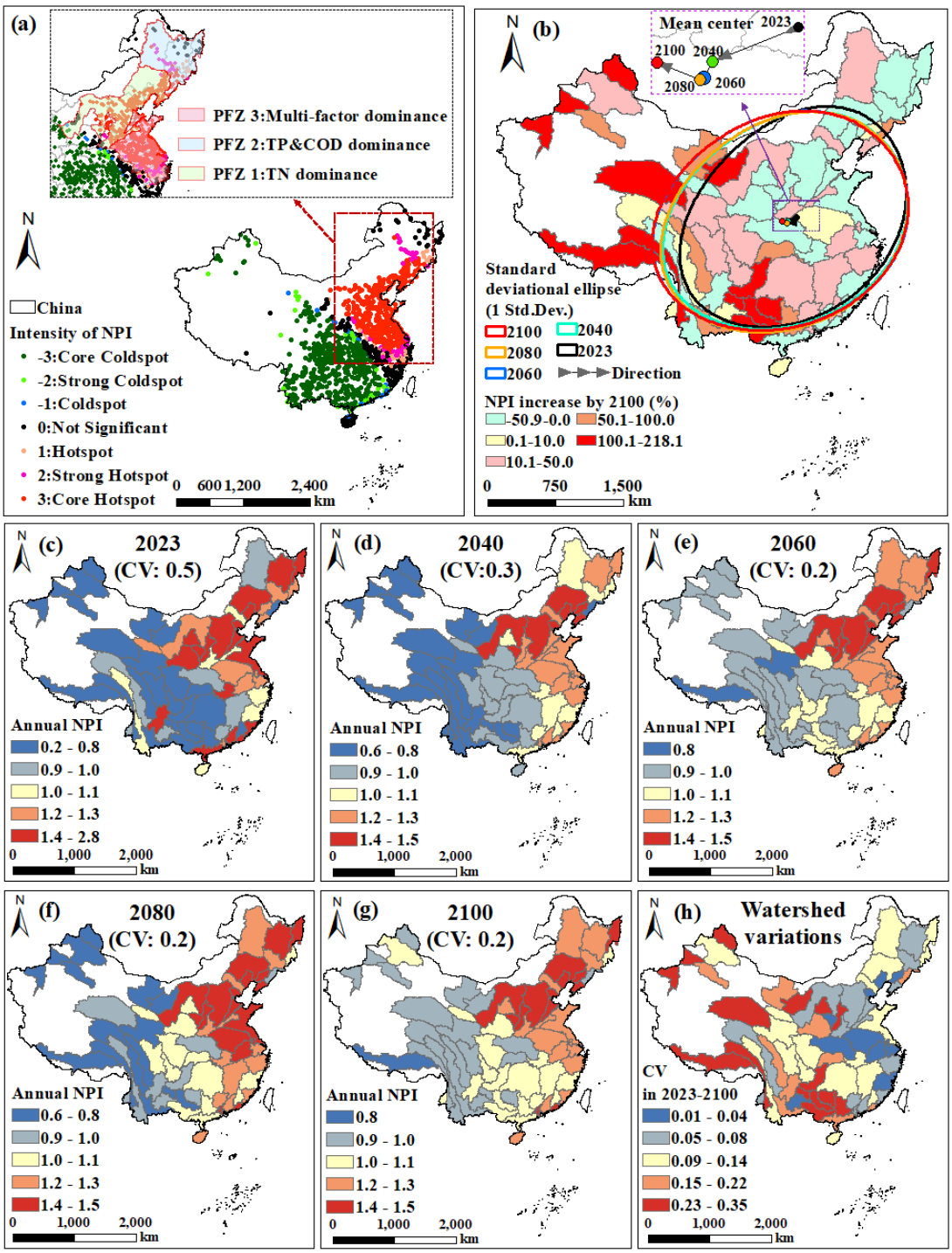

Fig. 3 Spatial patterns and future evolution of nutrient pollution. a: Spatial hotspot analysis and cluster patterns of nutrient pollution. b: Spatial trajectory of pollution center from 2023 to 2100. c-g: Spatial distribution of annual NPI from 2023 to 2100. h: Spatial evolution degree of NPI determined by the coefficient of variation (CV) from 2023 to 2100.

Driving mechanisms of spatiotemporal evolution in nutrient pollution

Region-specific dominant drivers, identified through SHAP-based relative importance, exhibited

strong contrasts across the three PFZs (Figs. 4a–c). In agriculturally intensive PFZ-1, climate variables and landscape fragmentation metrics were identified as the principal features in the model, with mean absolute SHAP contributions of 27.0% and 48.8%, respectively; followed by socioeconomic activities, accounting for 17.9% (Fig. S19). Climatic factors initiate nutrient release and transport events, while landscape structure governs the connectivity, retention, and efficiency of nutrient transfer [9, 16]. The mechanistic role of landscape configuration in modulating nutrient pollution is supported by a growing body of hydrological and biogeochemical research. For instance, landscape fragmentation can alter flow paths and residence times, affecting nutrient retention and transformation in riparian zones and wetlands; while connectivity metrics, such as patch density and interspersion, have been shown to govern the efficiency of nutrient transfer from terrestrial to aquatic systems [9, 15]. In agricultural settings, the configuration of cropland and natural vegetation influences both the source strength and the potential for in-stream processing [48, 49]. In PFZ-3, encompassing dense urban agglomerations and erosion-prone foothills, climatic factors showed a higher mean contribution of 35.5%, whereas the contribution from socioeconomic activities declined to 8.0% (Fig. S20). Notably, the modeled relationship between rainfall and nutrient pollution diverged from that in PFZ-1 (Figs. S21–22). Although PFZ-2 and PFZ-3 exhibited similar temporal variation patterns (Figs. 2g–h), the feature importance rankings from their respective models differed substantially. In PFZ-2, where extended frozen periods prevail, the model assigned lower relative importance to climatic variables (contributing 12.9%) while amplifying the importance of direct anthropogenic pressures (fertilizer application, cropland proportion, et al.; accounting for 15.8%) and landscape metrics (accounting for 55.5%). Additionally, the contribution from geological and geographical factors increased by 3.1%–6.3% compared to PFZs-1–2 (Figs. S23). Even in coldspot regions with minimal human activity, the model identified subtle anthropogenic influences, captured by forest patch density (For_PD), impervious surface complexity (Im_LSI) and cropland proportion, as highly important features, with a combined landscape contribution of 64.5% (Figs. 4d and S24). These distinctions emphasize that regional anthropogenic and biogeochemical contexts are identified as stronger predictors of nutrient pollution dynamics than uniform climate forcing (with feature importance of 7.2%-35.5%, Figs. S19-24). This finding does not imply that climate is mechanistically unimportant, but rather that, under the current and projected conditions captured by our models, landscape configuration explains a larger share of the spatial and temporal variance in nutrient pollution.

The SHAP dependence analysis revealed nonlinear relationships and identified points, referred to here as critical transition thresholds (CTTs) at which the direction or intensity of a driver's association with NPI, shifted (Figs. S21–22, S25–26 and Table S1). Several drivers displayed counterintuitive patterns within the model. Evaporation (CTT: 38.8 mm; 95% CI: 32.2–57.7 mm) and temperature (CTT: 11.0 °C; 95% CI: 6.0–11.7 °C) in PFZ-1 were negatively correlated with NPI, implying enhanced dilution or degradation of pollutants under drier atmospheric conditions [50]. Vegetation (NDVI) also exhibited a unimodal relationship in the model, associated with improving water quality only at intermediate levels [0.3–0.7 (95% CI for upper threshold: 0.3–0.8) in PFZ-2; 0.4–0.5 (95% CI for upper threshold: 0.5–0.7) in PFZ-3]. Beyond these thresholds, dense vegetation likely facilitated the mobilization of legacy nutrients stored in soils and biomass [51], suggesting that vegetation's role is hydrologically mediated rather than purely structural. Moreover, the model-derived CTTs for climate and NDVI factors were context-dependent. In PFZ-1 and PFZ-2, the rainfall-NPI relationship followed an inverted U-shaped pattern (Figs. S21 and S25), with positive associations within 34.4–198.5 mm (95% CI for upper threshold: 36.5–204.2 mm) and

92.2–217.7 mm (95% CI for upper threshold: 134.5–232.8 mm), respectively, switching to negative associations beyond these ranges. In contrast, in the urbanizing PFZ-3, the model showed a monotonic increase in NPI with rainfall above 129.8 mm (95% CI: 62.0–176.8 mm) (Figs. S22), a pattern that aligns with expectations from enhanced surface runoff and combined sewer overflows [18, 52]. Temperature effects also diverged across PFZs. In PFZ-2, predicted NPI intensified at both cold (<9.5 °C; 95% CI: 1.1–15.1 °C) and warm (>22.6 °C; 95% CI: 22.4–26.3 °C) extremes; while in PFZ-3, it rose steadily beyond 16.8 °C (95% CI: 14.9–19.9 °C), consistent with hypotheses around oxygen depletion and intensified urban heat island effects [22, 53]. Evaporation showed opposing behaviors between PFZ-1 and PFZ-3, switching from negative to positive correlation at 94.0 mm (95% CI: 88.0–108.7 mm) in the latter, which may reflect contrasting hydrological and land-use contexts. In agricultural basins (PFZ-1), elevated evapotranspiration may reduce leaching and enhanced plant uptake [12, 27, 54], potentially mitigating nutrient export; which is a hypothesized pathway consistent with our statistical findings but requiring direct hydrological validation using process-based models. In urban basins (PFZ-3), impervious surfaces accumulated pollutants during dry periods, amplifying first-flush effects and evaporation-concentration dynamics [9, 18, 42], which increased nutrient discharge. Similarly, NDVI's differentiated role across PFZs [e.g., as a sink in PFZ-1 above 0.3 (95% CI: 0.27–0.34), versus a potential mobilizer in PFZ-2 and PFZ-3] points to testable hypotheses about vegetation-mediated nutrient dynamics that dense vegetation in agricultural watersheds improved nutrient retention by reducing erosion and absorbing dissolved nutrients [12, 38], whereas occasionally facilitated pollutant release in urban systems. More importantly, landscape metrics exhibited pronounced source–sink transition in the model, including For_PD [0.02 (95% CI: 0.02–0.02), 0.04 (95% CI: 0.01–0.04), and 0.01 (95% CI: 0.01–0.01) in PFZ-1, PFZ-3, and coldspots, respectively], Gra_IJI [75.6 (95% CI: 73.9–75.6) in PFZ-2], and Im_LSI [7.2 (95% CI: 6.8–7.2) in coldspots]. The model identified these values as critical configurations where landscape connectivity or retention collapses.

The two-way SHAP interaction analysis revealed that synergistic interactions among drivers introduced additional complexity, frequently modulating or reversing the marginal effects of individual variables (Figs. 4a1–c4 and S27). In PFZ-1, moderate rainfall (12–41 mm) counteracted the dilution effect of high evaporation (>80 mm), likely by remobilizing nutrients accumulated during dry periods without generating sufficient runoff for dispersion. Conversely, in PFZ-3, simultaneous high evaporation and rainfall (>52–118 mm) or elevated temperatures (>20–22 °C) triggered strong positive SHAP values, a pattern consistent with the concept of monsoon-driven urban flushing [54, 55]. This contrast explains the divergent evaporation-pollution relationships observed between PFZ-1 and PFZ-3. In PFZ-2, the model indicated that high NDVI (>0.62) could transform the typically negative association of low rainfall (<74 mm) into a positive one, aligning with the hypothesis that dense vegetation might facilitate the flushing of sequestered nutrients [56]. Evaporation in PFZ-2, while showing a minor individual effect, emerged as a positive interactive contributor when combined with fertilizer inputs, Gra_IJI, NDVI, rainfall, or temperature, particularly beyond ~90 mm. This suggests a potential mechanism where evaporation concentrates soil nutrients, later mobilized by rainfall in cold-climate settings, though this remains a hypothesis. Landscape-climate interactions also induced pivotal functional shifts. In PFZ-3, high NDVI combined with high cropland division (>0.95) was associated with a shift from positive to negative contribution, suggesting that well-vegetated and hydrologically disconnected patches could act as nutrient sinks. Together, these model-derived interaction patterns highlight the irreducible complexity of coupled climate-landscape-anthropogenic systems and underscore the need for

mechanism-based and region-specific management frameworks.

The crossing of CTTs offers a potential explanatory mechanism for the forecasted spatiotemporal homogenization of nutrient pollution, a finding with broad implications for environmental equity. Our model suggests that as climate change increases compound extreme events [28, 46], historically distinct basins may converge toward a more uniformly polluted state. The dramatic increases predicted for historically pristine coldspots are not anomalies but the result of these systems crossing newly relevant anthropogenic thresholds. Our projections indicate that key anthropogenic drivers (e.g., GDP, impervious surface complexity, cropland expansion) in coldspots are likely to exceed the upper bounds of their 95% confidence intervals of their historically respective CTTs under the SSP5-8.5 scenario (Figs. 4e-f, S26 and Table S1), often exceeding the point estimates by 2.4–113.7%. This projected crossing of driver thresholds, as learned from historical data, underpins the associated NPI increases in our model (up to 218.1% in some basins), suggesting that these historically pristine systems may cross critical tipping points that trigger nonlinear pollution amplification. We emphasize that these estimates should be interpreted as early-warning signals highlighting the potential direction and magnitude of change under a high-emission future, rather than as precise numerical predictions. As indicated by Zhang et al., and consistent with our model's feature importance, future nutrient dynamics may be strongly influenced by anthropogenic forces [57]. This underscores that the pollution trajectory may be non-stationary and governed by threshold-crossing behavior, suggesting that future risks could be less dependent on past patterns and more on where and when these critical points are exceeded. This dynamic offers a hypothesis for the shifting geography of pollution liability, exemplified by the westward and southward movement of pollution centers in China, which mirrors global concerns about climate change redistributing environmental burdens toward vulnerable regions. Our framework, therefore, helps define a governance paradigm for a non-stationary world, where anticipating potential regime shifts is a crucial first step toward achieving equitable outcomes under global changes, though the underlying mechanisms require further causal confirmation.

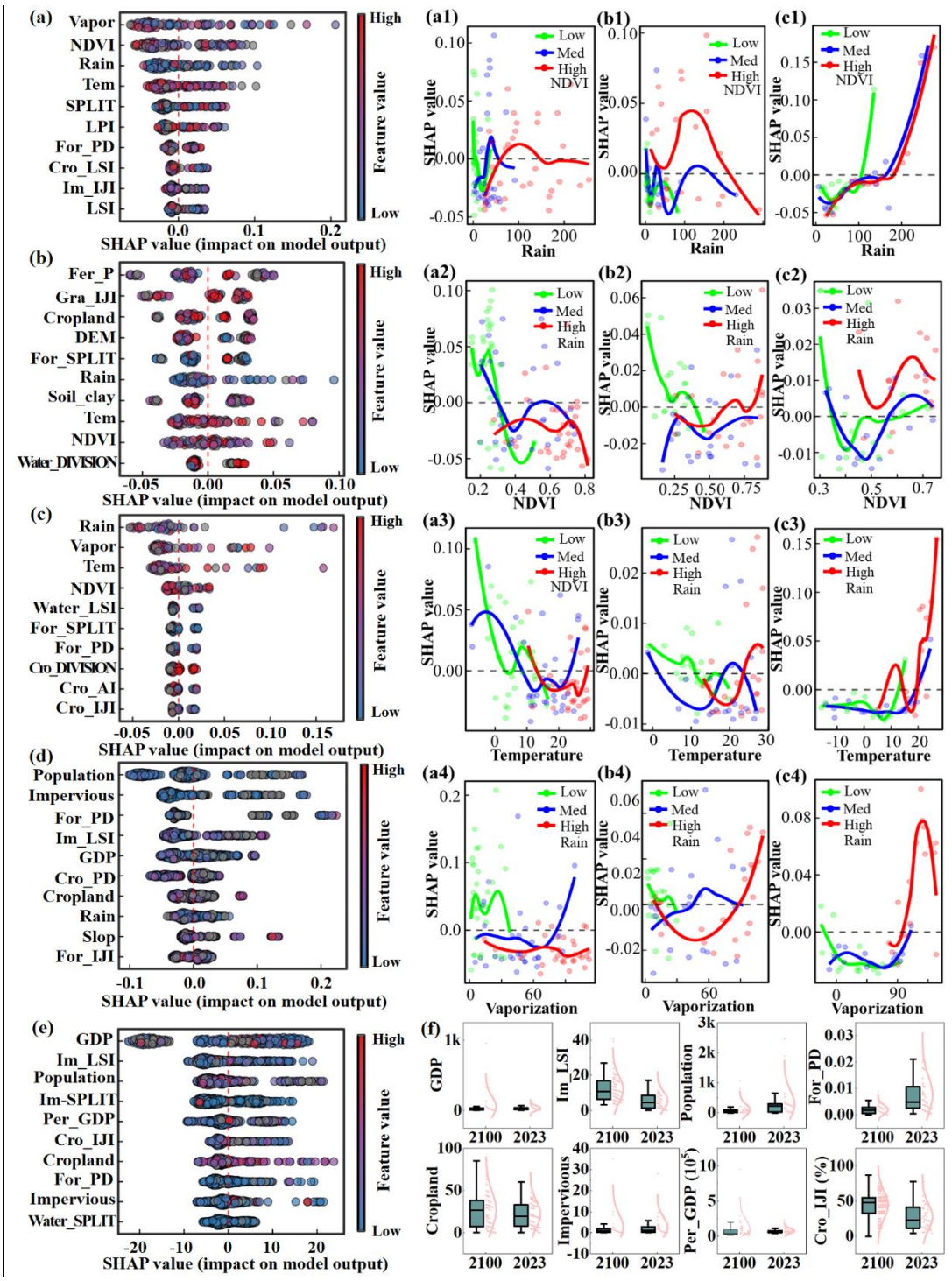

Fig. 4 Driving mechanisms of spatiotemporal evolution in nutrient pollution. a-d: Key drivers for nutrient pollution in the four PFZs, respectively (see Fig. S28 for the complete factor ranking). a1-c4: Interaction effects of key drivers in the three hotspots PFZs, respectively. e: Key drivers contributing to increased pollution in coldspots. f: Statistics on key drivers in 2023 and 2100 in coldspots.

Discussion

This study advances the predictive understanding of nutrient pollution by revealing and quantifying a disruptive spatiotemporal shift across China's river basins, through an explainable machine learning framework. It is a process that challenges established paradigms of static pollution geography and seasonal dynamics. Our analysis, integrating millions of daily records, moves beyond identifying current hotspots to project a fundamental reorganization of pollution risks. Our projections delineate a temporal restructuring, decoupling peak pollution from traditional summer maxima [22, 37], flattening seasonal cycles, and elevating risks in the transitional periods of spring and autumn. This aligns with Gudmundsson et al. [33], who projected a shift toward earlier seasonal runoff timing. Concurrently, we identify a westward and southward migration of pollution centers into historically pristine basins. Perhaps most critically, this shift manifests as a counterintuitive spatial homogenization, a convergence of historically distinct basins toward a uniformly degraded state that masks a deepening underlying inequity. This work provides the first integrated, process-aware projection of such systemic change, demonstrating that future water quality is governed not by linear extrapolation but by nonlinear and threshold-driven dynamics that may actively redraw the map of environmental risk. Conventionally, climate change is often regarded as the predominant future driver of aquatic environmental pressures [5, 16]. However, within our predictive framework, changes in landscape configuration are identified as the most dominant factors of future nutrient pollution patterns (with feature importance up to 64.5%), far exceeding the direct contribution of climatic factors (7.2%–35.5%) (Figs. S19–S20, S23–S24). This discrepancy likely arises from our inclusion of high-resolution landscape metrics typically omitted from broad-scale studies, and from machine learning's capture of nonlinear threshold effects that conventional models miss [17, 18]. This result is reinforced by our uncertainty quantification, which revealed that GCM structural uncertainty contributes minimally projection variability (e.g., the 95% confidence interval for the maximum projected increase in coldspot basins ranged from +217.7% to +219.1%) (Fig. S17). This pivotal finding, though correlational, calls for a reorientation of water resources management from a core focus on climate adaptation toward a foundation of strategic land-use planning, while acknowledging that the underlying mechanisms require further causal validation. We acknowledge that China's changing demographics, declining population (Fig. 4f), may indirectly influence future land use and agricultural intensity, with potential knock-on effects on nutrient runoff not fully captured by our current driver set. These complex socio-demographic interactions warrant further investigation in integrated modeling frameworks. The high-resolution NPI risk maps produced by our framework (Figs. 2–3) provide an operational tool for environmental agencies to prioritize basins and seasons for regulatory intervention, enabling targeted allocation of monitoring and mitigation resources where future pollution risks are highest.

This projected spatiotemporal shift transcends hydrological change, representing a significant realignment of socio-ecological risk with profound implications for environmental equity [44, 51]. The migration of pollution pressure systematically targets historically low-emission basins in western and southwestern China, regions marked by greater ecological fragility. This pattern extends previous hotspot-focused assessments [17, 43], which have predominantly emphasized pollution persistence in eastern China, suggesting a phenomenon that coarser-resolution global models might overlook. This shift creates a fundamental misalignment between historical responsibility and future impact, posing a direct challenge to the principle of Common but Differentiated Responsibility. These areas, historically minor contributors to national nutrient emissions due to lower socioeconomic development (Fig. S18), now face a compounding burden.

Crucially, the dramatic relative increases projected for these regions must be interpreted in the context of their exceptionally low baseline NPI values. In absolute terms, even after these increases, the NPI in such basins remains lower than current levels in eastern China's hotspots. Thus, while the relative change signals a significant loss of natural buffering capacity, it does not imply that these pristine areas will become the most polluted regions in absolute terms. Rather, the core concern is the erosion of their ecological integrity and the imposition of new risks on vulnerable communities. Climate-driven perturbations, such as permafrost thaw and altered precipitation regimes, may be simultaneously degrading the natural buffering capacity of these landscapes [15, 19, 33]. This could create a vicious cycle where rising pollutant loads converge with declining ecosystem resilience, disproportionately impacting communities with limited adaptive resources and infrastructure [35]. The resulting spatial homogenization of pollution metrics is therefore deceptive, masking a critical divergence in vulnerability and a potential escalation in associated human health risks from waterborne contaminants [20, 36]. This dynamic underscores an urgent need to reframe water quality governance within a distributive justice framework. This means that while continuing stringent mitigation efforts in eastern hotspots, the emerging risks in coldspot regions, where even modest absolute increases translate into large relative changes that signal a loss of natural buffering capacity, must not be overlooked. Effective management must look beyond static emission allocations based on historical responsibility, and proactively integrate forward-looking assessments of ecological sensitivity and social vulnerability. This proactive approach helps prevent systemic inequity [21] and safeguard ecological integrity before pressures intensify further.

The identification of critical transition thresholds (CTTs) from our model provides a potential explanation for this disruptive shift and offers a basis for its management. Our model suggests that the future reorganization of pollution is not a linear extrapolation of past trends but may be a consequence of crossing key nonlinear tipping points in climatic, landscape, and anthropogenic drivers. For instance, the dramatic pollution surges in coldspots are associated with previously minor anthropogenic pressures (e.g., cropland expansion, impervious surface growth) exceeding their basin-specific CTTs, which we interpret as potential triggers for fundamental regime shifts. This underscores a critical limitation of current management based on static water quality standards or uniform reduction targets. Our threshold-aware framework, grounded in data-driven insights, provides a model-derived basis for transitioning to dynamic and anticipatory governance. By identifying region-specific CTTs from the model, management can consider evolving from reactive clean-up to proactive early warning. These thresholds can be operationalized as early warning indicators in real-time monitoring systems. When key drivers (e.g., fertilizer application, impervious surface expansion) approach their CTTs, pre-defined management responses such as seasonal fertilizer restrictions, enhanced riparian buffer inspections, or targeted public outreach can be triggered before pollution regimes potentially shift irreversibly. This paradigm shift would involve monitoring for compound threshold exceedances, where combinations of drivers, rather than single factors, may precipitate nonlinear risk amplification, enabling pre-emptive interventions such as tailored fertilizer restrictions or runoff control activations before pollution regimes potentially shift irreversibly [20, 27]. While CTT values are region-specific and their mechanistic underpinnings require further testing, the principle of compound threshold management provides a universally applicable model for anticipatory governance, relevant for managing nutrient pulses in agricultural heartlands like the U.S. Midwest or urbanizing watersheds in Southeast Asia.

Globally, the disruptive spatiotemporal shift documented in China's river basins presents a critical test case for water governance under non-stationary conditions. This phenomenon challenges the foundational assumptions of conventional basin management, which often relies on static boundaries, historical pollutant relationships, and uniform standards. Our work provides a transferable analytical framework, whose core innovation lies in a paradigm shift from managing pollutant concentrations to governing the underlying drivers and thresholds that control their nonlinear mobilization and redistribution across landscapes and seasons. This approach is universally applicable to large, heterogeneous river systems facing climatic and land-use pressures, from the agricultural heartlands of North America to the monsoon-fed catchments of South Asia. It offers a scientifically robust basis for transcending political-administrative fragmentation (evident in the United States' focus on critical source areas [58] and the European Union's basin-scale monitoring [59], and transboundary system like the Ganga and Mississippi Rivers [60]), a persistent obstacle in transboundary basins, by defining management units based on shared hydrodynamic and biogeochemical processes. Moreover, by identifying critical transition thresholds from data-driven models, it replaces ambiguous and state-based targets with dynamic and mechanism-informed early-warning indicators. For global practice, this enables a shift from reactive and equity-blind remediation to proactive and resilience-focused investment. It argues for reorienting international cooperation around adaptive governance mechanisms, such as dynamic pollution budgets tied to ecological functional zones and forward-looking resource allocation that prioritizes future vulnerability hotspots [61]. Ultimately, this framework provides a vital tool for navigating the uncertainty of the Anthropocene, ensuring that water security strategies are robust to the profound spatial and temporal reorganizations of pollution risk that lie ahead.

Several limitations warrant consideration. First, the majority of stations are operated under the national regulatory network with standardized calibration procedures. We acknowledge that even high-frequency monitoring may miss very short-lived nutrient pulses, and that we lack independent field validation for all regions. Second, our use of monthly climate and hydrological predictors may under-represent event-scale flash pollution risks, particularly in basins where nutrient loads are dominated by extreme runoff events. Addressing this limitation will require coupling our framework with event-scale hydro-meteorological forcing and expanding validation using targeted high-frequency field campaigns during extreme events. Third, spatial coverage is uneven, with relatively sparse monitoring in western and high-altitude basins. Although our sensitivity analysis indicates that the projected homogenization and westward-southward shifts remain after excluding the sparsely monitored western stations (Fig.S29), we highlight the need for targeted field campaigns to validate and refine these projections. Fourth, while projections under a high-emission scenario (SSP5-8.5) provide a stringent risk assessment, uncertainties inherent in climate and socioeconomic models propagate into our estimates. On the one hand, our analysis is contingent on a single socioeconomic pathway; projections under more sustainable scenarios would likely exhibit less pronounced spatial homogenization. On the other hand, although we quantified GCM structural uncertainty via Monte Carlo propagation (Fig. S17), uncertainties inherent in socioeconomic and land-use projections are not fully propagated, meaning the confidence intervals are likely conservative. Fifth, the CTTs identified for individual drivers are data-driven constructs derived from a single baseline year, though bootstrap resampling confirmed their statistical robustness (Table S1). Applying these historically derived thresholds to project conditions in 2100

involves model extrapolation; thus, while the projected crossing of driver thresholds in coldspots underpins the associated NPI estimates (e.g., up to 218.1%), these estimates should be interpreted as scenario-based risk indicators rather than precise numerical predictions. Finally, the interpretability of our machine-learning models, though enhanced by SHAP, is correlational rather than causal; thus, our finding would benefit from causal validation through hybrid approaches that couple AI with established process-based watershed models. Furthermore, while we identify critical thresholds and synergistic interactions, the precise biogeochemical and hydrological mechanisms underlying these nonlinearities require targeted field experiments and finer-scale modeling. Future research should prioritize hybrid modeling frameworks that integrate process-based understanding with machine learning to better constrain long-term projections and improve extrapolation reliability under non-stationary conditions, alongside experimental validation of the proposed thresholds. Despite these limitations, our core contribution, a process-explicit and threshold-aware framework for diagnosing and projecting disruptive spatiotemporal shifts, provides a robust foundation for advancing precision water governance in an era of non-stationary environmental change across diverse global settings.

Methods

Construction of water quality and impact factor database

Daily water quality parameters, including total nitrogen (TN), total phosphorus (TP), ammonia nitrogen (NH_3-N), chemical oxygen demand (COD), water temperature (WT), and dissolved oxygen (DO), were sourced from the Ministry of Ecology and Environment of China (MEEC). These parameters were selected for their relevance in characterizing nutrient-related basin pollution [43, 62]. Our analysis used 2023 as the baseline, a strategic choice that ensured temporal alignment between water quality and potential driving factors (e.g., climatic, land-use, and socioeconomic variables for 2023) while capturing the most recent, representative state of nutrient pollution. There are 1606 monitoring stations nationwide, offering extensive spatial and temporal coverage (Fig. 1). Nevertheless, some areas, particularly in the west and at high altitudes, exhibit relatively sparse coverage (Fig. S30). According to MEEC, these indicators were measured at automatic monitoring stations according to the national standard methods for the examination of water and wastewater [26, 63]. Specifically, COD concentration was determined using the potassium permanganate titration method [63]. Most monitoring stations operate under the national network and yield continuous daily records with only sporadic missing values. To address these occasional gaps, we applied targeted time-series interpolation using the *imputeTS* package [52]. We visually inspected all imputed values and cross-validated them against nearby stations where possible to ensure consistency. Outliers were identified using the interquartile range rule and excluded if implausible. For basins without monitoring stations, we did not spatially interpolate water quality measurements. Instead, future projections rely on machine learning models trained on station data and gridded environmental drivers (see Methods: Future evolution of nutrient pollution). Fig. S31 shows the stepwise methodological flowchart of this study.

To quantify the spatiotemporal pollution patterns, we constructed the impact factor database as comprehensively as possible, guided by data availability and the literature worldwide [19, 31, 64]. Especially, the pronounced spatiotemporal disparities inherent to China's natural and anthropogenic characteristics, including air temperature, precipitation, terrain, land use and land cover, population, et al., may contribute significantly [65]. Besides, landscape patterns serve as a key

factor governing pollutant inputs to rivers, yet they are rarely addressed in nationwide studies [66]. For more information on factor selection, please refer to the Supplementary Text S1. This study included 65 driving variables initially, covering climate, hydrology, geology, geography, and anthropogenic factors. We then examined the statistical correlations between the variables to mitigate multi-collinearity effects (Fig. S32). Finally, 41 variables were selected by removing variables with $|R^2| > 0.8$, including temperature, precipitation, evaporation, runoff, soil properties, slope, population, gross domestic product (GDP), landscape indicators, et al. We used Fragstats 4.2 software to calculate landscape metrics, and ArcGIS 10.2 software to process data on climate, topography, etc. Detailed information on the data interpretation, sources, and statistics was provided in Supplementary Tables S2-S4 and Text S2.

Construction and validation of the Nutrient Pollution Index (NPI)

To capture the multifaceted nature of nutrient pollution, a nationwide Nutrient Pollution Index (NPI) was constructed. The NPI incorporates the aforementioned water quality indicators through a nonlinear weighted model, enabling for comparable assessments of pollution levels across stations and basins. Specifically, TN, TP, NH_3-N and COD serve as direct indicators of nutrient pollution [56], while WT and DO serve as indirect factors by influencing nutrient transformation and biogeochemical cycling processes [22]. Accordingly, WT and DO were introduced as nonlinear constraints to amplify the contribution of direct indicators, as shown in Eq. (1).

$$\mathrm{NPI} = \sum w_j \times {S_j}^{\alpha} \quad (1)$$

$$S_j = \frac{C_j}{C_{j\_base}} \quad (2)$$

where, S_j indicates the pollution index for indicator j; C_j and C_{j_base} are the monitored and base concentrations (spatiotemporal averages) of indicator j, respectively, mg L^{-1}; w_j denotes the weight of indicator j, determined by the entropy weighting method (Supplementary Text S3). α represents a nonlinear coefficient, when WT>20 °C, α=WT/20, serving as the correction factor for TP and TN [53]; when DO<4 mg L^{-1}, α=4/DO, serving as the correction factor for NH_3-N [43]. Note that if S_j<1, the reciprocal of α is used.

To validate the robustness of the constructed NPI, we performed a comprehensive sensitivity analysis. This analysis systematically examined the influence of key methodological choices, including alternative weighting schemes, and the inclusion or exclusion of indirect indicators. Details are provided in Supplementary Table S5 and Text S4. For each configuration, we compared spatiotemporal patterns and key derived metrics such as the spatial coefficient of variation and seasonal fluctuations.

Spatial and temporal patterns of nutrient pollution

To clarify spatial patterns of nutrient pollution, we defined 69 integrated basins using the Level 04 and 05 HydroBASINS dataset, following the approach of Liu *et al*. [65], of which 67 were used for analysis based on monitoring covering. Each basin exhibited similar geographical features and was named after the dominant river within it (Fig. S33). To examine the temporal dynamics of nutrient pollution, the daily NPI time series data were decomposed using the Seasonal-Trend decomposition procedure based on Loess (STL). STL decomposes the time series into seasonal, trend, and residual components, offering insights into long-term trends and periodic variations in nutrient pollution. Following STL decomposition, we applied segmented linear regression to the

trend component to detect and characterize change points in nutrient pollution trends. This approach allows for the identification of abrupt shifts or gradual changes in pollution patterns over time and was conducted in the R environment using packages including *imputeTS*, *stlplus*, *segmented*, et al. We employed a seasonal fluctuation index to analyze seasonal variations, as shown in Eq. (3). The four seasons were divided into spring (March-May), summer (June-August), autumn (September-November), and winter (December-February), based on China's climatic characteristics and relevant literature [41].

$$R = \frac{S_{max}}{S_{min}} \quad (3)$$

where, R represents seasonal fluctuation index, S_{max} and S_{min} indicate seasonal maximum and seasonal minimum of NPI, respectively. Seasonal NPI is calculated as the average of daily NPI values.

The Getis-Ord Gi* statistic was used to identify statistically significant spatial clusters of high (hotspots) and low (coldspots) NPI values, revealing the spatial distribution of nutrient pollution intensity. The method uses a fixed distance band to define neighboring features. Resultant z-scores and p-values determine the statistical significance of identified clusters, which were categorized as core hotspot ($Z \geqslant 2.58$, $P \leqslant 0.01$). strong hotspot ($1.96 \leqslant Z < 2.56$, $0.01 < P \leqslant 0.05$), hotspot ($1.65 \leqslant Z < 1.96$, $0.05 < P \leqslant 0.1$), not significant ($-1.65 < Z < 1.65$, $P > 0.1$), core coldspot (($Z \leqslant -2.58$, $P \leqslant 0.01$). strong coldspot ($-2.58 \leqslant Z < -1.96$, $0.01 < P \leqslant 0.05$), and coldspot ($-1.96 < Z \leqslant -1.65$, $0.05 < P \leqslant 0.1$).

Future evolution of nutrient pollution

Projecting the spatiotemporal evolution of nutrient pollution under future climate and socioeconomic change is critical for proactive watershed management. To this end, we adopted a spatially explicit, machine-learning-based approach. Given the pronounced spatial heterogeneity of NPI (Fig. 1), a single national-scale model would obscure critical regional differences. Therefore, we constructed separate models for each of the spatially coherent clusters. This ensured that the unique anthropogenic and biogeochemical context of each region was explicitly captured. We selected the Random Forest (RF) model for its widely reported robustness in watershed pollution modeling [13, 67]. For each basin within a cluster, the monthly averaged NPI (response variable) was paired with the corresponding monthly values of the 41 potential driving factors (predictor variables) across the study period. This temporal resolution was chosen to capture intra-annual dynamics while maintaining adequate signal strength for model training. Each regional RF model was trained using 80% of the basin-month observations from its respective cluster, with the remaining 20% withheld for testing. Model hyperparameters were optimized through 500 iterations of automated tuning to maximize predictive accuracy. It is worth noting that for regions with low site density in the west and at high altitudes, we performed a sensitivity analysis in which we excluded subsets of sparsely monitored western stations from model training and then compared the resulting spatial patterns and trend metrics with those from the full-data model. The performance of each final model was rigorously evaluated on the test set using the coefficient of determination (R^2) and Root Mean Square Error (RMSE), all of which demonstrated robust explanatory power (see Fig. S34 for detailed performance).

Future projections were conducted under the SSP5-8.5 scenario, a high-emission pathway characterized by intensive fossil-fuel development and limited environmental policy intervention.

We selected this scenario to serve as a stress test for exploring the upper bound of potential nutrient pollution risks and to reveal the maximum possible extent of spatiotemporal reorganization in the absence of aggressive climate and pollution mitigation policies. This approach provides a strong and policy-relevant warning signal for water resources management. The pre-trained Random Forest models served as the core predictive engine. These models were driven by future projections of key climatic (e.g., temperature, precipitation, evaporation) and anthropogenic (e.g., land use, GDP, population density) variables, which were derived from downscaled CMIP6 climate models and spatially explicit socioeconomic datasets for the periods centered on 2040, 2060, 2080, and 2100. The specific factors used for future projections were determined from the spatiotemporal mechanisms of nutrient pollution. To quantify the uncertainty arising from future climate projections, we explicitly accounted for inter-model variability within the GCM (General Circulation Model) ensemble. We propagated this climate uncertainty through the pre-trained Random Forest models using a Monte Carlo framework. All models demonstrated good performance with R^2 of 0.88–0.92; details are provided in Supplementary Table S6, Text S5.

All future driver data were preprocessed to align with the spatial resolution and normalization scheme of the historical training data. The modeling process thus involved applying the static, region-specific model structures to the dynamic future driver inputs, thereby isolating the impact of changing environmental conditions on the NPI. To quantify projection uncertainties arising from both model internals and scenario variability, we employed a bootstrap resampling approach (n=500), generating 95% confidence intervals for all future NPI estimates. The resultant projections were then analyzed to elucidate the dominant spatiotemporal trends of nutrient pollution across China's river basins.

Key driver identification and threshold quantification

To decipher the region-specific driving mechanisms of nutrient pollution, we used the region-based models established in the previous section. To interpret the trained models and quantify factor importance, we employed SHapley Additive exPlanations (SHAP) analysis, which provides a unified measure of the marginal contribution of each driver. To further investigate interactive effects, we extended this analysis to two-way SHAP interaction plots, where the SHAP dependence of one feature was conditioned on the values of another feature. These visualizations revealed synergistic or antagonistic interactions that would be obscured in univariate analyses. All analyses were conducted in the R platform (version 4.4). RF models were implemented using the *randomForest* and *ranger* packages, and SHAP values were computed with the *treeshap* package [68].

To move beyond identifying driver importance and quantify the nonlinear relationships between key drivers and nutrient pollution, we developed a systematic approach to detect critical transition thresholds (CTTs) from the trained Random Forest models. For each driver identified as important in the SHAP analysis, we generated SHAP dependence plots, which display the marginal effect of that driver on the model's predicted NPI output across its entire observed range. A CTT is defined as the driver value at which its SHAP contribution changes sign, specifically the point where the smoothed dependence curve crosses the zero line. To obtain a smooth and interpretable curve, we applied locally weighted scatterplot smoothing (LOESS) to the raw SHAP dependence scatter. The crossing point was then identified numerically. This sign change marks a transition in the driver's modeled relationship with NPI, from a net pollution-enhancing effect to a net pollution-mitigating

effect, or vice versa.

To evaluate the statistical robustness of the identified CTTs, we employed a bootstrap resampling procedure. From the original dataset for each region, we drew 500 bootstrap samples with replacement, each of the same size as the original. For every bootstrap sample, we retrained the Random Forest model (using the same hyperparameters as the original) and repeated the entire CTT identification process. This generated a distribution of threshold values for each driver, and we computed the 2.5th and 97.5th percentiles to obtain a 95% confidence interval. CTTs with narrow confidence intervals were interpreted as robust, while those with wide intervals were interpreted with caution. All bootstrap analyses were performed in R using the boot package, and results are presented in Table S1.

References

1. Wang M*, et al.* A triple increase in global river basins with water scarcity due to future pollution. *Nature Communications* **15**, 880 (2024).
2. Jones ER, Bierkens MFP, van Vliet MTH. Current and future global water scarcity intensifies when accounting for surface water quality. *Nature Climate Change* **14**, 629-635 (2024).
3. Nkwasa A*, et al.* One third of African rivers fail to meet the good ambient water quality nutrient targets. *Ecological Indicators* **166**, 112544 (2024).
4. Schulte-Uebbing LF, Beusen AHW, Bouwman AF, de Vries W. From planetary to regional boundaries for agricultural nitrogen pollution. *Nature* **610**, 507-512 (2022).
5. Yang Y*, et al.* Climate change exacerbates the environmental impacts of agriculture. *Science* **385**, eadn3747 (2024).
6. Dykes C, Pearson J, Bending GD, Abolfathi S. Hydraulic efficiency and mixing dynamics in surface flow constructed wetlands: Influence of design, vegetation phenology, and climate variability. *Water Research* **285**, 124110 (2025).
7. Smits AP*, et al.* Detecting signals of large-scale climate phenomena in discharge and nutrient loads in the Mississippi-Atchafalaya River basin. *Geophysical Research Letters* **46**, 3791-3801 (2019).
8. Royer TV, David MB, Gentry LE. Timing of riverine export of nitrate and phosphorus from agricultural watersheds in Illinois: Implications for reducing nutrient loading to the Mississippi River. *Environmental Science & Technology* **40**, 4126-4131 (2006).
9. Zhang X, Chen L, Yu Y, Shen Z. Water quality variability affected by landscape patterns and the associated temporal observation scales in the rapidly urbanizing watershed. *Journal of Environmental Management* **298**, 113523 (2021).
10. Ke Q, Zhang K. Scale issues in runoff and sediment delivery (SIRSD): A systematic review and bibliometric analysis. *Earth-Science Reviews* **251**, 104729 (2024).
11. Dykes C, Pearson J, Bending G, Abolfathi S. Impact of seasonal climate variability on constructed wetland treatment efficiency. *Journal of Water Process Engineering* **72**, 107350 (2025).
12. Schürings C*, et al.* River ecological status is shaped by agricultural land use intensity across Europe. *Water Research* **251**, 121136 (2024).
13. Mahdian M*, et al.* Linking hypolimnion to epilimnion in a stratified arctic lake: Machine learning-based estimation of hypolimnetic water quality from epilimnetic measurements. *Water Research* **294**, 125367 (2026).
14. Zhang H*, et al.* Natural and anthropogenic imprints on seasonal river water quality trends across China. *npj Clean Water* **8**, 49 (2025).
15. Salliou N, Urech P, Leitão JP, Fappiano F, Grêt-Regamey A. Urban water projects must consider landscape architecture. *Nature Water* **3**, 967-971 (2025).
16. Richardson CM*, et al.* Effects of climate change on river and groundwater nutrient inputs to the coastal ocean. *Communications Earth & Environment* **6**, 761 (2025).
17. Chen X*, et al.* Multi-scale modeling of nutrient pollution in the rivers of China. *Environmental Science & Technology* **53**, 9614-9625 (2019).
18. Bello A-AD, Haniffah MRM. Modelling the effects of urbanization on nutrients pollution for prospective management of a tropical watershed: A case study of Skudai River watershed. *Ecological Modelling* **459**,

109721 (2021).
19. Merz E*, et al.* Disruption of ecological networks in lakes by climate change and nutrient fluctuations. *Nature Climate Change* **13**, 389-396 (2023).
20. Gu B*, et al.* Cost-effective mitigation of nitrogen pollution from global croplands. *Nature* **613**, 77-84 (2023).
21. Hernández-Carrasco D, Tylianakis JM, Lytle DA, Tonkin JD. Ecological and evolutionary consequences of changing seasonality. *Science* **388**, eads4880 (2025).
22. Jansen J*, et al.* Climate-driven deoxygenation of northern lakes. *Nature Climate Change* **14**, 832-838 (2024).
23. Mozafari Z*, et al.* Impact of climatic factors on eutrophication in the World's largest lake. *Ecological Indicators* **175**, 113497 (2025).
24. Mohammadpour A*, et al.* Advanced water quality assessment using machine learning: Source identification and probabilistic health risk analysis. *Results in Engineering* **27**, 105421 (2025).
25. Liu H, Gao L, Yuan Z, Ma T. Contrasting seasonal variations in riverine nitrogen and phosphorus concentrations in China: implications for N/P imbalances. *Water Research* **287**, 124317 (2025).
26. MEEC. Environmental Quality Standards for Surface Water (GB 3838-2002). *Beijing*, (2002).
27. Rizzo A, Sarti C, Nardini A, Conte G, Masi F, Pistocchi A. Nature-based solutions for nutrient pollution control in European agricultural regions: A literature review. *Ecological Engineering* **186**, 106772 (2023).
28. Tong Y*, et al.* Lake warming intensifies the seasonal pattern of internal nutrient cycling in the eutrophic lake and potential impacts on algal blooms. *Water Research* **188**, 116570 (2021).
29. Yang Y*, et al.* Seasonal variations and hydrological management regulate nutrient transport in cascade damming: Insights from carbon and nitrogen isotopes. *Water Research* **271**, 122894 (2025).
30. Li Y*, et al.* The future of Chinese rivers: Increasing plastics, nutrients and Cryptosporidium pollution in half of the basins. *Resources, Conservation and Recycling* **205**, 107553 (2024).
31. Sinha E, Michalak AM, Balaji V. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science* **357**, 405-408 (2017).
32. Xia Q*, et al.* Effect and genesis of soil nitrogen loading and hydrogeological conditions on the distribution of shallow groundwater nitrogen pollution in the North China Plain. *Water Research* **243**, 120346 (2023).
33. Gudmundsson L*, et al.* Past and future change in global river flows. *Nature Reviews Earth & Environment*, https://doi.org/10.1038/s43017-43025-00745-z (2025).
34. Yu S*, et al.* Climate warming and nutrient enrichment destabilize plankton network stability over the past century. *Communications Earth & Environment* **6**, 247 (2025).
35. Hutchins DA, Tagliabue A. Feedbacks between phytoplankton and nutrient cycles in a warming ocean. *Nature Geoscience* **17**, 495-502 (2024).
36. Siirila-Woodburn ER*, et al.* A low-to-no snow future and its impacts on water resources in the western United States. *Nature Reviews Earth & Environment* **2**, 800-819 (2021).
37. Deng Y, Ling Z, Jiang W. Urban water bodies in China: Spatial distribution patterns, temporal change characteristics, and relationship with economic development. *Ecological Indicators* **157**, 111139 (2023).
38. Yang J*, et al.* NDVI variations of different terrestrial ecosystems and their response to major driving factors on two side regions of the Hu-Line. *Ecological Indicators* **159**, 111667 (2024).
39. Schmale J, Zieger P, Ekman AML. Aerosols in current and future Arctic climate. *Nature Climate Change* **11**, 95-105 (2021).
40. Li T*, et al.* Occurrence, distribution, and potential ecological risks of antibiotics in a seasonal freeze-thaw

basin. *Journal of Hazardous Materials* **459**, 132301 (2023).
41. Liu D, Du Y, Yu S, Luo J, Duan H. Human activities determine quantity and composition of dissolved organic matter in lakes along the Yangtze River. *Water Research* **168**, 115132 (2020).
42. Zhao Y-L*, et al.* Spatiotemporal drivers of urban water pollution: Assessment of 102 cities across the Yangtze River Basin. *Environmental Science and Ecotechnology* **20**, 100412 (2024).
43. Wu J*, et al.* Assessing water quality in the Pearl River for the last decade based on clustering: Characteristic, evolution and policy implications. *Water Research* **244**, 120492 (2023).
44. Li A, Yuan Q, Strokal M, Kroeze C, Ma L, Liu Y. Equality in river pollution control in China. *Science of The Total Environment* **777**, 146105 (2021).
45. Bertassello LE, Basu NB, Maes J, Grizzetti B, La Notte A, Feyen L. The important role of wetland conservation and restoration in nitrogen removal across European river basins. *Nature Water* **3**, 867-880 (2025).
46. Gao T*, et al.* Carbon dynamics shift in changing cryosphere and hydrosphere of the Third Pole. *Earth-Science Reviews* **250**, 104717 (2024).
47. Kurylyk BL. Engineering challenges of warming. *Nature Climate Change* **9**, 807-808 (2019).
48. Xu Q, Yan T, Wang C, Hua L, Zhai L. Managing landscape patterns at the riparian zone and sub-basin scale is equally important for water quality protection. *Water Research* **229**, 119280 (2023).
49. Yang L*, et al.* A landscape source–sink model to understanding the seasonal dynamics of antibiotics in soils at watershed scale. *Journal of Hazardous Materials* **465**, 133224 (2024).
50. Farrell KJ*, et al.* Ecosystem-scale nutrient cycling responses to increasing air temperatures vary with lake trophic state. *Ecological Modelling* **430**, 109134 (2020).
51. Zheng M*, et al.* Shifts in precipitation regimes exacerbate global inequality in grassland nitrogen cycles. *Nature Communications* **16**, 7888 (2025).
52. Tebong NK, Simo T, Takougang AN, Sandjon AT, Herve NP. Application of deep learning algorithms to confluent flow-rate forecast with multivariate decomposed variables. *Journal of Hydrology: Regional Studies* **46**, 101357 (2023).
53. Graham DJ, Bierkens MFP, van Vliet MTH. Impacts of droughts and heatwaves on river water quality worldwide. *Journal of Hydrology* **629**, 130590 (2024).
54. Tian Y, Shen J, Feng J, Wang T, Jiao Y, Wang X. Research advancements on agricultural non-point source pollution in major lake and reservoir watersheds of China: Status, sources, monitoring, and prospects. *Ecological Indicators* **178**, 113981 (2025).
55. Wang S, Wang A, Yang D, Gu Y, Tang L, Sun X. Understanding the spatiotemporal variability in nonpoint source nutrient loads and its effect on water quality in the upper Xin'an river basin, Eastern China. *Journal of Hydrology* **621**, 129582 (2023).
56. Choi M*, et al.* Assessing sources of nutrients in small watersheds with different land-use patterns using TN, TP, and NO3-N. *Journal of Hydrology: Regional Studies* **55**, 101958 (2024).
57. Zhang H*, et al.* Changes in China's river water quality since 1980: management implications from sustainable development. *npj Clean Water* **6**, 45 (2023).
58. Kirk L, Compton JE, Neale A, Sabo RD, Christensen J. Our national nutrient reduction needs: Applying a conservation prioritization framework to US agricultural lands. *Journal of Environmental Management* **351**, 119758 (2024).

59. Essex B, Koop SHA, Van Leeuwen CJ. Proposal for a National Blueprint Framework to Monitor Progress on Water-Related Sustainable Development Goals in Europe. *Environmental Management* **65**, 1-18 (2020).
60. Yu J, Xian Q, Cheng S, Chen J. Horizontal ecological compensation policy and water pollution governance: Evidence from cross-border cooperation in China. *Environmental Impact Assessment Review* **105**, 107367 (2024).
61. Naylor A, Ford J, Pearce T, Alstine JV. Conceptualizing Climate Vulnerability in Complex Adaptive Systems. *One Earth* **2**, 444-454 (2020).
62. Murphy RR*, et al.* Nutrient improvements in Chesapeake Bay: Direct effect of load reductions and implications for coastal management. *Environmental Science & Technology* **56**, 260-270 (2022).
63. MEEC. Methods for Monitoring and Analysis of Water and Wastewater. *Beijing: China Environmental Science Press*, (2002).
64. Woodward G*, et al.* Continental-scale effects of nutrient pollution on stream ecosystem functioning. *Science* **336**, 1438-1440 (2012).
65. Liu D, Jiang X, Duan M, Yu S, Bai Y. Human and natural activities regulate organic matter transport in Chinese rivers. *Water Research* **245**, 120622 (2023).
66. Zheng Y, Li C, Wang Q, Yu J, Xu S, Li S. New modeling framework for describing the effects of landscape pattern changes on nutrient pollution transport. *Science of The Total Environment* **959**, 178090 (2025).
67. Jinge M*, et al.* Projected response of algal blooms in global lakes to future climatic and land use changes: Machine learning approaches. *Water Research* **271**, 122889 (2025).
68. Komisarczyk K, Kozminski P, Maksymiuk S, Biecek P. treeshap: Compute SHAP Values for Your Tree-Based Models Using the 'TreeSHAP' Algorithm. *https://doiorg/1032614/CRANpackagetreeshap*, (2024).

Figure legends

Fig. 1 Quantification of nutrient pollution in China. a: Site-based nutrient pollution index (NPI). c-f: Exceedance rates for each site against the Class III water quality standards for surface water [26] and site counts in each level.

Fig. 2 Intra-annual temporal variations and future evolution of nutrient pollution. a-e: Seasonal NPI across sites and related fluctuations. f-i: Four identified variation types by daily NPI values and specific pollutants. j-n: Predicted monthly NPI for 2040–2100 under four variation types. L1-L5 in Figs. a-d indicate 5 levels integrated four-season data with quantile-defined boundaries. R1-R5 in Fig. e represent seasonal fluctuation rate calculated from the seasonal maximum and minimum values.

Fig. 3 Spatial patterns and future evolution of nutrient pollution. a: Spatial hotspot analysis and cluster patterns of nutrient pollution. b: Spatial trajectory of pollution center from 2023 to 2100. c-g: Spatial distribution of annual NPI from 2023 to 2100. h: Spatial evolution degree of NPI determined by the coefficient of variation (CV) from 2023 to 2100.

Fig. 4 Driving mechanisms of spatiotemporal evolution in nutrient pollution. a-d: Key drivers for nutrient pollution in the four PFZs, respectively (see Fig. S28 for the complete factor ranking). a1-c4: Interaction effects of key drivers in the three hotspots PFZs, respectively. e: Key drivers contributing to increased pollution in coldspots. f: Statistics on key drivers in 2023 and 2100 in coldspots.

Acknowledgements
This research was funded by the Basic Science Center Project of the Natural Science Foundation of China (NSFC) (52388101), National Natural Science Foundation of China (No. 42407068), National Science and Technology Major Project of China (No. 2025ZD1204701), Innovative Research Group of the National Natural Science Foundation of China (No. 52221003), and Joint Funds of the National Natural Science Foundation of China (U2340219).

Contributions
X.Z. led data acquisition, conceptualization, methodology, data analysis and the writing of the first draft. D.Y. contributed equally to data acquisition, assisted in data analysis. B.G. guided the overall project scientifically and logistically. L.C. guided the overall project scientifically and logistically, including contributions in writing. H.Z. guided the editing of multiple versions, assisted in data analysis. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing financial or non-financial interests.

Supplementary information
Supplementary Figs. S1-S34, Tables S1-S6, and Text S1-S5.

Data availability
All data used in our synthesis are publicly available. Temperature data is available at https://doi.org/10.11888/Meteoro.tpdc.270961. Precipitation data can be accessed via https://doi.org/10.11888/Atmos.tpdc.300523. Evaporation and runoff from ERA5-Land dataset can be accessed via https://doi.org/10.24381/cds.68d2bb30. Soil_SOC, Soil_pH, Soil_N and Soil_caly can be accessed via https://doi.org/10.5194/soil-7-217-2021. The DEM data can be found at https://doi.org/10.5281/zenodo.14511570. The NDVI can be found at https://doi.org/10.11888/Terre.tpdc.300330. The GDP and fertilizer data were obtained from Statistical Yearbook via https://www.stats.gov.cn/. Population data can be accessed via https://doi.org/10.48690/1531770. Land-use/land-cover data can be freely accessed at https://doi.org/10.5281/zenodo.12779975. Data for future projections: Temperature is available at https://doi.org/10.11866/db.loess.2021.003; Precipitation is available at https://doi.org/10.11866/db.loess.2021.002; Evaporation can be found at and https://doi.org/10.11888/Atmos.tpdc.300558; GDP is available at https://doi.org/10.5281/zenodo.7898409; Population can be accessed via https://doi.org/10.1038/s41597-022-01675-x; Land-use/land-cover data can be freely accessed at https://doi.org/10.25584/data.2020-07.1357/1644253.