A multiple spatial scales water use simulation for capturing its spatial heterogeneity through cellular automata model

Jiayu Zhang[a], Dedi Liu*[,a,b,c], Jiaoyang Wang[a], Feng Yue[a], Hanxu Liang[a], Zhengbo Peng[a], Wei Guan[a]

[a] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China

[b]Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University, Wuhan, China

[c] Department of Earth Science, University of the Western Cape, Robert Sobukwe Road, Bellville 7535, Republic of South Africa

* Correspondence to: Dedi Liu: dediliu@whu.edu.cn

Abstract: Reliable water use simulation is essential for sustainable water resource planning, especially under intensifying pressures from climate change, population growth, and socio-economic transitions. While previous studies have extensively explored water availability as supply side modeling across multiple spatial scales for its spatial heterogeneity, the water demand side remains relatively underdeveloped—often constrained by fixed spatial scales and coarse statistical data that assume spatial homogeneity. This mismatch between supply side and demand side limits the ability of existing models to accurately represent spatial heterogeneity in water use and brings uncertainty into water resource allocation strategies. To address this mismatch, we propose a novel multi-scale water use simulation framework by integrating cellular automata (CA) model with Generalized Likelihood Uncertainty Estimation (GLUE). The CA model captures the spatial heterogeneity of water use through the grid-based update rules. Two update rules are adopted—probability rule (i.e., capturing stochastic transitions via distribution fitting) and linear rule (i.e., modeling neighborhood-weighted evolution). To evaluate the impacts of spatial scale on water use heterogeneity, simulations are conducted at three spatial scales: 1 km, appropriate scale, and prefecture scale across 341 prefectures in China. Results show that both the update rule and spatial scale significantly affect spatial heterogeneity and uncertainty of water use. The probability rule can capture the broader variability but results in higher Root Mean Squared Error (*RMSE*) and Relative Error (*RE*) while the linear rule brings more stable performance with lower errors. While the 1 km scale increases uncertainty due to sensitivity to local fluctuations, and the prefecture scale suppresses spatial details, the appropriate scale offers the best trade-off between stability and spatial heterogeneity. The uncertainty quantified by GLUE, expresses as confidence intervals, varies across prefectures and spatial scales. Overall, the proposed framework offers a flexible tool for multi-scale water use simulation and highlights the critical role of spatial heterogeneity, thereby supporting adaptive water resource planning and management.

Key words: ~~demand-side~~ water ~~modeling~~use; spatial scale; ~~water resources management; water scarity assessment~~cellular automata; multi-scale simulation

1 Introduction

Water scarcity has become one of the most pressing global challenges, exacerbated by climate change, population growth, and unsustainable water use practices (Avargani et al. 2022, Huang et al. 2021, Kaewmai et al. 2019, Rosa et al. 2020). Nearly 2.3 billion people is living in regions facing water scarcity (Brunner et al. 2019, Dolan et al. 2021, Mekonnen and Hoekstra 2016). In this context, accurate and timely assessments of water scarcity are essential for effective water management, resource allocation, and policy-making (Avargani et al. 2022, Cao et al. 2018). A scientifically rigorous water scarcity assessment requires a comprehensive understanding of both available water resources and water use (Brunner et al. 2019, Ji et al. 2025, Sun et al. 2022). However, these two components are often represented at incompatible spatial scales, resulting in mismatches that complicate accurate evaluation and integrated water resources planning (Almino and Rufino 2021, Kang et al. 2017).

In the past few decades, significant advancements in hydrological modeling, satellite remote sensing, and reanalysis datasets have enabled researchers to simulate surface and ground available water resources across various temporal and spatial scales—from daily basin-scale runoff forecasts to long-term continental water balance projections (Horta et al. 2024, Su et al. 2024, Yang et al. 2021, Zhang and Long 2021). These advancements have laid the foundation for widely used tools such as SWAT, VIC, H08, and PCR-GLOBWB, which help water managers better understand available water resources under different spatial-temporal scales (Hersbach et al. 2020, Kovacevic et al. 2020, Noori and Kalin 2016, Sunkara and Singh 2022). Though there has been water use simulation progress, it still faces significant challenges and has yet to achieve the same spatial level of sophistication as the simulation of available water resources. The primary limitations of water use simulation lies in the available spatial scales of water use data (Hou et al. 2024, Zhang et al. 2023). Much

of the existing researches rely on a coarse, aggregated dataset such as national statistics, sectoral usage reports, or administrative boundaries (e.g., counties, provinces) (Carvalho et al. 2021, Wu et al. 2022, Zhang et al. 2023). These datasets inherently assume uniform water use within each administrative unit, and often overlook the spatial heterogeneity of water use. This oversimplification limits the ability of water use simulations to capture spatial variation, particularly in regions with pronounced heterogeneity (Brunner et al. 2019, Su et al. 2024).

Simulating water use at a grid scale provides a promising solution to address the spatial scale mismatch, allowing for a more accurate representation of water use dynamics and capturing regional variations (Su et al. 2024, Wu and Lu 2021). A grid-based approach allows for a more accurate representation of water use dynamics by capturing fine-scale spatial heterogeneity, which administrative survey data often overlook. Zhang et al. (2023) integrated the Iterative Input Selection algorithm with Convolutional Neural Networks to simulate annual irrigation water use, producing high-resolution grid maps with a spatial resolution of 1 km for mainland China. Hou et al. (2024) developed a monthly dataset on industrial water withdrawals, incorporating spatial resolutions of 0.1° and 0.25° by utilizing enterprise data, product yields, and water use statistics. These studies have demonstrated that the downscaled water use data from administrative level to grid scale can catch a more detailed and region-specific representation of water use patterns, thereby improving their applicability in water resource management and water scarcity assessment. However, these studies still rely on a fixed spatial grid scale, limiting their ability to capture the full spatial heterogeneity of water use. The effects of spatial resolution on the representation of water use heterogeneity remain insufficiently explored. In particular, finer scales (such as 1 km grids) can capture localized variations in water use, coarser scales (such as regional or administrative boundaries) tend to smooth over spatial differences, potentially obscuring critical patterns (Luo et al. 2020, Sun et al. 2022). Therefore, simulating water use across multiple spatial scales is essential

for capturing the full spectrum of spatial heterogeneity. By incorporating the results at different spatial scales, this simulation can account for the fine-scale dynamics of water use in urban, agricultural, and industrial areas more effectively, offering a comprehensive understanding of water use patterns and improving the accuracy of water scarcity assessments (Su et al. 2024, Sunkara and Singh 2022).

Since the spatial scale of water use simulation depends on the spatial scale of the input data (Horta et al. 2024, Sharifi et al. 2021, Zhang et al. 2023), achieving multi-scale water use simulation requires a model that can flexibly handle different input spatial scales. Such a model should be adaptable to varying spatial resolutions and ensure accurate simulation of water use, whether the focus is on fine-scale urban areas, intermediate regional levels, or broader national assessments. The cellular automata (CA) model, with its grid-based structure, offers a suitable framework for spatially explicit modeling across multiple scales by adjusting both the spatial resolution of input data and the design of updating rules (Al-Shaar et al. 2022, Sapino et al. 2023, Tariq et al. 2023, Wang et al. 2020). The CA model can thus be employed to support multi-scale water use simulation. And as uncertainties are inherently associated with the selection of spatial scales and update rules (Yin et al. 2020, Zhang and Long 2021), it is critical to quantify and address these uncertainties to ensure the reliability of simulation outcomes.

The aim of this research is to develop a multi-scale water use simulation framework that specifically addresses the impact of spatial scale on the spatial heterogeneity of water use. The framework is mainly composed of the CA model for simulating water use dynamics across various spatial scales and an uncertainty analysis technique for quantifying the uncertainties of the simulation. The proposed framework will not only facilitate a deeper understanding of the influences of spatial scale on water use heterogeneity in diverse regions, but also improves the accuracy of water scarcity assessments and supports more effective resource management. Ultimately, the proposed framework contributes to advancing sustainable water governance in

areas facing pronounced water stress.

2. Methodology

To develop a multi-scale water use simulation model, the dynamic spatial scale simulation capabilities of the CA model should be firstly leveraged. The particularly advantage of CA model is modeling complex global dynamics through simple local interactions and transition rules (Al-Shaar et al. 2022, Liu et al. 2021, Tariq et al. 2023). Moreover, the grid-based structure of the CA model allows it to flexibly accommodate various spatial resolutions, making it well-suited for modeling water use at multiple spatial scales (Sapino et al. 2023, Wang et al. 2020). In this framework, water use grid maps at different spatial scales are first prepared as inputs to the CA model. Each grid is treated as an individual cell in the CA model, with the water use amount at each cell representing its state. Based on the neighborhood configuration and the initial baseline year water use data, water use values for future years are simulated using predefined update rules for each cell at multiple spatial scales. The procedure of our proposed CA model for multi-scale water use simulation is depicted in Figure 1.

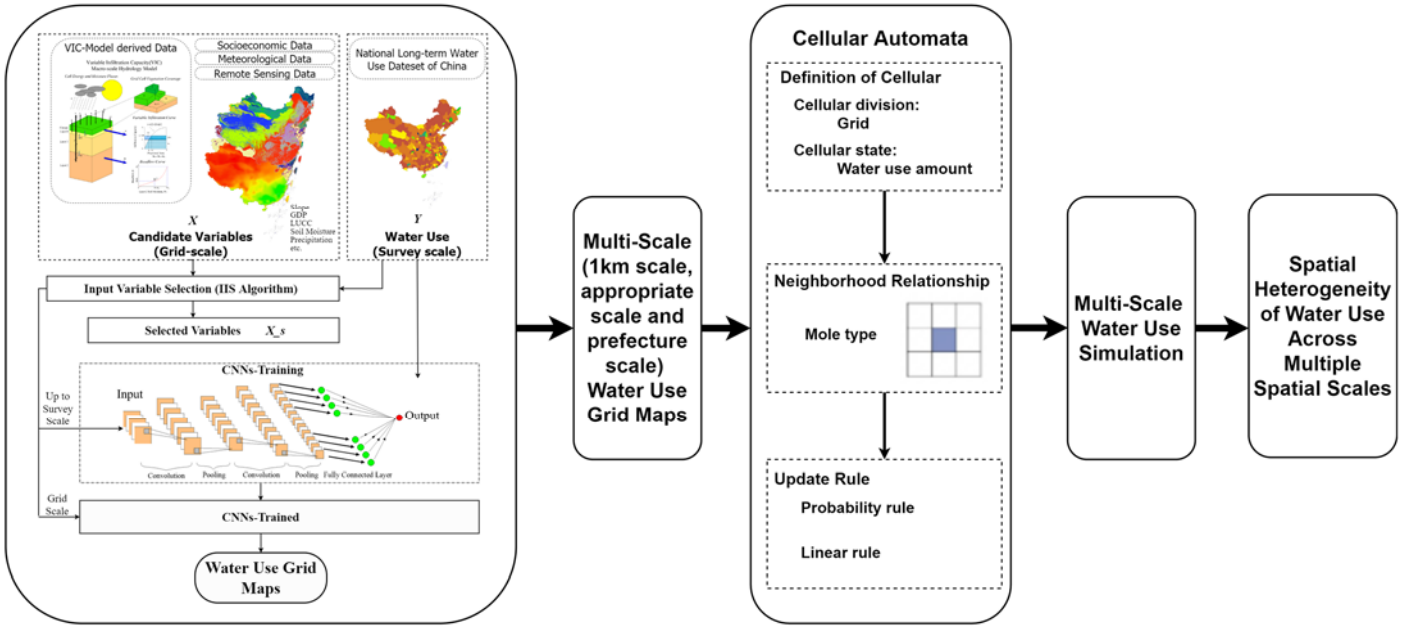


Figure 1. Procedure of the multi-scale water use simulation

To capture the spatial heterogeneity of water use dynamics at different scales, the Coefficient of Variation (CV) (Abdi 2010) and Moran's I spatial autocorrelation index (Tiefelsdorf et al. 1995) are employed here. The CV measures the relative variability of water use across regions, with higher values indicating greater spatial heterogeneity. Moran's I assesses the degree of spatial autocorrelation, identifying whether similar values are spatially clustered (positive values) or dispersed (negative values). These two indices are applied to evaluate the spatial heterogeneity of water use grid maps at different spatial scales (i.e., 1 km, appropriate scale, and prefecture scale), and figure out the influences of the spatial variations on water use dynamics and uncertainty in the simulation results.

To quantify the uncertainties of water use simulation for providing a more robust foundation for water scarcity assessment and policy-making (He et al. 2018, Knox et al. 2018, Sharifi et al. 2021), the Generalized Likelihood Uncertainty Estimation (GLUE) method is adopted here. GLUE provides a probabilistic framework for model evaluation by exploring a wide range of parameter sets and assigning likelihoods to each based on their performance. This approach effectively addresses model equifinality—where multiple parameter combinations yield similar outputs—and is particularly well-suited for complex, non-linear models such as cellular automata (Liu et al. 2015, Yin et al. 2020).

2.1 Water Use Grid Maps generating

The spatial scale of water use simulation is determined by the spatial scale of the input data, so water use grid maps at different spatial scales ~~should be~~were prepared as input to the simulation model. Here, water use refers to the total water consumption across three major sectors: irrigation, industrial, and urban domestic water use. The water use data considered in this study account for both groundwater and surface water sources (e.g., rivers, lakes, and reservoirs). These data were drawn from Zhou et al. (2020), which compiled water use

data across 341 Chinese prefectures. The dataset includes water consumption data for irrigation, industrial, and domestic uses, incorporating both groundwater and surface water sources. The water use data were sourced from two major nationally coordinated surveys: the First and Second National Water Resources Assessment Programs (1965–2000) and the Water Resources Bulletins published by 31 provincial governments (2001–2013). Both surveys were led by the Ministry of Water Resources of China, and followed consistent methodologies in terms of definitions, survey units, sector classifications, field measurements, and quality assurance. To obtain the water use grid maps, several steps should be done to convert the water use data at administrative survey scale into ~~the varying~~ spatially explicit ~~scales. First, the water use data at administrative survey scale is processed and downscaled to grid-based formats by the iterative input selection (Galelli and Castelletti 2013) and Convolutional Neural Networks (CNN) algorithms. The iterative input selection algorithm is used to select the most relevant variables for water use, while the CNN model captures the relationships between input variables and water use (Zhang et al. 2023). Next, the appropriate spatial scale for water use simulation is identified using an end-to-end deep learning-based spatial scale adaptive selection model (Liu et al. 2022, Zhang et al.~~ grids of varying resolutions~~2025). Finally, water use grid maps are obtained at small spatial scale (e.g., 1km), and then the appropriate spatial scale and prefecture scale~~.

The grid maps of irrigation, domestic, and industrial water use are generated from the prefecture-level statistical survey data and water use sector-specific predictor variables. For each sector, the most relevant input variables are identified through an iterative input variables selection algorithm (Zhang et al., 2023; Zhang et al. 2025). Specifically, irrigation water use was modeled by the potential evapotranspiration, normalized difference vegetation index (*NDVI*), rainfall and soil moisture; domestic water use was modeled by population, rainfall, temperature and night-light; industrial water use was modeled by *GDP*, night-light, population and rainfall. And then these sectoral gird maps were aggregated to form total water use grid maps for modeling.

This aggregation is done for two reasons: the first one is that the temporal trend of the total water use has become stable during the study period (1998–2013) and future. The average annual growth rate is only about 0.87% (from 505.53 billion m^3 in 1998 to 575.44 billion m^3 in 2013) due to the policy interventions, technological improvements, and industrial structure changes. Since the primary objective of our study is to examine the influence of spatial scale on the spatial heterogeneity of water use, a temporally stable indicator helps minimize the confounding effects of sector-specific temporal fluctuations; the second reason is that the total water use can figure out the scale effects across regions instead of the sector-level temporal variability while the sectoral differences are implicitly in the inputs before the aggregation.

Earlier studies often applied a fixed spatial resolution in different regions, which could not account for differences in land area, natural endowments, and water use structures, and leaded to the discrepancies in information density and potential over- or underestimation of water use. To address this issue, an appropriate spatial scale can be determined by the deep learning-based spatiotemporal scale adaptive selection model (Liu et al., 2021; Zhang et al., 2025). And the model can balance the accuracy of the simulation based on the spatial information density of gridded water use data, and its results vary across prefectures. The spatial scale selection module in the selection model figures out the appropriate spatial scale by maximizing information density while balancing simulation accuracy in terms of the Conditional Entropy, Kullback–Leibler Divergence Loss and Relative Error performance metrics. This selection module enables each prefecture to adopt its own appropriate spatial scale rather than a fix resolution. The detailed values of the appropriate spatial scale (in km) for each prefecture and water use sector (irrigation, industrial, and domestic) are provided in an accompanying Excel file, which can be accessed and downloaded via the data link: 10.6084/m9.figshare.30445157. This file allows users to examine the spatial heterogeneity of the appropriate scale across regions in detail. Finally, total water use grid maps are generated at three spatial resolutions: the

small scale (e.g., 1 km), the appropriate spatial scale as determined by the selection module, and the prefecture scale as the statistical survey water use data.

2.2 Cellular automata model for water use simulation

The CA model, grounded in complexity theory, is widely used in land use and urban growth modeling. It provides a robust platform for simulating spatial phenomena governed by local interactions and transition rules (Sapino et al. 2023, Tariq et al. 2023). Each cell in a CA model represents a discrete spatial unit that updates its state over time based on predefined rules and the states of its neighboring cells. It's decentralized, bottom-up modeling structure enables the simulation of complex global behaviors emerging from simple local dynamics (Al-Shaar et al. 2022, Wang et al. 2020). ~~This model often employs two distinct update rules to simulate water use over time: the probability rule and the linear rule. Both rules account for spatial interactions between neighboring cells and allow for the simulation of water use at multi-scale, capturing the heterogeneity of water use across different geographical regions. By applying both rules independently, the effectiveness of update rules in simulating water use at multi-scale is assessed~~The CA model is introduced to simulate the temporal evolution of water use at the grid scale, complementing the static spatial distribution obtained from the Convolutional Neural Network (CNN) downscaling. While the CNN model effectively reconstructs the spatial pattern of water use for each prefecture based on physical and socioeconomic predictors (Zhang et al., 2023), it does not explicitly account for the spatial dependence and interactions among adjacent grid cells. The CA model addresses this limitation by incorporating spatial adjacency effects and feedback mechanisms, allowing each cell's water use to be influenced by its neighbors. This enables the model to represent the diffusion and clustering behaviors of water use, which are essential for capturing the spatial heterogeneity and dynamic interactions of human water activities.

Both the probability and the linear update rules are designed and tested to capture the dual nature of water use dynamics. The probability rule has been widely applied in significant spatial and temporal variation areas in land use simulation and other fields. It will be designed here for the water use at different scales. Rather than assuming temporal stability, the probability rule explicitly incorporates the variations through calibrating the state transition matrix and probability distributions for each prefecture independently by the own historical water use record. This rule enables the simulation to capture both the structured temporal dependence and the inherent randomness in water use, ensuring adaptability to local conditions. The linear update rule assumes that changes in water use are more deterministic and can be approximated as a linear combination of the cell's own state and those of its neighbors. This rule is more appropriate for long-term, high spatial autocorrelation, and persistent patterns. After implementing and comparing the water use simulation results of the two rules in the CA framework, their results can assess the relative effectiveness of stochastic versus deterministic update mechanisms across different spatial scales. These two rules not only strengthen the robustness of the modeling framework but also provide insights into the dominant processes shaping water use dynamics in different regions.

2.2.1 Probability rule in CA

~~The probability rule in the CA model is designed to capture the stochastic nature of water use transitions.~~ ~~The~~The probability rule in the CA model is designed to represent the stochastic state transitions of water use over time. It abstracts the temporal dynamics of water use at the grid level into a probabilistic transition framework that can be applied consistently across different spatial scales and regions, while remaining adaptable to significant spatial and temporal variations. This adaptability is achieved by calibrating the update rule separately for each prefecture by its own historical water use record for appropriately capturing both long-

term trends and localized fluctuations. In this approach, the state of each grid cell (i.e., representing the amount of water use) is divided into k distinct intervals using equal-frequency categorization ~~according to its~~based on the cell's historical ~~data~~water use record. This categorization ensures that the intervals reflect the variations in water use over time. For each interval, the most suitable statistical distribution is selected using the Akaike Information Criterion (AIC). ~~The~~The selection process enables the model to represent the probabilistic characteristics of water use within each intensity class. And the distribution is chosen from a set of candidate distributions, including normal, lognormal, exponential, gamma, and uniform.~~–~~

Once the ~~most appropriate~~optimal distribution is found for each interval, a state transition matrix is ~~created~~constructed based on ~~the transition~~observed transitions of ~~states~~grid cells between intervals from one year to the next. The transition matrix ~~indicates~~captures the ~~probabilities~~likelihood of a grid ~~cell's~~ cell moving from its current water use state ~~transitioning~~ to another ~~state (i.e., another interval)~~ in the ~~following year. Using these transition probabilities, the model predicts which interval each grid cell's state will fall into at the next~~ subsequent time step. ~~A random sample is~~ The model first generates the next state probabilistically through the transition matrix, and then ~~drawn~~the water use samples are generated from the corresponding probability distribution ~~of~~. These two steps incorporate both the ~~predicted interval, providing a predicted water use value for the next time step. This approach introduces variability, reflecting~~structured temporal dependence and the inherent randomness ~~and uncertainty~~ in future water use patterns.

In the probability rule, the calibrated parameter is the number of state intervals and is denoted as k. The value of k directly affects the granularity of the state categorization and the accuracy of the state transition matrix. A larger k increases the resolution of the state representation and captures the finer variations in water use, but a larger k can also lead to overfitting. And a smaller k oversimplifies the demand pattern. To calibrate the parameter k, the historical and observed water use data is divided into a calibration and a validation sets.

And the performances of the model with different k values is then evaluated by the Root Mean Squared Error (*RMSE*) and Relative Error (*RE*) metrics. The optimal k can be calibrated by the minimums of the *RMSE* and *RE* in the validation period, ensuring a balance between model accuracy and generalizability.

2.2.2 Linear rule in CA

The linear rule in CA updates the water use of each cell according to the weighted average of its current state and the states of its neighboring cells. The linear rule assumes that the water use at a given grid cell is influenced by both its own previous state and the water use of adjacent cells. The linear rule is expressed as the Equation (1).

$$W(t+1) = \alpha W(t) + \beta \sum_{i=1}^{n} \omega_i W_i(t) \tag{1}$$

where $W(t+1)$ is the predicted water use of the central cell at time $t+1$, $W(t)$ is the current water use of the central cell at time t, $W_i(t+1)$ represents the water use at the i^{th} neighboring cell ($i=1,2,\ldots,8$) at time t, ω_i is the weight that is assigned to the i^{th} neighboring cell, α and β are coefficients that control the relative influence of the central cell's own water use and the neighboring cells' water use.

To accurately capture the influences of neighboring cells, the weight ω_i for each neighboring cell is determined by an inverse distance weighting scheme with an exponential decay and is expressed as the Equation (2).

$$\omega_i = \frac{1}{d_i^p} \tag{2}$$

where d_i is the Euclidean distance between the central cell and the i^{th} neighboring cell, p is an adjustable exponent that controls the rate of decay in the influence of neighboring cells.

To determine the parameters α, β, and p, the observed period is also divided into a calibration period and

a validation period. After the model runs with different parameter combinations, and their performances are assessed in terms of *RMSE* and *RE*, the optimal parameter set will be picked out by the minimum of the performance metrics in the validation period.

2.3 Generalized Likelihood Uncertainty Estimation for water use simulation

Generalized Likelihood Uncertainty (GLUE) is a probabilistic approach that evaluates a model's performance by considering a wide range of plausible parameter sets and quantifying the likelihood of each set in its ability to reproduce observed data (Chen et al. 2011, Sharifi et al. 2021, Taormina et al. 2016). Unlike traditional deterministic calibration methods, GLUE acknowledges the inherent equifinality in environmental modeling that also is the possibility of the acceptable results from the multiple parameters set. The GLUE is incorporated into the CA model to assess the uncertainty in water use simulations.

The GLUE is applied in both the probability rule and the linear rule of the CA model to evaluate the uncertainty of water use simulations. There are six steps. (1) Parameter Selection: the number of state intervals k for the probability rule**,** and the self-influence coefficient α, the neighboring influence coefficient β, and the distance decay exponent p for the linear rule; (2) Parameter Sampling: a large number of parameter sets are generated by the uniform sampling within specified ranges for each parameter; (3) Model Simulation: the CA model is executed by each parameter set to simulate water use for the target period; (4) Likelihood Estimation: *RMSE* and *RE* values are calculated for each simulation to evaluate how well the simulated results match the observed data; (5) Behavioral Parameter Identification: behavioral parameter sets can yield acceptable likelihood values according to the thresholds that is determined by the calibration data; (6) Uncertainty Quantification: from the range of outputs from the behavioral parameter sets, a prediction interval is constructed quantify the uncertainty associated with the water use projections.

2.4 Spatial Heterogeneity of water use

To comprehensively understand the variability and spatial relationships of water use patterns across different spatial scales, Coefficient of Variation (CV) (Abdi 2010) and Moran's I (Tiefelsdorf et al. 1995) are used to analyze the spatial heterogeneity of water use. CV is defined as the ratio of the standard deviation to the mean of water use values at each spatial scale, offering a normalized measure of dispersion. It is a primary indicator used to quantify variability in water use across grid cells (Canchola et al. 2017). A higher CV value indicates greater spatial variation in water use, while a lower CV suggests more uniform water use patterns across the region. Once the CV for each grid cell is calculated, the extent of water use spatial heterogeneity is assessed, and areas with high variability are identified as more susceptible to water stress and fluctuations in demand (Botta-Dukát 2023, Liu et al. 2020).

Moran's I can reflect the spatial autocorrelation by measuring the degree to which one grid cell's water use is similar to that of neighboring grid cells. Moran's I provides an overall measure of spatial dependence, where a positive Moran's I indicates a clustering of similar water use values (either high or low) in neighboring grid cells, while a negative Moran's I suggests a dispersed pattern. A value close to zero indicates a random spatial pattern with no significant clustering (Gedamu et al. 2024, Shortridge 2007). By calculating Moran's I across different spatial scales, we are able to detect whether water use patterns exhibit spatial clustering or they are more randomly distributed. Moran's I can therefore help us understand the regional patterns of water use and identify areas that require targeted management interventions (Fu et al. 2024, Yamada 2024).

Both the CV and Moran's I provide a robust framework to analyze spatial heterogeneity in water use. The CV provides a measure of variability, while Moran's I reveals the extent of spatial correlation in water use. They offer a more comprehensive understanding of the spatial dynamics across different regions.

3 Study area and datasets

3.1 Study Area

Situated on the northwestern shore of the Pacific Ocean, China boasts vast territory, a large population, and diverse climate conditions (Ji et al. 2025, Sun et al. 2022). Over the years, the volume of water use in China has surged significantly rising from 3305 km^3 in 1965 to 5925 km^3 in 2024 (Ji et al. 2025). Owing to the country's varied climate conditions and pronounced spatiotemporal heterogeneity of water resources, water scarcity has become a recurrent challenge, exacerbated by the looming threats of climate change and rapid socio-economic development (Hou et al. 2024, Wang et al. 2021). Notably, China is expected to experience greater impacts from climate change than the global average (Kang et al. 2017, Sun et al. 2022, Yan et al. 2019). Therefore, conducting a multi-scale water use simulation study in China is particularly meaningful due to the country’s complex water resource challenges, amplified by regional disparities, climate change, and rapid socio-economic development. China comprises 31 provincial-level administrative divisions, including provinces, autonomous regions, and municipalities. As outlined in the study by Zhou et al. (2020), the administrative divisions are categorized at the prefecture level, totaling 341 prefectures (Figure. 2).

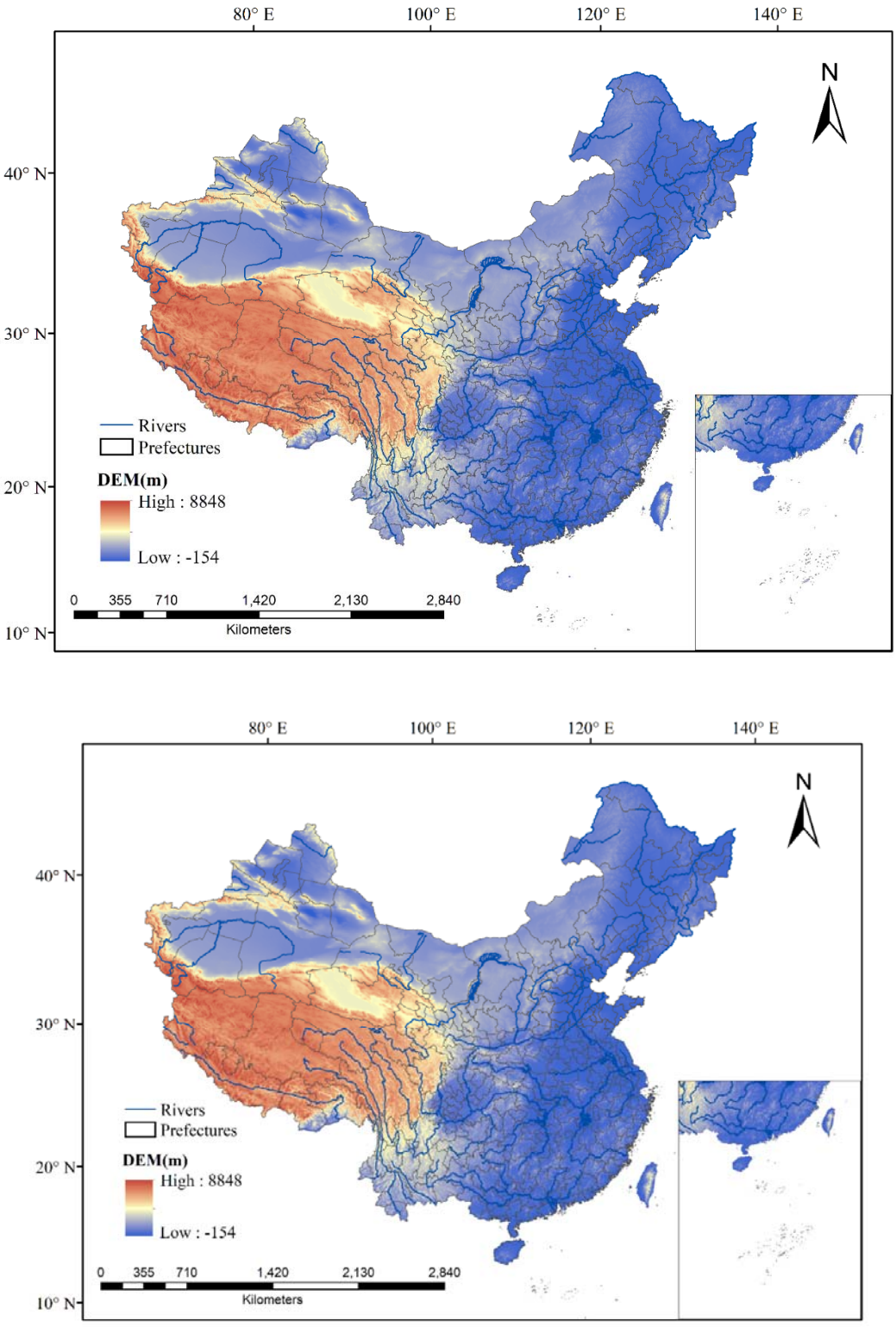


Figure. 2 Prefectures and major rivers in study area

3.2 Datasets and Preprocessing

To obtain water use grid maps, the observed datasets related to irrigation water use, domestic water use, and industrial water use were collected. The observed datasets are composed of the annual water use statistical survey data at the administrative scale, soil moisture data derived from hydrological models, socio-economic data like *GDP* (Gross Domestic Product) and population, meteorological data from point observations, and satellite remote sensing data including the normalized difference vegetation index (*NDVI*) and night light data.

Various spatial interpolation and downscaling methods are employed to transform the datasets into different spatial scales. To simulate water use at different spatial scales, we transform the spatial scale of the dataset to 1km, appropriate spatial scale (Zhang, et al., 2025) and prefecture scale. The details of the transform methods can be found in the reference of Zhang et al. (2023). Nighttime light remote sensing data is also combined with the land use information to perform upscaling or downscaling at spatial scales (Ye et al. 2021).All the adopted datasets and their corresponding preprocessing methods are listed in Table 1.

Table 1. Datasets and corresponding preprocessing methods

Variables	Data Sources	Origin Data Format	Origin Resolution	Processing Method
Precipitation (P)	CMA*	Point observation	\	Interpolation (ANUSPLIN)
Temperature (T)	CMA*	Point observation	\	Interpolation (IDW)
LUCC	RESDC**	Grid	30 m × 30 m	Upscale (Resample)
NDVI	RESDC**	Grid	1 km × 1 km	Upscale/downscale (Resample)
GDP	RESDC**	Grid	1 km × 1 km	Upscale/downscale (Resample)
Population	RESDC**	Grid	1 km × 1 km	Upscale/downscale (Resample)
DEM	SRTM	Grid	1 km × 1 km	Upscale/downscale (Resample)
Potential Evapotranspiration (PET)	\	\	\	P-M Equation/Interpolation (IDW)
Soil moisture (SM)	GDFC*** (He and Sheffield 2020)	NetCDF	0.25° × 0.25°	Downscale (Machine learning)(Guevara and Vargas 2019)
Night light (N-L)	DMSP-OLS	Grid	2.7 km × 2.7km	Upscale/downscale (Resample)
National Long-term Water Use Dataset of China	FigShare-PNAS(Zhou et al. 2020)	Excel	\	\

CMA*: China Meteorological Administration (http://data.cma.cn/); RESDC**: Resource and Environment Science and Data Center (https://www.resdc.cn/).

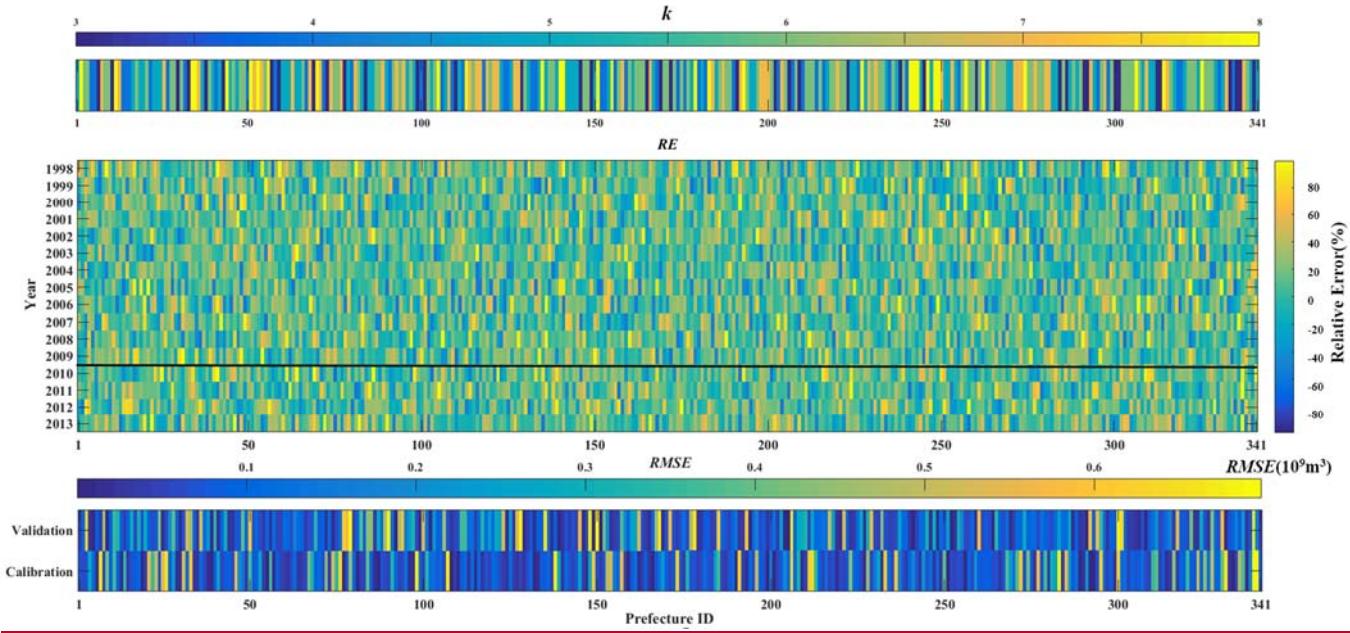
GDFC***: Global Drought and Flood Catalogue (http://hydrology.princeton.edu/data).

4. Results

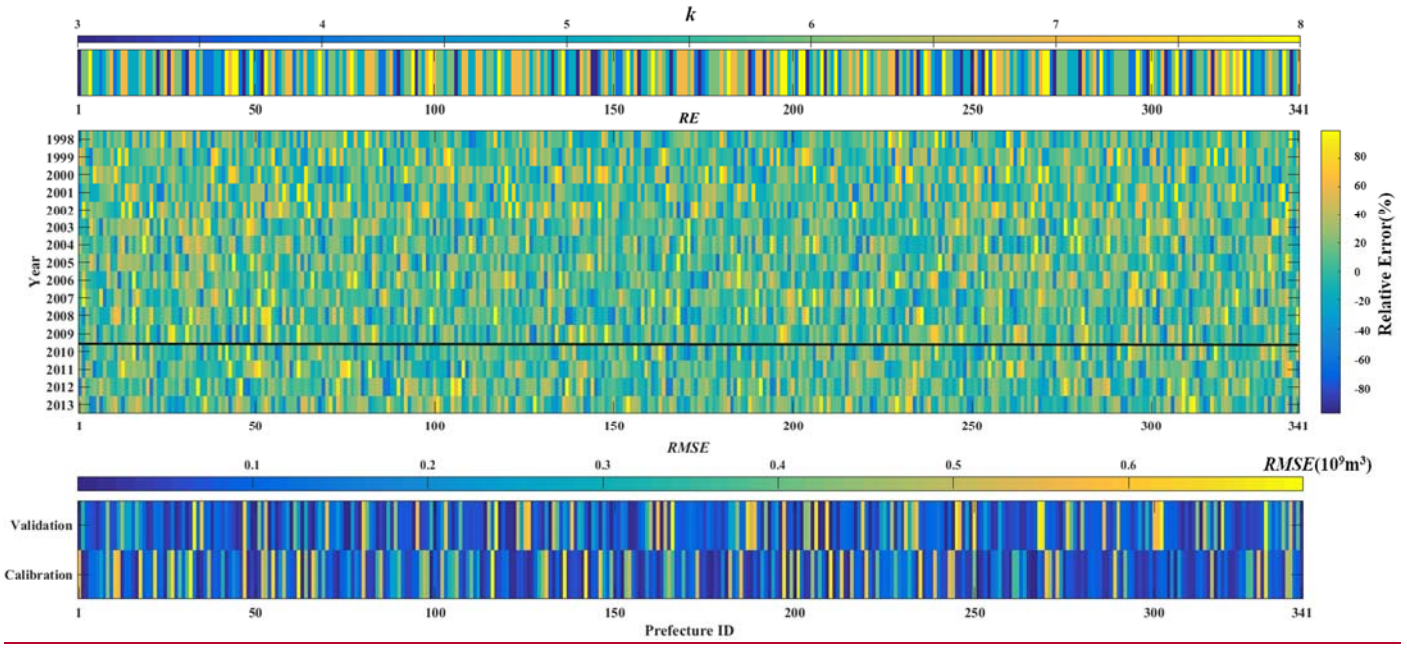
4.1 Water use simulation by CA model

4.1.1 Water use simulation from the probability rule CA model

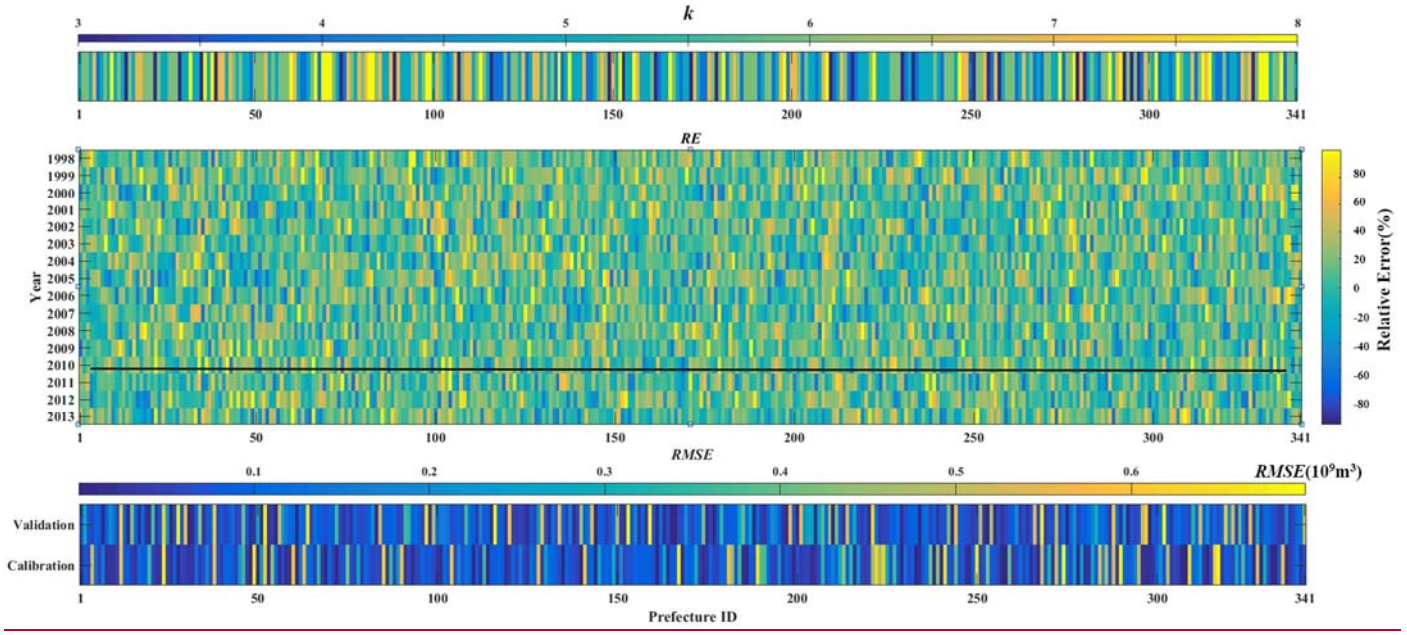
~~As~~In the CA model with the probability rule, the number of state intervals (k) is the only parameter to be calibrated. The dataset from 1998–2009 is used for calibration and 2010–2013 for validation, with *RMSE* and *RE* as performance metrics. The optimal value of k at three spatial scales (1 km, appropriate scale, prefecture scale) for each prefecture is determined by minimizing *RMSE* and *RE* in the validation period. The calibrated k values for each prefecture, along with the corresponding *RMSE* and *RE* during the calibration (1998–2009) and validation (2010–2013) periods at the three spatial scales, are presented in Figure 3.



(a)



(b)



(c)

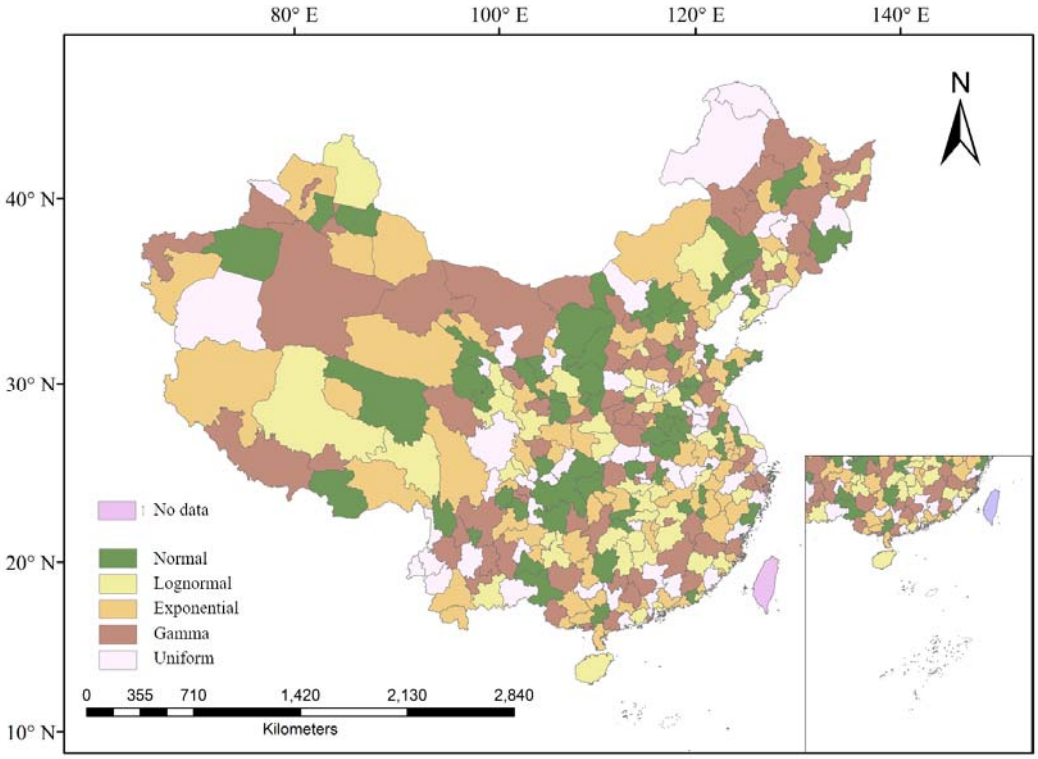
Figure .3 Optimal parameters of the probability rule CA model and the model performances at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

According to the results shown in Figure 3, the calibrated parameter *k* exhibits clear spatial heterogeneity across prefectures and varies with spatial scales. At the 1 km scale (Figure 3(a)), most prefectures show *k* values concentrated around 5–6, corresponding to relatively low *RMSE* and *RE* values. This suggests that a

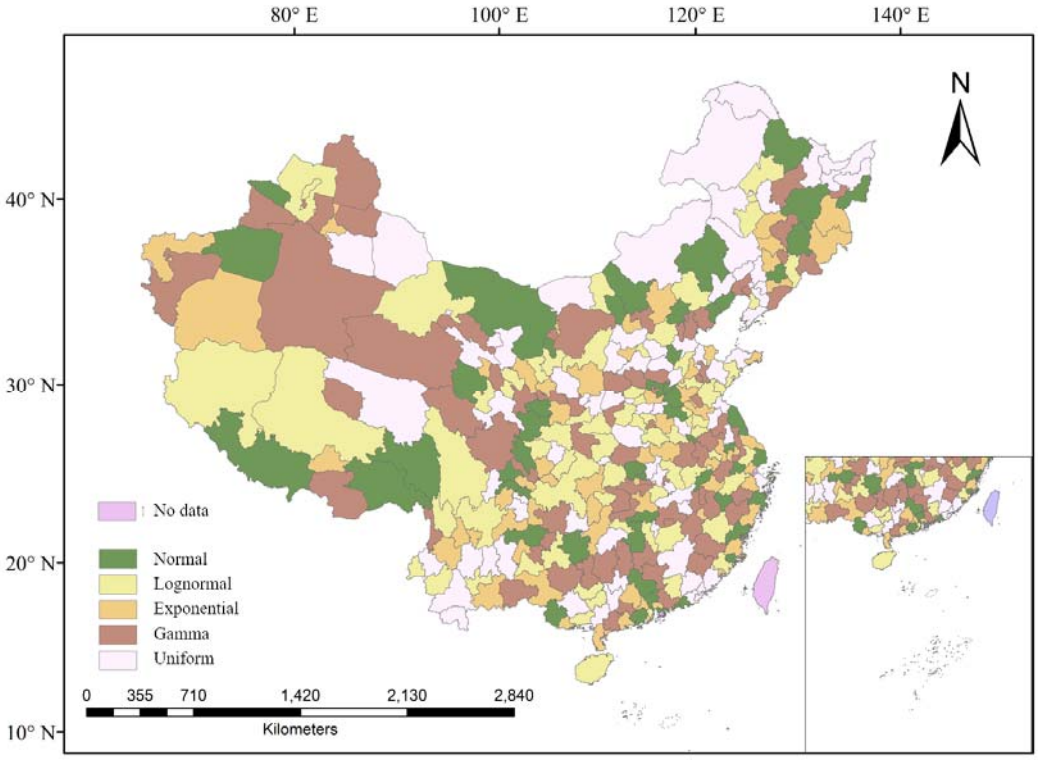
moderate number of state intervals can effectively capture local water use variability while avoiding overfitting. In these areas, the probability distributions and transition probabilities appear to reflect stable temporal patterns, resulting in more accurate simulations. At the appropriate spatial scale (Figure 3(b)), the distribution of *k* becomes more diversified among prefectures. Some regions require larger *k* values (>=7) to preserve finer distinctions in water use states, while others perform better with smaller *k* values (<=4) that smooth out excessive variability. Different from the results at the 1 km scale, the overall *RMSE* and *RE* values are slightly higher, indicating that while the appropriate scale balances detail and generalization, it may not fully capture abrupt local changes in some prefectures. At the prefecture scale (Figure 3(c)), *k* values are generally smaller (mostly 3–4), reflecting the reduced spatial detail at coarser resolution. Thus, their accuracies of the simulation decrease, with higher *RMSE* and *RE* values. When the input variability of small scale is strongly aggregated at large scales, fewer state intervals oversimplify the temporal transitions, leading to greater deviations from observed water use patterns.

After determining the optimal *k* values at each scale, the next step is to characterize the statistical nature of water use within each state interval. The Akaike Information Criterion (AIC) is taken as performance metric to select the most suitable probability distribution for each interval in every prefecture. The AIC can balance the model fitness and the complexity ~~by~~through penalizing ~~the number of parameters to prevent overfitting, it has been widely applied in probability distributions selection. Here, the AIC is employed to determine the most suitable probability~~ excessive parameters, it can reduce the risk of overfitting. The selected distribution types not only fit the historical data well but also is used to generate the future scenarios. The complete AIC values and corresponding best-fitting distribution types for each prefecture are provided in an open-access Excel file, which can be downloaded from: 10.6084/m9.figshare.30445157. The results of the optimal probability distributions for water use grids ~~across various prefectures. And the optimal probability~~

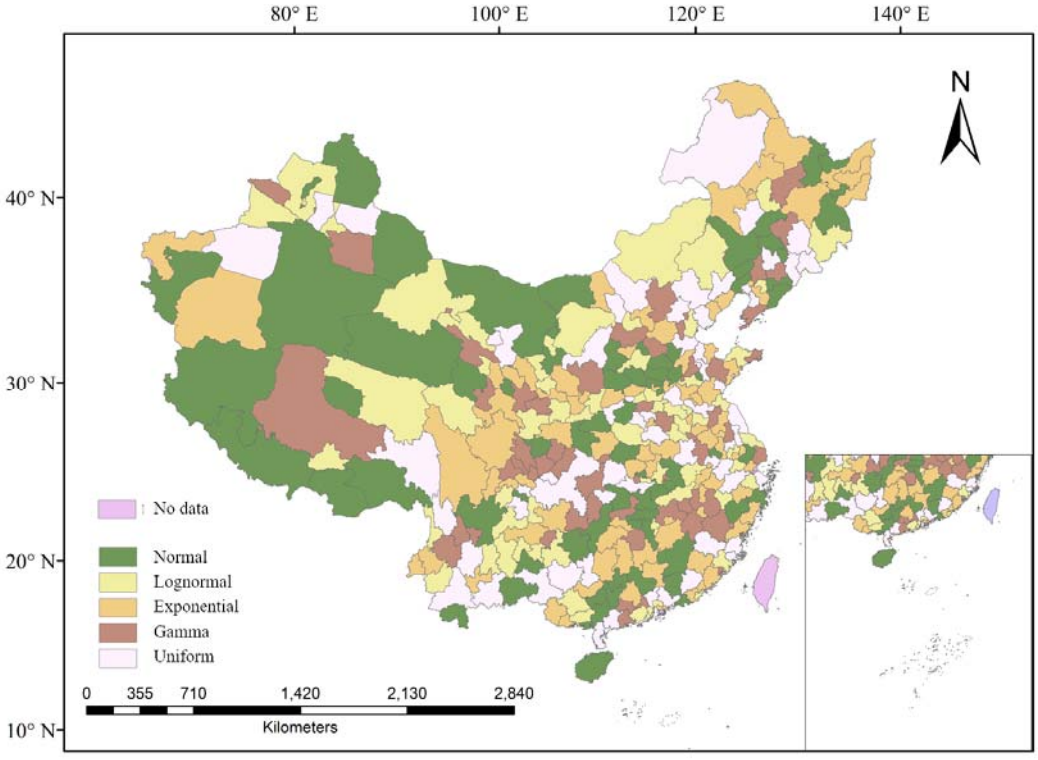
~~distributions of water use grid at three~~ at the three different spatial scales (i.e., ~~1km~~1 km scale, appropriate spatial scale, and prefecture scale) are ~~identified as illustrated~~shown in Figure ~~3~~4. These distributions, combined with the calibrated *k* values, form the basis of the probability rule CA model's ability to reproduce the spatial and temporal heterogeneity of water use.



(a)



(b)



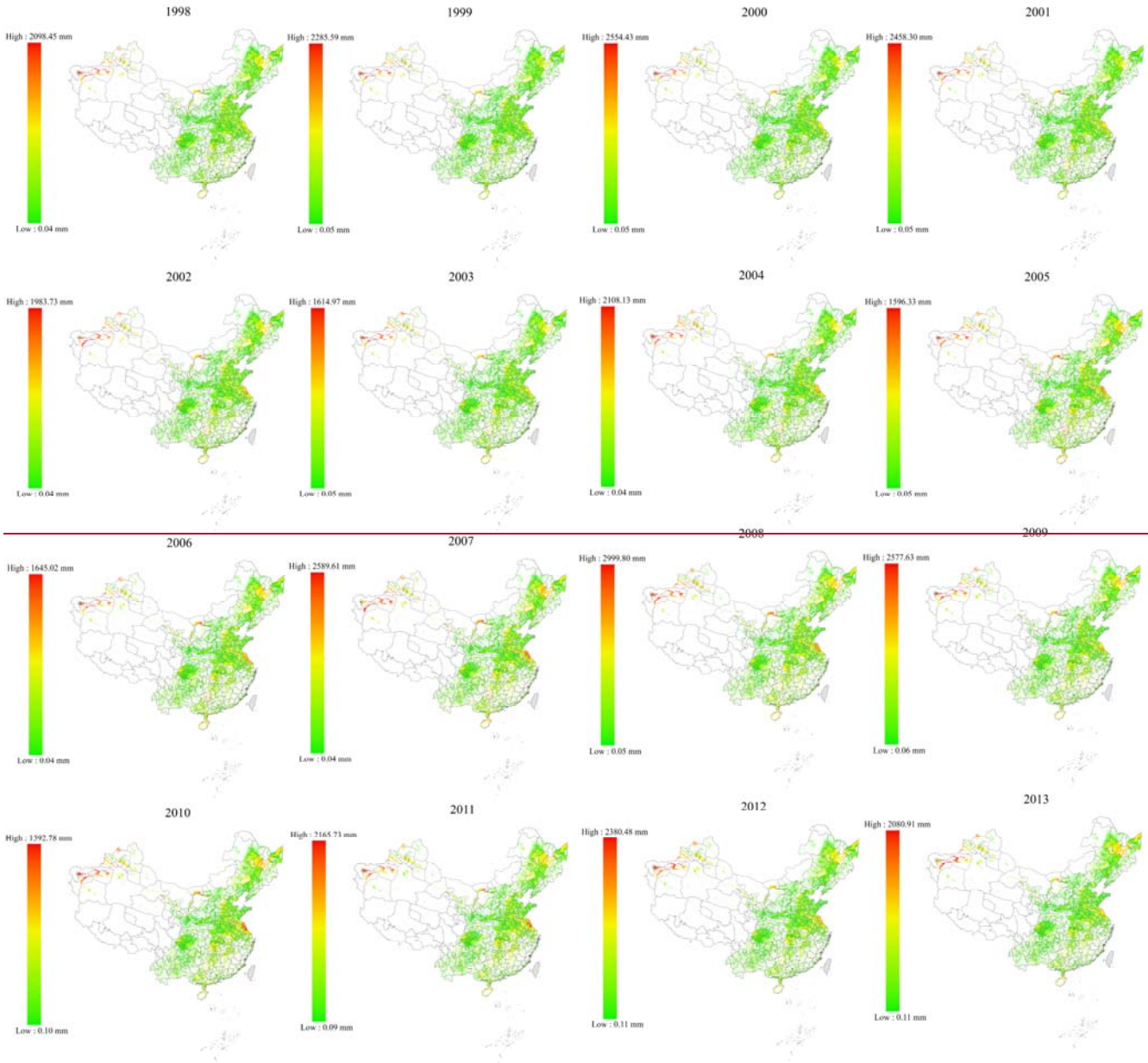
(c)

Figure. ~~34~~ Probability distribution types of water use at three spatial scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

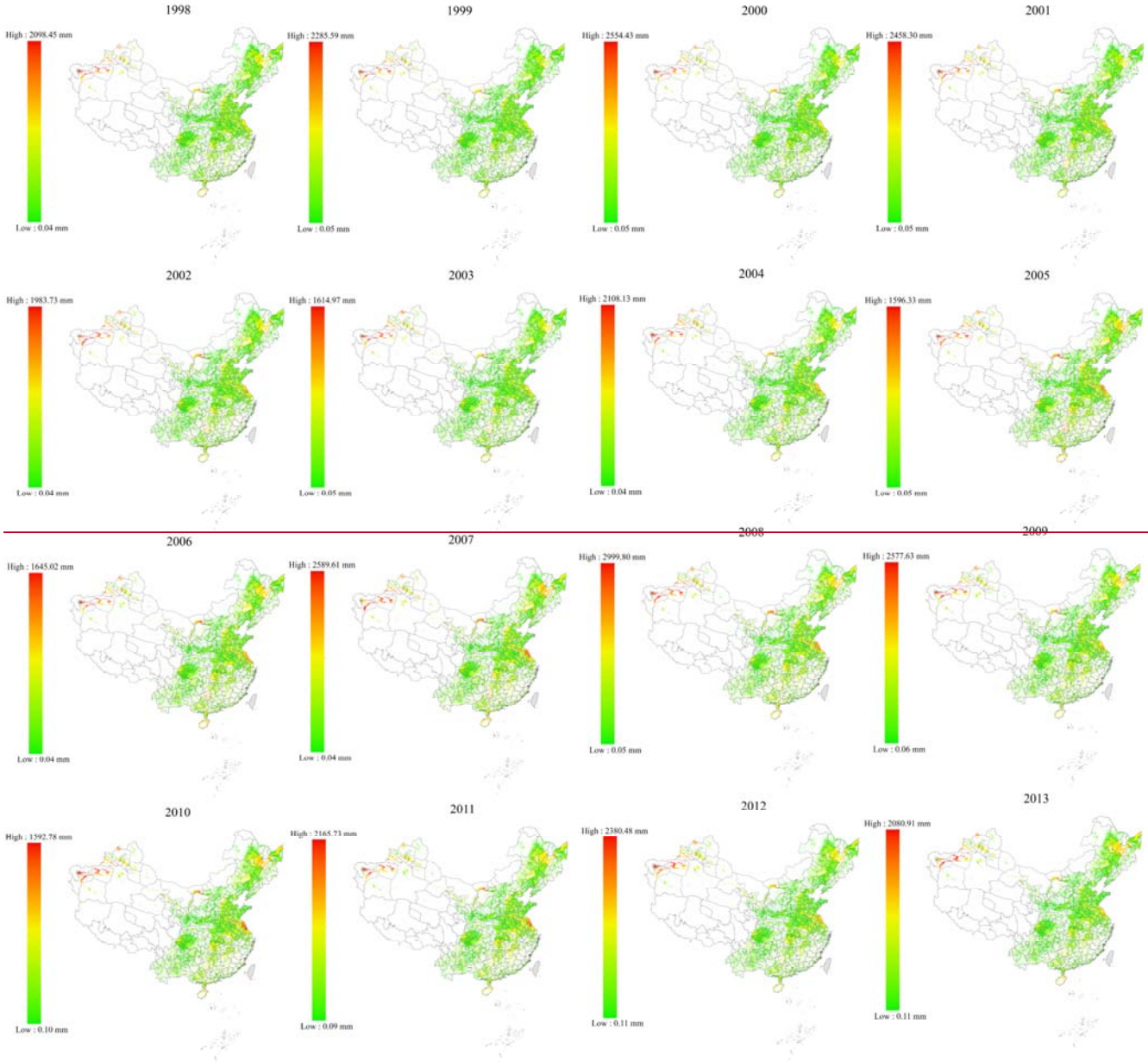
The optimal probability distributions of water use grid maps across three spatial scales (i.e., 1 km scale, appropriate spatial scale, and prefecture scale) were identified and shown in Figure ~~3~~4. The distribution types of water use reflect the underlying dynamics of water use at various scales. At the 1 km scale (as shown in Figure ~~3~~4 (a)), most areas show a predominance of normal and log-normal distributions (i.e., in green and light yellow), indicating more stable and symmetric water use patterns. Water use in these areas are consistent with relatively socio-economic and climatic conditions, where are less fluctuated. However, the water use distributions are more frequently exponential or gamma (i.e., in light orange and brown) in areas with rapidly urbanization or industrialization in the central and eastern parts of China. And thus, water use are higher variability and stochasticity. The distribution types of water use at the appropriate spatial scale (as shown in Figure ~~3~~4 (b)) are similar with those at 1 km scale as shown in Figure ~~3~~4 (a). More exponential and gamma distributions can be found in the significant agricultural activity areas at the appropriate spatial scale due to the irregular water use patterns driven by seasonal fluctuations and economic activities. When the spatial scale is up from 1 km to the appropriate ones, more nuanced understanding of water use dynamics can capture the impacts of local factors such as crop irrigation and industrial processes. Although the exponential distribution are found to be one of the prevalent types in regions with irregular or rapidly changing water use patterns at the prefecture scale as shown in Figure ~~3~~4 (c), uniform and gamma distributions have appeared more frequently at this bigger prefecture scale. The transition to coarser spatial resolution leads to reduced spatial variability in water use, and the heterogeneity of water use is potentially oversimplified.

Based on the historical water use record from 1998 to 2013, a transition matrix ~~at a~~for each grid cell ~~can be determined~~was constructed to represent the probability of transition from ~~the~~its current state to ~~its following time.~~ the subsequent year. According to ~~the~~these transition ~~matrixes for each grid cell~~matrices, the interval for the next state can be predicted~~. A~~, and a random sample is ~~then~~ drawn from the ~~predicted interval~~corresponding

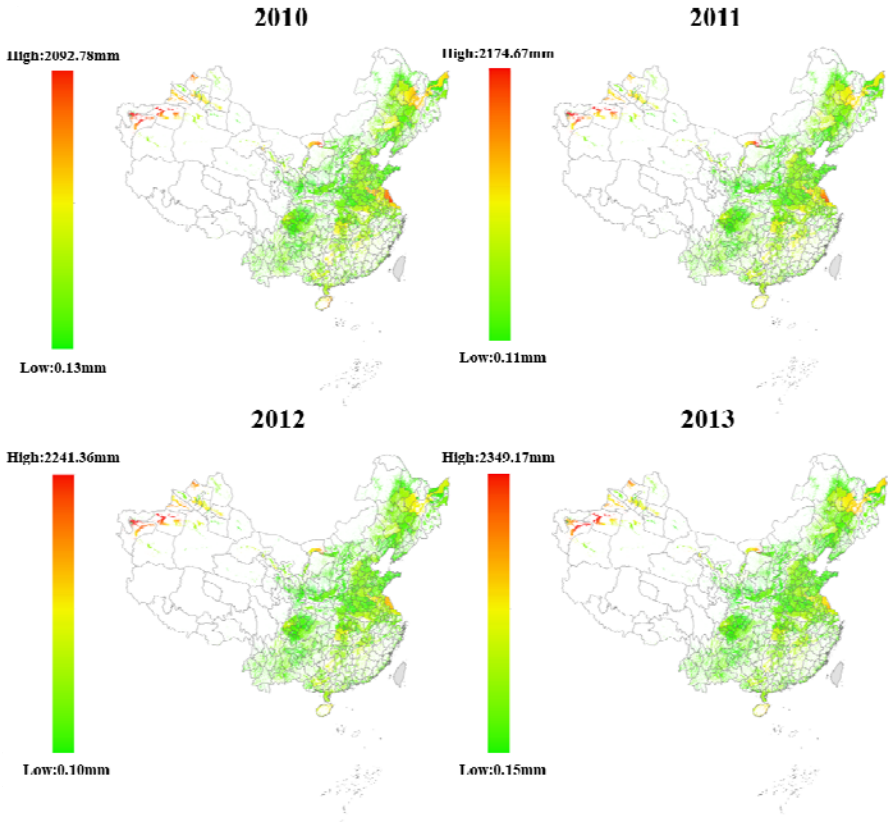
probability distribution to obtain the ~~predicted~~simulated water use value~~ for the following year. To generate the simulated water use, the~~. The probability rule is applied at three ~~different~~ spatial scales (i.e., 1 km, appropriate scale, and prefecture scale~~), and their~~). To improve clarity, only the simulated water use results for the validation period (2010–2013) are ~~shown as~~displayed in Figure ~~4.~~



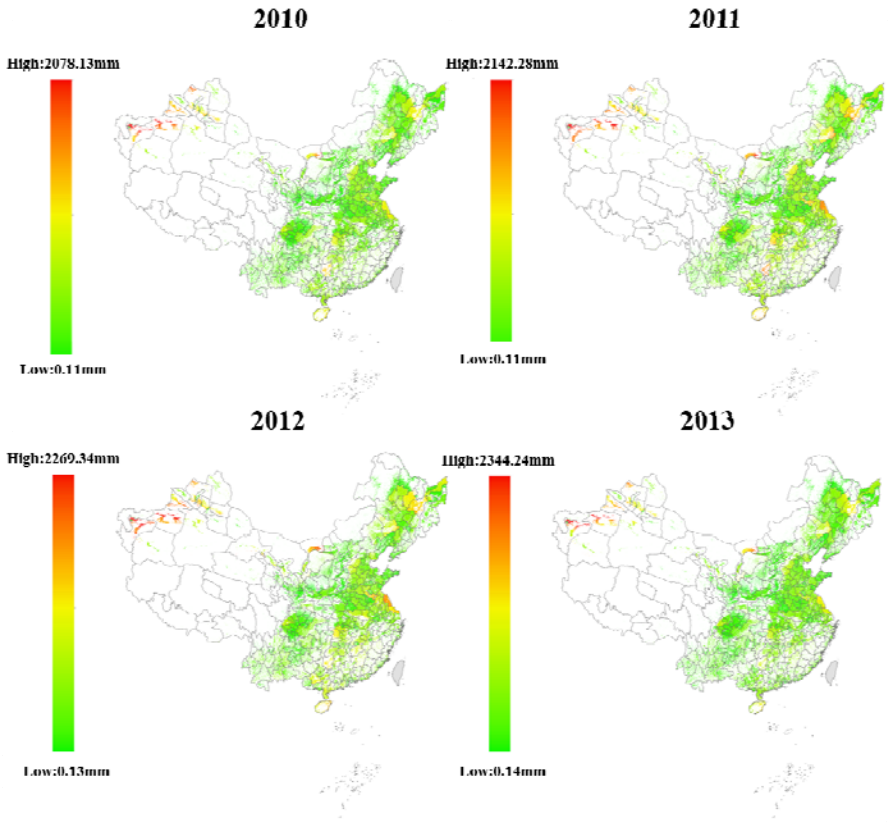
~~(~~5, enabling a~~)~~ direct comparison of spatial patterns across scales during independent validation. The complete annual simulation results for the entire study period (1998–2013)—including all three spatial scales and both update rules—are available for download at: 10.6084/m9.figshare.30445157.


1998
High : 2098.45 mm
Low : 0.04 mm
1999
High : 2285.59 mm
Low : 0.05 mm
2000
High : 2554.43 mm
Low : 0.05 mm
2001
High : 2458.30 mm
Low : 0.05 mm
2002
High : 1983.73 mm
Low : 0.04 mm
2003
High : 1614.97 mm
Low : 0.05 mm
2004
High : 2108.13 mm
Low : 0.04 mm
2005
High : 1596.33 mm
Low : 0.05 mm
2006
High : 1645.02 mm
Low : 0.04 mm
2007
High : 2589.61 mm
Low : 0.04 mm
2008
High : 2999.80 mm
Low : 0.05 mm
2009
High : 2577.63 mm
Low : 0.06 mm
2010
High : 1592.78 mm
Low : 0.10 mm
2011
High : 2165.73 mm
Low : 0.09 mm
2012
High : 2380.48 mm
Low : 0.11 mm
2013
High : 2080.91 mm
Low : 0.11 mm

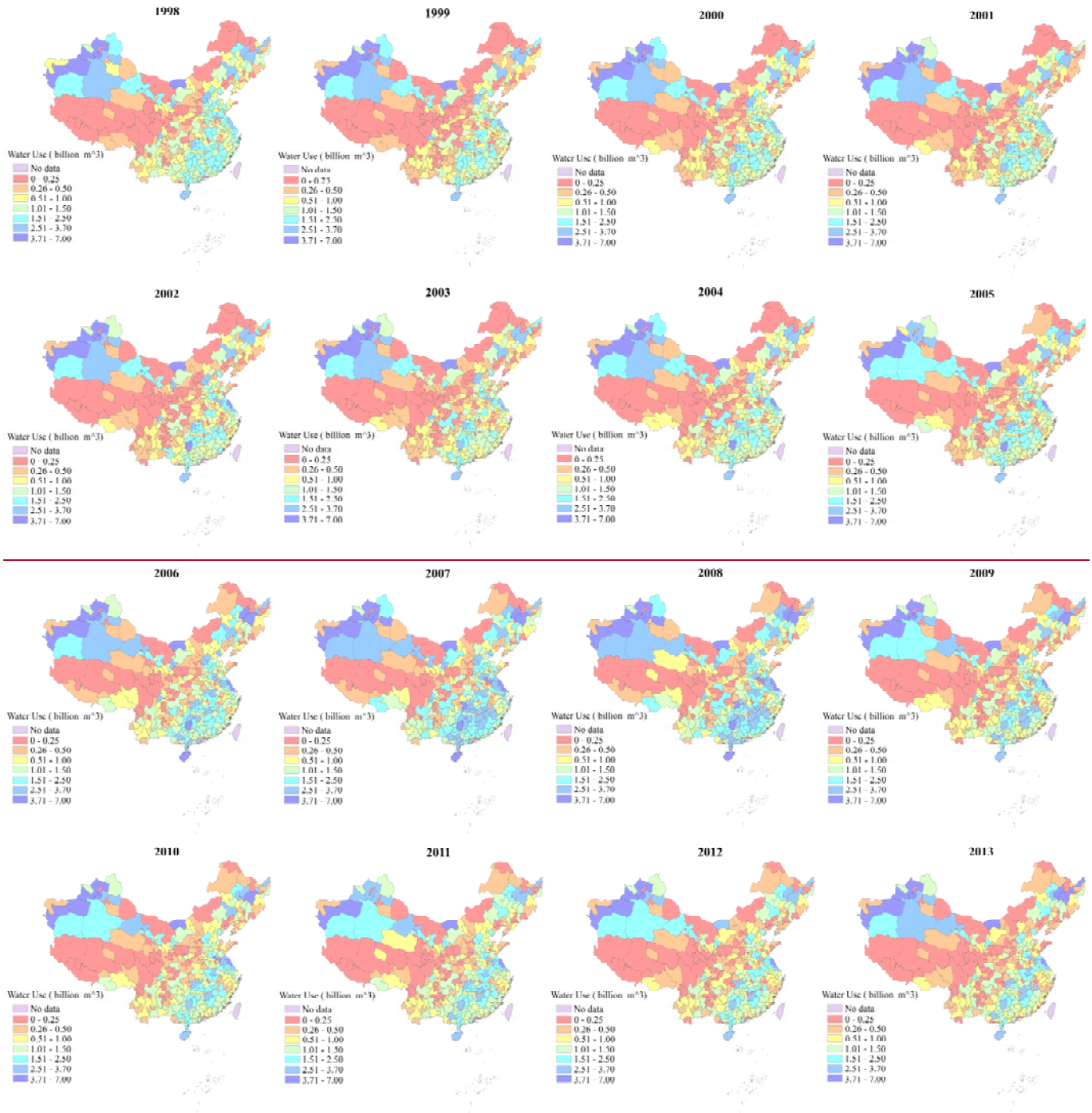
429


2010
High:2092.78mm
Low:0.13mm
2011
High:2174.67mm
Low:0.11mm
2012
High:2241.36mm
Low:0.10mm
2013
High:2349.17mm
Low:0.15mm

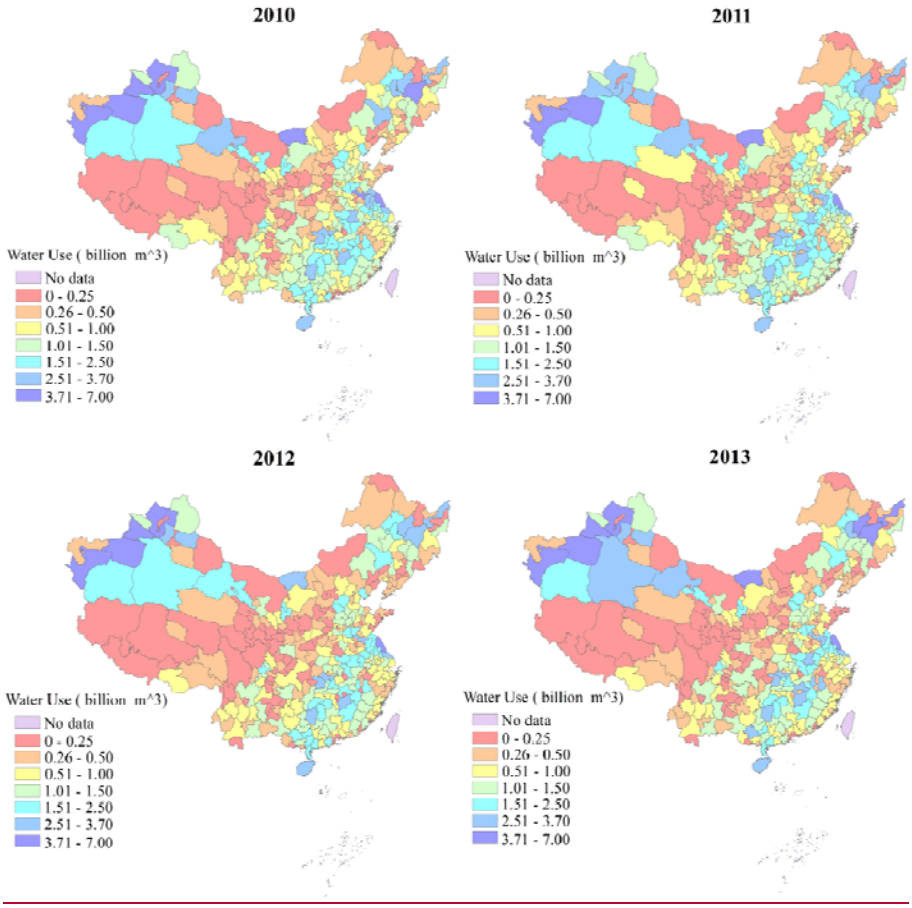
(a)


2010
High:2078.13mm
Low:0.11mm
2011
High:2142.28mm
Low:0.11mm
2012
High:2269.34mm
Low:0.13mm
2013
High:2344.24mm
Low:0.14mm

(b)


1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Water Use (billion m^3)
No data
0 - 0.25
0.26 - 0.50
0.51 - 1.00
1.01 - 1.50
1.51 - 2.50
2.51 - 3.70
3.71 - 7.00

434



(c)

Figure. 45 Water use simulation results from the probability rule at three different scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

Water use simulation maps from 1998 to 2013 through the probability rule, has clearly highlight the spatial and temporal heterogenic at different spatial scales. The maps reveal the local increases in water use at the 1 km scale (as shown in Figure 45 (a)), particularly in major metropolitan areas in eastern and central parts of China. These increases can be found across the whole time of the selected series driven by the population growth, industrial expansion, and urbanization. The results at the appropriate spatial scale (as shown in Figure 45 (b)) are different from those at the 1 km scale, their local variations become more homogeneous and stationary. It demonstrates that the model can capture the macro-level variations across different geographical areas. The water use trends are further found to be flatter with less spatial heterogeneity at the prefecture scale (Figure 45 (c)). The local variations of water use cannot be reflected at the prefecture scale due to the coarser

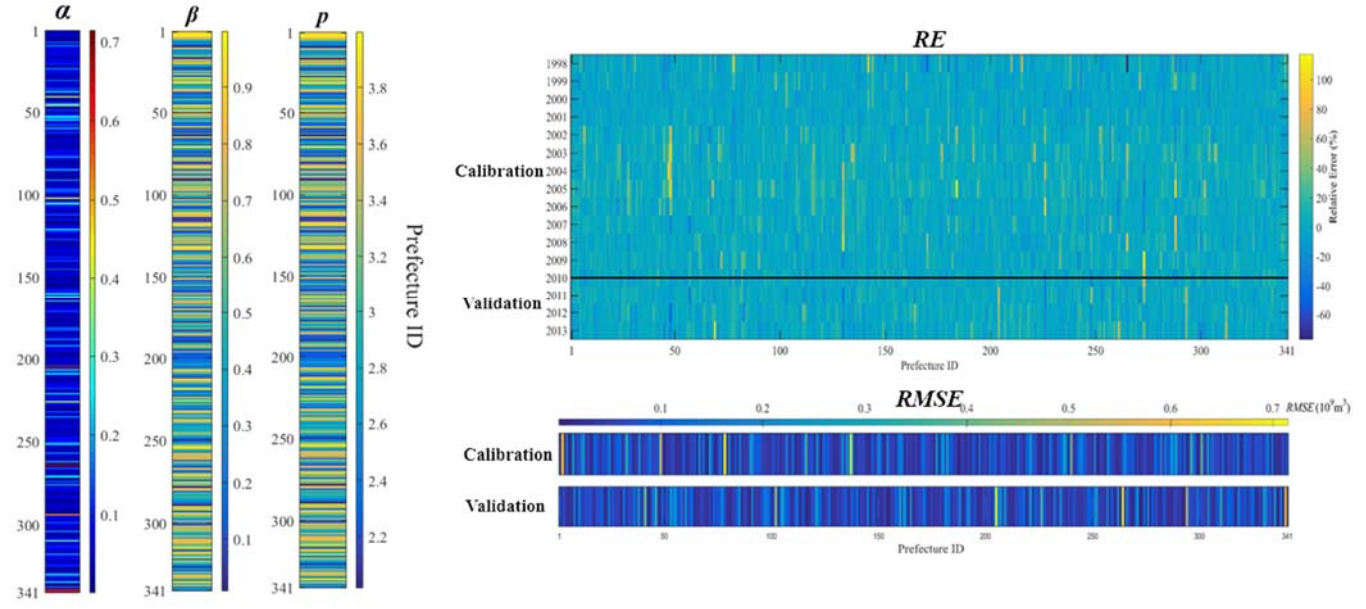
spatial resolution. Areas with stable water use, such as the northern provinces, show a more uniform distribution of water use, indicating that coarser resolution tend to mask these local variations.

The water use results at the three different spatial scales from probability rule show that the probability rule CA model can effectively capture the spatial heterogeneity in water use. The incorporation of transition probabilities and statistical distributions helps account for the temporal water use variations. However, water uses in the areas with experiencing rapid changes show great fluctuations and spatial fragmentation, and they are hard to be simulated by the probability rule. As the linear rule CA model can provide a more stable alternative in rapid change areas, it can be implemented in water use simulation, too.

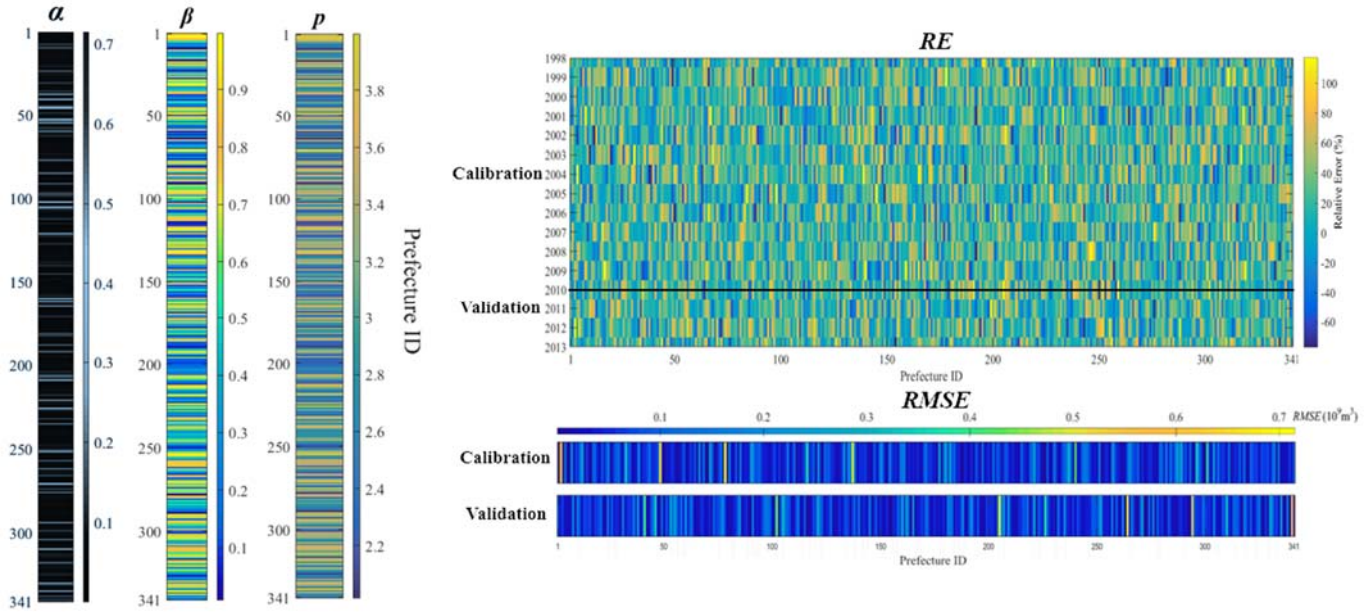
4.1.2 Water use simulation from linear rule CA model

There are three parameters to be calibrated in the linear rule CA model: the self-influence coefficient α, the neighboring influence coefficient β, and the spatial decay exponent p. ~~The dataset from 1998–2009 is for calibrating the model while the dataset from 2010–2013 is for its validation. And *RMSE* and *RE* are taken as the performance evaluation metrics. After minimizing the *RMSE* and *RE* metrics, the three optimal parameter values of the model at the three different scales~~The calibration and validation are taken the statistics water use at the prefecture-level as the reference (i.e., observed water use data). Specifically, for a given parameter set, the gridded water use is firstly simulated. The simulated grids are then aggregated into the total water use at each prefecture scale along their boundaries. These total water uses are assessed by the observed water use data from water resources bulletins and related statistical surveys. The calibration period covers 1998–2009 and the parameter values are determined by minimizing *RMSE* and *RE* between the simulated and observed total water use at the prefecture scale. The validation period covers 2010–2013 and the performance of the model is also assessed by *RMSE* and *RE*. The optimal parameters at three spatial scales during the calibration

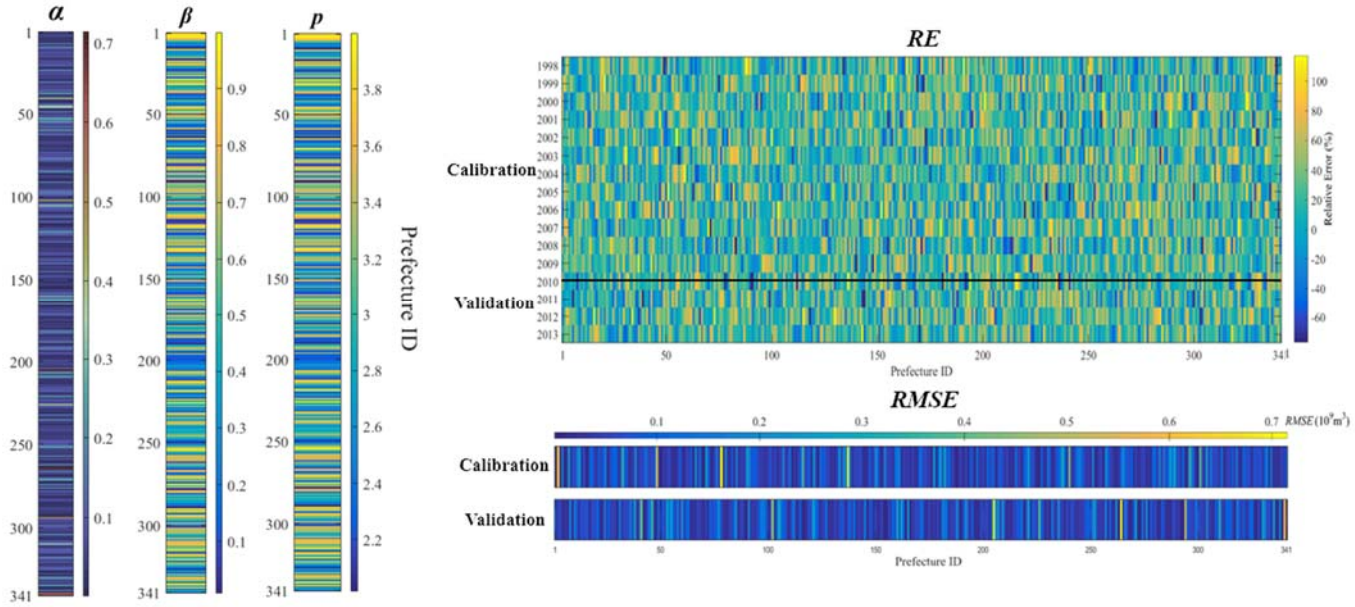
and validation periods, are illustrated in Figure 5.



(a)



(b)

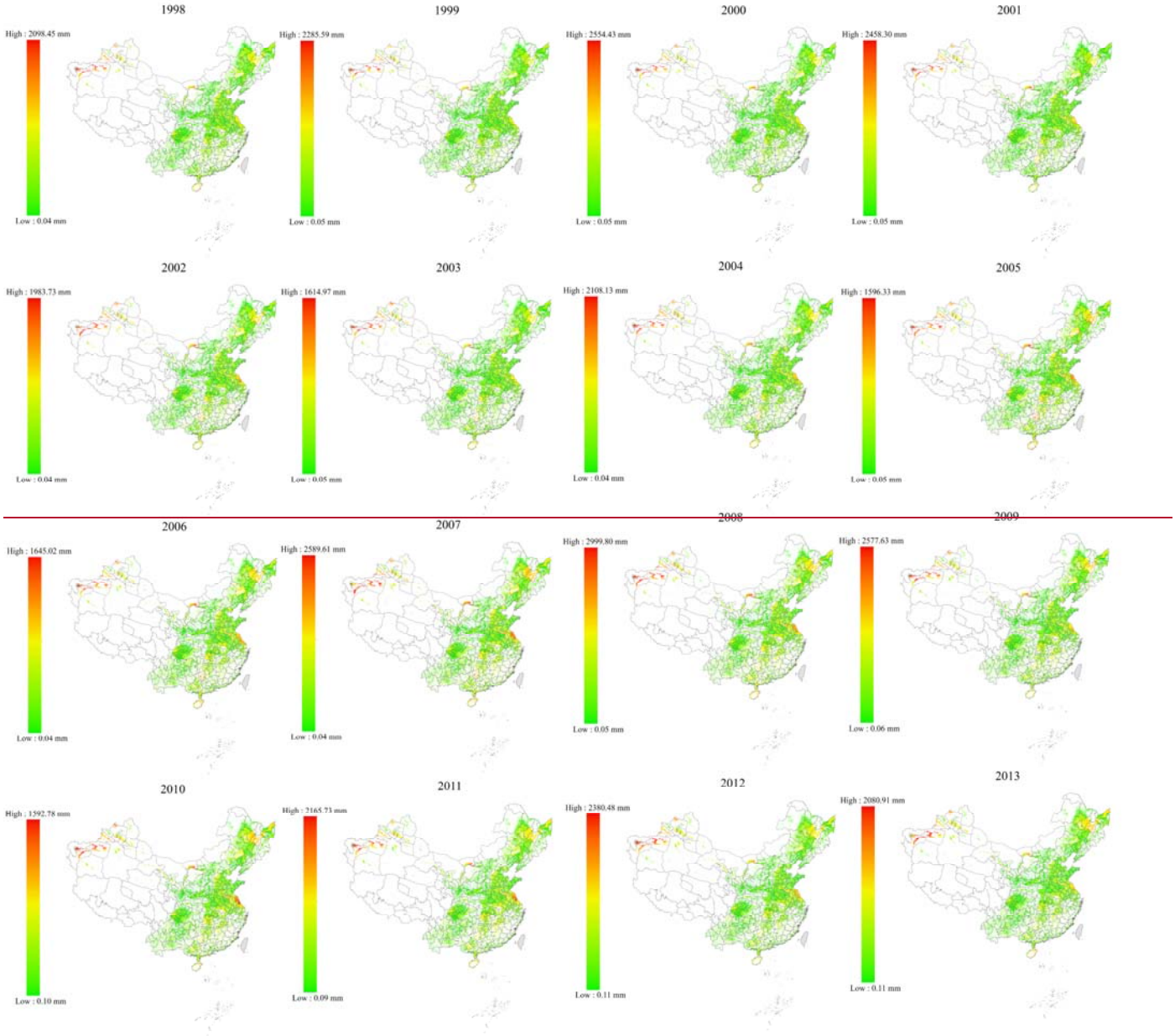


(c)

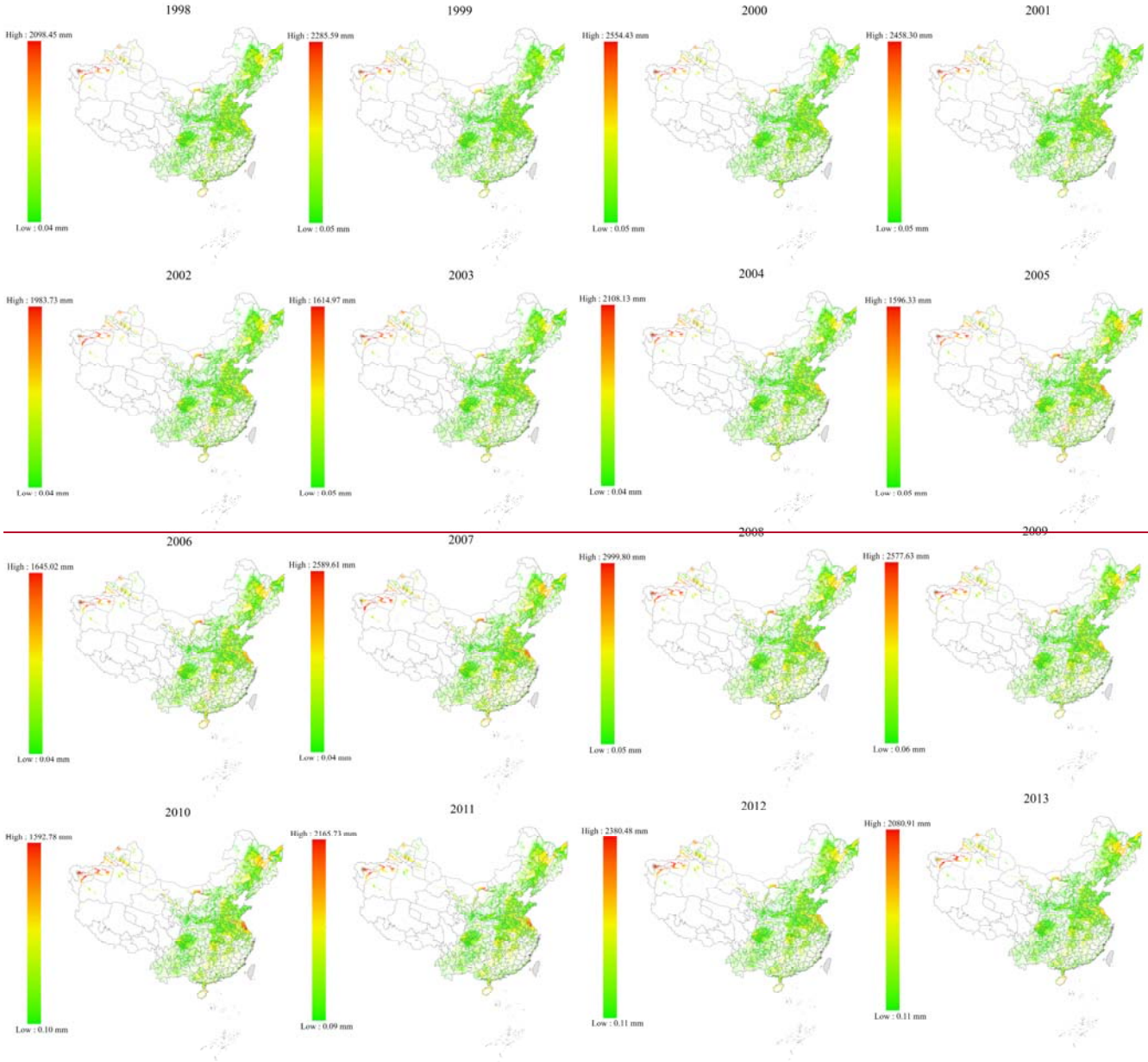
Figure ~~5~~6. Optimal parameters of the linear rule CA model and the model performances at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

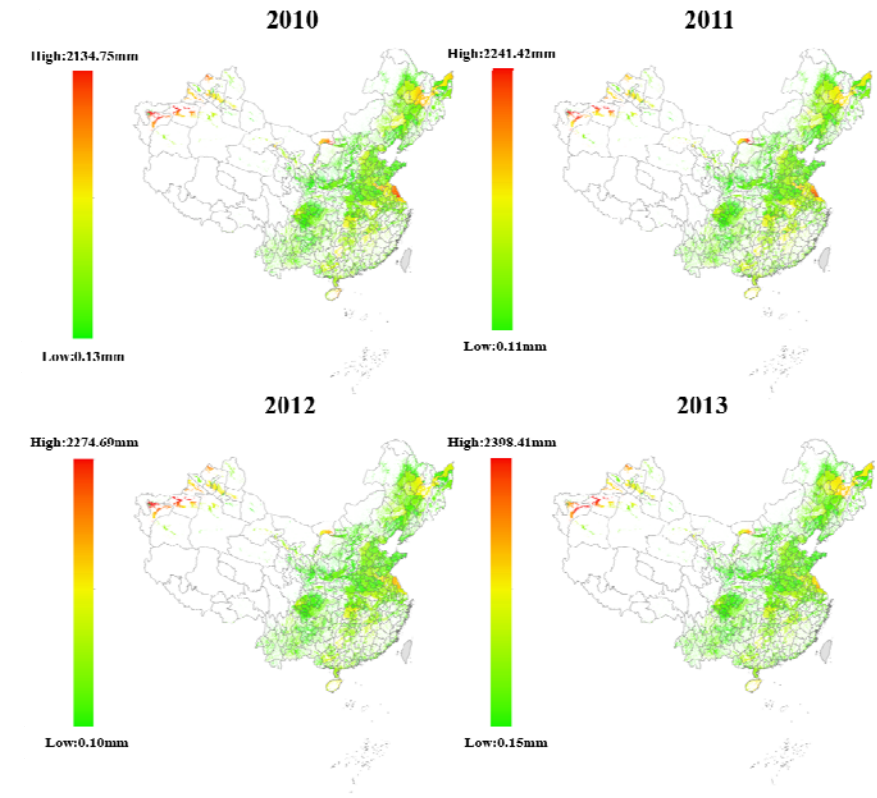
According to the results as shown in Figure ~~5~~6, the calibrated parameters are varied with scales and the performances of the models are acceptable. The spatial heterogeneity of the parameters α, β, and p tend to reflect local water use patterns. There is a balance between α and β in the areas with lower *RMSE* and *RE* values from the results at 1 km scale as shown in Figure 5(a). And both self-dependence and spatial diffusion jointly govern stable and reliable simulations. But at the appropriate spatial scale (Figure ~~5~~6 (b)), the parameter values show their deviations from those of the 1 km scale, and α and β values become more regionally specific. As β plays a more dominant role in every prefecture, and higher *RMSE* and *RE* values are found. So only heavily spatial diffusion may fail to capture the complexities of local water use dynamics. These areas often experience a more varied water use. At the prefecture scale (Figure ~~5~~6 (c)), the parameters are more likely to be dominated by α, especially in regions with large-scale infrastructure or agricultural variations. Their simulation errors are higher because the spatial variations are more difficult to be captured at this prefecture coarser resolutions. And the influence of β tends to be lower at this scale.

After calibrating the parameters for the linear rule of the CA model, the water use grid maps ~~for the period 1998–2013 are obtained and their simulated water use maps from 1998 to 2013 are also illustrated in Figure 6~~were generated at three spatial scales (i.e., ~~1km scale~~1 km, appropriate spatial scale, and prefecture scale).

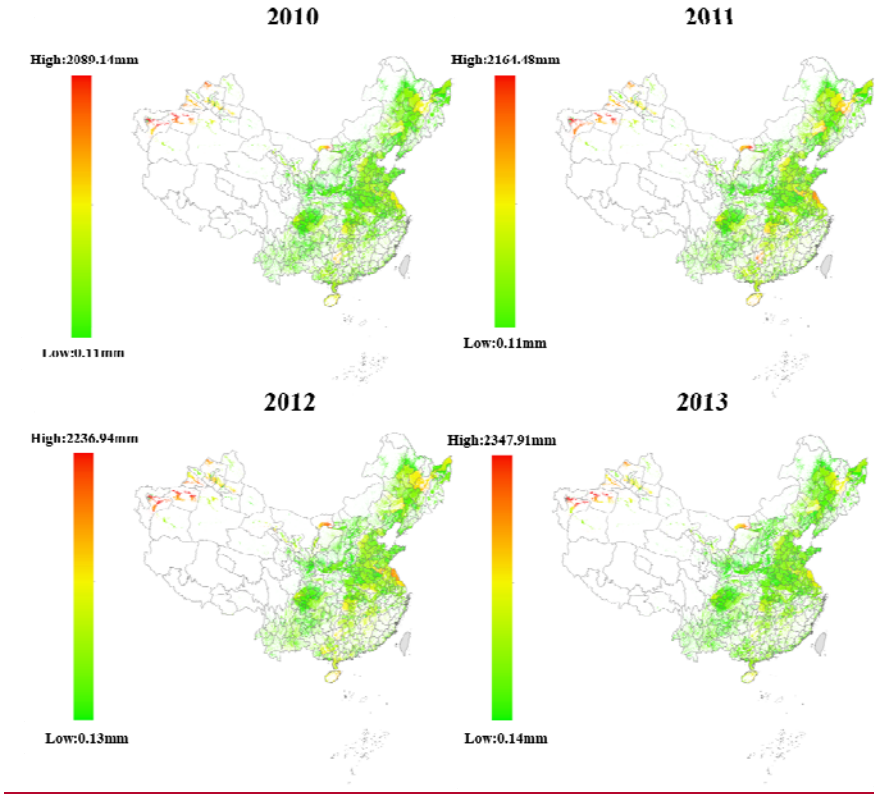


~~(~~For brevity and readability, Figure 7 presents only the simulation results for the validation years (2010–2013), which allow a~~)~~ more straightforward assessment of the model's predictive performance and spatial differences among scales. The full set of simulated water use maps from 1998–2013, can be accessed at: 10.6084/m9.figshare.30445157.

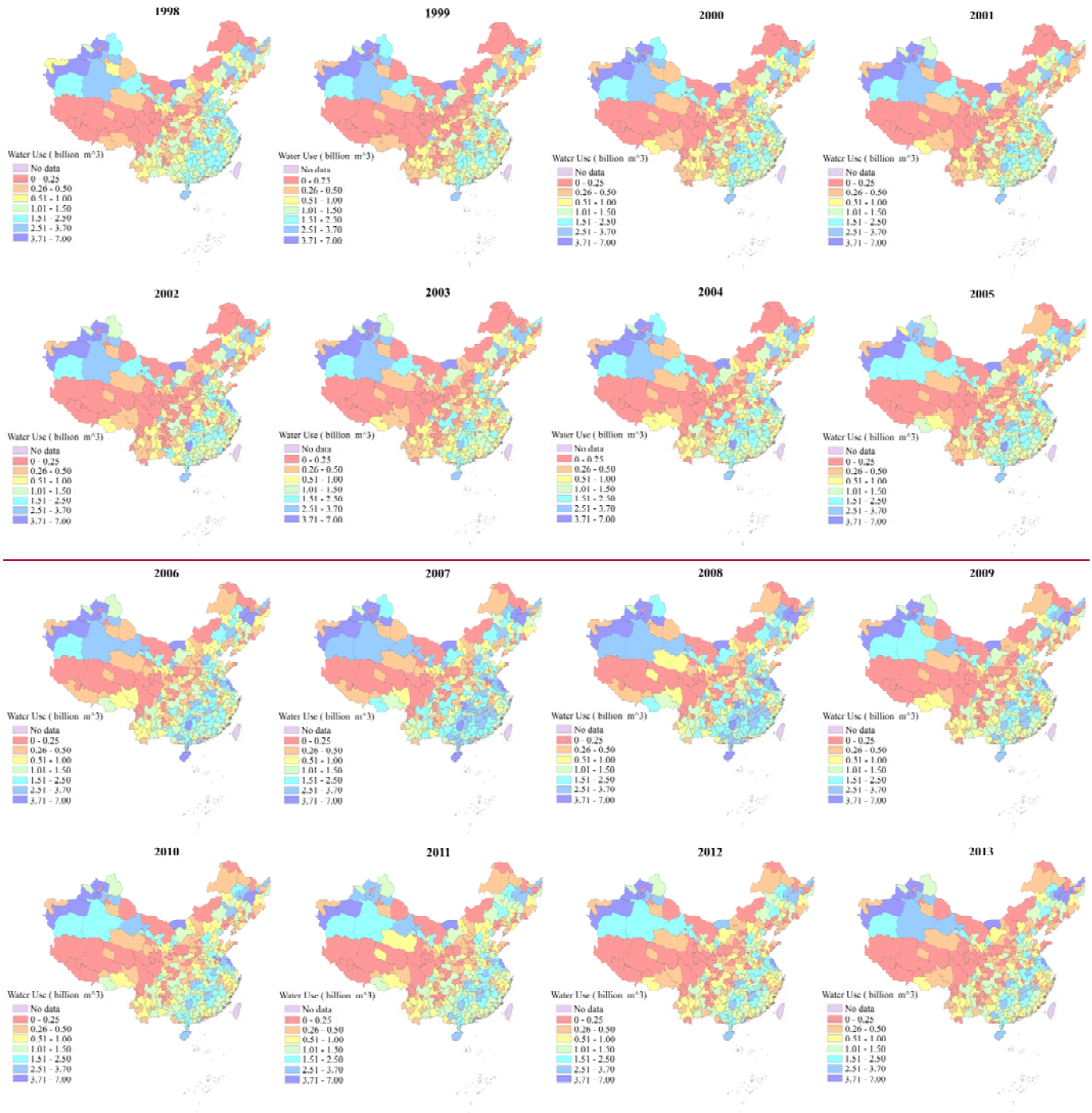

1998
High : 2098.45 mm
Low : 0.04 mm
1999
High : 2285.59 mm
Low : 0.05 mm
2000
High : 2554.43 mm
Low : 0.05 mm
2001
High : 2458.30 mm
Low : 0.05 mm
2002
High : 1983.73 mm
Low : 0.04 mm
2003
High : 1614.97 mm
Low : 0.05 mm
2004
High : 2108.13 mm
Low : 0.04 mm
2005
High : 1596.33 mm
Low : 0.05 mm
2006
High : 1645.02 mm
Low : 0.04 mm
2007
High : 2589.61 mm
Low : 0.04 mm
2008
High : 2999.80 mm
Low : 0.05 mm
2009
High : 2577.63 mm
Low : 0.06 mm
2010
High : 1592.78 mm
Low : 0.10 mm
2011
High : 2165.73 mm
Low : 0.09 mm
2012
High : 2380.48 mm
Low : 0.11 mm
2013
High : 2080.91 mm
Low : 0.11 mm


2010
High:2134.75mm
Low:0.13mm
2011
High:2241.42mm
Low:0.11mm
2012
High:2274.69mm
Low:0.10mm
2013
High:2398.41mm
Low:0.15mm

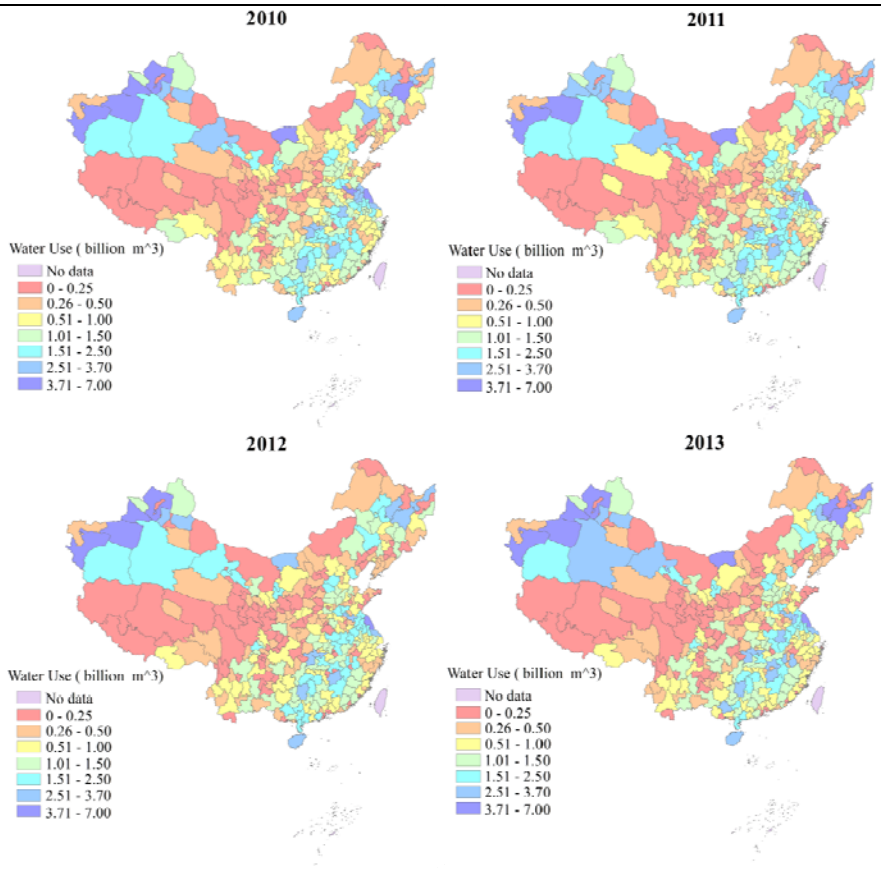
(a)


2010
High:2089.14mm
Low:0.11mm
2011
High:2164.48mm
Low:0.11mm
2012
High:2236.94mm
Low:0.13mm
2013
High:2347.91mm
Low:0.14mm

(b)


1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Water Use (billion m^3)
No data
0 - 0.25
0.26 - 0.50
0.51 - 1.00
1.01 - 1.50
1.51 - 2.50
2.51 - 3.70
3.71 - 7.00

503



(c)

Figure ~~6~~7. Water use simulation results from the linear rule at three different scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

The simulated water use maps of the linear rule from 1998 to 2013 are shown in Figure ~~6~~7, and they exhibit clear spatial dynamics at different spatial scales. At the 1 km scale (Figure ~~6~~7 (a)), the maps reveal the fine-scale variations in water use, highlighting their localized hotspots of high-water use, particularly in economically developed urban centers such as the Yangtze River Delta, the Pearl River Delta, and the Beijing–Tianjin–Hebei region. There are upward trends, which aligns with the patterns of population growth, industrial expansion, and urbanization. From the results at the appropriate spatial scale as shown in Figure ~~6~~7 (b), the localized variations in water use are captured with more clarity and precision. The appropriate scale provides a balanced representation that captures both the fine-scale dynamics and broader regional trends. At the prefecture scale (Figure ~~6~~7 (c)), the water use patterns become more smoothed. This bigger resolution limits

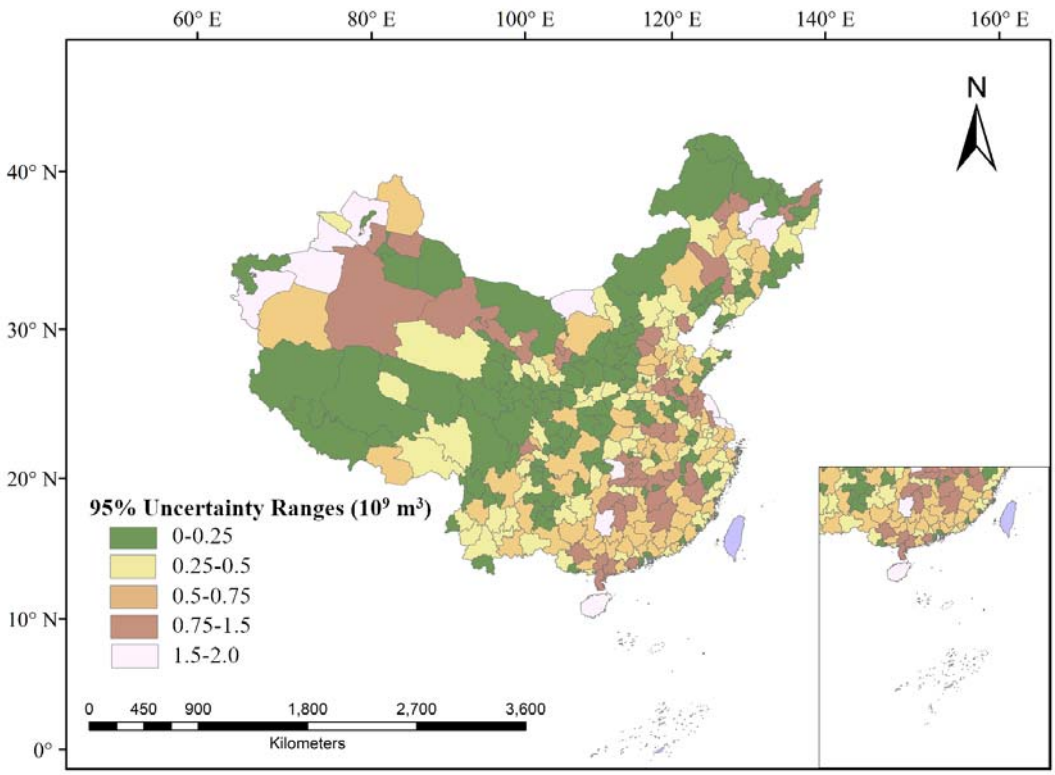
the model's ability to capture the dynamics even though the general upward trend in water use in urbanized regions is still observable. The localized fluctuations or abrupt changes driven by specific local policies or external factors are not fully represented at these bigger scales.

Moderate water use levels have been found in the North China Plain (NCP), compared with northeastern China, even though the NCP features intensive irrigation and high population density. Several factors contribute to this pattern. First, the calibration and validation rely on prefecture-level statistical survey data (e.g., water resources bulletins). In recent years, a decreasing trend of agricultural water use has been reported in many NCP prefectures, partly due to improvements in irrigation efficiency, the implementation of water-saving policies, and adjustments in cropping structures. Second, total water use in our study is an aggregate of irrigation, domestic, and industrial sectors. While total water use in the NCP remains dominated by irrigation, many northeastern prefectures show a growing contribution from industrial water use, particularly heavy industries and thermal power generation, leading to higher overall totals. Third, climatic and water availability differences also play a role. Despite a shorter growing season, northeastern China often supports water-intensive crops and benefits from relatively abundant local water resources, which result in higher water use per unit area. In addition, some previous studies that reported higher water use in the NCP were sector-specific (mainly irrigation) or based on different temporal baselines, which partly explains the discrepancy with our aggregated results.

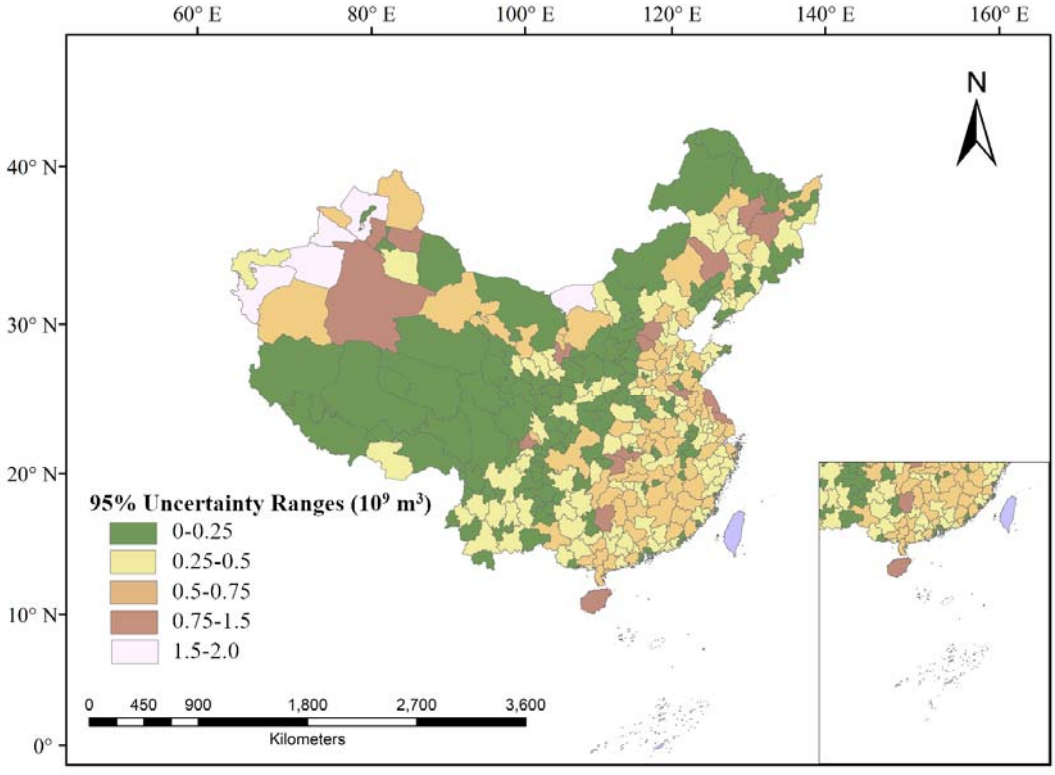
The simulation results across three different spatial scales can also indicate the impacts of scale on perceived water use. Finer-resolution maps reveal localized hotspots of high water use that can be averaged out in coarser resolution. It demonstrates that the CA model's flexibility and further confirms the importance of multi-scale analysis in understanding water use.

4.2 Uncertainty estimation of water use simulation from GLUE

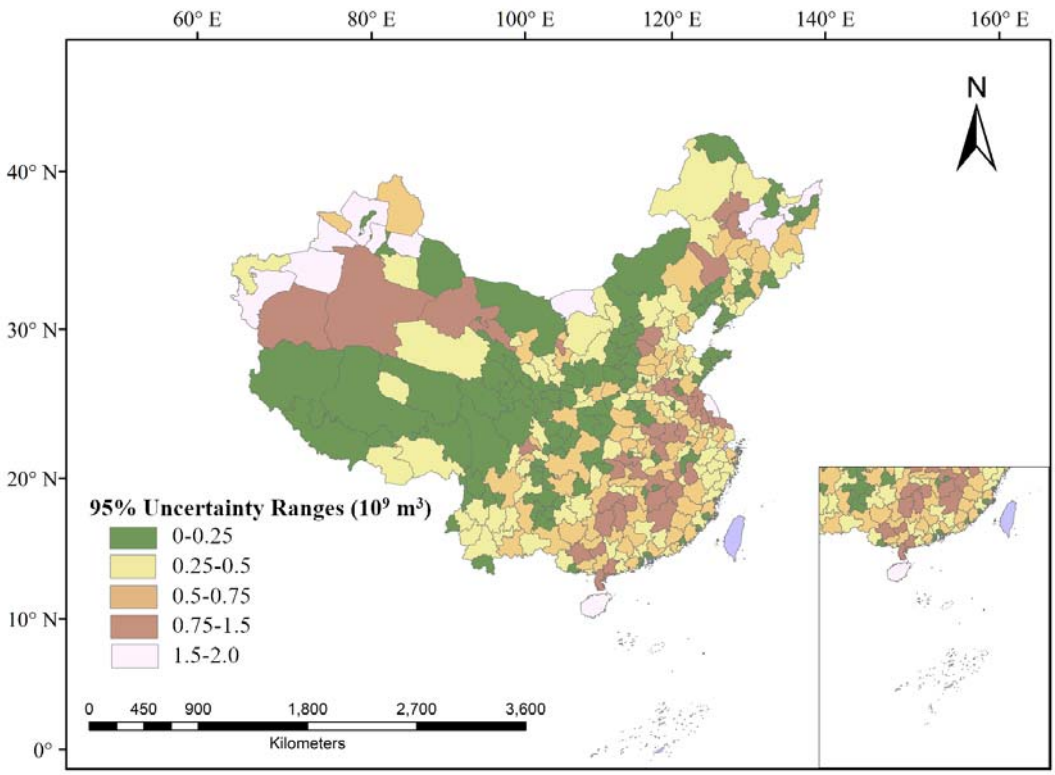
To further assess the reliability and robustness of the water use simulations from the CA model, the uncertainty is quantified by GLUE. According to the calibrated parameters for the probability and linear rules, the uncertainty ranges of the model outputs across different spatial scales are obtained through GLUE. To figure out both the accuracy of simulations and the confidence levels of model predictions from the parameter uncertainties, ensembles of the parameters combination were generated by the uniform sampling as the description in Section 2.3. Corresponding results from the CA model were obtained and their performances are quantified by *RMSE* and *RE* metrics. If the *RMSE* and *RE* metrics are acceptable according to their pre-defined thresholds, the maximum and minimum of the simulation results from the CA model are taken as their uncertainty range. These uncertainties across the three spatial scales are shown in Figure ~~7~~8 from the probability rule and in Figure ~~8~~9 from the linear rule at the 95% confidence level (i.e., pre-defined threshold). These uncertainties are from the parameter variabilities. A wider range of the uncertainties indicates lower stable and reliable simulation results, while narrower ones suggest stable and reliable simulation results. The spatial variation of uncertainty ranges can also reveal the significant regional differences of the simulations at different spatial resolutions.


60° E
80° E
100° E
120° E
140° E
160° E
40° N
30° N
20° N
10° N
0°
N
95% Uncertainty Ranges (10⁹ m³)
0-0.25
0.25-0.5
0.5-0.75
0.75-1.5
1.5-2.0
0 450 900 1,800 2,700 3,600
Kilometers

(a)

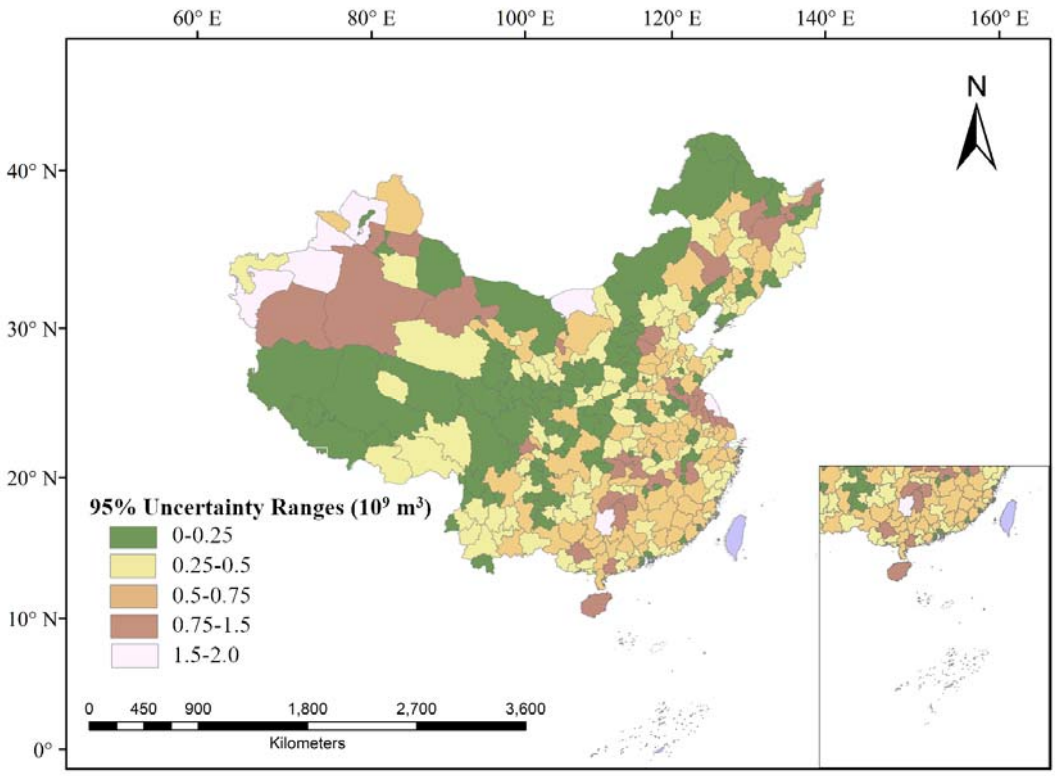

60° E
80° E
100° E
120° E
140° E
160° E
40° N
30° N
20° N
10° N
0°
N
95% Uncertainty Ranges (10⁹ m³)
0-0.25
0.25-0.5
0.5-0.75
0.75-1.5
1.5-2.0
0 450 900 1,800 2,700 3,600
Kilometers

(b)

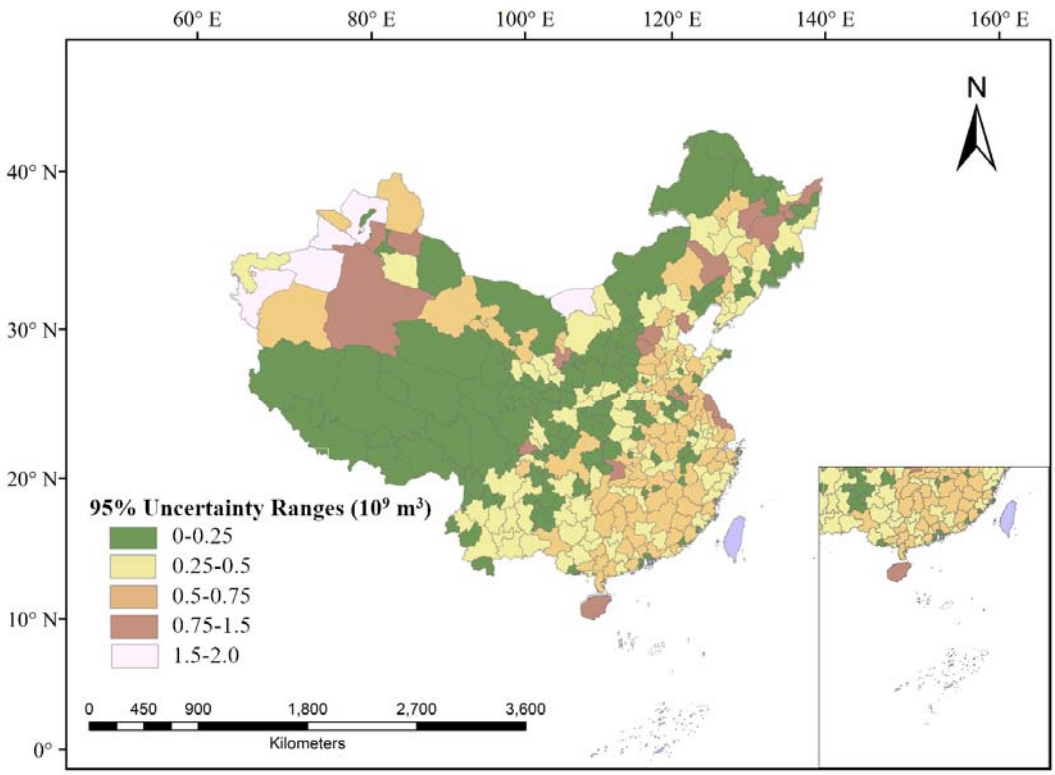


(c)

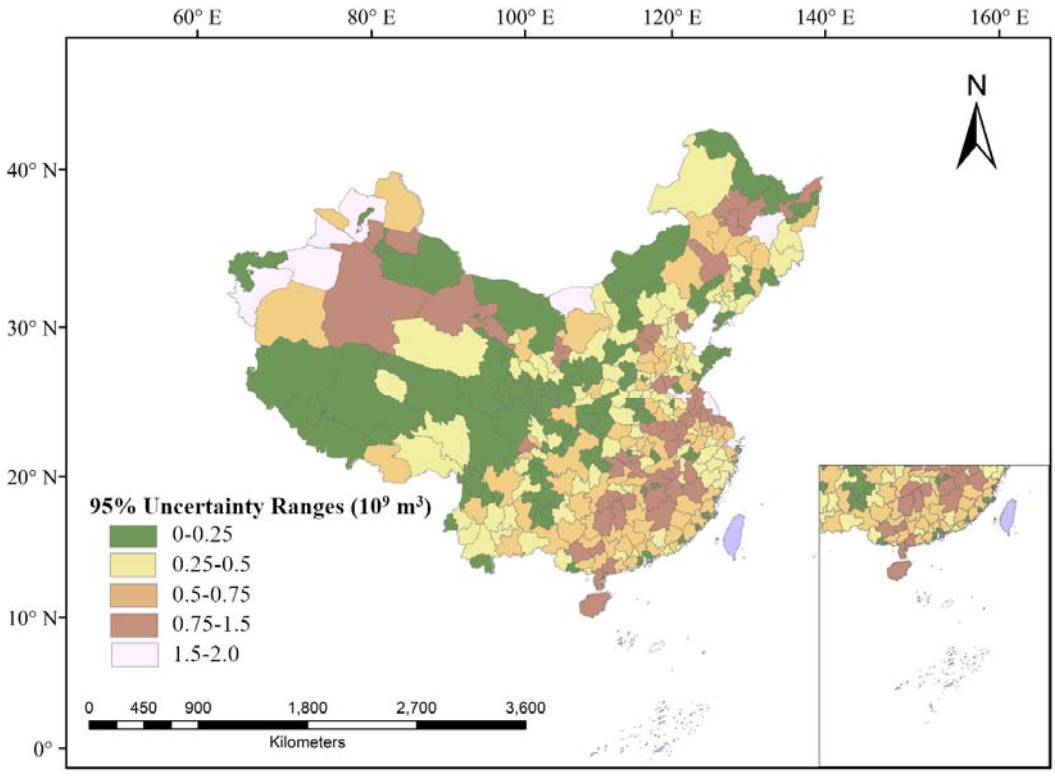
Figure ~~7~~8. The uncertainty ranges of water use simulations at the 95% confidence level at three spatial scales from the probability rule at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale



(a)



(b)



(c)

Figure ~~8~~9. The uncertainty ranges of water use simulations at the 95% confidence level at three spatial scales based on linear rule at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

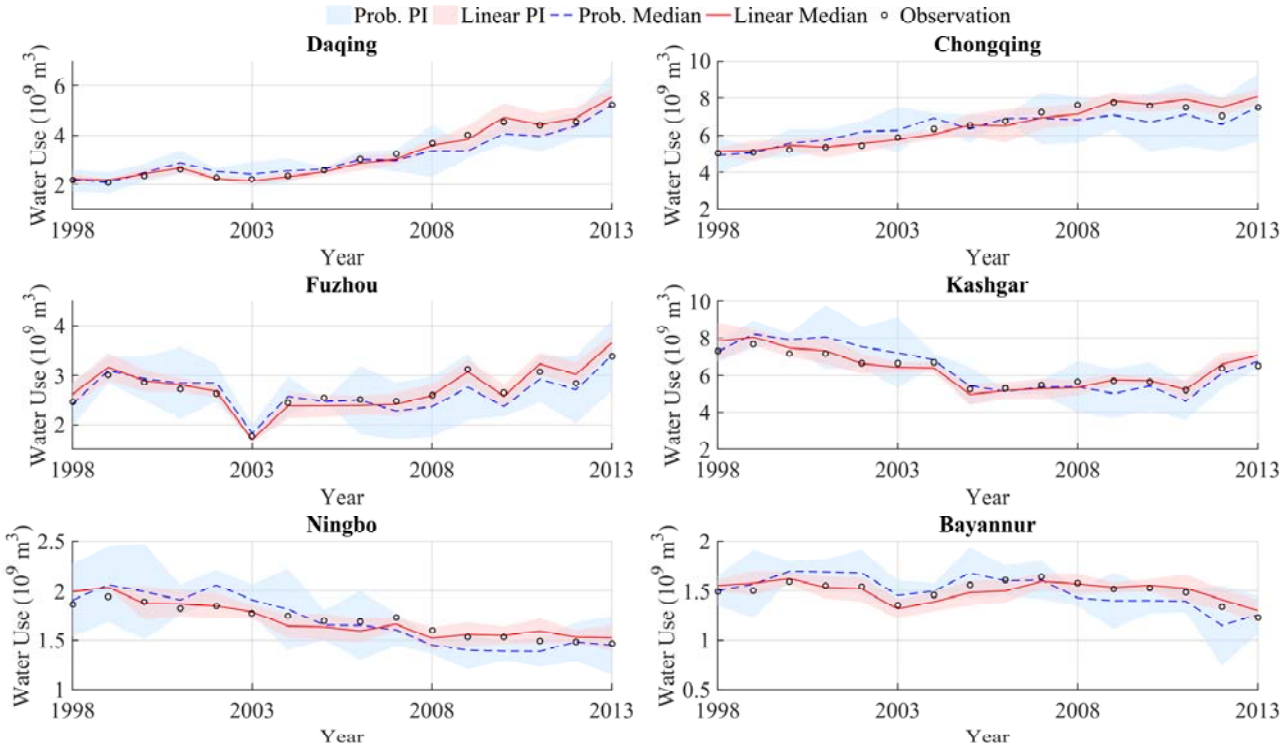
There are distinct regional patterns of simulation uncertainty from the spatial distribution of the uncertainty ranges at the 95% confidence level. Larger uncertainty ranges (i.e., 0.75–2.0*$10^9$$m^3$) are predominantly found in western and southwestern prefectures, such as Xinjiang, Qinghai, Gansu, and Chongqing, where data scarcity and complex local dynamics likely contribute to higher model uncertainty. In contrast, most eastern and northeastern regions, including Beijing, Jiangsu, Shandong, and Liaoning, exhibit relatively narrow uncertainty ranges (i.e., 0–0.5*$10^9$$m^3$), indicating more stable historical water use behavior and better alignment with CA model assumptions.

The model for the higher spatial resolutions has been found to be more sensitive to the local variations, and its uncertainty can be amplified in heterogeneous land use or socio-economic condition areas. For instance, large uncertainty ranges can be found in the densely populated urban centers or the regions undergoing rapid industrialization at the 1 km scale as shown in Figure ~~7~~8 (a) and Figure ~~8~~9 (a). But small uncertainty ranges can be found in the same areas from the coarser-scale simulations due to the averaging of the localized fluctuations as seen in Figure ~~7~~8 (c) and Figure ~~8~~9 (c) at the prefecture scale. However, the small uncertainty ranges are obtained at the cost of masking the sub-regional dynamics. There should be trade-off between capturing detail and maintaining stability. Thus, it is important to select an appropriate spatial scale for the specific planning or policy purposes.

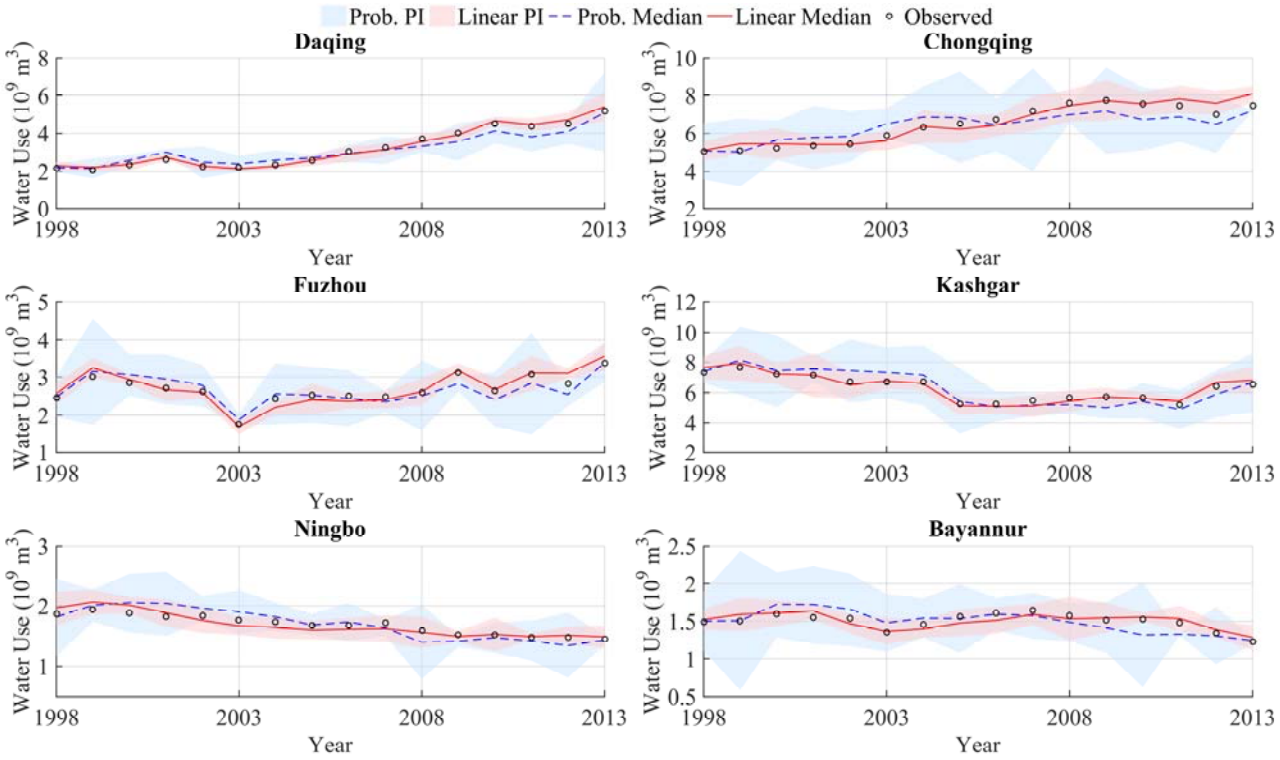
The larger uncertainty ranges are more common from the probability rule than those from the linear rule, particularly in regions with complex or highly variable water use patterns. This is largely attributed to the probability rule itself, because there are randomness in both state transitions and value sampling in fitting distributions. Although the probability rule can help the model to capture non-linear dynamics and abrupt changes, it brings higher uncertainty into outputs. In contrast, the linear rule generally brings less uncertainty, reflecting the deterministic structure of its update mechanism in the model. The amount of water use in each

cell is predicted by the weighted influences from itself and neighboring cells. In regions with relatively stable and spatially smooth development patterns, such as the eastern and northeastern parts of China, the linear rule is more effective. However, the linear rule's assumption of gradual spatial continuity fails to capture abrupt local changes in the rapidly changing or the weak spatial correlation water use areas—such as Fujian, Chongqing, and some parts of Guizhou and Sichuan. And the linear rule potentially leads to an underestimation of uncertainty.

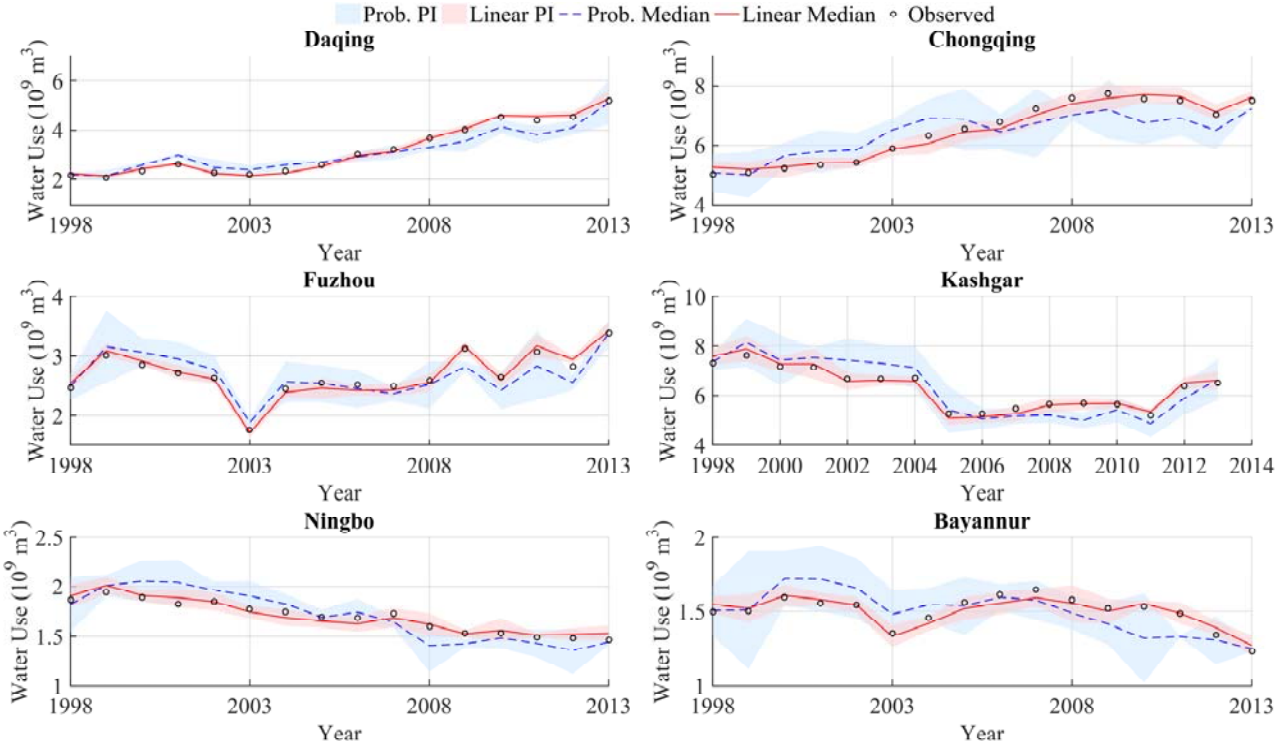
To further examine the temporal dynamics and regional differences in model uncertainty, Daqing, Chongqing, Fuzhou, Kashgar, Ningbo, and Bayannur are selected as six representative prefectures according to their geographic locations, socio-economic conditions, and water use behaviors. The time series of 95% uncertainty ranges from 1998 to 2013 for each representative prefecture is obtained from the results of the probability and linear update rules at different spatial scales, illustrated as Figure ~~9~~10.



(a)



(b)



(c)

Figure ~~9~~10. Time series of 95% uncertainty ranges from 1998 to 2013 for each representative prefecture from both the probability and linear update rules at different spatial scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

There is distinct the patterns across the selected prefectures as shown in Figure ~~9~~10. The results from the probability rule have been found to be consistently wider uncertainty intervals than that from the linear rule, particularly in prefectures with more unstable water use conditions. For instance, Kashgar and Bayannur that are located in arid and less-developed western regions exhibit larger and more variable uncertainty intervals from the probability rule. And they are results of the combined effects of high inter-annual variability and limited input data. In contrast, Ningbo and Daqing that are located in the more stable water use trends and better data coverage areas have been found to show relatively narrow and consistent uncertainty from both two rules. It is also interesting to find that the moderate uncertainty levels with obvious differences between the two rules have been in Chongqing and Fuzhou with too many transitional urban areas. As the rapid urbanization and economic shifts bring the localized instability, the probability rule can capture the shifts whereas the linear rule can only underestimate uncertainty during periods of structural change due to its lower sensitivity.

Beside the temporal and rule-based variability, the spatial scale size also plays a crucial role in shaping simulation uncertainty. With the spatial resolution of the model increases, the simulation uncertainty tends to increase, especially in regions with high spatial heterogeneity, such as arid or rapidly developing areas. The finer resolutions can effectively capture the localized variations in water use, there are higher levels of uncertainty due to the increased sensitivity to local fluctuations. In contrast, the results at coarser resolutions tend to be smoothed these local variations, there will be narrower uncertainty but potentially overlooking the sub-regional dynamics.

The uncertainty results have highlighted the importance of the spatial and temporal variations for evaluating the simulation model performance and uncertainty. The results of the representative six prefectures suggest that there is no one single update rule can outperform the others. Instead, the most appropriate rule

depends on the local water use dynamics and the purpose of the water use simulation.

5 Discussion

5.1 Impacts of Update Rules on Water Use Simulation

To assess the performance of ~~different~~the two update rules, ~~we upscaled~~ the simulation results at ~~the~~ three spatial scales were upscaled to the provincial administrative level, and ~~obtained~~ the *RMSE* and *RE* metrics ~~from~~were calculated for both the probability and linear ~~update~~ rules. An additional indicator, $\Delta signed$ was introduced to quantify both the magnitude and direction of the differences between the two simulations (probability vs. linear). $\Delta signed$ represents the mean signed grid-level difference normalized by the mean simulated water use for each province. A positive $\Delta signed$ indicates that the probability rule yields higher simulated water use than the linear rule, while a negative $\Delta signed$ indicates lower simulated values. There are notable differences ~~of~~between the results ~~from~~produced by the two update rules, as summarized in Table 2.

Table 2. Model performance at the provincial administrative level from different update rules

Provinces	*RMSE* (billion m^3)		*RE* (%)		$\Delta signed$ (%)
	Linear rule	Probability rule	Linear rule	Probability rule	
Beijing	0.03	0.05	+12	+19	+2.8
Tianjin	0.02	0.02	-11	-17	-0.6
Shanghai	0.02	0.04	-13	-25	+3.1
Chongqing	0.04	0.07	+16	+17	+1.5
Anhui	0.12	0.17	+23	+32	+2.2
Fujian	0.17	0.21	-23	+37	+3.4

Gansu	0.35	0.43	+27	+29	-1.9
Guangdong	0.26	0.36	+21	+32	+3.9
Guizhou	0.31	0.40	-26	-34	-1.2
Hainan	0.11	0.21	-12	+19	+3.0
Heilongjiang	0.42	0.37	-27	+24	+2.1
Hunan	0.21	0.11	-16	+10	+1.6
Jilin	0.36	0.43	+28	+25	+2.5
Jiangsu	0.12	0.17	-12	-26	-0.7
Jiangxi	0.22	0.26	+24	-31	+1.3
Inner Mongolia	0.19	0.29	-23	-35	-2.1
Qinghai	0.31	0.41	+43	+36	+1.4
Ningxia	0.11	0.20	-31	-41	-2.4
Shandong	0.21	0.36	+29	-36	+4.1
Shanxi	0.42	0.61	+31	+42	+3.3
Shannxi	0.27	0.43	-32	-39	-1.5
Sichuan	0.39	0.41	+31	+46	+2.7
Xizang	0.56	0.61	-62	-79	-0.9
Xinjiang	0.68	0.72	+69	+83	+3.5
Yunnan	0.25	0.34	-37	-47	-1.1
Zhejiang	0.21	0.29	+18	+12	+1.9
Guangxi	0.34	0.42	-32	-39	-1.7
Hubei	0.29	0.39	+27	+26	+2.2
Liaoning	0.35	0.41	+33	+48	+3.7

~~The results from the linear rule have been found to performance better than that from the probability rule in terms of its lower *RMSE* and *RE* values. The differences between the results from the two update rules are based on their update mechanisms. The design distribution based state transitions are stochasticity in the probability rule. It can capture uncertainty in terms of the abrupt or nonlinear changes in water use while it can also bring the potential divergences from the observed trends, especially when data are sparse or when~~

~~local variability is high. As a result, several higher *RMSE* and *RE* values had been found, particularly those characterized by complex topography or heterogeneous socio-economic development patterns. In contrast, the linear rule employs a deterministic update process based on local spatial averages, which tends to smooth out their fluctuations and enhance stability. In other words, the heterogeneous socio-economic development patterns have inherently been smoothed for reducing the local-level variability by the linear rule at the provincial scale. The differences of their performance from the two rules are particularly noticeable in inland or in less urbanized provinces, where the probability rule's flexibility amplifies the uncertainties.~~

The linear rule generally exhibits lower *RMSE* and *RE* values, indicating higher stability and consistency with observed prefecture-level statistics. This is because the linear rule updates each grid deterministically based on spatial averages of its neighbors, which smooths fluctuations and captures persistent long-term patterns. Consequently, it tends to reduce local variability and enhance regional stability, especially at coarser spatial scales. By contrast, the probability rule explicitly incorporates stochasticity through distribution-based state transitions. This enables it to capture local irregularities, abrupt changes, and nonlinear water use dynamics driven by variations in industrial structure, irrigation demand, or climatic conditions. However, such stochastic behavior can also amplify uncertainty in regions with sparse data or complex spatial heterogeneity, resulting in slightly higher *RMSE* and *RE* values.

Quantitatively, the mean *RMSE* of the linear rule during the validation period (2010–2013) was 0.28 billion m^3, compared with 0.36 billion m^3 for the probability rule. The corresponding mean relative errors were ±22.4% and ±29.8%, respectively. At the national scale, the simulated total water use was 570.6 billion m^3 for the linear rule and 583.2 billion m^3 for the probability rule, differing by +1.1% and +3.5% from observed national statistics. Therefore, the linear rule is identified as the best-performing estimation framework for reproducing the observed spatiotemporal water use distribution in China, whereas the probability rule provides

valuable complementary insights for representing uncertainty and local heterogeneity.

The $\Delta signed$ results indicate that the probability rule tends to simulate higher water use in industrially intensive provinces (e.g., Shandong, Guangdong, and Liaoning) and slightly lower values in water-scarce inland regions (e.g., Ningxia, Gansu, and Inner Mongolia). The results confirm that the two rules emphasize different aspects of the underlying processes—the probability rule better reflects stochastic local variability, while the linear rule offers greater stability and smoother large-scale consistency.

5.2 Impact of Spatial Scales on Water Use Simulation

To investigate how spatial resolution influences the accuracy of the water use simulation, their performances at the three spatial scales (i.e., 1 km scale, appropriate intermediate scale, and prefecture scale) are evaluated and are aggregated to the provincial level to ensure comparability across scales. Their results are summarized in Table 3 and indicate their notable differences in simulating accuracy depending on its spatial scales.

Table 3. Model performance at the provincial administrative level at different spatial scales

Provinces	*RMSE* (billion m^3)			*RE* (%)		
	1 km scale	Appropriate scale	Prefecture scale	1 km scale	Appropriate scale	Prefecture scale
Beijing	0.04	0.03	0.05	+22	+12	+19
Tianjin	0.04	0.02	0.06	+26	-11	-17
Shanghai	0.03	0.02	0.04	-23	-13	-25
Chongqing	0.06	0.04	0.07	-26	+16	+27
Anhui	0.17	0.12	0.19	+35	+23	+32

Fujian	0.23	0.17	0.21	-31	-23	+37
Gansu	0.41	0.35	0.43	+31	+27	+29
Guangdong	0.31	0.26	0.36	-29	+21	+32
Guizhou	0.36	0.31	0.40	+31	-26	-34
Hainan	0.19	0.11	0.21	+16	-12	+19
Heilongjiang	0.49	0.42	0.47	+31	-27	+34
Hunan	0.26	0.21	0.31	-19	-16	+20
Jilin	0.42	0.36	0.43	-31	+28	+35
Jiangsu	0.19	0.12	0.17	-21	-12	-26
Jiangxi	0.31	0.22	0.36	-26	+24	-31
Inner Mongolia	0.26	0.19	0.29	-31	-23	-35
Qinghai	0.40	0.31	0.41	+52	+43	+56
Ningxia	0.21	0.11	0.30	-39	-31	-41
Shandong	0.29	0.21	0.36	-32	+29	-36
Shanxi	0.53	0.42	0.61	+36	+31	+42
Shannxi	0.34	0.27	0.43	-36	-32	-39
Sichuan	0.42	0.39	0.41	+41	+31	+46
Xizang	0.67	0.56	0.61	-74	-62	-79
Xinjiang	0.79	0.68	0.72	+72	+69	+83
Yunnan	0.31	0.25	0.34	-49	-37	-47
Zhejiang	0.24	0.21	0.29	+23	+18	+32
Guangxi	0.39	0.34	0.42	-36	-32	-39
Hubei	0.35	0.29	0.39	+31	+27	+36
Liaoning	0.37	0.35	0.41	+42	+33	+48

The relatively lower accuracy of the 1 km scale simulations is attributed to the increased sensitivity to local heterogeneity. At the 1km scale, small errors from the input variables or local noise can accumulate and amplify to larger discrepancies. Although the results from1kmscale take more spatial detail to reflect the variations in water use, it also brings the uncertainty, especially in the regions with the fragmented data or the

high socio-economic diversity. The simulation results from the prefecture level tend to oversimplify the spatial variation. Generally, the results from the coarse resolution have smoothed the intra-regional differences in water use patterns, and resulted in under or over-predictions at the provincial administrative level. The reduced spatial granularity leads to reduction in uncertainty of simulation but can mask the disparities that are critical to policy implementation and resource allocation.

The results from the appropriate scale can balance the spatial sensitivity and model stability. It is fine enough to capture meaningful spatial heterogeneity, and it is yet coarse enough to mitigate excessive noise and data sparsity effects. The most reliable simulating water use should be based on this scale. And it should be noted that that it is crucial for improving simulation reliability to select an appropriate spatial resolution that aligns with both the model structure and the scale of decision-making.

5.3 Spatial Heterogeneity of Water Use Grid Maps

To better understand the ~~variation of the~~ spatial heterogeneity of simulated water use, the Coefficient of Variation (CV) and ~~Moran's~~Moran's I were ~~used~~applied to quantify the ~~degree of~~ variability and spatial autocorrelation of water use across ~~at~~ three ~~different spatial scales across the study area.~~ spatial scales — 1 km, the appropriate spatial scale, and the prefecture scale. These two metrics together reveal how spatial scale and model design influence the representation of water-use heterogeneity.

~~The~~ At the 1 km scale, the highest CV values ~~have been found in the results of water use simulation at the 1 km scale~~are observed among the three scales~~. There are the most significant regional~~, indicating the greatest variability in water use across grid cells. This fine resolution captures the most detailed local differences ~~at the 1 km scale. As Moran's I can reveal the positive autocorrelation that suggests the clustering of high or low water use. The results of Moran's I analysis at the 1 km scale show that there is a strong pattern~~

~~of spatial dependence~~ , especially in ~~urban~~ areas~~, particularly in regions~~ with ~~intense~~intensive agricultural or industrial ~~water use~~activities. Moran's I results also reveal strong positive spatial autocorrelation, suggesting distinct clustering of high- or low-water-use regions, particularly around large urban centers and irrigation districts.

~~Different from~~At the ~~results from~~appropriate spatial scale, both the CV and Moran's I values indicate a moderated heterogeneity pattern. Compared with the 1 km scale, the ~~results from the appropriate scale show the decreasing water use heterogeneity with lower CV values. As more local variations are captured by the smaller~~ variability decreases as small-scale ~~grid, the bigger named appropriate~~ noise is smoothed out, while the spatial clustering remains evident but less fragmented. This scale ~~here can balance the~~ provides a balanced representation by capturing regional heterogeneity ~~and the reducing~~without introducing excessive spatial ~~details for water us simulation. And there are also positive autocorrelations with Moran's I values at the appropriate scale even with less degree of clustering~~detail or instability. The appropriate spatial scale therefore achieves the optimal trade-off between capturing local patterns and maintaining spatial coherence.

~~It is no doubt that the lowest spatial heterogeneity of water use simulation have been found with~~At the prefecture scale, the lowest CV values ~~at the prefecture scale. The~~ are recorded, showing that much of the local variability has been smoothed. Moran's I values remain moderately positive, reflecting that some regional ~~variations of water use data at larger spatial scales are often smoothed out. The moderate positive autocorrelation of Moran's I values indicate that while there is still~~clustering persists but overall spatial dependence~~, it is~~ becomes less ~~distinct compared to finer~~pronounced. At this coarser scale, water-use patterns become generalized, reducing the granularity of spatial differences.

To further quantify these relationships, Table 4 presents the average CV and Moran's I values under different spatial scales and update rules.

Table 4 Comparison of CV and Moran's I under different spatial scales. and update rules

It is suggested that the spatial scale

Spatial scale	Update rule	*CV* (mean)	Moran's I (mean)	Interpretation
1 km	Probability rule	0.82	0.59	Highest heterogeneity; strong clustering in high-use regions.
1 km	Linear rule	0.78	0.56	Slightly smoother than the probability rule; still high local variation.
Appropriate scale	Probability rule	0.71	0.72	Balanced heterogeneity and spatial dependence; optimal trade-off between detail and stability.
Appropriate scale	Linear rule	0.68	0.70	Slightly lower variability, indicating smoother transitions among neighboring cells.
Prefecture scale	Probability rule	0.60	0.74	Variability largely smoothed; moderate clustering remains.
Prefecture scale	Linear rule	0.57	0.73	Lowest heterogeneity; strong regional aggregation of water-use patterns.

In addition to the influence of spatial scale, the choice of update rule also affects the spatial heterogeneity of simulated water use. Across all scales, the probability rule yields slightly higher CV values than the linear rule, indicating that it better preserves local variability and stochastic fluctuations of water use. This difference stems from the probabilistic state-transition mechanism, which allows grid-level water use to fluctuate around its expected trajectory, capturing year-to-year uncertainty that is smoothed out in the deterministic linear formulation.

Conversely, Moran's I values under the linear rule are comparable to or slightly lower than those from the probability rule, suggesting smoother and more spatially continuous water-use patterns. This indicates a stronger spatial-averaging effect of the water use simulation significantly impactsdeterministic update mechanism, which may be advantageous for long-term or regionally aggregated assessments but less effective in representing local variability.

Overall, both spatial scale and update rule jointly shape the representation of spatial heterogeneity. in water-use simulation. The results from 1 km scale can capturecaptures the most detailed variability, while the appropriate spatial scale provides a more balanced viewand realistic depiction of spatial heterogeneity,

regional patterns, and the prefecture scale ~~tends to generalize water use patterns, reducing the granularity of~~generalizes spatial differences. Meanwhile, the probability rule emphasizes local randomness and uncertainty, whereas the linear rule accentuates deterministic spatial continuity.

There had been some water use simulation results at previous studies. Fox example, Huang et al. (2018) produced a global-scale monthly water withdrawal dataset at 0.5° resolution, distinguishing six sectors (e.g., irrigation, domestic, electricity generation, livestock, mining, manufacturing) over the period 1971–2010; Hou et al. (2024) developed China's industrial water withdrawal dataset (CIWW), providing gridded monthly data from 1965 to 2020 at 0.1° and 0.25° resolutions; Zhang et al. (2025) presented a high-resolution sectoral water use dataset (HSWUD) for mainland China, covering irrigation, manufacturing, thermal power cooling, and domestic use at 0.1° × 0.1° resolution, with strong consistency to prefecture-level statistics (R^2≈0.88). As the results shown in the previous sections, our dataset is generated at different spatial resolutions (e.g., 1 km×1 km, appropriate spatial scale), enabling detailed representation of spatial heterogeneity within prefectures. Due to the substantial differences in spatial resolutions between these datasets, it is not easy to compare the differences of the spatial distribution patterns. But the relative values of performance metrics such as *RMSE* and *RE* can figure out the better one among them. The values of *RMSE* within 0.1 (i.e., normalized by mean water use) and a *RE* within −20 % to +30 % are found across all prefectures according to the results of our simulation. And all these results are consistent.

Thus, the results show that, relative to the three reference datasets, our model's prefecture-level water use estimates achieve a *RMSE* within 0.1 (normalized by mean water use) and a *RE* within −20 % to +30 % across all prefectures. These results within the range generally are considered acceptable for large-scale water use modeling, indicating that our estimates are consistent with these previous studies while offering finer spatial details.

6 Conclusion

A multi-scale water use simulation framework has been proposed through integrating a CA model with GLUE to address spatial heterogeneity and uncertainty in this study. Both probability rule and linear rule across three spatial scales (i.e., 1 km, appropriate scale, and prefecture scale) have been applied over 341 prefectures in China as a case study. The impacts of model structure and spatial scale on the spatial heterogeneity and uncertainty have been figured out.

It is interesting to find that the probability rule effectively captures stochastic variations and abrupt transitions in water use but brings larger uncertainty due to its random sampling nature. And the linear rule brings more stable and accurate simulations, particularly in regions with smoother water use patterns. The local noise and uncertainty tend to be amplified in the results of the water use simulation at the 1km scale , while the essential spatial heterogeneity tend to be oversimplified and suppressed in the results at the prefecture scale The most reliable simulation are found from the appropriate scale due its trade-off between capturing spatial heterogeneity and maintaining model stability.

Future improvements for the water use simulation should involve the adaptive update rules that respond to external drivers such as policy shifts or climate shocks, and extend the simulation to finer temporal scales (e.g., seasonal or monthly) for improved short-term decision making. More uncertainty quantification methods, such as Bayesian inference or Markov Chain Monte Carlo, are recommended to enhance performance in high-dimensional settings. Overall, our study can contribute to the water use simulation at a multi-scale and with uncertainty-aware. Our proposed framework will be helpful for the integrated water management, infrastructure planning, and environmental policy under changing socio-economic and climatic conditions.

Author contributions

JZ designed the model architecture, performed the computations, conducted the statistical analysis, and drafted the manuscript. DL acquired funding, contributed to the study design, provided research data, supervised the project, and guided the manuscript revision. JW contributed to manuscript revision discussions and provided advice on submission procedures. FU, LX, ZP, and WG participated in revision discussions and contributed to figure and chart preparation. All authors reviewed and approved the final version of the manuscript for submission.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgement

The authors gratefully acknowledge the financial support from National Key R&D Program of China (2024YFC3012402 and 2022YFC3202803), the National Natural Science Foundation of China (Nos.52379022, and 51879194).

References

Abdi H. Coefficient of variation. Encycl. Res. Des., 1(5), 169–171, https://doi.org/10.4135/9781412961288.n56, 2010.

Al-Shaar, W., Nehme, N., Haidar, H. and Lakiss, H. Forecasted water demand using Extended Cellular Automata Markov Chain Model: case of Saida and Jezzine regions in Lebanon, Sustain. Water Resour. Manag., 8(3), 71, https://doi.org/10.1007/s40899-022-00656-7, 2022.

Almino, L.M.D. and Rufino, I.A.A. Dynamic modeling and urban water demand scenarios: simulations in Campina Grande-PB, Eng. Sanit. Ambient., 26, 915–925, https://doi.org/10.1590/S1413-415220190015, 2021.

Avargani, H.K., Shahdany, S.M.H., Kamrani, K., Maestre, J.M., Garmdareh, S.E.H. and Liaghat, A. Prioritization of surface water distribution in irrigation districts to mitigate crop yield reduction during water scarcity, Agric. Water Manag. 269, 107653, https://doi.org/10.1016/j.agwat.2022.107653, 2022.

Botta-Dukát, Z.J.S.R. Quartile coefficient of variation is more robust than CV for traits calculated as a ratio, Sci. Rep., 13(1), 4671, https://doi.org/10.1038/s41598-023-31306-9, 2023.

Brunner, M.I., Zappa, M. and Stähli, M. Scale matters: Effects of temporal and spatial data resolution on water scarcity assessments, Adv. Water Resour., 123, 134–144, https://doi.org/10.1016/j.advwatres.2018.12.001, 2019.

Canchola, J., Tang, S., Hemyari, P., Paxinos, E. and Marins, E.J.M.P.B. Correct use of percent coefficient of variation (CV) formula for log-transformed data, MOJ Proteom. Bioinform., 6(4), 316–317, https://doi.org/10.15406/mojpb.2017.06.00200, 2017.

Cao, X., Wu, M., Zheng, Y., Guo, X., Chen, D. and Wang, W. Can China achieve food security through the development of irrigation ? , Reg. Environ. Change, 18(2), 465–475, https://doi.org/10.1007/s10113-017-1214-5, 2018.

Carvalho, T.M.N., Filho, F.d.A.d.S. and Porto, V.C. Urban Water Demand Modeling Using Machine Learning Techniques: Case Study of Fortaleza, Brazil, J. Water Resour. Plan. Manag., 147(1), 05020026, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001196, 2021.

Chen, J., Brissette, F.P. and Leconte, R. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, J. Hydrol., 401(3), 190–202, https://doi.org/10.1016/j.jhydrol.2011.02.020, 2011.

Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P. and Edmonds, J. Evaluating the economic impact of water scarcity in a changing world, Nat. Commun., 12(1), , https://doi.org/10.1038/s41467-021-21668-4, 2021.

Fu, Q., Zhou, M., Li, Y., Ye, X., Yang, M. and Wang, Y.J.G.A. Flow spatiotemporal Moran's I: Measuring the spatiotemporal autocorrelation of flow data, Geogr. Anal., 56(4), 799–824, https://doi.org/10.1111/gean.12397, 2024.

Galelli, S. and Castelletti, A. Tree-based iterative input variable selection for hydrological modeling, Water Resour. Res., 49(7), 4295–4310, https://doi.org/10.1002/wrcr.20481, 2013.

Gedamu, W.T., Plank-Wiedenbeck, U., Wodajo, B.T.J.A.A. and Prevention. A spatial autocorrelation analysis of road traffic crash by severity using Moran’s I spatial statistics: A comparative study of Addis Ababa and Berlin cities, Accid. Anal. Prev., 200, 107535, https://doi.org/10.1016/j.aap.2024.107535, 2024.

Guevara, M. and Vargas, R. Downscaling satellite soil moisture using geomorphometry and machine learning, PLoS ONE, 14(9), e0222590, https://doi.org/10.1371/journal.pone.0222590, 2019.

He, S.K., Guo, S.L., Liu, Z.J., Yin, J.B., Chen, K.B. and Wu, X.S. Uncertainty analysis of hydrological multi-model ensembles based on CBP-BMA method, Hydrol. Res., 49(5), 1636–1651, https://doi.org/10.2166/nh.2017.248, 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146(730), 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Horta, A., Oliveira, A.R., Azevedo, L. and Ramos, T.B.J.G.R. Assessing the use of digital soil maps in hydrological modeling for soil-water budget simulations-implications for water management plans in southern Portugal, Geoderma Reg., 36, e00741, https://doi.org/10.1016/j.geodrs.2023.e00741, 2024.

Hou, C., Li, Y., Sang, S., Zhao, X., Liu, Y., Liu, Y. and Zhao, F. High-resolution mapping of monthly industrial water withdrawal in China from 1965 to 2020, Earth Syst. Sci. Data, 16(5), 2449–2464,

https://doi.org/10.5194/essd-16-2449-2024, 2024.
Huang, Z., Hejazi, M., Li X., Tang Q., Vernon, C., Leng, G., Liu, Y., Doll P., Eisner, S., Gerten, D., Hanasaki N.and Wada Y. Reconstrcution of global gridded monthly sectoral water withdrawals for 1970-2010 and analysis of their spatiotemporal patterns. Hydrol. Earth Syst. Sci. 22(4):2117-2133, http://doi.org/10.5194/hess-22-2117-2018, 2018.
Huang, Z., Yuan, X. and Liu, X. The key drivers for the changes in global water scarcity: Water withdrawal versus water availability, J. Hydrol., 601, 126658, https://doi.org/10.1016/j.jhydrol.2021.126658, 2021.
Ji, Y., Zuo, Q., Zhao, C., Zhang, Z. and Wu, Q.J.E.I.A.R. Measurement and decomposition of green water resource utilization efficiency across multiple water use sectors in China: A perspective on water-saving potential, Environ. Impact Assess. Rev., 112, 107806, https://doi.org/10.1016/j.eiar.2025.107806, 2024.
Kaewmai, R., Grant, T., Eady, S., Mungkalasiri, J. and Musikavong, C. Improving regional water scarcity footprint characterization factors of an available water remaining (AWARE) method, Sci. Total Environ., 681, 444–455, https://doi.org/10.1016/j.scitotenv.2019.05.006, 2019.
Kang, S., Hao, X., Du, T., Tong, L., Su, X., Lu, H., Li, X., Huo, Z., Li, S. and Ding, R. Improving agricultural water productivity to ensure food security in China under changing environment: From research to practice, Agric. Water Manag., 179, 5–17, https://doi.org/10.1016/j.agwat.2016.05.019, 2017.
Knox, J.W., Haro-Monteagudo, D., Hess, T.M. and Morris, J. Identifying Trade-Offs and Reconciling Competing Demands for Water: Integrating Agriculture Into a Robust Decision-Making Framework, Earth's Future, 6(10), 1457–1470, https://doi.org/10.1029/2018EF000905, 2018.
Kovacevic, J., Cvijetinovic, Z., Stancic, N., Brodic, N. and Mihajlovic, D. New Downscaling Approach Using ESA CCI SM Products for Obtaining High Resolution Surface Soil Moisture, Remote Sens., 12(7), https://doi.org/10.3390/rs12071119, 2020.
Liu, D., Li, X., Guo, S., Rosbjerg, D. and Chen, H. Using a Bayesian Probabilistic Forecasting Model to Analyze the Uncertainty in Real-Time Dynamic Control of the Flood Limiting Water Level for Reservoir Operation, J. Hydrol. Eng., 20(2), 04014036, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000849, 2015.
Liu, D., Zhang, Y., Zhang, J., Xiong, L., Liu, P., Chen, H. and Yin, J. Rainfall estimation using measurement report data from time-division long term evolution networks, J. Hydrol., 600, 126530, https://doi.org/10.1016/j.jhydrol.2021.126530, 2021.
Liu, Y., Wang, N., Jiang, C., Archer, L. and Wang, Y. Temporal and spatial distribution of soil water and nitrate content affected by surface irrigation and fertilizer rate in silage corn fields, Sci. Rep., 10(1), 8317, https://doi.org/10.1038/s41598-020-65119-5, 2020.
Liu, Y., Zhang, J., Chen, L., Chu, H., Wang, J.Z. and Ma, L. SSAS: Spatiotemporal Scale Adaptive Selection for Improving Bias Correction on Precipitation, IEEE Trans. Cybern., 52(11), 12175–12188, doi: 10.1109/TCYB.2021.3072483, 2021.
Luo, Y., Zhang, Z., Li, Z., Chen, Y., Zhang, L., Cao, J. and Tao, F. Identifying the spatiotemporal changes of annual harvesting areas for three staple crops in China by integrating multi-data sources, Environ. Res. Lett., 15(7), 074003, https://doi.org/10.1088/1748-9326/ab9754, 2020.
Mekonnen, M.M. and Hoekstra, A.Y. Four billion people facing severe water scarcity, Sci. Adv., 2(2), e1500323, https://doi.org/10.1126/sciadv.1500323, 2016.
Noori, N. and Kalin, L. Coupling SWAT and ANN models for enhanced daily streamflow prediction, J. Hydrol., 533, 141–151, https://doi.org/10.1016/j.jhydrol.2015.12.020, 2016.
Rosa, L., Chiarelli, D.D., Rulli, M.C., Dell'Angelo, J. and D'Odorico, P. Global agricultural economic water scarcity, Sci. Adv., 6(18), eaay5552, https://doi.org/10.1126/sciadv.aay5552, 2020.
Sapino, F., Haer, T., Saiz-Santiago, P. and Pérez-Blanco, C.D. A multi-agent cellular automata model to explore water trading potential under information transaction costs, J. Hydrol., 618, 128669, https://doi.org/10.1016/j.jhydrol.2023.128669, 2023.
Sharifi, H., Roozbahani, A. and Hashemy Shahdany, S.M. Evaluating the Performance of Agricultural Water Distribution Systems Using FIS, ANN and ANFIS Intelligent Models, Water Resour. Manag., 35(6), 1797–1816, https://doi.org/10.1007/s11269-020-02685-3, 2021.
Shortridge, A. Practical limits of Moran's autocorrelation index for raster class maps. Computers Environment and Urban Systems. Computers, Comput. Environ. Urban Syst., 31(3), 362–371, https://doi.org/10.1016/j.compenvurbsys.2006.04.002, 2007.
Su, Y., Liao, S., Ren, J. and Zhao, Z.J.F.i.E.S. Research on the decoupling relationship between water resources utilization and economic development at the county scale in Qian'nan Prefecture, Guizhou Province, Front. Environ. Sci., 12, 1347652, https://doi.org/10.3389/fenvs.2024.1347652, 2024.
Sun, S., Tang, Q.H., Konar, M., Huang, Z.W., Gleeson, T., Ma, T., Fang, C.L. and Cai, X.M. Domestic Groundwater Depletion Supports China's Full Supply Chains, Water Resour. Res., 58(5), e2021WR031555, https://doi.org/10.1029/2021WR031555, 2022.

Sunkara, S.V. and Singh, R. Assessing the impact of the temporal resolution of performance indicators on optimal decisions of a water resources system, J. Hydrol., 612, 128185, https://doi.org/10.1016/j.jhydrol.2022.128185, 2022.
Taormina, R., Galelli, S., Karakaya, G. and Ahipasaoglu, S.D. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models, J. Hydrol., 542, 18–34, https://doi.org/10.1016/j.jhydrol.2016.08.034, 2016.
Tariq, A., Mumtaz, F., Majeed, M. and Zeng, X. Spatio-temporal assessment of land use land cover based on trajectories and cellular automata Markov modelling and its impact on land surface temperature of Lahore district Pakistan, Environ. Monit. Assess., 195(1), 195, https://doi.org/10.1007/s10661-022-09942-9, 2023.
Tiefelsdorf, M., Boots, B.J.E. and A, P. The exact distribution of Moran's I, Environ. Plan. A, 27(6), 985–999, https://doi.org/10.1068/a270985, 1995.
Wang, C., Tang, G., Xiong, W., Ma, Z. and Zhu, S. Infrared precipitation estimation using convolutional neural network for FengYun satellites, J. Hydrol., 603, 127113, https://doi.org/10.1016/j.jhydrol.2021.127113, 2021.
Wang, T., Guo, Z., Shen, Y., Cui, Z. and Goodwin, A. Accumulation mechanism of biofilm under different water shear forces along the networked pipelines in a drip irrigation system, Sci. Rep., 10(1), 6960, https://doi.org/10.1038/s41598-020-63602-1, 2020.
Wu, B., Tian, F., Zhang, M., Piao, S., Zeng, H., Zhu, W., Liu, J., Elnashar, A. and Lu, Y. Quantifying global agricultural water appropriation with data derived from earth observations, J. Clean. Prod., 358, 131891, https://doi.org/10.1016/j.jclepro.2022.131891, 2022.
Wu, J. and Lu, J. Spatial scale effects of landscape metrics on stream water quality and their seasonal changes, Water Res., 191, 116811, https://doi.org/10.1016/j.watres.2021.116811, 2021.
Yamada, H.J.M. A new perspective on Moran's coefficient: Revisited, Math., 12(2), 253, https://doi.org/10.3390/math12020253, 2024.
Yan, J., Jia, S., Lv, A. and Zhu, W. Water Resources Assessment of China's Transboundary River Basins Using a Machine Learning Approach, Water Resour. Res., 55(1), 632–655, https://doi.org/10.1029/2018WR023274, 2019.
Yang, Y.R., Xiong, Q.Y., Wu, C., Zou, Q.H., Yu, Y., Yi, H.L. and Gao, M. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism, Environ. Sci. Pollut. Res., 28(39), 55129–55139, https://doi.org/10.1007/s11356-021-14134-8, 2021.
Ye, Y., Huang, L., Zheng, Q., Liang, C., Dong, B., Deng, J. and Han, X. A feasible framework to downscale NPP-VIIRS nighttime light imagery using multi-source spatial variables and geographically weighted regression, Int. J. Appl. Earth Obs. Geoinf., 104, 102513, https://doi.org/10.1016/j.jag.2021.102513, 2021.
Yin, J., Guo, S., Gu, L., He, S., Ba, H., Tian, J., Li, Q. and Chen, J. Projected changes of bivariate flood quantiles and estimation uncertainty based on multi-model ensembles over China, J. Hydrol., 585, 124760, https://doi.org/10.1016/j.jhydrol.2019.124760, 2020.
Zhang, C. and Long, D. Estimating Spatially Explicit Irrigation Water Use Based on Remotely Sensed Evapotranspiration and Modeled Root Zone Soil Moisture, Water Resour. Res., 57(12), e2021WR031382, https://doi.org/10.1029/2021WR031382, 2021.
Zhang, J.Y., Liu, D.D., Guo, S.L., Xiong, L.H., Liu, P., Chen, J. and Yin, J.B. High resolution annual irrigation water use maps in China based-on input variables selection and convolutional neural networks, J. Clean. Prod., 405, 136175, https://doi.org/10.1016/j.jclepro.2023.136175, 2023.
Zhang, J., Liu, D., Xu, Y., Xiong, L., Chen, J., Chen, H., & Yin, J. Appropriate spatiotemporal scale selection for water use simulation in China, J. Hydrol., 2025, 133502, https://doi.org/_10.1016/j.jhydrol.2024.133502, 2025.
Zhang, Y., Yin, Y., Yin, M., Zhang X.. A High-Resolution Gridded Dataset for China's Monthly Sectoral Water Use. Sci Data 12, 1157. https://doi.org/10.1038/s41597-025-05400-2, 2025.
Zhou, F., Bo, Y., Ciais, P., Dumas, P., Tang, Q.H., Wang, X.H., Liu, J.G., Zheng, C.M., Polcher, J., Yin, Z., Guimberteau, M., Peng, S.S., Ottle, C., Zhao, X.N., Zhao, J.S., Tan, Q., Chen, L., Shen, H.Z., Yang, H., Piao, S.L., Wang, H. and Wada, Y. Deceleration of China's human water use and its key drivers, Proc. Natl. Acad. Sci. U.S.A., 117(14), 7702–7711, https://doi.org/10.1073/pnas.1913587117, 2020.

A multiple spatial scales water use simulation for capturing its spatial heterogeneity through cellular automata model

Jiayu Zhang[a], Dedi Liu*[,a,b,c], Jiaoyang Wang[a], Feng Yue[a], Hanxu Liang[a], Zhengbo Peng[a], Wei Guan[a]

[a] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China

[b]Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University, Wuhan, China

[c] Department of Earth Science, University of the Western Cape, Robert Sobukwe Road, Bellville 7535, Republic of South Africa

* Correspondence to: Dedi Liu: dediliu@whu.edu.cn

Abstract: Reliable water use simulation is essential for sustainable water resource planning, especially under intensifying pressures from climate change, population growth, and socio-economic transitions. While previous studies have extensively explored water availability as supply side modeling across multiple spatial scales for its spatial heterogeneity, the water demand side remains relatively underdeveloped—often constrained by fixed spatial scales and coarse statistical data that assume spatial homogeneity. This mismatch between supply side and demand side limits the ability of existing models to accurately represent spatial heterogeneity in water use and brings uncertainty into water resource allocation strategies. To address this mismatch, we propose a novel multi-scale water use simulation framework by integrating cellular automata (CA) model with Generalized Likelihood Uncertainty Estimation (GLUE). The CA model captures the spatial heterogeneity of water use through the grid-based update rules. Two update rules are adopted—probability rule (i.e., capturing stochastic transitions via distribution fitting) and linear rule (i.e., modeling neighborhood-weighted evolution). To evaluate the impacts of spatial scale on water use heterogeneity, simulations are conducted at three spatial scales: 1 km, appropriate scale, and prefecture scale across 341 prefectures in China. Results show that both the update rule and spatial scale significantly affect spatial heterogeneity and uncertainty of water use. The probability rule can capture the broader variability but results in higher Root Mean Squared Error (*RMSE*) and Relative Error (*RE*) while the linear rule brings more stable performance with lower errors. While the 1 km scale increases uncertainty due to sensitivity to local fluctuations, and the prefecture scale suppresses spatial details, the appropriate scale offers the best trade-off between stability and spatial heterogeneity. The uncertainty quantified by GLUE, expresses as confidence intervals, varies across prefectures and spatial scales. Overall, the proposed framework offers a flexible tool for multi-scale water use simulation and highlights the critical role of spatial heterogeneity, thereby supporting adaptive water resource planning and management.

Key words: water use; spatial scale; cellular automata; multi-scale simulation

1 Introduction

Water scarcity has become one of the most pressing global challenges, exacerbated by climate change, population growth, and unsustainable water use practices (Avargani et al. 2022, Huang et al. 2021, Kaewmai et al. 2019, Rosa et al. 2020). Nearly 2.3 billion people is living in regions facing water scarcity (Brunner et al. 2019, Dolan et al. 2021, Mekonnen and Hoekstra 2016). In this context, accurate and timely assessments of water scarcity are essential for effective water management, resource allocation, and policy-making (Avargani et al. 2022, Cao et al. 2018). A scientifically rigorous water scarcity assessment requires a comprehensive understanding of both available water resources and water use (Brunner et al. 2019, Ji et al. 2025, Sun et al. 2022). However, these two components are often represented at incompatible spatial scales, resulting in mismatches that complicate accurate evaluation and integrated water resources planning (Almino and Rufino 2021, Kang et al. 2017).

In the past few decades, significant advancements in hydrological modeling, satellite remote sensing, and reanalysis datasets have enabled researchers to simulate surface and ground available water resources across various temporal and spatial scales—from daily basin-scale runoff forecasts to long-term continental water balance projections (Horta et al. 2024, Su et al. 2024, Yang et al. 2021, Zhang and Long 2021). These advancements have laid the foundation for widely used tools such as SWAT, VIC, H08, and PCR-GLOBWB, which help water managers better understand available water resources under different spatial-temporal scales (Hersbach et al. 2020, Kovacevic et al. 2020, Noori and Kalin 2016, Sunkara and Singh 2022). Though there has been water use simulation progress, it still faces significant challenges and has yet to achieve the same spatial level of sophistication as the simulation of available water resources. The primary limitations of water use simulation lies in the available spatial scales of water use data (Hou et al. 2024, Zhang et al. 2023). Much of the existing researches rely on a coarse, aggregated dataset such as national statistics, sectoral usage reports,

or administrative boundaries (e.g., counties, provinces) (Carvalho et al. 2021, Wu et al. 2022, Zhang et al. 2023). These datasets inherently assume uniform water use within each administrative unit, and often overlook the spatial heterogeneity of water use. This oversimplification limits the ability of water use simulations to capture spatial variation, particularly in regions with pronounced heterogeneity (Brunner et al. 2019, Su et al. 2024).

Simulating water use at a grid scale provides a promising solution to address the spatial scale mismatch, allowing for a more accurate representation of water use dynamics and capturing regional variations (Su et al. 2024, Wu and Lu 2021). A grid-based approach allows for a more accurate representation of water use dynamics by capturing fine-scale spatial heterogeneity, which administrative survey data often overlook. Zhang et al. (2023) integrated the Iterative Input Selection algorithm with Convolutional Neural Networks to simulate annual irrigation water use, producing high-resolution grid maps with a spatial resolution of 1 km for mainland China. Hou et al. (2024) developed a monthly dataset on industrial water withdrawals, incorporating spatial resolutions of 0.1° and 0.25° by utilizing enterprise data, product yields, and water use statistics. These studies have demonstrated that the downscaled water use data from administrative level to grid scale can catch a more detailed and region-specific representation of water use patterns, thereby improving their applicability in water resource management and water scarcity assessment. However, these studies still rely on a fixed spatial grid scale, limiting their ability to capture the full spatial heterogeneity of water use. The effects of spatial resolution on the representation of water use heterogeneity remain insufficiently explored. In particular, finer scales (such as 1 km grids) can capture localized variations in water use, coarser scales (such as regional or administrative boundaries) tend to smooth over spatial differences, potentially obscuring critical patterns (Luo et al. 2020, Sun et al. 2022). Therefore, simulating water use across multiple spatial scales is essential for capturing the full spectrum of spatial heterogeneity. By incorporating the results at different spatial scales,

this simulation can account for the fine-scale dynamics of water use in urban, agricultural, and industrial areas more effectively, offering a comprehensive understanding of water use patterns and improving the accuracy of water scarcity assessments (Su et al. 2024, Sunkara and Singh 2022).

Since the spatial scale of water use simulation depends on the spatial scale of the input data (Horta et al. 2024, Sharifi et al. 2021, Zhang et al. 2023), achieving multi-scale water use simulation requires a model that can flexibly handle different input spatial scales. Such a model should be adaptable to varying spatial resolutions and ensure accurate simulation of water use, whether the focus is on fine-scale urban areas, intermediate regional levels, or broader national assessments. The cellular automata (CA) model, with its grid-based structure, offers a suitable framework for spatially explicit modeling across multiple scales by adjusting both the spatial resolution of input data and the design of updating rules (Al-Shaar et al. 2022, Sapino et al. 2023, Tariq et al. 2023, Wang et al. 2020). The CA model can thus be employed to support multi-scale water use simulation. And as uncertainties are inherently associated with the selection of spatial scales and update rules (Yin et al. 2020, Zhang and Long 2021), it is critical to quantify and address these uncertainties to ensure the reliability of simulation outcomes.

The aim of this research is to develop a multi-scale water use simulation framework that specifically addresses the impact of spatial scale on the spatial heterogeneity of water use. The framework is mainly composed of the CA model for simulating water use dynamics across various spatial scales and an uncertainty analysis technique for quantifying the uncertainties of the simulation. The proposed framework will not only facilitate a deeper understanding of the influences of spatial scale on water use heterogeneity in diverse regions, but also improves the accuracy of water scarcity assessments and supports more effective resource management. Ultimately, the proposed framework contributes to advancing sustainable water governance in areas facing pronounced water stress.

2. Methodology

To develop a multi-scale water use simulation model, the dynamic spatial scale simulation capabilities of the CA model should be firstly leveraged. The particularly advantage of CA model is modeling complex global dynamics through simple local interactions and transition rules (Al-Shaar et al. 2022, Liu et al. 2021, Tariq et al. 2023). Moreover, the grid-based structure of the CA model allows it to flexibly accommodate various spatial resolutions, making it well-suited for modeling water use at multiple spatial scales (Sapino et al. 2023, Wang et al. 2020). In this framework, water use grid maps at different spatial scales are first prepared as inputs to the CA model. Each grid is treated as an individual cell in the CA model, with the water use amount at each cell representing its state. Based on the neighborhood configuration and the initial baseline year water use data, water use values for future years are simulated using predefined update rules for each cell at multiple spatial scales. The procedure of our proposed CA model for multi-scale water use simulation is depicted in Figure 1.

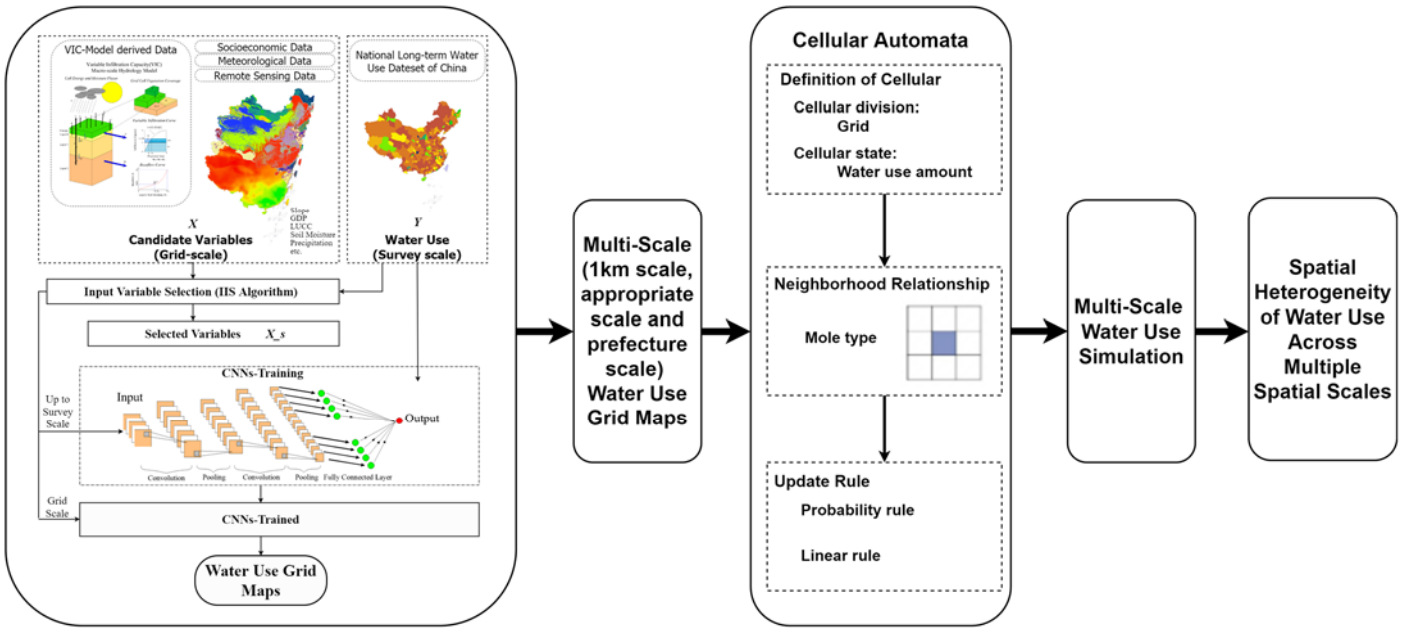


Figure 1. Procedure of the multi-scale water use simulation

To capture the spatial heterogeneity of water use dynamics at different scales, the Coefficient of Variation

(CV) (Abdi 2010) and Moran's I spatial autocorrelation index (Tiefelsdorf et al. 1995) are employed here. The CV measures the relative variability of water use across regions, with higher values indicating greater spatial heterogeneity. Moran's I assesses the degree of spatial autocorrelation, identifying whether similar values are spatially clustered (positive values) or dispersed (negative values). These two indices are applied to evaluate the spatial heterogeneity of water use grid maps at different spatial scales (i.e., 1 km, appropriate scale, and prefecture scale), and figure out the influences of the spatial variations on water use dynamics and uncertainty in the simulation results.

To quantify the uncertainties of water use simulation for providing a more robust foundation for water scarcity assessment and policy-making (He et al. 2018, Knox et al. 2018, Sharifi et al. 2021), the Generalized Likelihood Uncertainty Estimation (GLUE) method is adopted here. GLUE provides a probabilistic framework for model evaluation by exploring a wide range of parameter sets and assigning likelihoods to each based on their performance. This approach effectively addresses model equifinality—where multiple parameter combinations yield similar outputs—and is particularly well-suited for complex, non-linear models such as cellular automata (Liu et al. 2015, Yin et al. 2020).

2.1 Water Use Grid Maps generating

The spatial scale of water use simulation is determined by the spatial scale of the input data, so water use grid maps at different spatial scales were prepared as input to the simulation model. Here, water use refers to the total water consumption across three major sectors: irrigation, industrial, and urban domestic water use. The water use data considered in this study account for both groundwater and surface water sources (e.g., rivers, lakes, and reservoirs). These data were drawn from Zhou et al. (2020), which compiled water use data across 341 Chinese prefectures. The dataset includes water consumption data for irrigation, industrial, and domestic uses, incorporating both groundwater and surface water sources. The water use data were sourced

from two major nationally coordinated surveys: the First and Second National Water Resources Assessment Programs (1965–2000) and the Water Resources Bulletins published by 31 provincial governments (2001–2013). Both surveys were led by the Ministry of Water Resources of China, and followed consistent methodologies in terms of definitions, survey units, sector classifications, field measurements, and quality assurance. To obtain the water use grid maps, several steps should be done to convert the water use data at administrative survey scale into spatially explicit grids of varying resolutions.

The grid maps of irrigation, domestic, and industrial water use are generated from the prefecture-level statistical survey data and water use sector-specific predictor variables. For each sector, the most relevant input variables are identified through an iterative input variables selection algorithm (Zhang et al., 2023; Zhang et al. 2025). Specifically, irrigation water use was modeled by the potential evapotranspiration, normalized difference vegetation index (*NDVI*), rainfall and soil moisture; domestic water use was modeled by population, rainfall, temperature and night-light; industrial water use was modeled by *GDP*, night-light, population and rainfall. And then these sectoral gird maps were aggregated to form total water use grid maps for modeling. This aggregation is done for two reasons: the first one is that the temporal trend of the total water use has become stable during the study period (1998–2013) and future. The average annual growth rate is only about 0.87% (from 505.53 billion m^3 in 1998 to 575.44 billion m^3 in 2013) due to the policy interventions, technological improvements, and industrial structure changes. Since the primary objective of our study is to examine the influence of spatial scale on the spatial heterogeneity of water use, a temporally stable indicator helps minimize the confounding effects of sector-specific temporal fluctuations; the second reason is that the total water use can figure out the scale effects across regions instead of the sector-level temporal variability while the sectoral differences are implicitly in the inputs before the aggregation.

Earlier studies often applied a fixed spatial resolution in different regions, which could not account for

differences in land area, natural endowments, and water use structures, and leaded to the discrepancies in information density and potential over- or underestimation of water use. To address this issue, an appropriate spatial scale can be determined by the deep learning-based spatiotemporal scale adaptive selection model (Liu et al., 2021; Zhang et al., 2025). And the model can balance the accuracy of the simulation based on the spatial information density of gridded water use data, and its results vary across prefectures. The spatial scale selection module in the selection model figures out the appropriate spatial scale by maximizing information density while balancing simulation accuracy in terms of the Conditional Entropy, Kullback–Leibler Divergence Loss and Relative Error performance metrics. This selection module enables each prefecture to adopt its own appropriate spatial scale rather than a fix resolution. The detailed values of the appropriate spatial scale (in km) for each prefecture and water use sector (irrigation, industrial, and domestic) are provided in an accompanying Excel file, which can be accessed and downloaded via the data link: 10.6084/m9.figshare.30445157. This file allows users to examine the spatial heterogeneity of the appropriate scale across regions in detail. Finally, total water use grid maps are generated at three spatial resolutions: the small scale (e.g., 1 km), the appropriate spatial scale as determined by the selection module, and the prefecture scale as the statistical survey water use data.

2.2 Cellular automata model for water use simulation

The CA model, grounded in complexity theory, is widely used in land use and urban growth modeling. It provides a robust platform for simulating spatial phenomena governed by local interactions and transition rules (Sapino et al. 2023, Tariq et al. 2023). Each cell in a CA model represents a discrete spatial unit that updates its state over time based on predefined rules and the states of its neighboring cells. It's decentralized, bottom-up modeling structure enables the simulation of complex global behaviors emerging from simple local dynamics (Al-Shaar et al. 2022, Wang et al. 2020). The CA model is introduced to simulate the temporal

evolution of water use at the grid scale, complementing the static spatial distribution obtained from the Convolutional Neural Network (CNN) downscaling. While the CNN model effectively reconstructs the spatial pattern of water use for each prefecture based on physical and socioeconomic predictors (Zhang et al., 2023), it does not explicitly account for the spatial dependence and interactions among adjacent grid cells. The CA model addresses this limitation by incorporating spatial adjacency effects and feedback mechanisms, allowing each cell's water use to be influenced by its neighbors. This enables the model to represent the diffusion and clustering behaviors of water use, which are essential for capturing the spatial heterogeneity and dynamic interactions of human water activities.

Both the probability and the linear update rules are designed and tested to capture the dual nature of water use dynamics. The probability rule has been widely applied in significant spatial and temporal variation areas in land use simulation and other fields. It will be designed here for the water use at different scales. Rather than assuming temporal stability, the probability rule explicitly incorporates the variations through calibrating the state transition matrix and probability distributions for each prefecture independently by the own historical water use record. This rule enables the simulation to capture both the structured temporal dependence and the inherent randomness in water use, ensuring adaptability to local conditions. The linear update rule assumes that changes in water use are more deterministic and can be approximated as a linear combination of the cell's own state and those of its neighbors. This rule is more appropriate for long-term, high spatial autocorrelation, and persistent patterns. After implementing and comparing the water use simulation results of the two rules in the CA framework, their results can assess the relative effectiveness of stochastic versus deterministic update mechanisms across different spatial scales. These two rules not only strengthen the robustness of the modeling framework but also provide insights into the dominant processes shaping water use dynamics in different regions.

2.2.1 Probability rule in CA

The probability rule in the CA model is designed to represent the stochastic state transitions of water use over time. It abstracts the temporal dynamics of water use at the grid level into a probabilistic transition framework that can be applied consistently across different spatial scales and regions, while remaining adaptable to significant spatial and temporal variations. This adaptability is achieved by calibrating the update rule separately for each prefecture by its own historical water use record for appropriately capturing both long-term trends and localized fluctuations. In this approach, the state of each grid cell (i.e., representing the amount of water use) is divided into k distinct intervals using equal-frequency categorization based on the cell's historical water use record. This categorization ensures that the intervals reflect the variations in water use over time. For each interval, the most suitable statistical distribution is selected using the Akaike Information Criterion (AIC). The selection process enables the model to represent the probabilistic characteristics of water use within each intensity class. And the distribution is chosen from a set of candidate distributions, including normal, lognormal, exponential, gamma, and uniform.

Once the optimal distribution is found for each interval, a state transition matrix is constructed based on observed transitions of grid cells between intervals from one year to the next. The transition matrix captures the likelihood of a grid cell moving from its current water use state to another in the subsequent time step. The model first generates the next state probabilistically through the transition matrix, and then the water use samples are generated from the corresponding probability distribution. These two steps incorporate both the structured temporal dependence and the inherent randomness in future water use patterns.

In the probability rule, the calibrated parameter is the number of state intervals and is denoted as k. The value of k directly affects the granularity of the state categorization and the accuracy of the state transition matrix. A larger k increases the resolution of the state representation and captures the finer variations in water

use, but a larger k can also lead to overfitting. And a smaller k oversimplifies the demand pattern. To calibrate the parameter k, the historical and observed water use data is divided into a calibration and a validation sets. And the performances of the model with different k values is then evaluated by the Root Mean Squared Error (*RMSE*) and Relative Error (*RE*) metrics. The optimal k can be calibrated by the minimums of the *RMSE* and *RE* in the validation period, ensuring a balance between model accuracy and generalizability.

2.2.2 Linear rule in CA

The linear rule in CA updates the water use of each cell according to the weighted average of its current state and the states of its neighboring cells. The linear rule assumes that the water use at a given grid cell is influenced by both its own previous state and the water use of adjacent cells. The linear rule is expressed as the Equation (1).

$$W(t+1) = \alpha W(t) + \beta \sum_{i=1}^{n} \omega_i W_i(t) \quad (1)$$

where $W(t+1)$ is the predicted water use of the central cell at time $t+1$, $W(t)$ is the current water use of the central cell at time t, $W_i(t+1)$ represents the water use at the i^{th} neighboring cell ($i=1,2,\ldots,8$) at time t, ω_i is the weight that is assigned to the i^{th} neighboring cell, α and β are coefficients that control the relative influence of the central cell's own water use and the neighboring cells' water use.

To accurately capture the influences of neighboring cells, the weight ω_i for each neighboring cell is determined by an inverse distance weighting scheme with an exponential decay and is expressed as the Equation (2).

$$\omega_i = \frac{1}{d_i^p} \quad (2)$$

where d_i is the Euclidean distance between the central cell and the i^{th} neighboring cell, p is an adjustable exponent that controls the rate of decay in the influence of neighboring cells.

To determine the parameters α, β, and p, the observed period is also divided into a calibration period and a validation period. After the model runs with different parameter combinations, and their performances are assessed in terms of *RMSE* and *RE*, the optimal parameter set will be picked out by the minimum of the performance metrics in the validation period.

2.3 Generalized Likelihood Uncertainty Estimation for water use simulation

Generalized Likelihood Uncertainty (GLUE) is a probabilistic approach that evaluates a model's performance by considering a wide range of plausible parameter sets and quantifying the likelihood of each set in its ability to reproduce observed data (Chen et al. 2011, Sharifi et al. 2021, Taormina et al. 2016). Unlike traditional deterministic calibration methods, GLUE acknowledges the inherent equifinality in environmental modeling that also is the possibility of the acceptable results from the multiple parameters set. The GLUE is incorporated into the CA model to assess the uncertainty in water use simulations.

The GLUE is applied in both the probability rule and the linear rule of the CA model to evaluate the uncertainty of water use simulations. There are six steps. (1) Parameter Selection: the number of state intervals k for the probability rule**,** and the self-influence coefficient α, the neighboring influence coefficient β, and the distance decay exponent p for the linear rule; (2) Parameter Sampling: a large number of parameter sets are generated by the uniform sampling within specified ranges for each parameter; (3) Model Simulation: the CA model is executed by each parameter set to simulate water use for the target period; (4) Likelihood Estimation: *RMSE* and *RE* values are calculated for each simulation to evaluate how well the simulated results match the observed data; (5) Behavioral Parameter Identification: behavioral parameter sets can yield acceptable likelihood values according to the thresholds that is determined by the calibration data; (6) Uncertainty Quantification: from the range of outputs from the behavioral parameter sets, a prediction interval is constructed quantify the uncertainty associated with the water use projections.

2.4 Spatial Heterogeneity of water use

To comprehensively understand the variability and spatial relationships of water use patterns across different spatial scales, Coefficient of Variation (CV) (Abdi 2010) and Moran's I (Tiefelsdorf et al. 1995) are used to analyze the spatial heterogeneity of water use. CV is defined as the ratio of the standard deviation to the mean of water use values at each spatial scale, offering a normalized measure of dispersion. It is a primary indicator used to quantify variability in water use across grid cells (Canchola et al. 2017). A higher CV value indicates greater spatial variation in water use, while a lower CV suggests more uniform water use patterns across the region. Once the CV for each grid cell is calculated, the extent of water use spatial heterogeneity is assessed, and areas with high variability are identified as more susceptible to water stress and fluctuations in demand (Botta-Dukát 2023, Liu et al. 2020).

Moran's I can reflect the spatial autocorrelation by measuring the degree to which one grid cell's water use is similar to that of neighboring grid cells. Moran's I provides an overall measure of spatial dependence, where a positive Moran's I indicates a clustering of similar water use values (either high or low) in neighboring grid cells, while a negative Moran's I suggests a dispersed pattern. A value close to zero indicates a random spatial pattern with no significant clustering (Gedamu et al. 2024, Shortridge 2007). By calculating Moran's I across different spatial scales, we are able to detect whether water use patterns exhibit spatial clustering or they are more randomly distributed. Moran's I can therefore help us understand the regional patterns of water use and identify areas that require targeted management interventions (Fu et al. 2024, Yamada 2024).

Both the CV and Moran's I provide a robust framework to analyze spatial heterogeneity in water use. The CV provides a measure of variability, while Moran's I reveals the extent of spatial correlation in water use. They offer a more comprehensive understanding of the spatial dynamics across different regions.

3 Study area and datasets

3.1 Study Area

Situated on the northwestern shore of the Pacific Ocean, China boasts vast territory, a large population, and diverse climate conditions (Ji et al. 2025, Sun et al. 2022). Over the years, the volume of water use in China has surged significantly rising from 3305 km^3 in 1965 to 5925 km^3 in 2024 (Ji et al. 2025). Owing to the country's varied climate conditions and pronounced spatiotemporal heterogeneity of water resources, water scarcity has become a recurrent challenge, exacerbated by the looming threats of climate change and rapid socio-economic development (Hou et al. 2024, Wang et al. 2021). Notably, China is expected to experience greater impacts from climate change than the global average (Kang et al. 2017, Sun et al. 2022, Yan et al. 2019). Therefore, conducting a multi-scale water use simulation study in China is particularly meaningful due to the country’s complex water resource challenges, amplified by regional disparities, climate change, and rapid socio-economic development. China comprises 31 provincial-level administrative divisions, including provinces, autonomous regions, and municipalities. As outlined in the study by Zhou et al. (2020), the administrative divisions are categorized at the prefecture level, totaling 341 prefectures (Figure. 2).

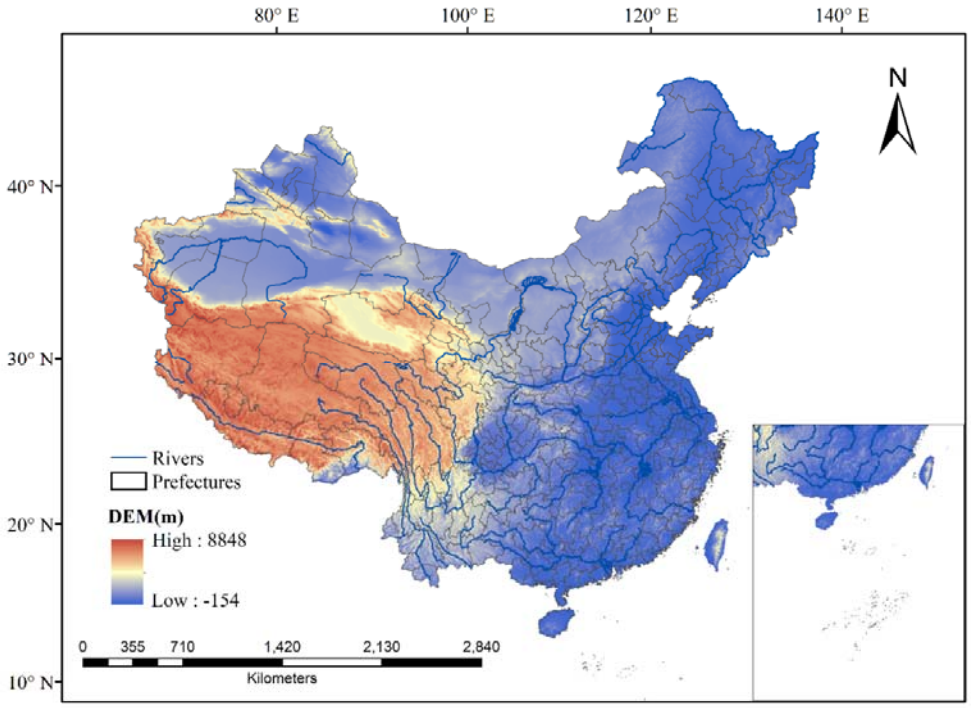


Figure. 2 Prefectures and major rivers in study area

3.2 Datasets and Preprocessing

To obtain water use grid maps, the observed datasets related to irrigation water use, domestic water use, and industrial water use were collected. The observed datasets are composed of the annual water use statistical survey data at the administrative scale, soil moisture data derived from hydrological models, socio-economic data like *GDP* (Gross Domestic Product) and population, meteorological data from point observations, and satellite remote sensing data including the normalized difference vegetation index (*NDVI*) and night light data.

Various spatial interpolation and downscaling methods are employed to transform the datasets into different spatial scales. To simulate water use at different spatial scales, we transform the spatial scale of the dataset to 1km, appropriate spatial scale (Zhang, et al., 2025) and prefecture scale. The details of the transform methods can be found in the reference of Zhang et al. (2023). Nighttime light remote sensing data is also combined with the land use information to perform upscaling or downscaling at spatial scales (Ye et al. 2021).All the adopted datasets and their corresponding preprocessing methods are listed in Table 1.

Table 1. Datasets and corresponding preprocessing methods

Variables	Data Sources	Origin Data Format	Origin Resolution	Processing Method
Precipitation (P)	CMA*	Point observation	\	Interpolation (ANUSPLIN)
Temperature (T)	CMA*	Point observation	\	Interpolation (IDW)
LUCC	RESDC**	Grid	30 m × 30 m	Upscale (Resample)
NDVI	RESDC**	Grid	1 km × 1 km	Upscale/downscale (Resample)
GDP	RESDC**	Grid	1 km × 1 km	Upscale/downscale (Resample)
Population	RESDC**	Grid	1 km × 1 km	Upscale/downscale (Resample)
DEM	SRTM	Grid	1 km × 1 km	Upscale/downscale (Resample)
Potential Evapotranspiration (PET)	\	\	\	P-M Equation/Interpolation (IDW)
Soil moisture (SM)	GDFC*** (He and Sheffield 2020)	NetCDF	0.25° × 0.25°	Downscale (Machine learning)(Guevara and Vargas 2019)
Night light (N-L)	DMSP-OLS	Grid	2.7 km × 2.7km	Upscale/downscale (Resample)
National Long-term Water Use Dataset of China	FigShare-PNAS(Zhou et al. 2020)	Excel	\	\

CMA*: China Meteorological Administration (http://data.cma.cn/); RESDC**: Resource and Environment Science and Data Center (https://www.resdc.cn/).

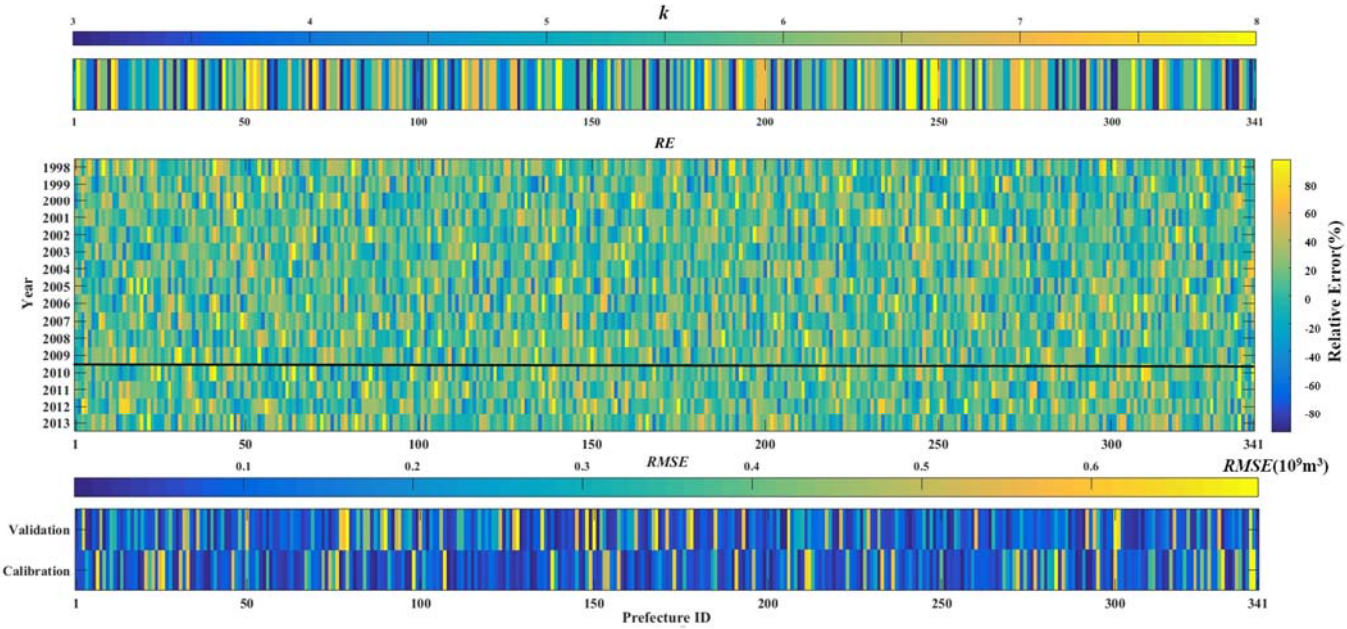
GDFC***: Global Drought and Flood Catalogue (http://hydrology.princeton.edu/data).

4. Results

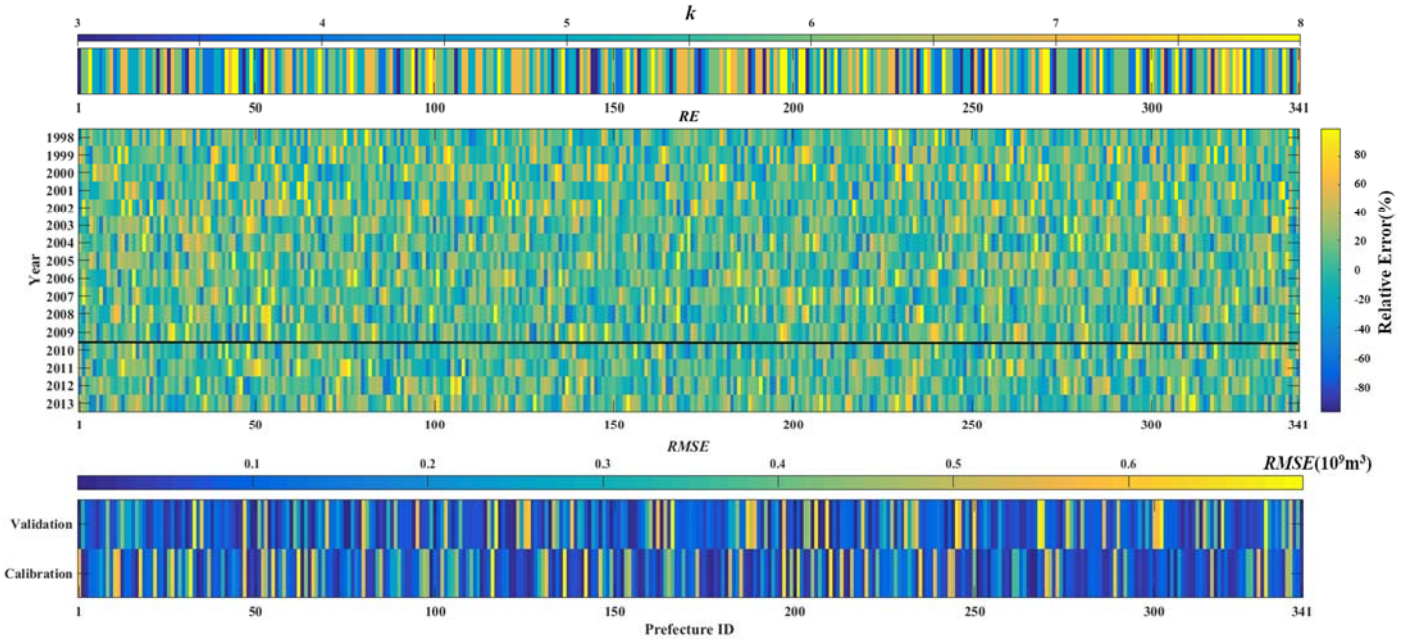
4.1 Water use simulation by CA model

4.1.1 Water use simulation from the probability rule CA model

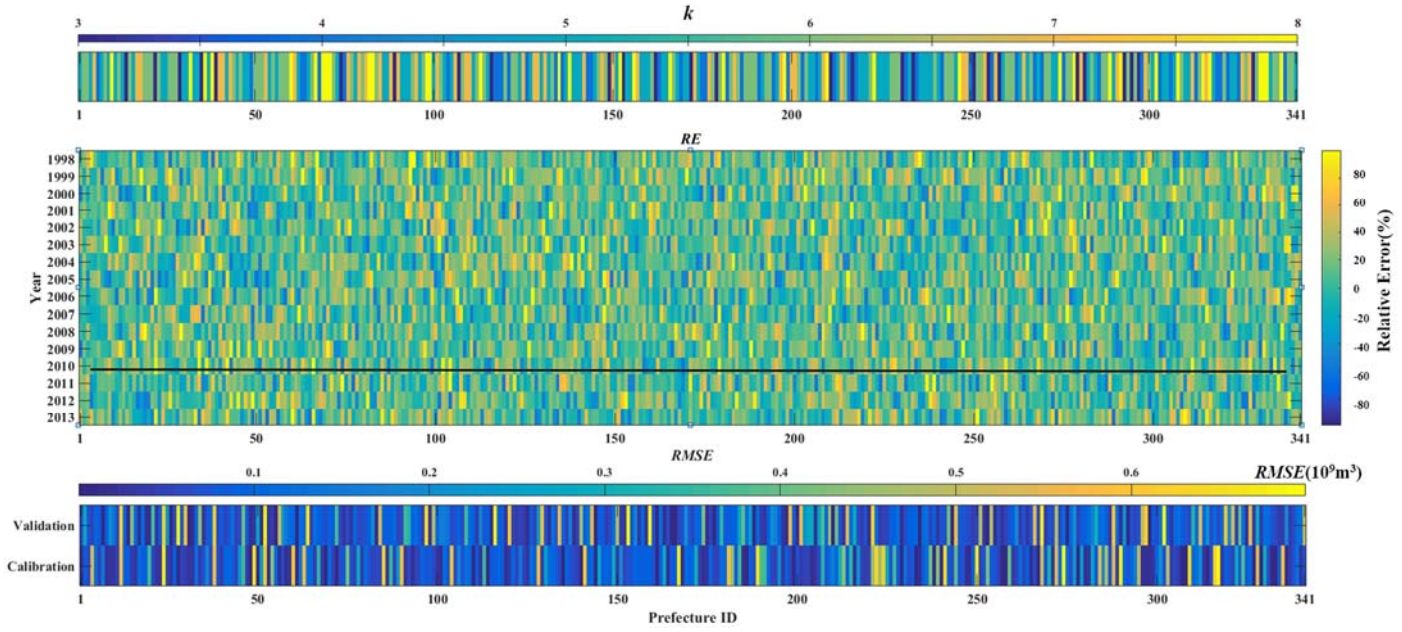
In the CA model with the probability rule, the number of state intervals (k) is the only parameter to be calibrated. The dataset from 1998–2009 is used for calibration and 2010–2013 for validation, with *RMSE* and *RE* as performance metrics. The optimal value of k at three spatial scales (1 km, appropriate scale, prefecture scale) for each prefecture is determined by minimizing *RMSE* and *RE* in the validation period. The calibrated k values for each prefecture, along with the corresponding *RMSE* and *RE* during the calibration (1998–2009) and validation (2010–2013) periods at the three spatial scales, are presented in Figure 3.



(a)



(b)



(c)

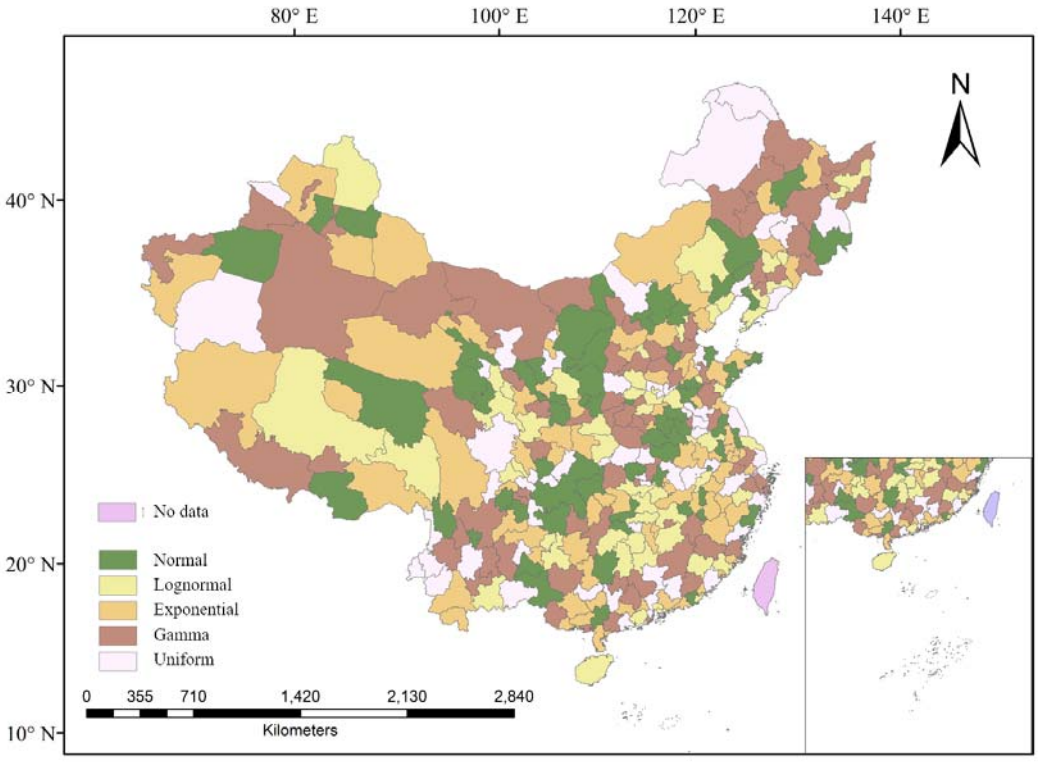
Figure .3 Optimal parameters of the probability rule CA model and the model performances at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

According to the results shown in Figure 3, the calibrated parameter *k* exhibits clear spatial heterogeneity across prefectures and varies with spatial scales. At the 1 km scale (Figure 3(a)), most prefectures show *k* values concentrated around 5–6, corresponding to relatively low *RMSE* and *RE* values. This suggests that a

moderate number of state intervals can effectively capture local water use variability while avoiding overfitting. In these areas, the probability distributions and transition probabilities appear to reflect stable temporal patterns, resulting in more accurate simulations. At the appropriate spatial scale (Figure 3(b)), the distribution of *k* becomes more diversified among prefectures. Some regions require larger *k* values (>=7) to preserve finer distinctions in water use states, while others perform better with smaller *k* values (<=4) that smooth out excessive variability. Different from the results at the 1 km scale, the overall *RMSE* and *RE* values are slightly higher, indicating that while the appropriate scale balances detail and generalization, it may not fully capture abrupt local changes in some prefectures. At the prefecture scale (Figure 3(c)), *k* values are generally smaller (mostly 3–4), reflecting the reduced spatial detail at coarser resolution. Thus, their accuracies of the simulation decrease, with higher *RMSE* and *RE* values. When the input variability of small scale is strongly aggregated at large scales, fewer state intervals oversimplify the temporal transitions, leading to greater deviations from observed water use patterns.

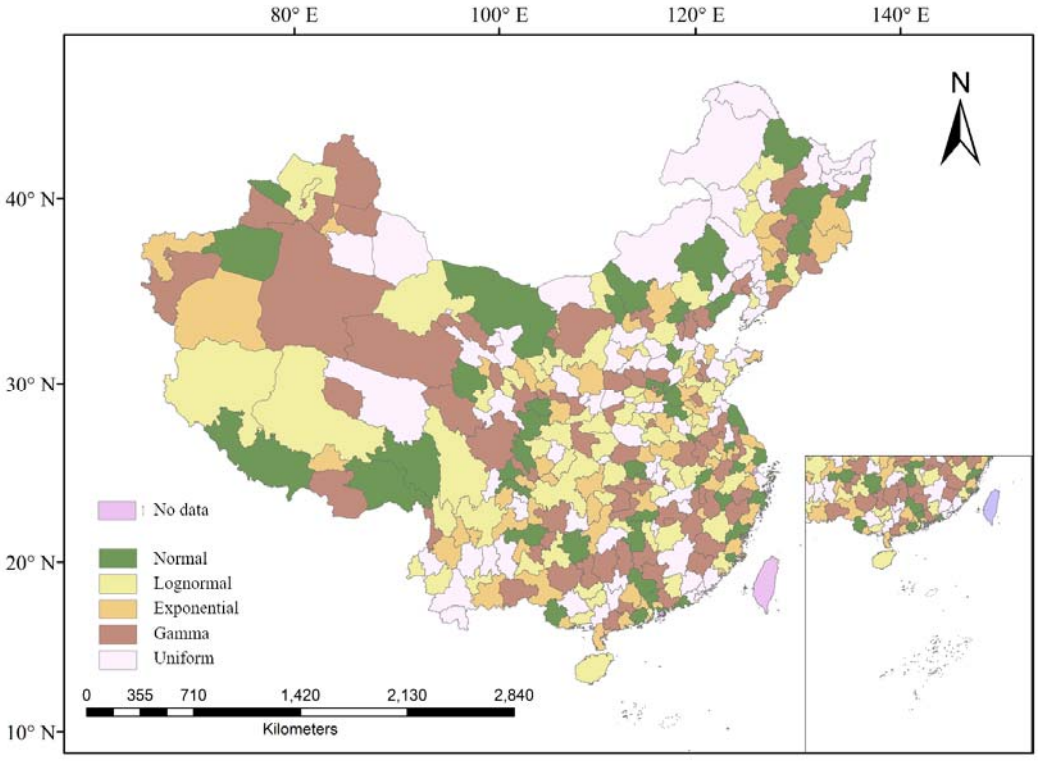
After determining the optimal *k* values at each scale, the next step is to characterize the statistical nature of water use within each state interval. The Akaike Information Criterion (AIC) is taken as performance metric to select the most suitable probability distribution for each interval in every prefecture. The AIC can balance the model fitness and the complexity through penalizing excessive parameters, it can reduce the risk of overfitting. The selected distribution types not only fit the historical data well but also is used to generate the future scenarios. The complete AIC values and corresponding best-fitting distribution types for each prefecture are provided in an open-access Excel file, which can be downloaded from: 10.6084/m9.figshare.30445157. The results of the optimal probability distributions for water use grids at the three different spatial scales (i.e., 1 km scale, appropriate spatial scale, and prefecture scale) are shown in Figure 4. These distributions, combined with the calibrated *k* values, form the basis of the probability rule CA model's ability to reproduce

369 the spatial and temporal heterogeneity of water use.



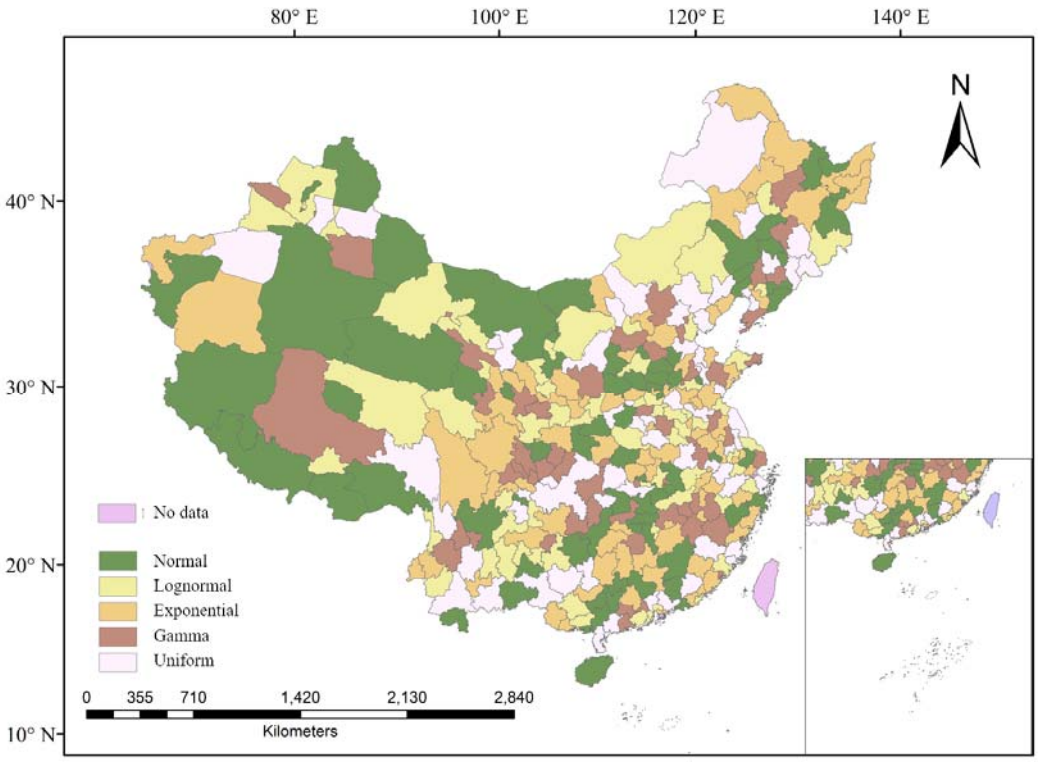
370

371 (a)



372

373 (b)



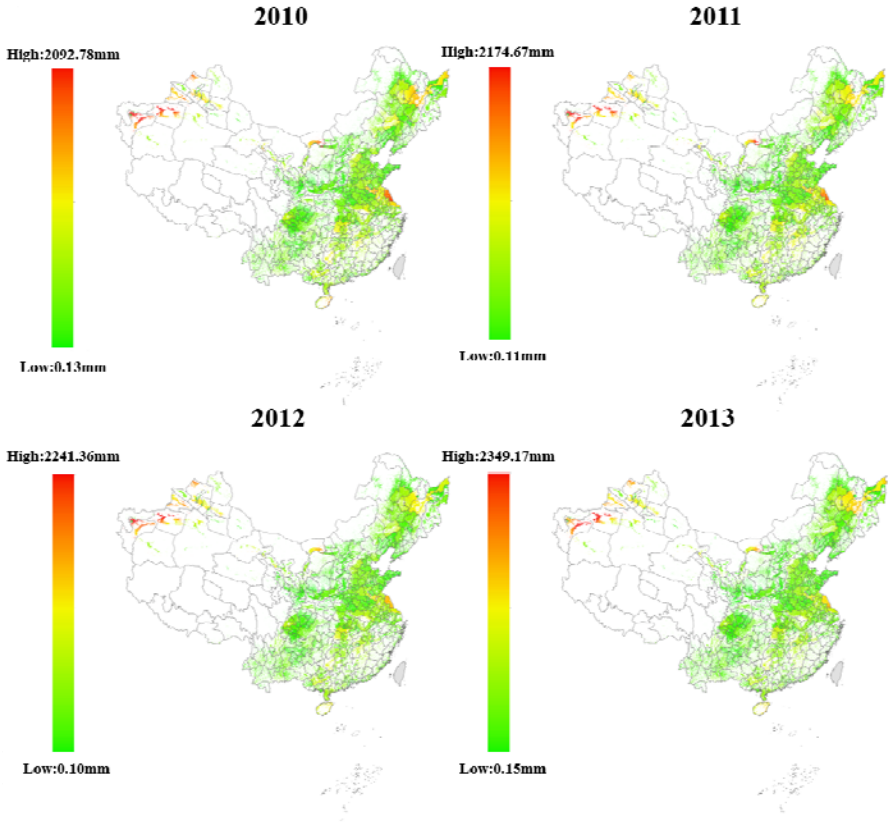
(c)

Figure. 4 Probability distribution types of water use at three spatial scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

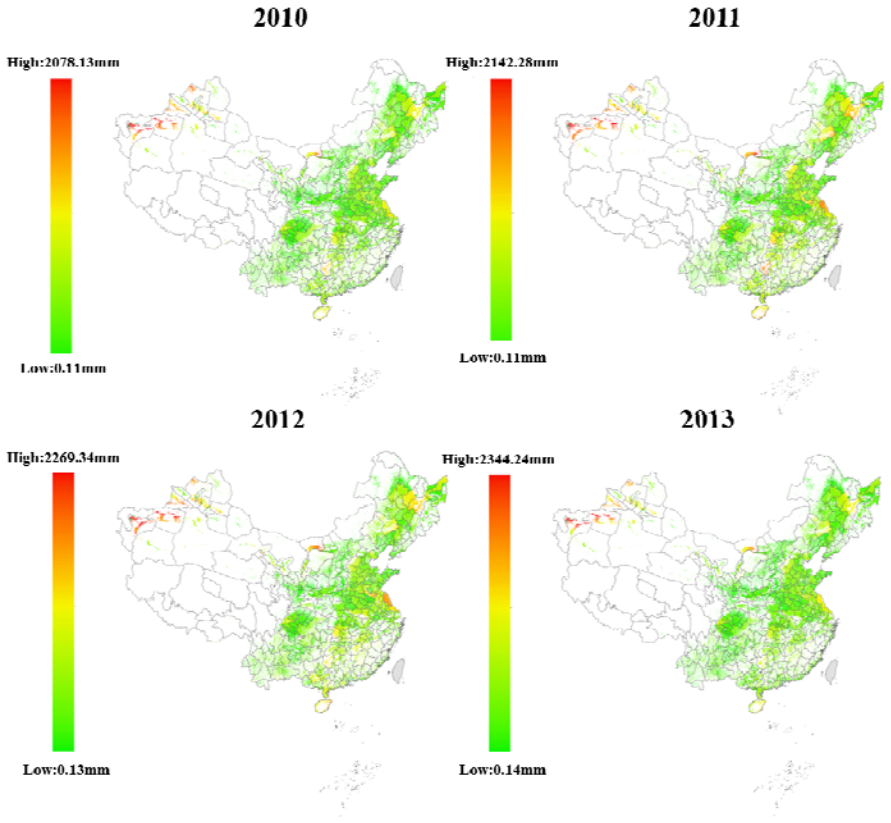
The optimal probability distributions of water use grid maps across three spatial scales (i.e., 1 km scale, appropriate spatial scale, and prefecture scale) were identified and shown in Figure 4. The distribution types of water use reflect the underlying dynamics of water use at various scales. At the 1 km scale (as shown in Figure 4 (a)), most areas show a predominance of normal and log-normal distributions (i.e., in green and light yellow), indicating more stable and symmetric water use patterns. Water use in these areas are consistent with relatively socio-economic and climatic conditions, where are less fluctuated. However, the water use distributions are more frequently exponential or gamma (i.e., in light orange and brown) in areas with rapidly urbanization or industrialization in the central and eastern parts of China. And thus, water use are higher variability and stochasticity. The distribution types of water use at the appropriate spatial scale (as shown in Figure 4 (b)) are similar with those at 1 km scale as shown in Figure 4 (a). More exponential and gamma

distributions can be found in the significant agricultural activity areas at the appropriate spatial scale due to the irregular water use patterns driven by seasonal fluctuations and economic activities. When the spatial scale is up from 1 km to the appropriate ones, more nuanced understanding of water use dynamics can capture the impacts of local factors such as crop irrigation and industrial processes. Although the exponential distribution are found to be one of the prevalent types in regions with irregular or rapidly changing water use patterns at the prefecture scale as shown in Figure 4 (c), uniform and gamma distributions have appeared more frequently at this bigger prefecture scale. The transition to coarser spatial resolution leads to reduced spatial variability in water use, and the heterogeneity of water use is potentially oversimplified.

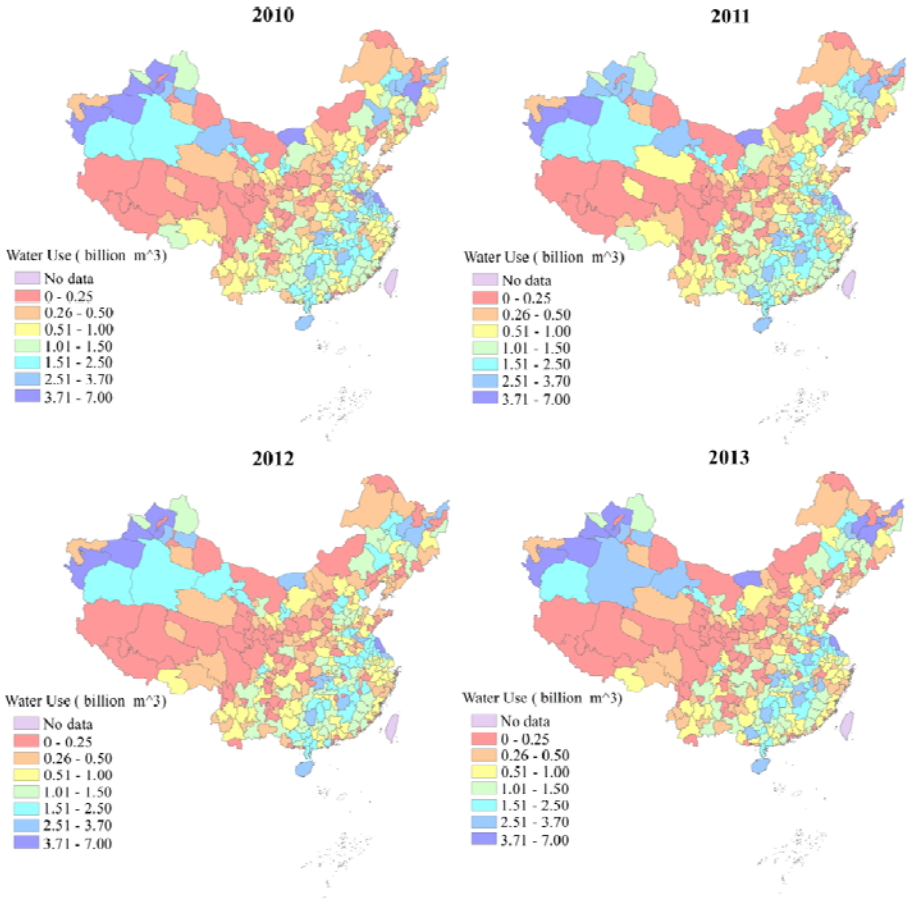
Based on the historical water use record from 1998 to 2013, a transition matrix for each grid cell was constructed to represent the probability of transition from its current state to the subsequent year. According to these transition matrices, the interval for the next state can be predicted, and a random sample is drawn from the corresponding probability distribution to obtain the simulated water use value. The probability rule is applied at three spatial scales (i.e., 1 km, appropriate scale, and prefecture scale). To improve clarity, only the simulated water use results for the validation period (2010–2013) are displayed in Figure 5, enabling a direct comparison of spatial patterns across scales during independent validation. The complete annual simulation results for the entire study period (1998–2013)—including all three spatial scales and both update rules—are available for download at: 10.6084/m9.figshare.30445157.


2010
High:2092.78mm
Low:0.13mm
2011
High:2174.67mm
Low:0.11mm
2012
High:2241.36mm
Low:0.10mm
2013
High:2349.17mm
Low:0.15mm

(a)


2010
High:2078.13mm
Low:0.11mm
2011
High:2142.28mm
Low:0.11mm
2012
High:2269.34mm
Low:0.13mm
2013
High:2344.24mm
Low:0.14mm

(b)



(c)

Figure. 5 Water use simulation results from the probability rule at three different scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

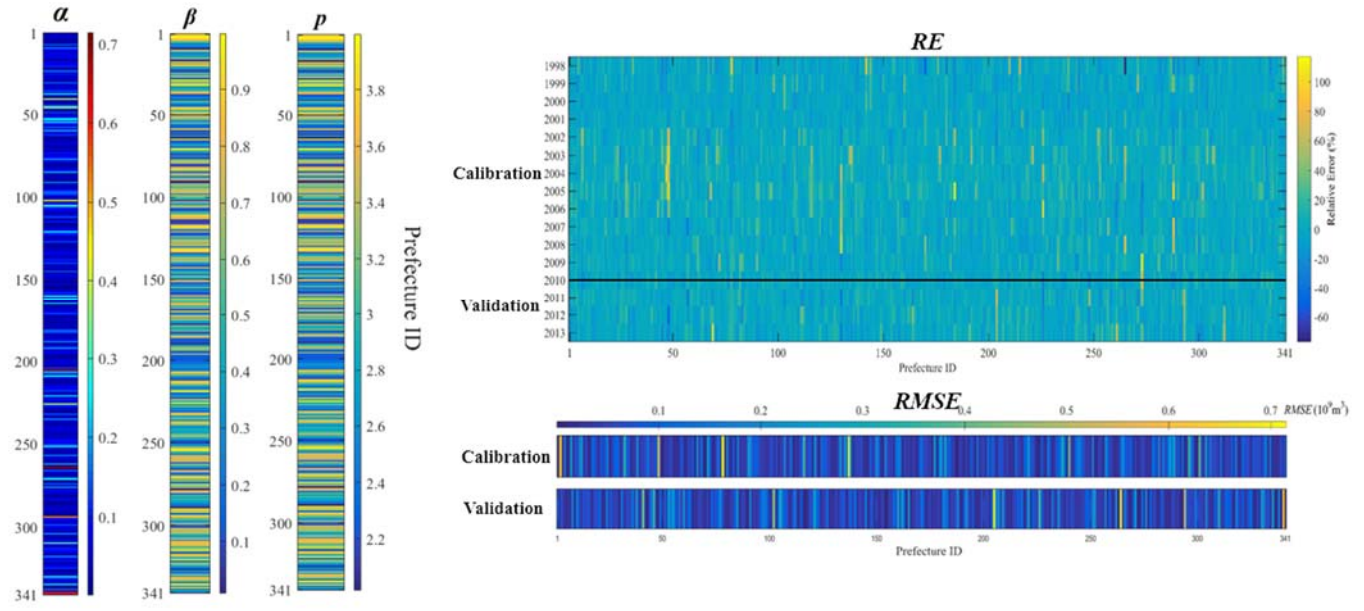
Water use simulation maps from 1998 to 2013 through the probability rule, has clearly highlight the spatial and temporal heterogenic at different spatial scales. The maps reveal the local increases in water use at the 1 km scale (as shown in Figure 5 (a)), particularly in major metropolitan areas in eastern and central parts of China. These increases can be found across the whole time of the selected series driven by the population growth, industrial expansion, and urbanization. The results at the appropriate spatial scale (as shown in Figure 5 (b)) are different from those at the 1 km scale, their local variations become more homogeneous and stationary. It demonstrates that the model can capture the macro-level variations across different geographical areas. The water use trends are further found to be flatter with less spatial heterogeneity at the prefecture scale (Figure 5 (c)). The local variations of water use cannot be reflected at the prefecture scale due to the coarser

spatial resolution. Areas with stable water use, such as the northern provinces, show a more uniform distribution of water use, indicating that coarser resolution tend to mask these local variations.

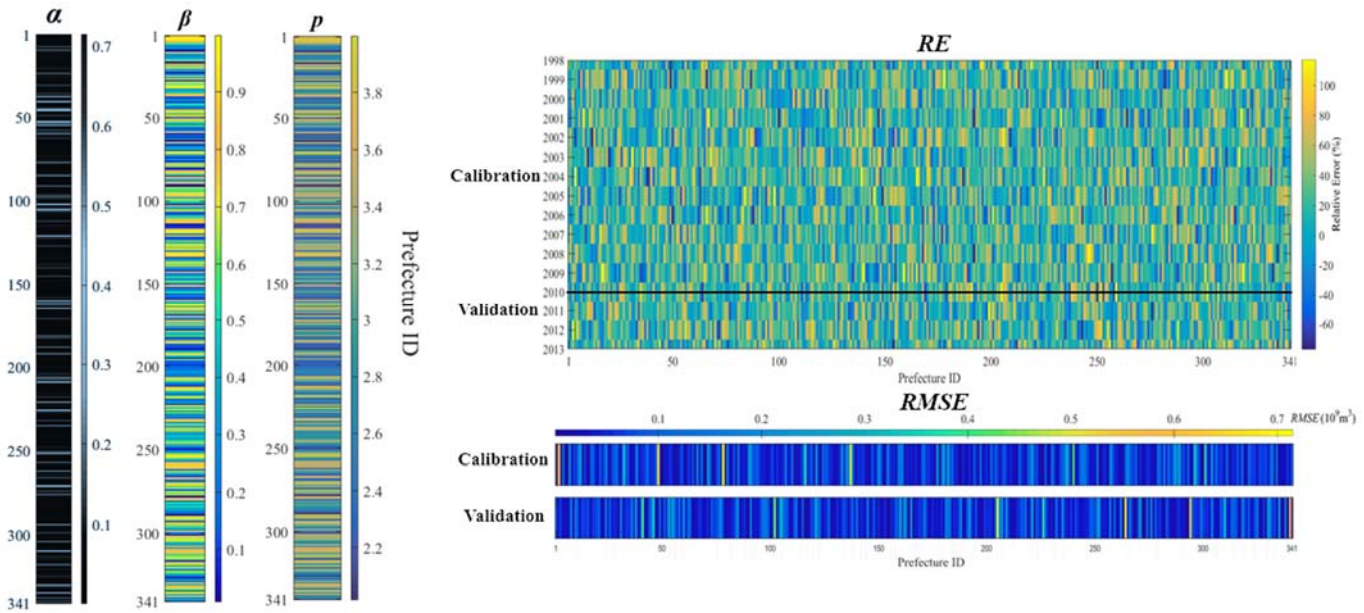
The water use results at the three different spatial scales from probability rule show that the probability rule CA model can effectively capture the spatial heterogeneity in water use. The incorporation of transition probabilities and statistical distributions helps account for the temporal water use variations. However, water uses in the areas with experiencing rapid changes show great fluctuations and spatial fragmentation, and they are hard to be simulated by the probability rule. As the linear rule CA model can provide a more stable alternative in rapid change areas, it can be implemented in water use simulation, too.

4.1.2 Water use simulation from linear rule CA model

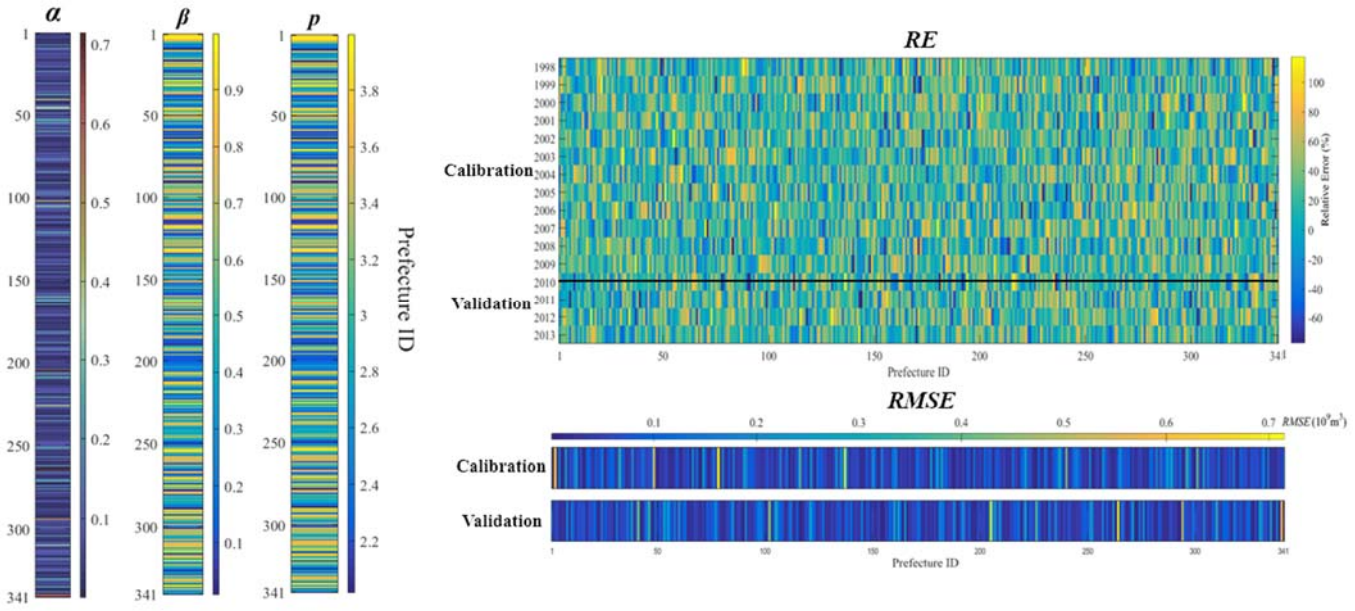
There are three parameters to be calibrated in the linear rule CA model: the self-influence coefficient α, the neighboring influence coefficient β, and the spatial decay exponent p. The calibration and validation are taken the statistics water use at the prefecture-level as the reference (i.e., observed water use data). Specifically, for a given parameter set, the gridded water use is firstly simulated. The simulated grids are then aggregated into the total water use at each prefecture scale along their boundaries. These total water uses are assessed by the observed water use data from water resources bulletins and related statistical surveys. The calibration period covers 1998–2009 and the parameter values are determined by minimizing *RMSE* and *RE* between the simulated and observed total water use at the prefecture scale. The validation period covers 2010–2013 and the performance of the model is also assessed by *RMSE* and *RE*. The optimal parameters at three spatial scales during the calibration and validation periods, are illustrated in Figure 5.


α
β
p
Prefecture ID
RE
Calibration
Validation
Relative Error (%)
Prefecture ID
RMSE
RMSE (10^9 m^3)
Calibration
Validation
Prefecture ID

(a)


α
β
p
Prefecture ID
RE
Calibration
Validation
Relative Error (%)
Prefecture ID
RMSE
RMSE (10^9 m^3)
Calibration
Validation
Prefecture ID

(b)

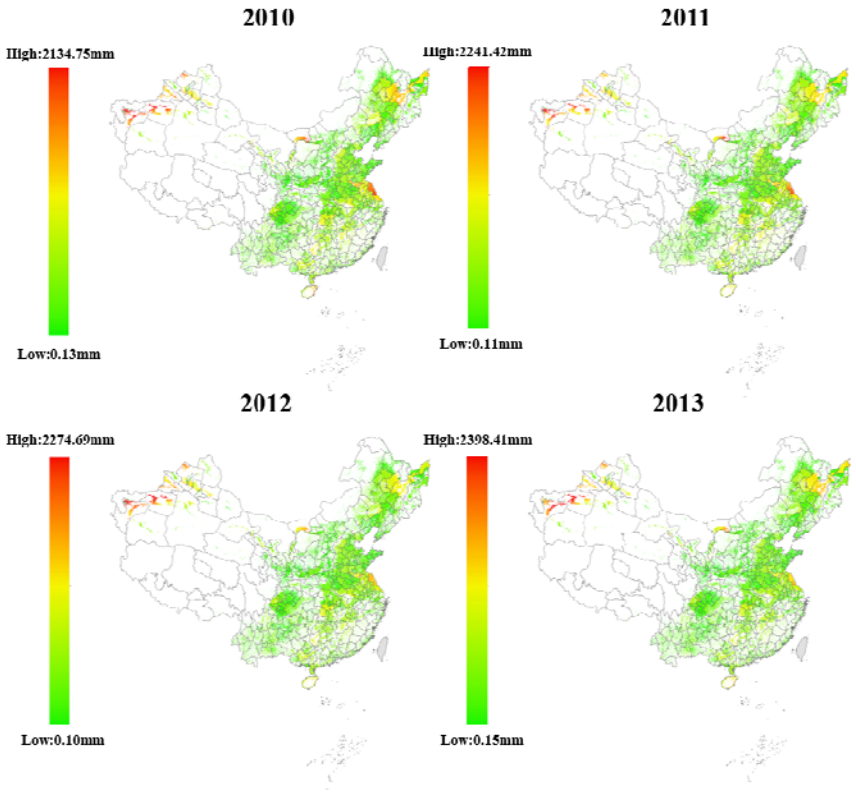


(c)

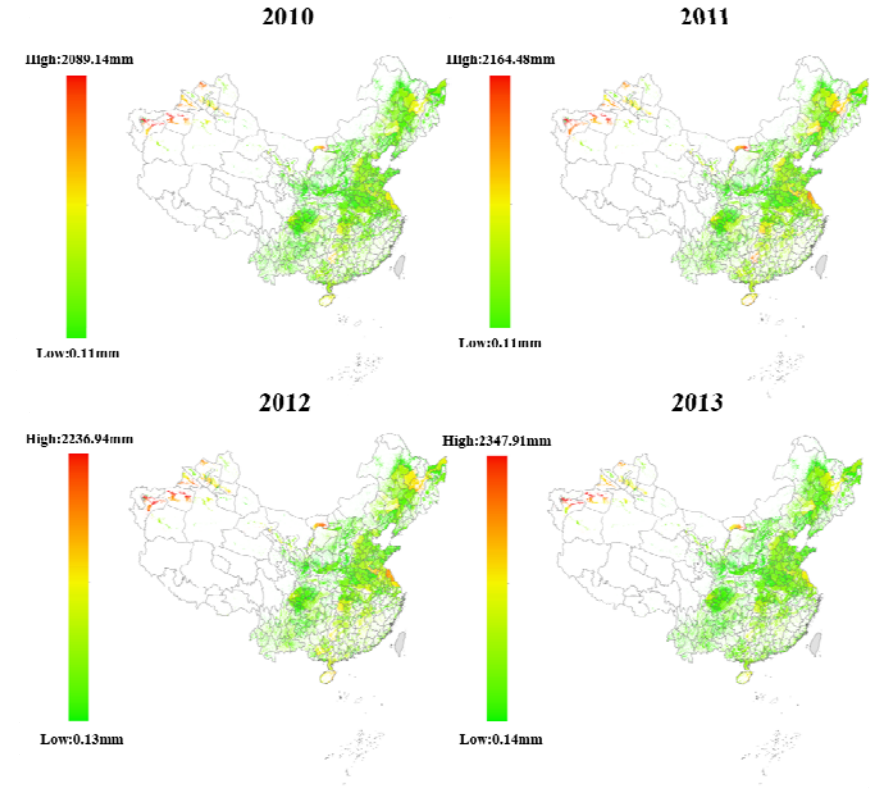
Figure 6. Optimal parameters of the linear rule CA model and the model performances at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

According to the results as shown in Figure 6, the calibrated parameters are varied with scales and the performances of the models are acceptable. The spatial heterogeneity of the parameters α, β, and p tend to reflect local water use patterns. There is a balance between α and β in the areas with lower *RMSE* and *RE* values from the results at 1 km scale as shown in Figure 5(a). And both self-dependence and spatial diffusion jointly govern stable and reliable simulations. But at the appropriate spatial scale (Figure 6 (b)), the parameter values show their deviations from those of the 1 km scale, and α and β values become more regionally specific. As β plays a more dominant role in every prefecture, and higher *RMSE* and *RE* values are found. So only heavily spatial diffusion may fail to capture the complexities of local water use dynamics. These areas often experience a more varied water use. At the prefecture scale (Figure 6 (c)), the parameters are more likely to be dominated by α, especially in regions with large-scale infrastructure or agricultural variations. Their simulation errors are higher because the spatial variations are more difficult to be captured at this prefecture coarser resolutions. And the influence of β tends to be lower at this scale.

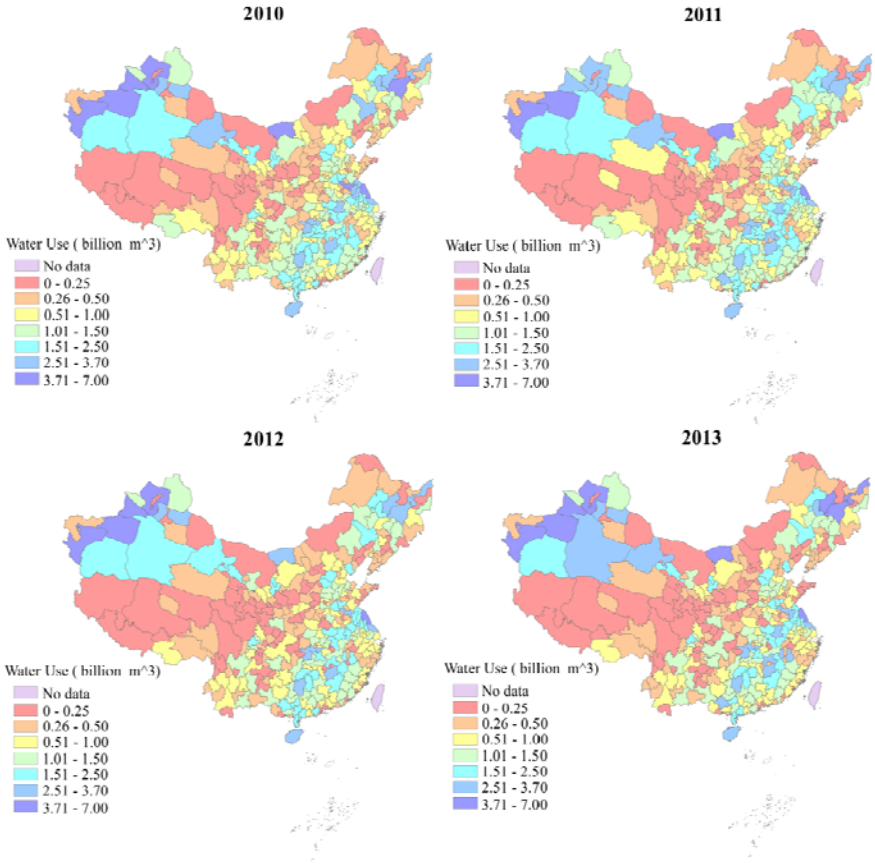
After calibrating the parameters for the linear rule of the CA model, the water use grid maps were generated at three spatial scales (i.e., 1 km, appropriate spatial scale, and prefecture scale). For brevity and readability, Figure 7 presents only the simulation results for the validation years (2010–2013), which allow a more straightforward assessment of the model's predictive performance and spatial differences among scales. The full set of simulated water use maps from 1998–2013, can be accessed at: 10.6084/m9.figshare.30445157.



(a)



(b)



(c)

Figure 7. Water use simulation results from the linear rule at three different scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

The simulated water use maps of the linear rule from 1998 to 2013 are shown in Figure 7, and they exhibit clear spatial dynamics at different spatial scales. At the 1 km scale (Figure 7 (a)), the maps reveal the fine-scale variations in water use, highlighting their localized hotspots of high-water use, particularly in economically developed urban centers such as the Yangtze River Delta, the Pearl River Delta, and the Beijing–Tianjin–Hebei region. There are upward trends, which aligns with the patterns of population growth, industrial expansion, and urbanization. From the results at the appropriate spatial scale as shown in Figure 7 (b), the localized variations in water use are captured with more clarity and precision. The appropriate scale provides a balanced representation that captures both the fine-scale dynamics and broader regional trends. At the prefecture scale (Figure 7 (c)), the water use patterns become more smoothed. This bigger resolution limits the model's ability to capture the dynamics even though the general upward trend in water use in urbanized regions is still observable. The localized fluctuations or abrupt changes driven by specific local policies or external factors are not fully represented at these bigger scales.

Moderate water use levels have been found in the North China Plain (NCP), compared with northeastern China, even though the NCP features intensive irrigation and high population density. Several factors contribute to this pattern. First, the calibration and validation rely on prefecture-level statistical survey data (e.g., water resources bulletins). In recent years, a decreasing trend of agricultural water use has been reported in many NCP prefectures, partly due to improvements in irrigation efficiency, the implementation of water-saving policies, and adjustments in cropping structures. Second, total water use in our study is an aggregate of irrigation, domestic, and industrial sectors. While total water use in the NCP remains dominated by irrigation, many northeastern prefectures show a growing contribution from industrial water use, particularly heavy industries and thermal power generation, leading to higher overall totals. Third, climatic and water availability differences also play a role. Despite a shorter growing season, northeastern China often supports

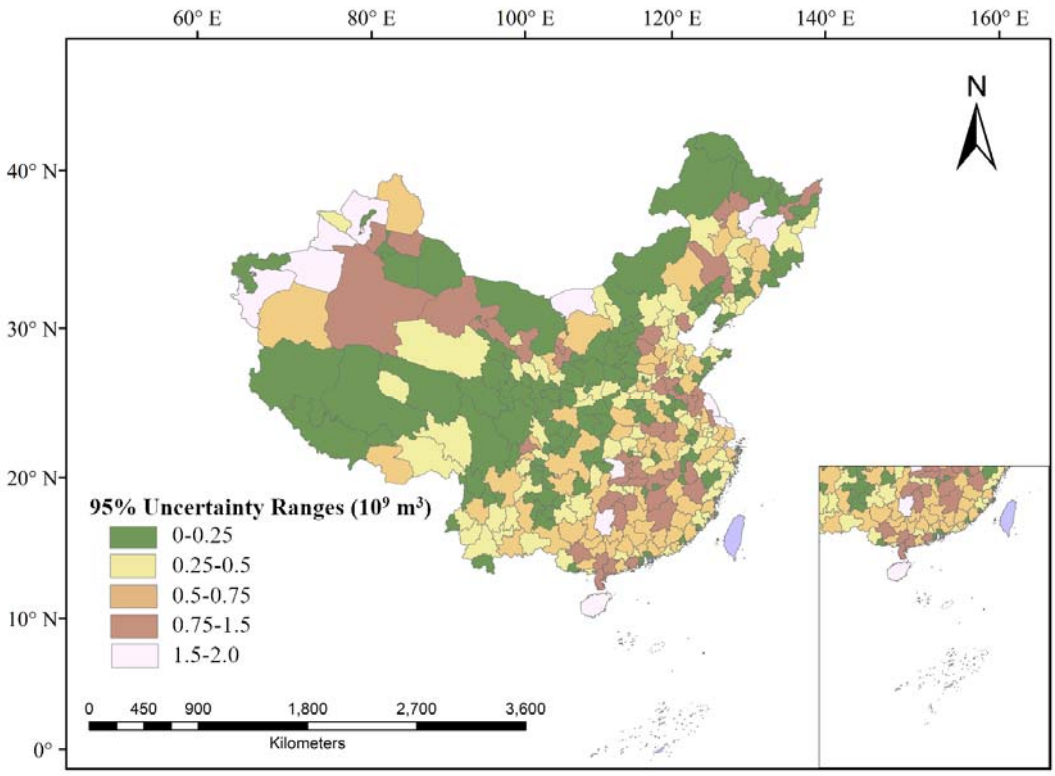
water-intensive crops and benefits from relatively abundant local water resources, which result in higher water use per unit area. In addition, some previous studies that reported higher water use in the NCP were sector-specific (mainly irrigation) or based on different temporal baselines, which partly explains the discrepancy with our aggregated results.

The simulation results across three different spatial scales can also indicate the impacts of scale on perceived water use. Finer-resolution maps reveal localized hotspots of high water use that can be averaged out in coarser resolution. It demonstrates that the CA model's flexibility and further confirms the importance of multi-scale analysis in understanding water use.

4.2 Uncertainty estimation of water use simulation from GLUE

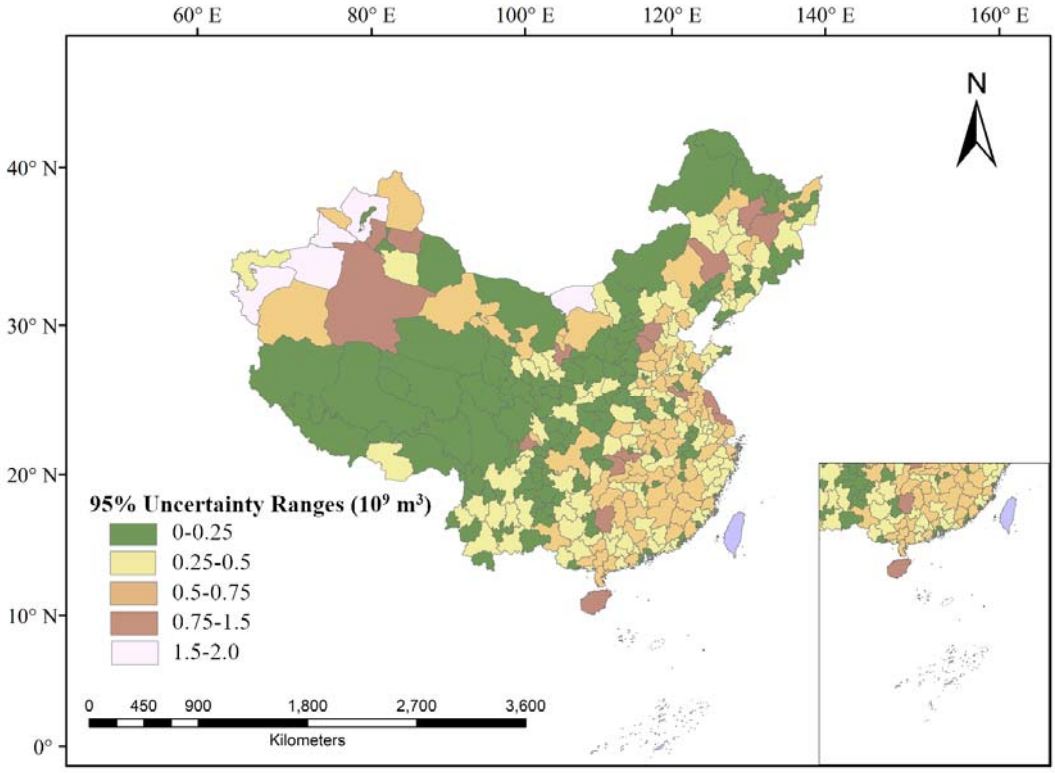
To further assess the reliability and robustness of the water use simulations from the CA model, the uncertainty is quantified by GLUE. According to the calibrated parameters for the probability and linear rules, the uncertainty ranges of the model outputs across different spatial scales are obtained through GLUE. To figure out both the accuracy of simulations and the confidence levels of model predictions from the parameter uncertainties, ensembles of the parameters combination were generated by the uniform sampling as the description in Section 2.3. Corresponding results from the CA model were obtained and their performances are quantified by *RMSE* and *RE* metrics. If the *RMSE* and *RE* metrics are acceptable according to their pre-defined thresholds, the maximum and minimum of the simulation results from the CA model are taken as their uncertainty range. These uncertainties across the three spatial scales are shown in Figure 8 from the probability rule and in Figure 9 from the linear rule at the 95% confidence level (i.e., pre-defined threshold). These uncertainties are from the parameter variabilities. A wider range of the uncertainties indicates lower stable and reliable simulation results, while narrower ones suggest stable and reliable simulation results. The spatial variation of uncertainty ranges can also reveal the significant regional differences of the simulations at

518 different spatial resolutions.



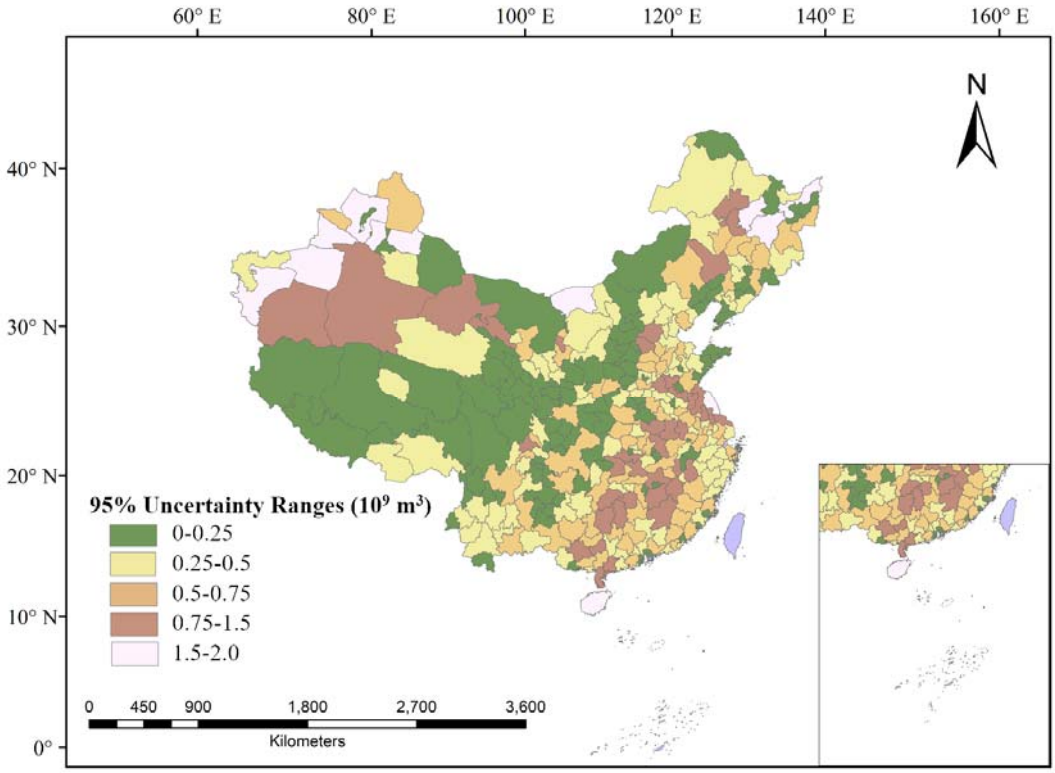
519

520 (a)



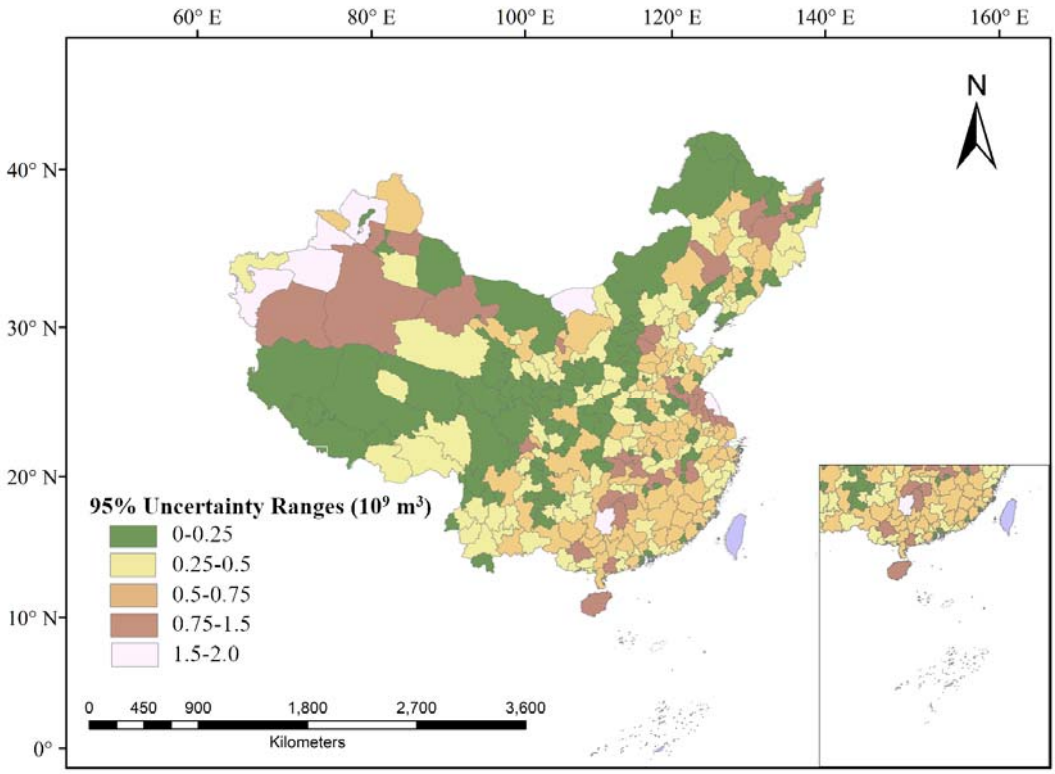
521

522 (b)

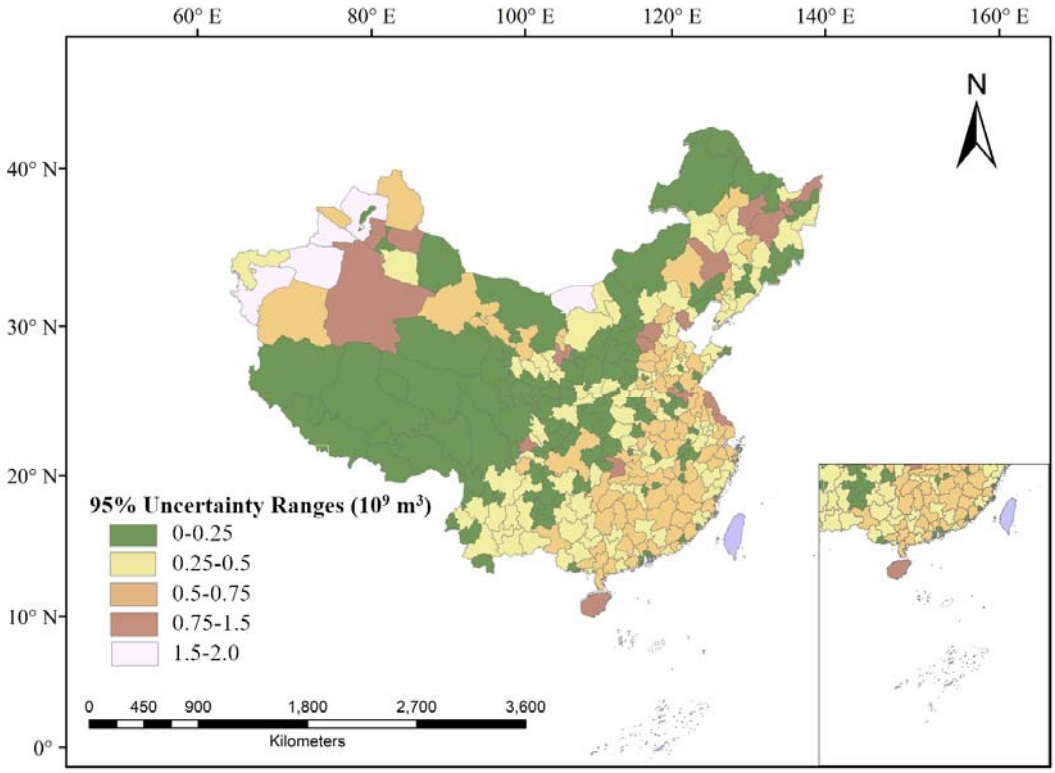


(c)

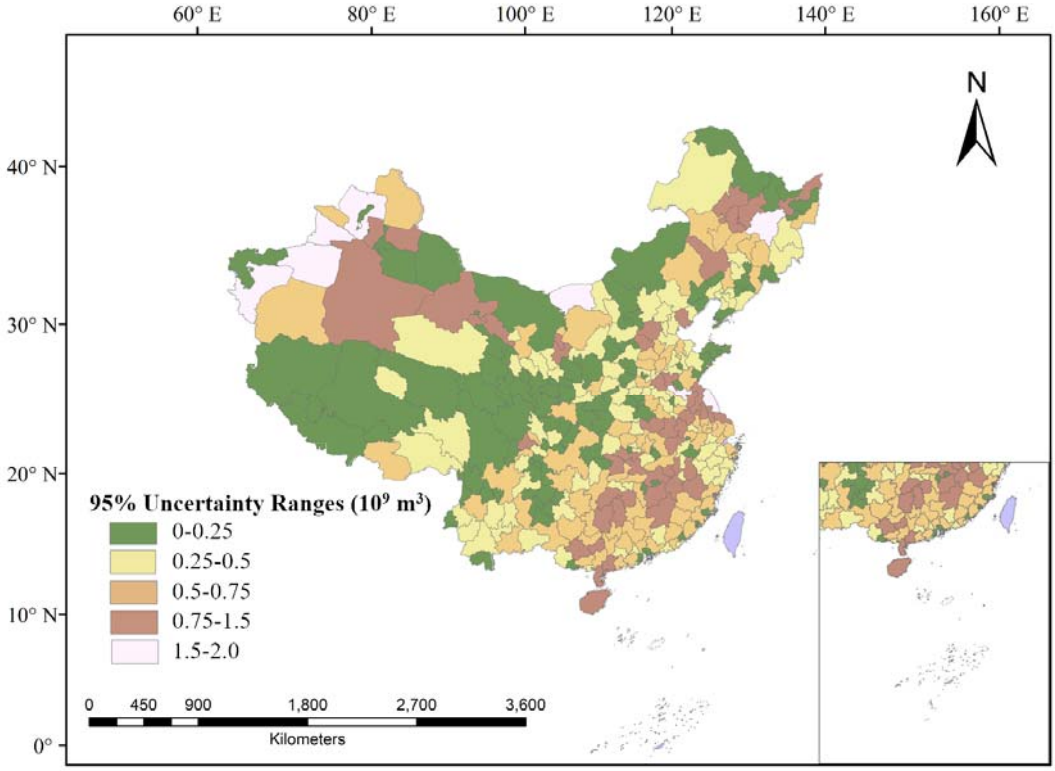
Figure 8. The uncertainty ranges of water use simulations at the 95% confidence level at three spatial scales from the probability rule at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale



(a)



(b)



(c)

Figure 9. The uncertainty ranges of water use simulations at the 95% confidence level at three spatial scales based on linear rule at: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

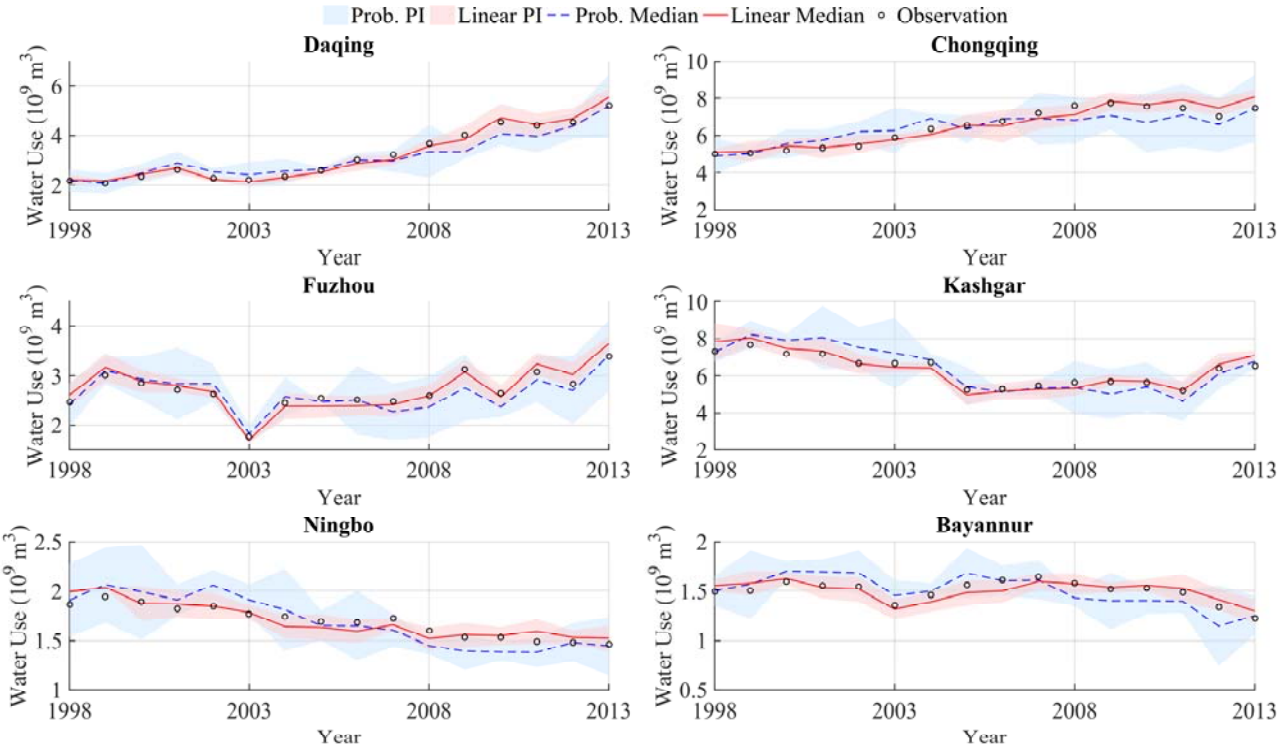
There are distinct regional patterns of simulation uncertainty from the spatial distribution of the uncertainty ranges at the 95% confidence level. Larger uncertainty ranges (i.e., 0.75–2.0*$10^9$$m^3$) are predominantly found in western and southwestern prefectures, such as Xinjiang, Qinghai, Gansu, and Chongqing, where data scarcity and complex local dynamics likely contribute to higher model uncertainty. In contrast, most eastern and northeastern regions, including Beijing, Jiangsu, Shandong, and Liaoning, exhibit relatively narrow uncertainty ranges (i.e., 0–0.5*$10^9$$m^3$), indicating more stable historical water use behavior and better alignment with CA model assumptions.

The model for the higher spatial resolutions has been found to be more sensitive to the local variations, and its uncertainty can be amplified in heterogeneous land use or socio-economic condition areas. For instance, large uncertainty ranges can be found in the densely populated urban centers or the regions undergoing rapid industrialization at the 1 km scale as shown in Figure 8 (a) and Figure 9 (a). But small uncertainty ranges can be found in the same areas from the coarser-scale simulations due to the averaging of the localized fluctuations as seen in Figure 8 (c) and Figure 9 (c) at the prefecture scale. However, the small uncertainty ranges are obtained at the cost of masking the sub-regional dynamics. There should be trade-off between capturing detail and maintaining stability. Thus, it is important to select an appropriate spatial scale for the specific planning or policy purposes.

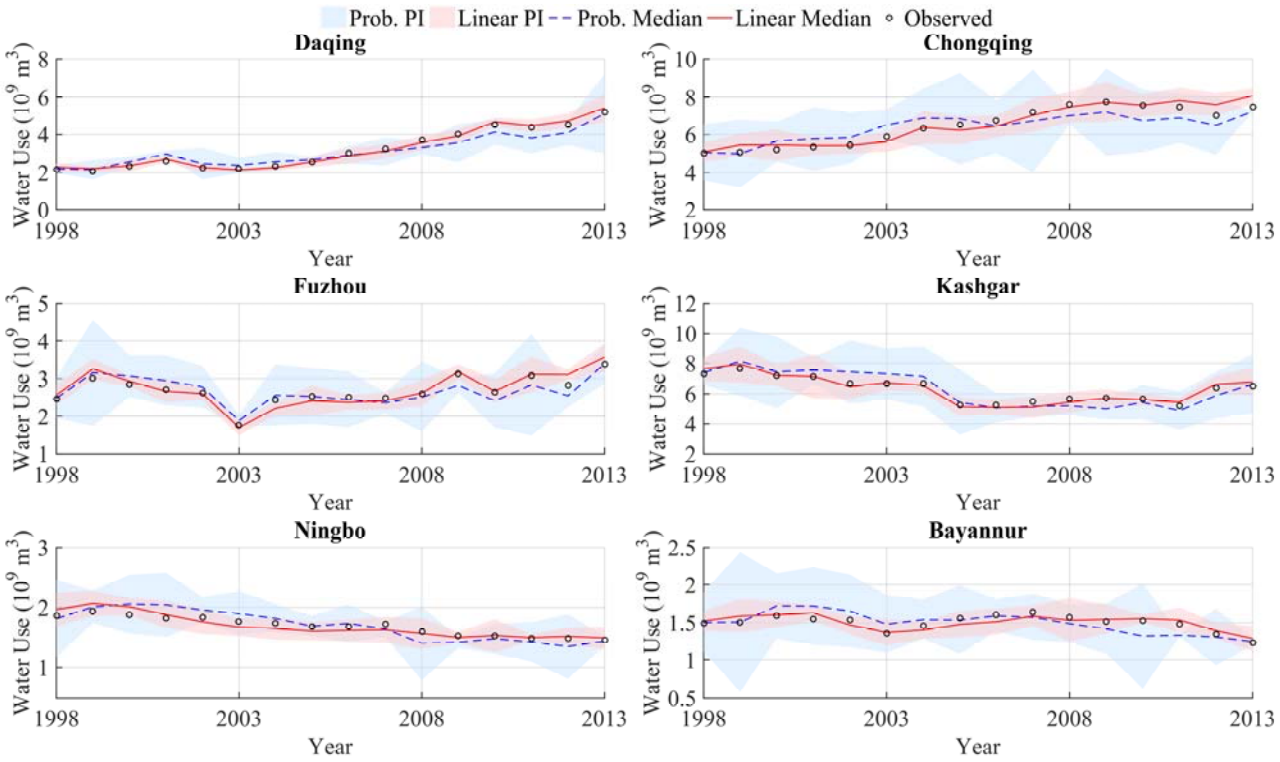
The larger uncertainty ranges are more common from the probability rule than those from the linear rule, particularly in regions with complex or highly variable water use patterns. This is largely attributed to the probability rule itself, because there are randomness in both state transitions and value sampling in fitting distributions. Although the probability rule can help the model to capture non-linear dynamics and abrupt changes, it brings higher uncertainty into outputs. In contrast, the linear rule generally brings less uncertainty, reflecting the deterministic structure of its update mechanism in the model. The amount of water use in each

cell is predicted by the weighted influences from itself and neighboring cells. In regions with relatively stable and spatially smooth development patterns, such as the eastern and northeastern parts of China, the linear rule is more effective. However, the linear rule's assumption of gradual spatial continuity fails to capture abrupt local changes in the rapidly changing or the weak spatial correlation water use areas—such as Fujian, Chongqing, and some parts of Guizhou and Sichuan. And the linear rule potentially leads to an underestimation of uncertainty.

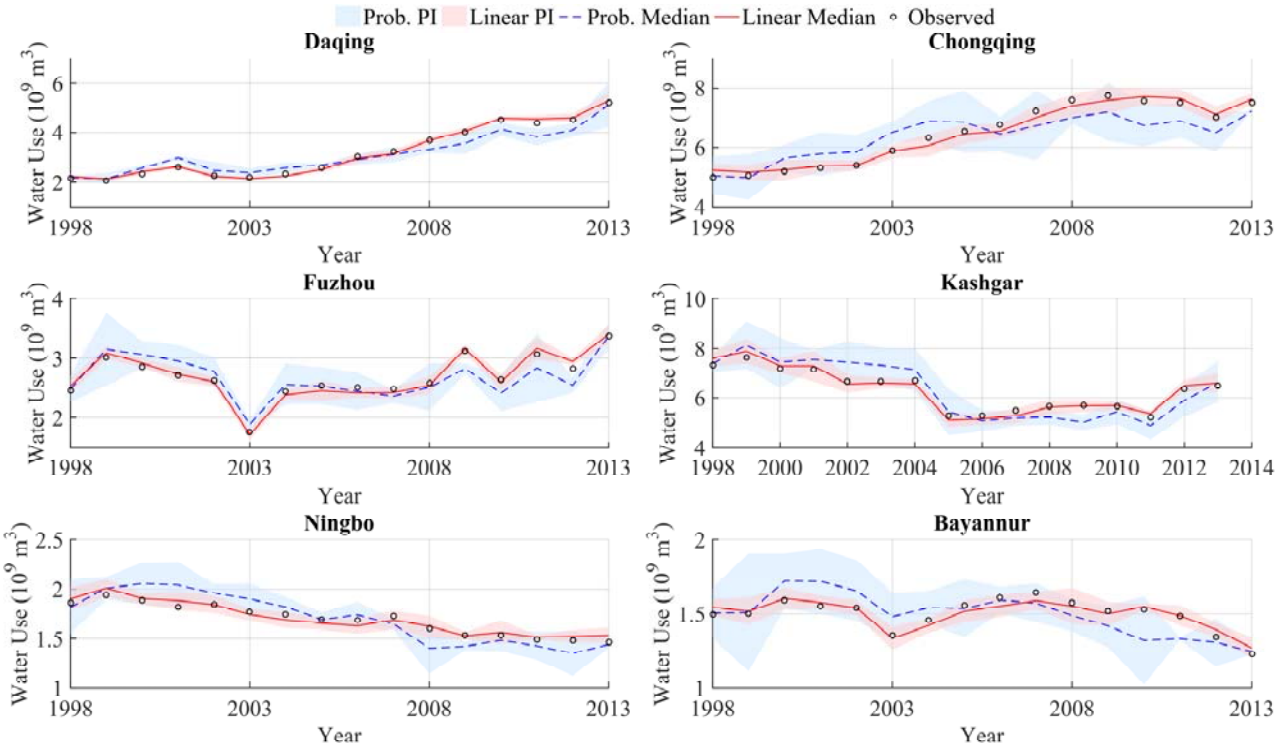
To further examine the temporal dynamics and regional differences in model uncertainty, Daqing, Chongqing, Fuzhou, Kashgar, Ningbo, and Bayannur are selected as six representative prefectures according to their geographic locations, socio-economic conditions, and water use behaviors. The time series of 95% uncertainty ranges from 1998 to 2013 for each representative prefecture is obtained from the results of the probability and linear update rules at different spatial scales, illustrated as Figure 10.



(a)



(b)



(c)

Figure 10. Time series of 95% uncertainty ranges from 1998 to 2013 for each representative prefecture from both the probability and linear update rules at different spatial scales: (a) 1km scale; (b) appropriate spatial scale; (c) prefecture scale

There is distinct the patterns across the selected prefectures as shown in Figure 10. The results from the probability rule have been found to be consistently wider uncertainty intervals than that from the linear rule, particularly in prefectures with more unstable water use conditions. For instance, Kashgar and Bayannur that are located in arid and less-developed western regions exhibit larger and more variable uncertainty intervals from the probability rule. And they are results of the combined effects of high inter-annual variability and limited input data. In contrast, Ningbo and Daqing that are located in the more stable water use trends and better data coverage areas have been found to show relatively narrow and consistent uncertainty from both two rules. It is also interesting to find that the moderate uncertainty levels with obvious differences between the two rules have been in Chongqing and Fuzhou with too many transitional urban areas. As the rapid urbanization and economic shifts bring the localized instability, the probability rule can capture the shifts whereas the linear rule can only underestimate uncertainty during periods of structural change due to its lower sensitivity.

Beside the temporal and rule-based variability, the spatial scale size also plays a crucial role in shaping simulation uncertainty. With the spatial resolution of the model increases, the simulation uncertainty tends to increase, especially in regions with high spatial heterogeneity, such as arid or rapidly developing areas. The finer resolutions can effectively capture the localized variations in water use, there are higher levels of uncertainty due to the increased sensitivity to local fluctuations. In contrast, the results at coarser resolutions tend to be smoothed these local variations, there will be narrower uncertainty but potentially overlooking the sub-regional dynamics.

The uncertainty results have highlighted the importance of the spatial and temporal variations for evaluating the simulation model performance and uncertainty. The results of the representative six prefectures suggest that there is no one single update rule can outperform the others. Instead, the most appropriate rule

599 depends on the local water use dynamics and the purpose of the water use simulation.

600 # 5 Discussion

601 ## 5.1 Impacts of Update Rules on Water Use Simulation

602 To assess the performance of the two update rules, the simulation results at three spatial scales were
603 upscaled to the provincial administrative level, and the *RMSE* and *RE* metrics were calculated for both the
604 probability and linear rules. An additional indicator, $\Delta signed$ was introduced to quantify both the magnitude
605 and direction of the differences between the two simulations (probability vs. linear). $\Delta signed$ represents the
606 mean signed grid-level difference normalized by the mean simulated water use for each province. A positive
607 $\Delta signed$ indicates that the probability rule yields higher simulated water use than the linear rule, while a
608 negative $\Delta signed$ indicates lower simulated values. There are notable differences between the results
609 produced by the two update rules, as summarized in Table 2.

610 Table 2. Model performance at the provincial administrative level from different update rules

Provinces	*RMSE* (billion m^3)		*RE* (%)		$\Delta signed$ (%)
	Linear rule	Probability rule	Linear rule	Probability rule	
Beijing	0.03	0.05	+12	+19	+2.8
Tianjin	0.02	0.02	-11	-17	-0.6
Shanghai	0.02	0.04	-13	-25	+3.1
Chongqing	0.04	0.07	+16	+17	+1.5
Anhui	0.12	0.17	+23	+32	+2.2
Fujian	0.17	0.21	-23	+37	+3.4
Gansu	0.35	0.43	+27	+29	-1.9
Guangdong	0.26	0.36	+21	+32	+3.9
Guizhou	0.31	0.40	-26	-34	-1.2
Hainan	0.11	0.21	-12	+19	+3.0

Heilongjiang	0.42	0.37	-27	+24	+2.1
Hunan	0.21	0.11	-16	+10	+1.6
Jilin	0.36	0.43	+28	+25	+2.5
Jiangsu	0.12	0.17	-12	-26	-0.7
Jiangxi	0.22	0.26	+24	-31	+1.3
Inner Mongolia	0.19	0.29	-23	-35	-2.1
Qinghai	0.31	0.41	+43	+36	+1.4
Ningxia	0.11	0.20	-31	-41	-2.4
Shandong	0.21	0.36	+29	-36	+4.1
Shanxi	0.42	0.61	+31	+42	+3.3
Shannxi	0.27	0.43	-32	-39	-1.5
Sichuan	0.39	0.41	+31	+46	+2.7
Xizang	0.56	0.61	-62	-79	-0.9
Xinjiang	0.68	0.72	+69	+83	+3.5
Yunnan	0.25	0.34	-37	-47	-1.1
Zhejiang	0.21	0.29	+18	+12	+1.9
Guangxi	0.34	0.42	-32	-39	-1.7
Hubei	0.29	0.39	+27	+26	+2.2
Liaoning	0.35	0.41	+33	+48	+3.7

The linear rule generally exhibits lower *RMSE* and *RE* values, indicating higher stability and consistency with observed prefecture-level statistics. This is because the linear rule updates each grid deterministically based on spatial averages of its neighbors, which smooths fluctuations and captures persistent long-term patterns. Consequently, it tends to reduce local variability and enhance regional stability, especially at coarser spatial scales. By contrast, the probability rule explicitly incorporates stochasticity through distribution-based state transitions. This enables it to capture local irregularities, abrupt changes, and nonlinear water use dynamics driven by variations in industrial structure, irrigation demand, or climatic conditions. However, such stochastic behavior can also amplify uncertainty in regions with sparse data or complex spatial heterogeneity,

resulting in slightly higher *RMSE* and *RE* values.

Quantitatively, the mean *RMSE* of the linear rule during the validation period (2010–2013) was 0.28 billion m^3, compared with 0.36 billion m^3 for the probability rule. The corresponding mean relative errors were ±22.4% and ±29.8%, respectively. At the national scale, the simulated total water use was 570.6 billion m^3 for the linear rule and 583.2 billion m^3 for the probability rule, differing by +1.1% and +3.5% from observed national statistics. Therefore, the linear rule is identified as the best-performing estimation framework for reproducing the observed spatiotemporal water use distribution in China, whereas the probability rule provides valuable complementary insights for representing uncertainty and local heterogeneity.

The $\Delta signed$ results indicate that the probability rule tends to simulate higher water use in industrially intensive provinces (e.g., Shandong, Guangdong, and Liaoning) and slightly lower values in water-scarce inland regions (e.g., Ningxia, Gansu, and Inner Mongolia). The results confirm that the two rules emphasize different aspects of the underlying processes—the probability rule better reflects stochastic local variability, while the linear rule offers greater stability and smoother large-scale consistency.

5.2 Impact of Spatial Scales on Water Use Simulation

To investigate how spatial resolution influences the accuracy of the water use simulation, their performances at the three spatial scales are evaluated and are aggregated to the provincial level to ensure comparability across scales. Their results are summarized in Table 3 and indicate their notable differences in simulating accuracy depending on its spatial scales.

Table 3. Model performance at the provincial administrative level at different spatial scales

Provinces	*RMSE* (billion m^3)			*RE* (%)		
	1 km scale	Appropriate scale	Prefecture scale	1 km scale	Appropriate scale	Prefecture scale
Beijing	0.04	0.03	0.05	+22	+12	+19

Tianjin	0.04	0.02	0.06	+26	-11	-17
Shanghai	0.03	0.02	0.04	-23	-13	-25
Chongqing	0.06	0.04	0.07	-26	+16	+27
Anhui	0.17	0.12	0.19	+35	+23	+32
Fujian	0.23	0.17	0.21	-31	-23	+37
Gansu	0.41	0.35	0.43	+31	+27	+29
Guangdong	0.31	0.26	0.36	-29	+21	+32
Guizhou	0.36	0.31	0.40	+31	-26	-34
Hainan	0.19	0.11	0.21	+16	-12	+19
Heilongjiang	0.49	0.42	0.47	+31	-27	+34
Hunan	0.26	0.21	0.31	-19	-16	+20
Jilin	0.42	0.36	0.43	-31	+28	+35
Jiangsu	0.19	0.12	0.17	-21	-12	-26
Jiangxi	0.31	0.22	0.36	-26	+24	-31
Inner Mongolia	0.26	0.19	0.29	-31	-23	-35
Qinghai	0.40	0.31	0.41	+52	+43	+56
Ningxia	0.21	0.11	0.30	-39	-31	-41
Shandong	0.29	0.21	0.36	-32	+29	-36
Shanxi	0.53	0.42	0.61	+36	+31	+42
Shannxi	0.34	0.27	0.43	-36	-32	-39
Sichuan	0.42	0.39	0.41	+41	+31	+46
Xizang	0.67	0.56	0.61	-74	-62	-79
Xinjiang	0.79	0.68	0.72	+72	+69	+83
Yunnan	0.31	0.25	0.34	-49	-37	-47
Zhejiang	0.24	0.21	0.29	+23	+18	+32
Guangxi	0.39	0.34	0.42	-36	-32	-39
Hubei	0.35	0.29	0.39	+31	+27	+36
Liaoning	0.37	0.35	0.41	+42	+33	+48

The relatively lower accuracy of the 1 km scale simulations is attributed to the increased sensitivity to local heterogeneity. At the 1km scale, small errors from the input variables or local noise can accumulate and

amplify to larger discrepancies. Although the results from1kmscale take more spatial detail to reflect the variations in water use, it also brings the uncertainty, especially in the regions with the fragmented data or the high socio-economic diversity. The simulation results from the prefecture level tend to oversimplify the spatial variation. Generally, the results from the coarse resolution have smoothed the intra-regional differences in water use patterns, and resulted in under or over-predictions at the provincial administrative level. The reduced spatial granularity leads to reduction in uncertainty of simulation but can mask the disparities that are critical to policy implementation and resource allocation.

The results from the appropriate scale can balance the spatial sensitivity and model stability. It is fine enough to capture meaningful spatial heterogeneity, and it is yet coarse enough to mitigate excessive noise and data sparsity effects. The most reliable simulating water use should be based on this scale. And it should be noted that that it is crucial for improving simulation reliability to select an appropriate spatial resolution that aligns with both the model structure and the scale of decision-making.

5.3 Spatial Heterogeneity of Water Use Grid Maps

To better understand the spatial heterogeneity of simulated water use, the Coefficient of Variation (CV) and Moran's I were applied to quantify the variability and spatial autocorrelation of water use across three spatial scales — 1 km, the appropriate spatial scale, and the prefecture scale. These two metrics together reveal how spatial scale and model design influence the representation of water-use heterogeneity.

At the 1 km scale, the highest CV values are observed among the three scales, indicating the greatest variability in water use across grid cells. This fine resolution captures the most detailed local differences, especially in areas with intensive agricultural or industrial activities. Moran's I results also reveal strong positive spatial autocorrelation, suggesting distinct clustering of high- or low-water-use regions, particularly around large urban centers and irrigation districts.

At the appropriate spatial scale, both the CV and Moran's I values indicate a moderated heterogeneity pattern. Compared with the 1 km scale, the variability decreases as small-scale noise is smoothed out, while the spatial clustering remains evident but less fragmented. This scale provides a balanced representation by capturing regional heterogeneity without introducing excessive spatial detail or instability. The appropriate spatial scale therefore achieves the optimal trade-off between capturing local patterns and maintaining spatial coherence.

At the prefecture scale, the lowest CV values are recorded, showing that much of the local variability has been smoothed. Moran's I values remain moderately positive, reflecting that some regional clustering persists but overall spatial dependence becomes less pronounced. At this coarser scale, water-use patterns become generalized, reducing the granularity of spatial differences.

To further quantify these relationships, Table 4 presents the average CV and Moran's I values under different spatial scales and update rules.

Table 4 Comparison of CV and Moran's I under different spatial scales and update rules

Spatial scale	**Update rule**	***CV* (mean)**	**Moran's I (mean)**	**Interpretation**
1 km	Probability rule	0.82	0.59	Highest heterogeneity; strong clustering in high-use regions.
1 km	Linear rule	0.78	0.56	Slightly smoother than the probability rule; still high local variation.
Appropriate scale	Probability rule	0.71	0.72	Balanced heterogeneity and spatial dependence; optimal trade-off between detail and stability.
Appropriate scale	Linear rule	0.68	0.70	Slightly lower variability, indicating smoother transitions among neighboring cells.
Prefecture scale	Probability rule	0.60	0.74	Variability largely smoothed; moderate clustering remains.
Prefecture scale	Linear rule	0.57	0.73	Lowest heterogeneity; strong regional aggregation of water-use patterns.

In addition to the influence of spatial scale, the choice of update rule also affects the spatial heterogeneity of simulated water use. Across all scales, the probability rule yields slightly higher CV values than the linear rule, indicating that it better preserves local variability and stochastic fluctuations of water use. This difference

stems from the probabilistic state-transition mechanism, which allows grid-level water use to fluctuate around its expected trajectory, capturing year-to-year uncertainty that is smoothed out in the deterministic linear formulation.

Conversely, Moran's I values under the linear rule are comparable to or slightly lower than those from the probability rule, suggesting smoother and more spatially continuous water-use patterns. This indicates a stronger spatial-averaging effect of the deterministic update mechanism, which may be advantageous for long-term or regionally aggregated assessments but less effective in representing local variability.

Overall, both spatial scale and update rule jointly shape the representation of spatial heterogeneity in water-use simulation. The 1 km scale captures the most detailed variability, the appropriate spatial scale provides a balanced and realistic depiction of regional patterns, and the prefecture scale generalizes spatial differences. Meanwhile, the probability rule emphasizes local randomness and uncertainty, whereas the linear rule accentuates deterministic spatial continuity.

There had been some water use simulation results at previous studies. Fox example, Huang et al. (2018) produced a global-scale monthly water withdrawal dataset at 0.5° resolution, distinguishing six sectors (e.g., irrigation, domestic, electricity generation, livestock, mining, manufacturing) over the period 1971–2010; Hou et al. (2024) developed China's industrial water withdrawal dataset (CIWW), providing gridded monthly data from 1965 to 2020 at 0.1° and 0.25° resolutions; Zhang et al. (2025) presented a high-resolution sectoral water use dataset (HSWUD) for mainland China, covering irrigation, manufacturing, thermal power cooling, and domestic use at 0.1° × 0.1° resolution, with strong consistency to prefecture-level statistics (R^2≈0.88). As the results shown in the previous sections, our dataset is generated at different spatial resolutions (e.g., 1 km×1 km, appropriate spatial scale), enabling detailed representation of spatial heterogeneity within prefectures. Due to the substantial differences in spatial resolutions between these datasets, it is not easy to compare the

differences of the spatial distribution patterns. But the relative values of performance metrics such as *RMSE* and *RE* can figure out the better one among them. The values of *RMSE* within 0.1 (i.e., normalized by mean water use) and a *RE* within −20 % to +30 % are found across all prefectures according to the results of our simulation. And all these results are consistent.

Thus, the results show that, relative to the three reference datasets, our model's prefecture-level water use estimates achieve a *RMSE* within 0.1 (normalized by mean water use) and a *RE* within −20 % to +30 % across all prefectures. These results within the range generally are considered acceptable for large-scale water use modeling, indicating that our estimates are consistent with these previous studies while offering finer spatial details.

6 Conclusion

A multi-scale water use simulation framework has been proposed through integrating a CA model with GLUE to address spatial heterogeneity and uncertainty in this study. Both probability rule and linear rule across three spatial scales (i.e., 1 km, appropriate scale, and prefecture scale) have been applied over 341 prefectures in China as a case study. The impacts of model structure and spatial scale on the spatial heterogeneity and uncertainty have been figured out.

It is interesting to find that the probability rule effectively captures stochastic variations and abrupt transitions in water use but brings larger uncertainty due to its random sampling nature. And the linear rule brings more stable and accurate simulations, particularly in regions with smoother water use patterns. The local noise and uncertainty tend to be amplified in the results of the water use simulation at the 1km scale , while the essential spatial heterogeneity tend to be oversimplified and suppressed in the results at the prefecture scale The most reliable simulation are found from the appropriate scale due its trade-off between capturing spatial heterogeneity and maintaining model stability.

Future improvements for the water use simulation should involve the adaptive update rules that respond to external drivers such as policy shifts or climate shocks, and extend the simulation to finer temporal scales (e.g., seasonal or monthly) for improved short-term decision making. More uncertainty quantification methods, such as Bayesian inference or Markov Chain Monte Carlo, are recommended to enhance performance in high-dimensional settings. Overall, our study can contribute to the water use simulation at a multi-scale and with uncertainty-aware. Our proposed framework will be helpful for the integrated water management, infrastructure planning, and environmental policy under changing socio-economic and climatic conditions.

Author contributions

JZ designed the model architecture, performed the computations, conducted the statistical analysis, and drafted the manuscript. DL acquired funding, contributed to the study design, provided research data, supervised the project, and guided the manuscript revision. JW contributed to manuscript revision discussions and provided advice on submission procedures. FU, LX, ZP, and WG participated in revision discussions and contributed to figure and chart preparation. All authors reviewed and approved the final version of the manuscript for submission.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgement

The authors gratefully acknowledge the financial support from National Key R&D Program of China (2024YFC3012402 and 2022YFC3202803), the National Natural Science Foundation of China (Nos.52379022, and 51879194).

References

Abdi H. Coefficient of variation. Encycl. Res. Des., 1(5), 169–171, https://doi.org/10.4135/9781412961288.n56, 2010.

Al-Shaar, W., Nehme, N., Haidar, H. and Lakiss, H. Forecasted water demand using Extended Cellular Automata Markov Chain Model: case of Saida and Jezzine regions in Lebanon, Sustain. Water Resour. Manag., 8(3), 71, https://doi.org/10.1007/s40899-022-00656-7, 2022.

Almino, L.M.D. and Rufino, I.A.A. Dynamic modeling and urban water demand scenarios: simulations in Campina Grande-PB, Eng. Sanit. Ambient., 26, 915–925, https://doi.org/10.1590/S1413-415220190015, 2021.

Avargani, H.K., Shahdany, S.M.H., Kamrani, K., Maestre, J.M., Garmdareh, S.E.H. and Liaghat, A. Prioritization of surface water distribution in irrigation districts to mitigate crop yield reduction during water scarcity, Agric. Water Manag. 269, 107653, https://doi.org/10.1016/j.agwat.2022.107653, 2022.

Botta-Dukát, Z.J.S.R. Quartile coefficient of variation is more robust than CV for traits calculated as a ratio, Sci. Rep., 13(1), 4671, https://doi.org/10.1038/s41598-023-31306-9, 2023.

Brunner, M.I., Zappa, M. and Stähli, M. Scale matters: Effects of temporal and spatial data resolution on water scarcity assessments, Adv. Water Resour., 123, 134–144, https://doi.org/ 10.1016/j.advwatres. 2018.12.001, 2019.

Canchola, J., Tang, S., Hemyari, P., Paxinos, E. and Marins, E.J.M.P.B. Correct use of percent coefficient of variation (CV) formula for log-transformed data, MOJ Proteom. Bioinform., 6(4), 316–317, https://doi.org/10.15406/mojpb.2017.06.00200, 2017.

Cao, X., Wu, M., Zheng, Y., Guo, X., Chen, D. and Wang, W. Can China achieve food security through the development of irrigation ? , Reg. Environ. Change, 18(2), 465–475, https://doi.org/10.1007/s10113-017-1214-5, 2018.

Carvalho, T.M.N., Filho, F.d.A.d.S. and Porto, V.C. Urban Water Demand Modeling Using Machine Learning Techniques: Case Study of Fortaleza, Brazil, J. Water Resour. Plan. Manag., 147(1), 05020026, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001196, 2021.

Chen, J., Brissette, F.P. and Leconte, R. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, J. eare., 401(3), 190–202, https://doi.org/10.1016/j.jhydrol.2011.02.020, 2011.

Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P. and Edmonds, J. Evaluating the economic impact of water scarcity in a changing world, Nat. Commun., 12(1), , https://doi.org/10.1038/s41467-021-21668-4, 2021.

Fu, Q., Zhou, M., Li, Y., Ye, X., Yang, M. and Wang, Y.J.G.A. Flow spatiotemporal Moran's I: Measuring the spatiotemporal autocorrelation of flow data, Geogr. Anal., 56(4), 799–824, https://doi.org/10.1111/gean.12397, 2024.

Galelli, S. and Castelletti, A. Tree-based iterative input variable selection for hydrological modeling, Water Resour. Res., 49(7), 4295–4310, https://doi.org/10.1002/wrcr.20481, 2013.

Gedamu, W.T., Plank-Wiedenbeck, U., Wodajo, B.T.J.A.A. and Prevention. A spatial autocorrelation analysis of road traffic crash by severity using Moran's I spatial statistics: A comparative study of Addis Ababa and Berlin cities, Accid. Anal. Prev., 200, 107535, https://doi.org/10.1016/j.aap.2024.107535, 2024.

Guevara, M. and Vargas, R. Downscaling satellite soil moisture using geomorphometry and machine learning, PLoS ONE, 14(9), e0222590, https://doi.org/10.1371/journal.pone.0222590, 2019.

He, S.K., Guo, S.L., Liu, Z.J., Yin, J.B., Chen, K.B. and Wu, X.S. Uncertainty analysis of hydrological multi-model ensembles based on CBP-BMA method, Hydrol. Res., 49(5), 1636–1651, https://doi.org/10.2166/nh.2017.248, 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146(730), 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Horta, A., Oliveira, A.R., Azevedo, L. and Ramos, T.B.J.G.R. Assessing the use of digital soil maps in hydrological modeling for soil-water budget simulations-implications for water management plans in southern Portugal, Geoderma Reg., 36, e00741, https://doi.org/10.1016/j.geodrs.2023.e00741, 2024.

Hou, C., Li, Y., Sang, S., Zhao, X., Liu, Y., Liu, Y. and Zhao, F. High-resolution mapping of monthly industrial water withdrawal in China from 1965 to 2020, Earth Syst. Sci. Data, 16(5), 2449–2464, https://doi.org/10.5194/essd-16-2449-2024, 2024.

Huang, Z., Hejazi, M., Li X., Tang Q., Vernon, C., Leng, G., Liu, Y., Doll P., Eisner, S., Gerten, D., Hanasaki N.and Wada Y. Reconstrcution of global gridded monthly sectoral water withdrawals for 1970-2010 and analysis of their spatiotemporal patterns. Hydrol. Earth Syst. Sci. 22(4):2117-2133, http://doi.org/10.5194/hess-22-2117-2018, 2018.

Huang, Z., Yuan, X. and Liu, X. The key drivers for the changes in global water scarcity: Water withdrawal versus water availability, J. Hydrol., 601, 126658, https://doi.org/10.1016/j.jhydrol.2021.126658, 2021.

Ji, Y., Zuo, Q., Zhao, C., Zhang, Z. and Wu, Q.J.E.I.A.R. Measurement and decomposition of green water resource utilization efficiency across multiple water use sectors in China: A perspective on water-saving potential, Environ. Impact Assess. Rev., 112, 107806, https://doi.org/10.1016/j.eiar.2025.107806, 2024.

Kaewmai, R., Grant, T., Eady, S., Mungkalasiri, J. and Musikavong, C. Improving regional water scarcity footprint characterization factors of an available water remaining (AWARE) method, Sci. Total Environ., 681, 444–455, https://doi.org/10.1016/j.scitotenv.2019.05.006, 2019.

Kang, S., Hao, X., Du, T., Tong, L., Su, X., Lu, H., Li, X., Huo, Z., Li, S. and Ding, R. Improving agricultural water productivity to ensure food security in China under changing environment: From research to practice, Agric. Water Manag., 179, 5–17, https://doi.org/10.1016/j.agwat.2016.05.019, 2017.

Knox, J.W., Haro-Monteagudo, D., Hess, T.M. and Morris, J. Identifying Trade-Offs and Reconciling Competing Demands for Water: Integrating Agriculture Into a Robust Decision-Making Framework, Earth's Future, 6(10), 1457–1470, https://doi.org/10.1029/2018EF000905, 2018.

Kovacevic, J., Cvijetinovic, Z., Stancic, N., Brodic, N. and Mihajlovic, D. New Downscaling Approach Using ESA CCI SM Products for Obtaining High Resolution Surface Soil Moisture, Remote Sens., 12(7), https://doi.org/10.3390/rs12071119, 2020.

Liu, D., Li, X., Guo, S., Rosbjerg, D. and Chen, H. Using a Bayesian Probabilistic Forecasting Model to Analyze the Uncertainty in Real-Time Dynamic Control of the Flood Limiting Water Level for Reservoir Operation, J. Hydrol. Eng., 20(2), 04014036, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000849, 2015.

Liu, D., Zhang, Y., Zhang, J., Xiong, L., Liu, P., Chen, H. and Yin, J. Rainfall estimation using measurement report data from time-division long term evolution networks, J. Hydrol., 600, 126530, https://doi.org/10.1016/j.jhydrol.2021.126530, 2021.

Liu, Y., Wang, N., Jiang, C., Archer, L. and Wang, Y. Temporal and spatial distribution of soil water and nitrate content affected by surface irrigation and fertilizer rate in silage corn fields, Sci. Rep., 10(1), 8317, https://doi.org/10.1038/s41598-020-65119-5, 2020.

Liu, Y., Zhang, J., Chen, L., Chu, H., Wang, J.Z. and Ma, L. SSAS: Spatiotemporal Scale Adaptive Selection for Improving Bias Correction on Precipitation, IEEE Trans. Cybern., 52(11), 12175–12188, doi: 10.1109/TCYB.2021.3072483, 2021.

Luo, Y., Zhang, Z., Li, Z., Chen, Y., Zhang, L., Cao, J. and Tao, F. Identifying the spatiotemporal changes of annual harvesting areas for three staple crops in China by integrating multi-data sources, Environ. Res. Lett., 15(7), 074003, https://doi.org/10.1088/1748-9326/ab9754, 2020.

Mekonnen, M.M. and Hoekstra, A.Y. Four billion people facing severe water scarcity, Sci. Adv., 2(2), e1500323, https://doi.org/10.1126/sciadv.1500323, 2016.

Noori, N. and Kalin, L. Coupling SWAT and ANN models for enhanced daily streamflow prediction, J. Hydrol., 533, 141–151, https://doi.org/10.1016/j.jhydrol.2015.12.020, 2016.

Rosa, L., Chiarelli, D.D., Rulli, M.C., Dell'Angelo, J. and D'Odorico, P. Global agricultural economic water scarcity, Sci. Adv., 6(18), eaay5552, https://doi.org/10.1126/sciadv.aay5552, 2020.

Sapino, F., Haer, T., Saiz-Santiago, P. and Pérez-Blanco, C.D. A multi-agent cellular automata model to explore water trading potential under information transaction costs, J. Hydrol., 618, 128669, https://doi.org/10.1016/j.jhydrol.2023.128669, 2023.

Sharifi, H., Roozbahani, A. and Hashemy Shahdany, S.M. Evaluating the Performance of Agricultural Water Distribution Systems Using FIS, ANN and ANFIS Intelligent Models, Water Resour. Manag., 35(6), 1797–1816, https://doi.org/10.1007/s11269-020-02685-3, 2021.

Shortridge, A. Practical limits of Moran's autocorrelation index for raster class maps. Computers Environment and Urban Systems. Computers, Comput. Environ. Urban Syst., 31(3), 362–371, https://doi.org/10.1016/j.compenvurbsys.2006.04.002, 2007.

Su, Y., Liao, S., Ren, J. and Zhao, Z.J.F.i.E.S. Research on the decoupling relationship between water resources utilization and economic development at the county scale in Qian'nan Prefecture, Guizhou Province, Front. Environ. Sci., 12, 1347652, https://doi.org/10.3389/fenvs.2024.1347652, 2024.

Sun, S., Tang, Q.H., Konar, M., Huang, Z.W., Gleeson, T., Ma, T., Fang, C.L. and Cai, X.M. Domestic Groundwater Depletion Supports China's Full Supply Chains, Water Resour. Res., 58(5), e2021WR031555, https://doi.org/10.1029/2021WR031555, 2022.

Sunkara, S.V. and Singh, R. Assessing the impact of the temporal resolution of performance indicators on optimal decisions of a water resources system, J. Hydrol., 612, 128185, https://doi.org/10.1016/j.jhydrol.2022.128185, 2022.

Taormina, R., Galelli, S., Karakaya, G. and Ahipasaoglu, S.D. An information theoretic approach to select alternate subsets of predictors for data-driven hydrological models, J. Hydrol., 542, 18–34, https://doi.org/10.1016/j.jhydrol.2016.08.034, 2016.

Tariq, A., Mumtaz, F., Majeed, M. and Zeng, X. Spatio-temporal assessment of land use land cover based on trajectories and cellular automata Markov modelling and its impact on land surface temperature of Lahore district Pakistan, Environ. Monit. Assess., 195(1), 195, https://doi.org/10.1007/s10661-022-09942-9, 2023.

Tiefelsdorf, M., Boots, B.J.E. and A, P. The exact distribution of Moran's I, Environ. Plan. A, 27(6), 985–999, https://doi.org/10.1068/a270985, 1995.

Wang, C., Tang, G., Xiong, W., Ma, Z. and Zhu, S. Infrared precipitation estimation using convolutional neural network for FengYun satellites, J. Hydrol., 603, 127113, https://doi.org/10.1016/j.jhydrol.2021.127113, 2021.

Wang, T., Guo, Z., Shen, Y., Cui, Z. and Goodwin, A. Accumulation mechanism of biofilm under different water shear forces along the networked pipelines in a drip irrigation system, Sci. Rep., 10(1), 6960, https://doi.org/10.1038/s41598-020-63602-1, 2020.

Wu, B., Tian, F., Zhang, M., Piao, S., Zeng, H., Zhu, W., Liu, J., Elnashar, A. and Lu, Y. Quantifying global agricultural water appropriation with data derived from earth observations, J. Clean. Prod., 358, 131891, https://doi.org/10.1016/j.jclepro.2022.131891, 2022.

Wu, J. and Lu, J. Spatial scale effects of landscape metrics on stream water quality and their seasonal changes, Water Res., 191, 116811, https://doi.org/10.1016/j.watres.2021.116811, 2021.

Yamada, H.J.M. A new perspective on Moran’s coefficient: Revisited, Math., 12(2), 253, https://doi.org/10.3390/math12020253, 2024.

Yan, J., Jia, S., Lv, A. and Zhu, W. Water Resources Assessment of China's Transboundary River Basins Using a Machine Learning Approach, Water Resour. Res., 55(1), 632–655, https://doi.org/10.1029/2018WR023274, 2019.

Yang, Y.R., Xiong, Q.Y., Wu, C., Zou, Q.H., Yu, Y., Yi, H.L. and Gao, M. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism, Environ. Sci. Pollut. Res., 28(39), 55129–55139, https://doi.org/10.1007/s11356-021-14134-8, 2021.

Ye, Y., Huang, L., Zheng, Q., Liang, C., Dong, B., Deng, J. and Han, X. A feasible framework to downscale NPP-VIIRS nighttime light imagery using multi-source spatial variables and geographically weighted regression, Int. J. Appl. Earth Obs. Geoinf., 104, 102513, https://doi.org/10.1016/j.jag.2021.102513, 2021.

Yin, J., Guo, S., Gu, L., He, S., Ba, H., Tian, J., Li, Q. and Chen, J. Projected changes of bivariate flood quantiles and estimation uncertainty based on multi-model ensembles over China, J. Hydrol., 585, 124760, https://doi.org/10.1016/j.jhydrol.2019.124760, 2020.

Zhang, C. and Long, D. Estimating Spatially Explicit Irrigation Water Use Based on Remotely Sensed Evapotranspiration and Modeled Root Zone Soil Moisture, Water Resour. Res., 57(12), e2021WR031382, https://doi.org/10.1029/2021WR031382, 2021.

Zhang, J.Y., Liu, D.D., Guo, S.L., Xiong, L.H., Liu, P., Chen, J. and Yin, J.B. High resolution annual irrigation water use maps in China based-on input variables selection and convolutional neural networks, J. Clean. Prod., 405, 136175, https://doi.org/10.1016/j.jclepro.2023.136175, 2023.

Zhang, J., Liu, D., Xu, Y., Xiong, L., Chen, J., Chen, H., & Yin, J. Appropriate spatiotemporal scale selection for water use simulation in China, J. Hydrol., 2025, 133502, https://doi.org/ 10.1016/j.jhydrol. 2024.133502, 2025.

Zhang, Y., Yin, Y., Yin, M., Zhang X.. A High-Resolution Gridded Dataset for China’s Monthly Sectoral Water Use. Sci Data 12, 1157. https://doi.org/10.1038/s41597-025-05400-2, 2025.

Zhou, F., Bo, Y., Ciais, P., Dumas, P., Tang, Q.H., Wang, X.H., Liu, J.G., Zheng, C.M., Polcher, J., Yin, Z., Guimberteau, M., Peng, S.S., Ottle, C., Zhao, X.N., Zhao, J.S., Tan, Q., Chen, L., Shen, H.Z., Yang, H., Piao, S.L., Wang, H. and Wada, Y. Deceleration of China's human water use and its key drivers, Proc. Natl. Acad. Sci. U.S.A., 117(14), 7702–7711, https://doi.org/10.1073/pnas.1913587117, 2020.