Modelling Escherichia coli concentrations

45.6 %–78.1 % of China's rivers show poor microbial water quality

Mei, Songtao; Wang, Kai; Schmitt, Heike; Hofstra, Nynke

**Citation (APA)**
Mei, S., Wang, K., Schmitt, H., & Hofstra, N. (2025). Modelling Escherichia coli concentrations: 45.6 %–78.1 % of China's rivers show poor microbial water quality. *Environmental Pollution*, *374*, Article 126184. https://doi.org/10.1016/j.envpol.2025.126184

**Important note**
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

# Modelling *Escherichia coli* concentrations: 45.6 %–78.1 % of China's rivers show poor microbial water quality☆

Songtao Mei [a,b,*] , Kai Wang [b,**] , Heike Schmitt [c,d] , Nynke Hofstra [a]

[a] *Earth Systems and Global Change Group, Wageningen University & Research, P.O. Box 47, 6700 AA, Wageningen, the Netherlands*
[b] *State Key Laboratory of Nutrient Use and Management, College of Resources and Environmental Sciences, National Academy of Agriculture Green Development, Key Laboratory of Plant-Soil Interactions of Ministry of Education, National Observation and Research Station of Agriculture Green Development (Quzhou, Hebei), China Agricultural University, Beijing, PR China*
[c] *National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA, Bilthoven, the Netherlands*
[d] *Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ, Delft, the Netherlands*

## ABSTRACT

Water quality is under threat due to the presence of pathogenic and antibiotic-resistant bacteria. *Escherichia coli* (*E. coli*) serves as an indicator of faecal contamination and the potential presence of other harmful pathogens. Understanding *E. coli* concentrations helps in assessing the overall health risks associated with waterborne diseases and developing effective water management strategies. Therefore, we developed the first large-scale model, GloWPa-Ecoli C1 to simulate *E. coli* loads and concentrations in rivers and apply this model to China. The model provides the first comprehensive overview of microbial water quality across China's rivers. The model simulates *E. coli* concentrations in 2020 to range from $10^{-1.2}$ to $10^{6.3}$ CFU/L, with 45.6 %–78.1 % of rivers exhibiting poor microbial water quality. Major hotspots of *E. coli* pollution are Haihe, Huaihe and Pearl River Basins. Direct discharge of human faecal waste contributes 80.2 % of the total *E. coli* load, while directly discharged livestock waste accounts for 13.1 %. To mitigate *E. coli* pollution in rivers in China, we recommend increasing human faecal waste collection rates, expanding wastewater treatment plant (WWTP) coverage, phasing out primary treatment WWTPs and eliminating direct livestock faecal waste discharge, particularly from smallholder farms. The study underscores the urgent need to improve microbial water quality in China's rivers. The findings provide actionable insights to inform policy development aimed at safeguarding water quality and public health. Furthermore, the modelling approach is applicable to other regions and microorganisms, offering a foundation for developing models to address antibiotic-resistant bacteria and other emerging water quality challenges.

## 1. Introduction

Water quality is at risk due to the presence of pathogenic and antibiotic-resistant bacteria, as these can pose serious risks to public health (Mbanga et al., 2020; Rose et al., 2023). One-third of river stretches in Latin America, Africa and Asia are affected by severe pathogen pollution (United Nations Environment Programme (UNEP), 2016). To assess water quality and safeguard human health, *Escherichia coli* (*E. coli*) has been widely applied as faecal indicator organism (Jang et al., 2017). According to the World Health Organization (World Health Organization (WHO), 2022), *E. coli* detection in water indicates recent faecal contamination and the potential presence of pathogens such as

Salmonella, Shigella, and viruses like norovirus and hepatitis A. *E. coli* resistant to third-generation cephalosporins and carbapenem is identified as bacterial priority pathogen of public health importance (World Health Organization (WHO), 2024). Additionally, certain *E. coli* strains are pathogenic, such as *E. coli* O157:H7 (Kaper et al., 2004). As estimated by Murray et al. (2022), *E. coli* is responsible for 900,000 resistance-associated deaths in 2019 globally. This underscores the importance of monitoring *E. coli* levels to preserve microbial water quality, reduce the potential risk of development and transfer of antibiotic resistance and eventually, protect human health.

Surface waters in China are widely used for various domestic purposes, including irrigation, recreation, fishing and as sources for

---

drinking water after treatment (Han et al., 2020; Zhu et al., 2023). Nevertheless, there is a lack of overview of *E. coli* concentrations in rivers as well as the overall microbial water quality in China. Despite that China's surface water quality standards (Ministry of Environmental Protection (MEP), 2002) have included faecal coliforms (FC) as a pathogenic microorganism indicator, the current surface water quality in China is still assessed solely by physicochemical indicators. Consequently, environmental authorities either do not monitor faecal coliform concentrations or withhold the data when monitoring is conducted, leading to a scarcity of information on both FC and *E. coli* concentrations.

Given these circumstances and considering that regular monitoring of the microbial water quality of rivers is costly and time-consuming, process-based modelling can be an efficient way to understand existing pollution dynamics. There are many models available for the simulation of *E. coli* concentrations: the Soil and Water Assessment Tool (SWAT) model has been applied in several catchments around the world (Coffey et al., 2010; Coffey et al., 2013; Iqbal & Hofstra, 2018); the Watershed Assessment Model (WAM) has been used in New Zealand (Collins & Rutherford, 2004); and the MIKE 21 FM model with ECOLab module has also been applied, for example in Bangladesh (Islam et al., 2018), to simulate *E. coli* concentrations. However, these models are small-scale, focusing on specific watersheds or river basins and requiring detailed input data, making them unsuitable for assessing microbial water quality in a large scale. Models for a larger scale have also been developed. Reder et al. (2015) adapted the WorldQual model to quantify historical FC concentrations in European rivers, and Jones et al. (2023) and United Nations Environment Programme (UNEP), 2016 simulated global FC concentrations. FC concentrations in China are included in these models, but a detailed source analysis specific to China is lacking. This underscores the need for a water quality model that incorporates local data and a more refined calculation methodology, along with a comprehensive source assessment to simulate *E. coli* in rivers in China.

Therefore, the objective of this study is to simulate spatially explicit *E. coli* concentrations in China's rivers for 2020, identify critical hotspots and sources, and explore priorities for *E. coli* pollution mitigation. We adapted and advanced the Global Waterborne Pathogen (GloWPa) model, which simulates waterborne pathogen load and concentrations at global, regional and national scales (Okaali et al., 2021; Vermeulen et al., 2017; Vermeulen et al., 2019), into the GloWPa-Ecoli C1 model. The model calculates *E. coli* load to rivers from humans and 14 livestock species and *E. coli* concentrations in rivers. It accounts for the treatment of human faecal waste in wastewater treatment plants (WWTPs) and onsite sanitation systems, storage and treatment of livestock manure in different breeding systems, the seasonality of livestock manure application to land and the *E. coli* concentration attenuation when travelling in rivers. Two scenarios were developed to evaluate the impact of direct discharge of faecal waste into rivers on in-stream *E. coli* concentrations. Modelled *E. coli* concentrations in rivers were compared with measured data to evaluate the model performance. Model outcomes were used to identify hotspots of high *E. coli* concentrations, evaluate the microbial water quality and attribute *E. coli* to different sources for all of China and its nine main river basins. In addition, priorities and opportunities for *E. coli* pollution management were explored. Our findings provide a comprehensive understanding of *E. coli* pollution dynamics in China's rivers and facilitate the development of effective microbial water quality control policies.

## 2. Methodology

The GloWPa-Ecoli C1 model is developed and advanced based on the GloWPa-Crypto C1 model (Vermeulen et al., 2019). Fig. S1 presents a schematic diagram of the modelling framework. After being excreted, *E. coli* in human faeces reach rivers in four ways: 1) with effluent from sewers connected to wastewater treatment plants, 2) with discharge from onsite sanitation systems, 3) through direct discharge of untreated

human faeces and 4) with runoff carrying faeces from land to rivers. For livestock, four manure reservoirs are accounted for: manure 1) directly applied on land, 2) dropped directly on land during livestock grazing, 3) applied on land after storage and 4) directly discharged into rivers after treatment. From land, *E. coli* is brought to rivers by runoff. The decay of *E. coli* during wastewater treatment and manure storage is taken into account. *E. coli* decay in rivers affected by temperature, solar radiation and sedimentation is incorporated as well. From the *E. coli* load in the rivers, the concentration of the bacteria in rivers is estimated by dividing the *E. coli* load by river discharge. The GloWPa-Ecoli C1 model has a spatial resolution of $0.5° \times 0.5°$ and represents the year 2020 with a monthly time step. The load and concentrations are evaluated for nine major river basins in China: Songliao, Haihe, Huaihe, Continental, Pearl, Yellow, Yangtze, Southeast and Southwest River Basins. Locations of the river basins are shown in Fig. S2. Key input variables and data sources are listed in Supplementary Material S1. Detailed calculation processes are presented in the following sections.

### 2.1. E. coli from humans to land and rivers

*E. coli* excretion from humans within faeces is calculated as follows:

$$E_{humanE,i} = C_{faeces} \times W_w \times P_i \times 365 \tag{1}$$

where $E_{humanE,i}$ is the annual *E. coli* excretion in province i (CFU); $C_{faeces}$ is the mean *E. coli* concentration in faeces; $W_w$ is the daily mean faeces wet weight and $P_i$ is the population in province i (cap). $C_{faeces}$ is $3.83 \times 10^7$ CFU/g, calculated from (Farnleitner et al., 2010; Islam et al., 2019; Masters et al., 2015) and $W_w$ is 243 g/cap/day (Rose et al., 2015). Population data of China is from WorldPop population density of 2020 (WorldPop, 2020).

After excretion, faeces enter the environment through various sanitation systems as in Musaazi (2020). These differ between urban and rural areas. The method to identify urban and rural areas in China is described in Supplementary Material S2. Fractions of human sanitation systems in China for 2018 from the WHO/UNICEF Joint Monitoring Program (JMP) are applied. The sanitation systems include the following categories: "flush to piped sewer system", "flush to elsewhere", pit latrine with slab, composting toilet, open defecation and other unimproved sanitation (World Health Organization and United Nations Children's Fund (WHO/UNICEF), 2023). In some of these sanitation systems, *E. coli* removal occurs, as explained in Sections 2.1.1 and 2.1.2. The *E. coli* load from each sanitation system is normalized by the province's population and multiplied by the population of each grid cell to estimate the human *E. coli* load for that cell. The monthly human *E. coli* load to land/rivers is calculated by dividing the annual load by 12.

#### 2.1.1. E. coli from WWTPs

Faeces flushed to piped sewer systems are treated in WWTPs and discharged to rivers. Associated *E. coli* load discharged from WWTPs to rivers is calculated as follows:

$$E_{WWTP} = \left( \left( E_{humanE,i} \times T_{WWTP,i} \times \left( f_{p,i} \times (1 - RE_p) \times f_{Ep} + f_{s,i} \times (1 - RE_s) \times f_{Es} + f_{t,i} \times (1 - RE_t) \times f_{Et} \right) \right) \Big/ P_i \right) \times P_{g,i} \tag{2}$$

where $E_{WWTP}$ is the annual *E. coli* load from WWTPs to rivers in a grid cell (CFU); $T_{WWTP,i}$ is wastewater treatment rate in province i; $f_{p,i}$, $f_{s,i}$, and $f_{t,i}$ is the fractions of WWTPs practising primary, secondary or tertiary treatment; $RE_p$, $RE_s$, $RE_t$ are the removal efficiencies of *E. coli* by primary, secondary and tertiary treatment; $f_{Ep}$, $f_{Es}$ and $f_{Et}$ are percentages of *E. coli* in the effluent as opposed to sludge after primary, secondary and tertiary treatment; $P_{g,i}$ is the population in a grid cell in province i (cap). It is assumed that all the WWTPs in China include a disinfection step, according to the Code for Design of Outdoor Wastewater Engineering (Ministry of Housing and Urban-Rural Construction

of the People's Republic of China (MHURC), 2016). Therefore, an additional 1.5 log-reduction of *E. coli* during disinfection is summarized from the literature and incorporated into the removal efficiencies. For details regarding *E. coli* load from WWTPs, see Supplementary Material S3.

### 2.1.2. E. coli from onsite sanitation systems and other sanitation types

Onsite sanitation systems (referred to as "toilets" hereafter) include pit latrines with slab and composting toilets. Human faeces are initially contained in toilets, where *E. coli*, associated with the solid matter, undergoes decay. Part of the waste is treated ($f_{tt,i}$) or covered and buried in the subsurface ($f_{tcb,i}$), consequently the associated *E. coli* cannot reach rivers. Part of the waste ($R_{t,dump,i}$) is dumped on site (Equation (6)), and the *E. coli* associated can be brought to the rivers by runoff (in Section 2.3). The remaining ($R_{t,dis,i}$) is discharged to surface waters due to the flushing of pit latrines or illegal toilet emptying (Musaazi, 2020):

$$E_{toilet} = \left( \left( E_{humanE,i} \times f_{t,i} \times \left( 1 - f_{tt,i} - f_{tcb,i} \right) \times RE_t \times R_{t,dis,i} \right) \Big/ P_i \right) \times P_{g,i}$$

$$\tag{3}$$

where $E_{toilet}$ is the annual *E. coli* emission from toilets to rivers in a grid cell (CFU); $f_{t,i}$ is the fraction of people using toilets in province i; $f_{tt,i}$ is the fraction of waste from toilets treated in province i; $f_{tcb,i}$ is the fraction of waste from toilets covered and buried in province i; $RE_t$ is the removal rate of *E. coli* in toilets (derived from (Musaazi, 2020), $R_{t,dis,i}$ is the fraction of human waste discharged to rivers from toilets in province i. $f_{tt,i}$, $f_{tcb,i}$ is taken from the input dataset (GWPP, 2020b) of Pathogen Flow & Mapping Tool, which is an interactive online platform developed by the Global Water Pathogens Project for understanding and visualization of pathogen fate and transport across different sanitation systems (GWPP, 2020a).

*E. coli* from "flush to elsewhere" is assumed to be directly discharged into rivers in both urban and rural areas, as this sanitation system is categorized as unimproved (World Health Organization and United Nations Children's Fund (WHO/UNICEF), 2018):

$$E_{flush\ to\ elsewhere} = \left( \left( E_{humanE,i} \times f_{flush\ to\ elsewhere,i} \Big/ P_i \right) \right) \times P_{g,i} \tag{4}$$

where $E_{flush\ to\ elsewhere}$ is the annual *E. coli* emission from people practising "flush to elsewhere" to rivers in a grid cell (CFU); $f_{flush\ to\ elsewhere,i}$ is the fraction of people applying "flush to elsewhere" in province i.

Open defecation in urban areas is defined as direct discharge due to the lack of space to drop faeces on land (Equation (5)), and in rural areas, *E. coli* from open defecation ends up on land (Equation (6)) (Vermeulen et al., 2015). Other unimproved sanitation is defined as direct discharge in both urban and rural areas as it is unknown how faeces are treated. *E. coli* load from other direct discharge sources (open defecation in urban areas and other unimproved sanitation) to the rivers is calculated as follows:

$$E_{other\ direct\ discharge} = \left( \left( E_{humanE,i} \times \left( f_{open,u,i} + f_{o,i} \right) \Big/ P_i \right) \right) \times P_{g,i} \tag{5}$$

where $E_{other\ direct\ discharge}$ is the annual *E. coli* emission from other direct discharge sources to rivers in a grid cell (CFU); $f_{open,u,i}$ is the fraction of people practising open defecation in urban areas in province i; $f_{o,i}$ is the fraction of people using other unimproved sanitation in province i.

*E. coli* load from toilets and open defecation in rural areas to land constitutes human diffuse sources:

where $E_{human\ diffuse\ land}$ is the annual *E. coli* emission from human diffuse sources to land in a grid cell (CFU); $f_{open,r,i}$ is the fraction of people practising open defecation in rural areas in province i; $R_{t,dump,i}$ is the fraction of human waste dumped on site from toilets in province i, taken from (GWPP, 2020b). The transportation of *E. coli* on land to rivers from human diffuse sources is calculated in Section 2.3.

### 2.2. E. coli from livestock to land and rivers

The calculation of *E. coli* from livestock to land and rivers is adapted from Vermeulen et al. (2017). In brief, manure production and *E. coli* excretion with manure is calculated for 14 main livestock species in China: pigs, sheep, goats, dairy cows, beef cattle, horses, donkeys, mules, camels, laying hens, broilers, meat ducks, laying ducks and rabbits in three livestock breeding systems: smallholder, medium and industrial farms. Livestock manure fate depends on storage versus grazing. In the grazing seasons, manure from livestock grazing on grassland (dairy cows, beef cattle, goats and sheep) is directly dropped on land. In other months, livestock is kept in house where the manure is collected and stored. In areas where livestock are not usually grazing, it is assumed that grazing only happens in smallholder farms and happens all year around. The remainder of the manure enters storage systems. Manure treatment in different storage systems for the 14 livestock species in the three livestock breeding systems is considered. To capture the seasonality of the *E. coli* load and concentrations, the timing of the application of stored manure on land is accounted for. It is assumed that manure is applied on land during the planting season of the three main staple crops in China: rice, maize and wheat. Moreover, a portion of the manure from livestock is directly discharged into the rivers after treatment in Southeast, Pearl, Haihe, Huaihe, Yellow and Yangtze River Basins, as it is assumed only in these river basins, livestock farms could access rivers to directly discharge livestock waste after treatment. Detailed calculation procedures of *E. coli* from livestock to land ($E_{livestock,\ land,\ m}$) and directly discharged into the rivers ($E_{livestockD}$) are elaborated in Supplementary Material S4.

### 2.3. E. coli transport from land to rivers

After rainfall, *E. coli* is released from human faeces and livestock manure applied on the field, and is subsequentially brought to rivers via runoff (Cho et al., 2016). For simulating the process of *E. coli* releasing from manure, the approach from the WAM model (Collins & Rutherford, 2004) is applied, which considers the relationship of *E. coli* released from manure with the runoff amount:

$$f_{re} = \frac{O}{K_t}, when\ O < K_t$$

$$f_{re} = 1, when\ O > K_t \tag{7}$$

where $f_{re}$ is the *E. coli* release rate (%); O is the monthly averaged surface run-off (mm/d) from the Variable Infiltration Capacity (VIC) model, a grid-based macroscale hydrological model (Liang et al., 1994), version 4.1.2; $K_t$ is the run-off threshold coefficient (mm/d). When O is smaller than $K_t$, the fraction of *E. coli* released from manure is proportional to the run-off amount. When O is greater than K, it is assumed that all *E. coli* are washed out. $K_t$ is set to 25 mm/day as in Collins and Rutherford (2004). After *E. coli* is released from the manure, the residual *E. coli* remains in the manure and dies off and is subsequently flushed away by runoff

$$E_{human\ diffuse\ land} = \left( \left( E_{humanE,i} \times \left( f_{t,i} \times \left( 1 - f_{tt,i} - f_{tcb,i} \right) \times RE_t \times R_{t,dump,i} + f_{open,r,i} \right) \Big/ P_i \right) \right) \times P_{g,i} \tag{6}$$

during the following months. Therefore, the *E. coli* released from manure ($E_{release, m}$) to runoff and left in manure ($E_{leftover, m}$) in month m is calculated as:

$$E_{release,m} = \left(\frac{E_{human\ diffuse\ land}}{12} + E_{livestock,land,m}\right) \times f_{re} \qquad (8)$$

$$E_{leftover,m} = \left(\frac{E_{human\ diffuse\ land}}{12} + E_{livestock,land,m}\right) \times (1 - f_{re}) \qquad (9)$$

Following the release, part of the *E. coli* is retained in soil, while the rest travels with surface run-off and ends up in surface waters. Blaustein et al. (2015) found that after the release, *E. coli* concentrations in surface runoff and leachate infiltrating into the soil are similar. Therefore, the fraction of *E. coli* in runoff is calculated as:

$$f_{runoff} = \frac{C_{runoff} \times V_{runoff}}{C_{runoff} \times V_{runoff} + C_{soil} \times V_{soil}} \qquad (10)$$

where $f_{runoff}$ is the fraction of *E. coli* in surface runoff; $C_{runoff}$ is the concentration of *E. coli* in surface runoff; $V_{runoff}$ is the amount of runoff; $C_{soil}$ is the *E. coli* concentration in leachate going into soil; $V_{soil}$ is the amount of rainfall that penetrates the soil. As $C_{runoff}$ is equivalent to $C_{soil}$, $f_{runoff}$ is equivalent to the runoff coefficient. Therefore, runoff coefficients from California Water Boards (2011) and land use data in China in 2020 (Xu et al., 2018) are combined to calculate $f_{runoff}$ in each grid cell. The handling of runoff coefficients and land use data is described in Supplementary Material S5. The release of *E. coli* from the soil reservoir is not considered, as it contributes a relatively small load to runoff compared to manure (Muirhead & Monaghan, 2012).

Total *E. coli* load to rivers in a grid cell from land in month m ($E_{diffuse, m}$) is calculated as:

$$E_{diffuse,m} = \left(E_{release,m} + E_{leftover,m-1} \times R_{a,m-1}\right) \times f_{runoff} \qquad (11)$$

where $R_{a, m-1}$ is the survival of *E. coli* in month m-1, calculated with Equation S(5) while $t_s$ in Equation S(5) is the number of days in month m-1. The calculation of $E_{leftover,m-1}$ for December 2019 is described in Supplementary Material S6.

### 2.4. Total E. coli load to the rivers

The total *E. coli* load to the rivers in a grid cell in month m ($E_m$, CFU/month) is calculated as:

$$E_m = \frac{\left(E_{WWTP} + E_{toilet} + E_{flush\ to\ elsewhere} + E_{other\ direct\ discharge}\right) + E_{livestockD}}{12} + E_{diffuse,m} \qquad (12)$$

### 2.5. Survival in stream

*E. coli* survival during transport in the rivers followed the approach from GloWPa-Cryptal C1 (Vermeulen et al., 2019), calculated as:

$$C_t = C_0 \times e^{-K_{r,m} \times t_{r,m}} \qquad (13)$$

where $C_t$ is the *E. coli* concentration in rivers (CFU/L) after travel time $t_{r,m}$ in month m (days), $C_0$ is the initial *E. coli* concentration (CFU/L), $K_{r,m}$ is the loss rate coefficient in month m (day$^{-1}$). The common pathways for *E. coli* decay in rivers include temperature-dependent decay (Blaustein et al., 2013), solar radiation-induced decay (Dean & Mitchell, 2022) and sedimentation (Pachepsky & Shelton, 2011). Accordingly, $K_{r,m}$ consists of three components:

$$K_{r,m} = K_{T,m} + K_{R,m} + K_{S,m} \qquad (14)$$

where $K_{T,m}$ is the temperature-dependent decay rate (day$^{-1}$), $K_{R,m}$ is the solar radiation-dependent decay rate (day$^{-1}$), and $K_{s,m}$ is the loss rate due to sedimentation (day$^{-1}$) in month m. Detailed calculation of $K_{T,m}$

$K_{R,m}$ $K_{s,m}$ see Supplementary Material S7.

Resuspension of *E. coli* from riverbed sediments, driven by shear stress, can be an important source of *E. coli* in rivers (Cho et al., 2010). However, modelling this process requires accurate data on sediment properties and *E. coli* concentrations in sediments (Pandey et al., 2012), which remain scarce in China. To evaluate the impact of incorporating resuspension process with limited *E. coli* sediment concentration data on the model, we tested two scenarios in our sensitivity analysis: one assuming low *E. coli* concentrations in sediments and another assuming high concentrations. Detailed methods are provided in Supplementary Material S8.

### 2.6. River geometry, water residence time and routing

The *E. coli* concentrations in rivers are calculated as:

$$C_m = \frac{(E_m + L_m) \times e^{-K_{r,m} \times t_{r,m}}}{Q_m} \qquad (15)$$

where $C_m$ is the *E. coli* concentrations in month m in a grid cell; $L_m$ is the *E. coli* load in month m from upstream grid cells that drain into the current grid cell; $Q_m$ is the river discharge (m$^3$/s). Grid cells with a monthly average discharge <1 m$^3$/s are excluded from the calculation, as the model was found to produce unrealistically high *E. coli* concentrations for locations with extremely low discharge. River geometry, water residence time, routing of *E. coli* in the rivers are calculated the same way as in the GloWPa-Crypto C1 model (Vermeulen et al., 2019).

### 2.7. Scenario development

Sections 2.1.2 and 2.2 assume that a portion of faecal waste, originating from both humans and livestock, is directly released into the environment. However, the presence and fractions of direct discharge from humans and livestock are highly uncertain (discussed in Section 4.1). Given the uncertainty and their potentially significant impact, two scenarios have been developed to explore the influence of direct discharge on *E. coli* load and concentrations. Under Scenario 1 (S1), both direct discharge of faecal waste from "flush to elsewhere" and livestock existed as described in the above sections, while in Scenario 2 (S2), no direct discharge of faecal waste from "flush to elsewhere" and livestock is assumed. The remaining fractions of human sanitation types and manure storage systems in S2 are normalized accordingly.

### 2.8. Model performance

To assess the model performance, simulated *E. coli* concentrations are compared with measurements obtained from published literature and environmental agency monitoring reports. An overview of the collected observational data is shown in Table S17. A map displaying the locations of the collected observational data is provided in Fig. S3. Some of the observational data are concentrations of faecal coliforms (FC). To maintain consistency, FC concentrations are transformed to *E. coli* concentrations based on the finding that 77 % of FC in river waters are *E. coli* (Garcia-Armisen et al., 2007). Root mean square error (RMSE) is calculated with log values of modelled concentrations for both scenarios to assess the model performance. Sensitivity analysis is performed to test how the model outcome is affected by input data. S2 is the baseline scenario for the sensitivity analysis. An overview of the changes to the variables used during the sensitivity analysis is provided in Table S18.

## 3. Results

### 3.1. Model performance

When comparing modelled concentration distribution with annual average observed concentration distribution, the model results show a
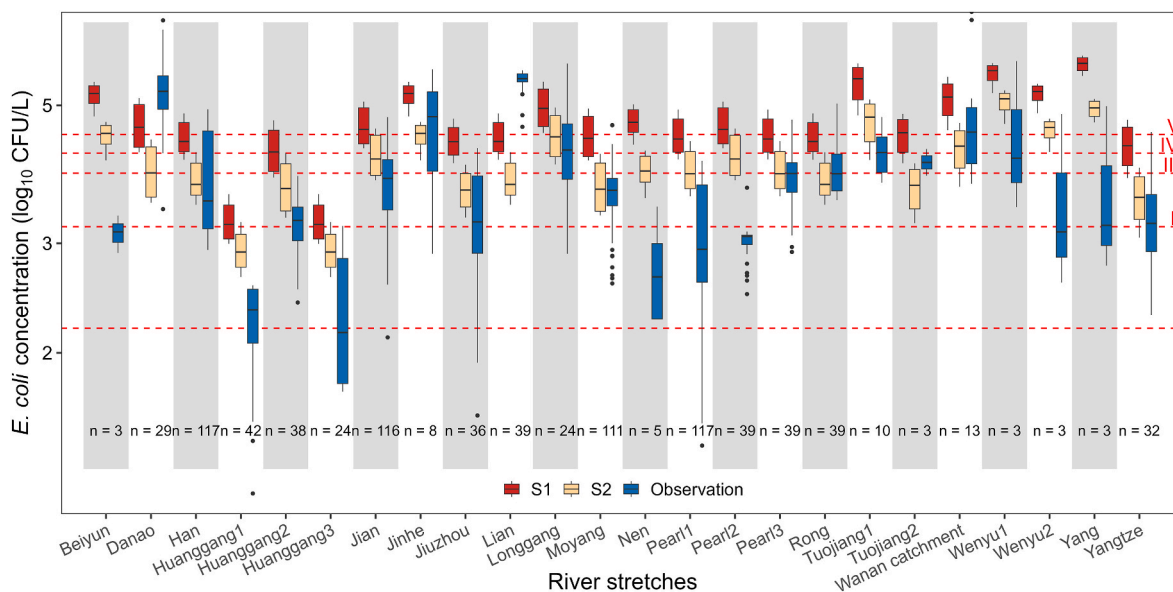
**Fig. 1.** Comparison of modelled *E. coli* concentrations (each river stretch has 12 modelled concentrations) from S1 and S2 with observed *E. coli* concentrations (collected FC concentrations are converted to *E. coli* concentrations). The n indicates the number of observations per river stretch (grid cell). Red dashed lines represent water quality thresholds from the Chinese Environmental Quality Standards for Surface Water (detailed in 3.2). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
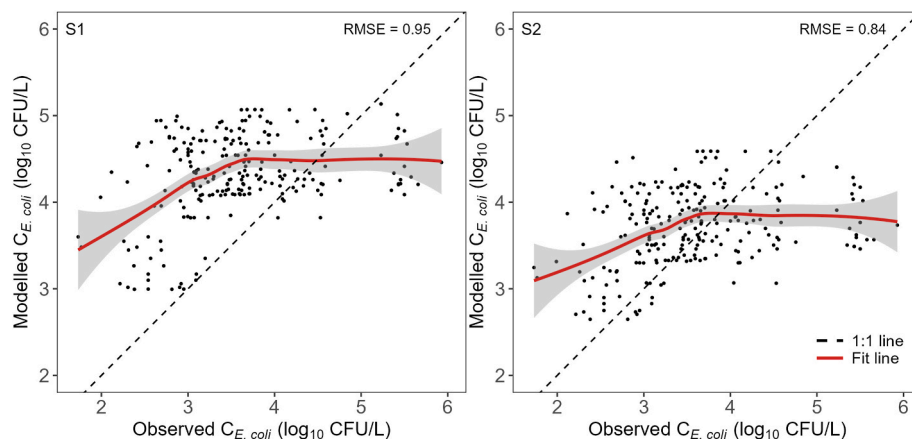


**Fig. 2.** Point-to-point comparison of monthly observed and modelled *E. coli* concentrations for scenarios S1 and S2. Each scatter plot represents the relationship between observed and modelled *E. coli* concentrations, with the black points indicating individual data points from the same river stretch (grid cell). The dashed black line represents the 1:1 line and the red line represents the smoothed fit line. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

generally good performance, while exhibiting both overestimations and underestimations, as presented in Fig. 1. Between the two scenarios, S1 results in higher concentrations compared with S2. S2 displays a closer alignment with the observed concentrations, with the majority of the modelled concentrations falling within the same grades of water quality as the observations. Fig. 2 further demonstrates a reasonably good alignment between the model results and the observations, supported by the RMSE values: 0.95 for S1 and 0.84 for S2. In both scenarios, the model tends to overestimate at low *E. coli* concentrations but struggles to accurately simulate conditions with elevated *E. coli* levels.

Temporally, monthly variations in *E. coli* concentrations in rivers are both modelled and observed, as depicted in Fig. 3. The monthly modelled concentrations vary within 1.5 log10 units between warm seasons (spring to summer) and cold seasons (autumn to winter). Modelled *E. coli* concentrations are typically lower in warm seasons and higher in cold seasons (Fig. 3 and Fig. S4), with some grid cells showing the opposite pattern (Fig. S5). Observations show higher *E. coli*

concentrations during warm seasons or cold seasons at some sampling sites, while others remain relatively constant or fluctuate without a clear trend throughout the year. The *E. coli* load to rivers generally remains relatively constant throughout the year (Fig. S6), while the *E. coli* load in some river stretches, such as Huanggang and Yangtze, displays a minor increase in spring and summer, correlating with the application of livestock manure (Fig. 3).

### 3.2. E. coli hotspots in rivers in China

Simulated *E. coli* concentrations in rivers in China in 2020 range from $10^{-0.4}$ CFU/L to $10^{6.3}$ CFU/L in S1 and $10^{-1.2}$ CFU/L to $10^{6.0}$ CFU/L in S2. Hotspots are defined as river sections where the *E. coli* concentrations exceed 7700 CFU/L, corresponding to water quality of Grade IV to V- as defined in the Chinese environmental quality standards for surface water (Ministry of Environmental Protection (MEP), 2002). The threshold for the water quality standards is shown in Table S19.
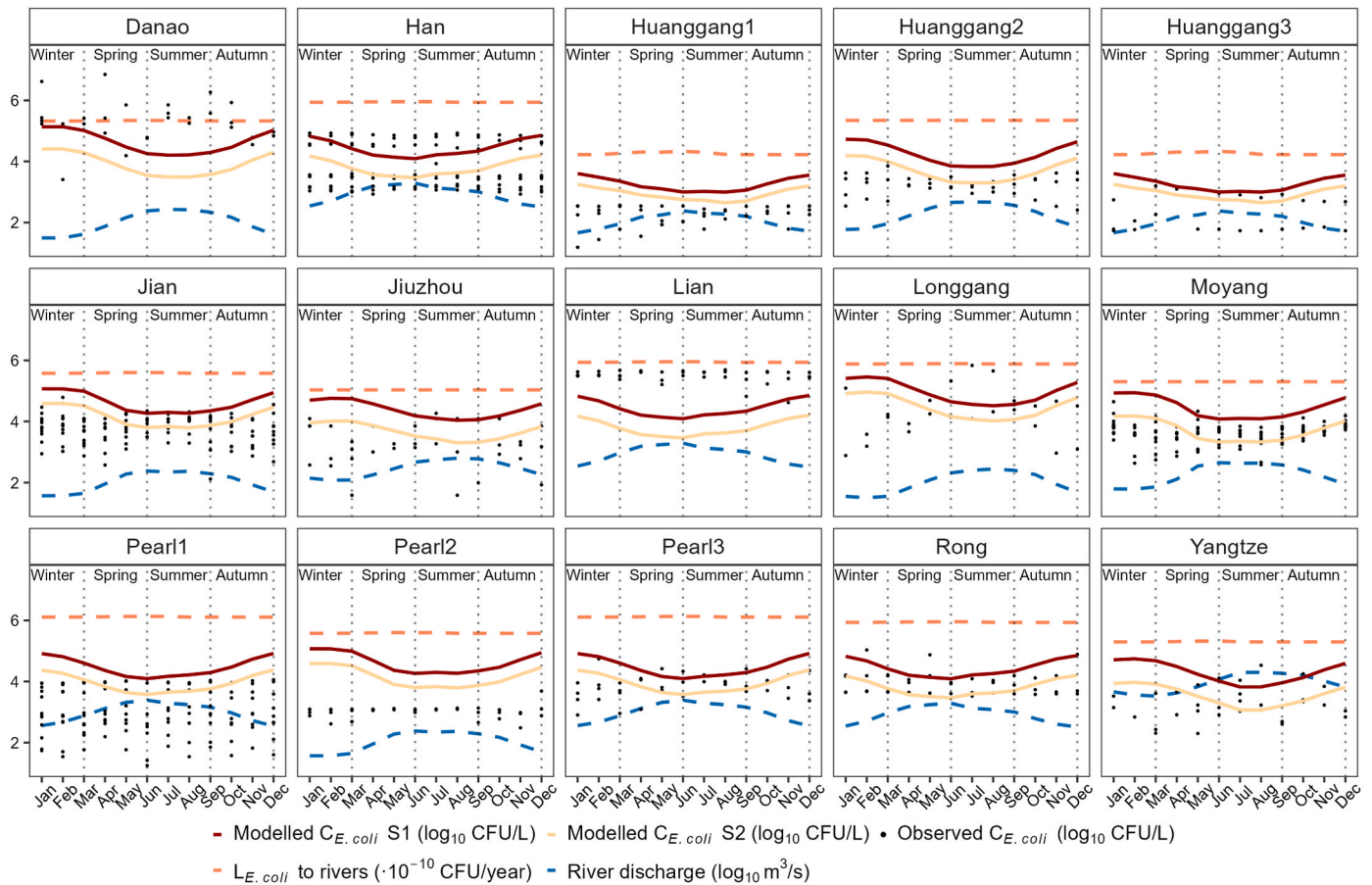
**Fig. 3.** Comparison of monthly modelled *E. coli* concentrations (S1 and S2) with observed *E. coli* concentrations (collected FC concentrations are converted to *E. coli* concentrations), monthly *E. coli* load of S2 to the rivers (*E. coli* load is divided by $10^{10}$ to make it fit in same ranges as *E. coli* concentrations and river discharge) and monthly river discharge. Each plot represents a grid cell. Spring lasts from March to May, summer from June to August, autumn from September to November, and winter from December to February.
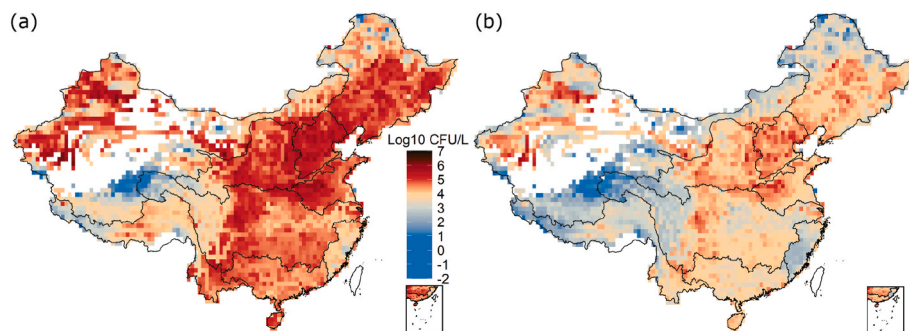


**Fig. 4.** (a): annual average *E. coli* concentrations in rivers in S1; (b): annual average *E. coli* concentrations in rivers in S2 (grid cells in white are either the discharge is below the threshold as in Section 2.6 or no data). The small map on the bottom right side represent the Nine-Dash Line in the South China Sea.

Spatially, hotspots of high *E. coli* concentrations mainly reside in Haihe, Huaihe, Pearl, middle and downstream areas of Yellow and Yangtze River Basins, central and south of Songliao and northwest of Continental River Basins in both scenarios, while S1 exhibits higher concentrations than S2 (Fig. 4). Northern river basins, particularly the Haihe and Huaihe River Basins, tend to exhibit higher concentrations compared with southern river basins. This is primarily due to lower river discharge in northern China (Fig. S7), which results in reduced dilution capacity, leading to elevated *E. coli* concentrations. It is also found that large rivers with high discharge such as Yangtze River and Yellow River, are less polluted by *E. coli* than small rivers around them. This could be attributed to the typically high dilution capacities in these large rivers

(Fig. S7) and this pattern is also found in Jones et al. (2023).

### 3.3. Water quality assessment

Water quality in nine main river basins under two scenarios in 2020 is evaluated using the annual averaged modelled *E. coli* concentrations. This evaluation followed the procedure outlined in the Surface water environment quality assessment methods (Ministry of Environmental Protection (MEP), 2011), while using *E. coli* as the sole indicator. The water quality assessment reveals that 78.1 % of rivers in S1 exhibit poor microbial water quality (Grade IV or worse), compared to 45.6 % in S2. In the Haihe, Huaihe, Pearl, Yellow, Yangtze, Songliao, and Southeast
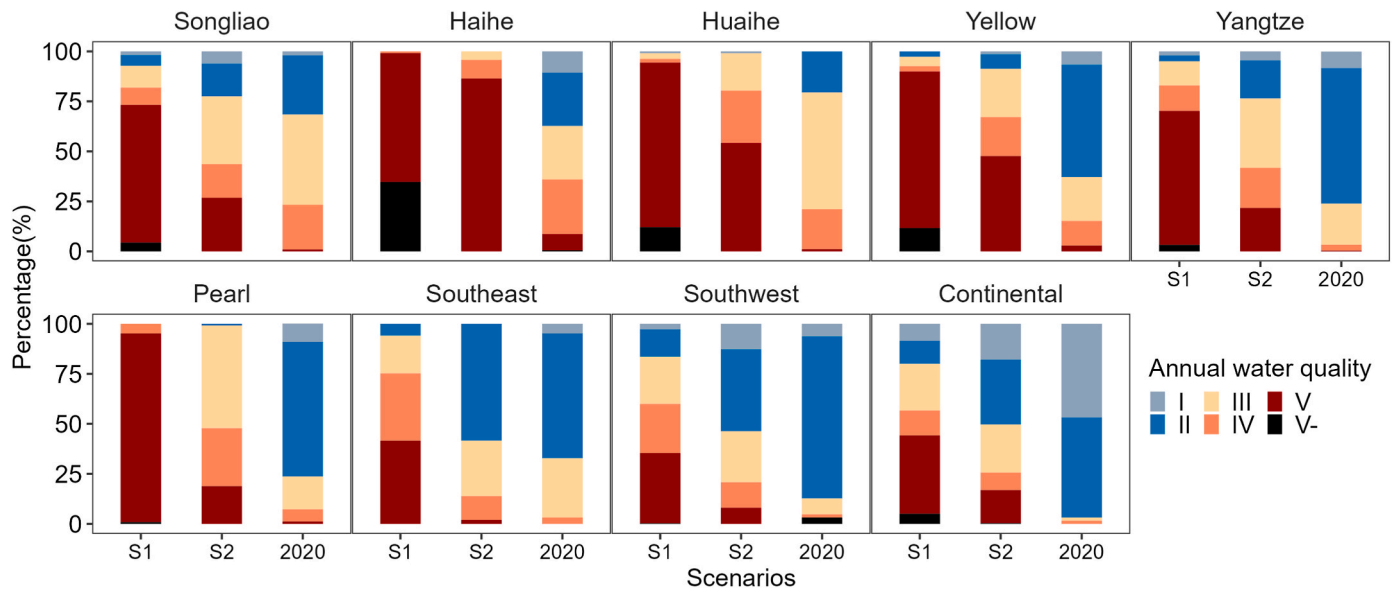
**Fig. 5.** Water quality assessed by *E. coli* in S1 and S2, and water quality assessment results of 2020 from China Ecological Environment Quality Bulletin (Ministry of Ecology and Environment of People's Republic of China (MEE), 2021).

River Basins, 75 %–100 % of the river stretches are classified as Grade IV or worse in S1 (Fig. 5). S2 predicts better water quality across all river basins compared with S1. The 2020 water quality assessment result from the China Ecological Environment Quality Bulletin (Ministry of Ecology and Environment of People's Republic of China (MEE), 2021), which assesses water quality using 21 chemical and physical parameters

without considering microorganisms, reports better water quality as compared to the predictions based on modelled *E. coli* concentrations from both S1 and S2. S2 predicted water quality in the Southeast, Songliao, and Southwest River Basins aligns roughly with the Bulletin's findings.
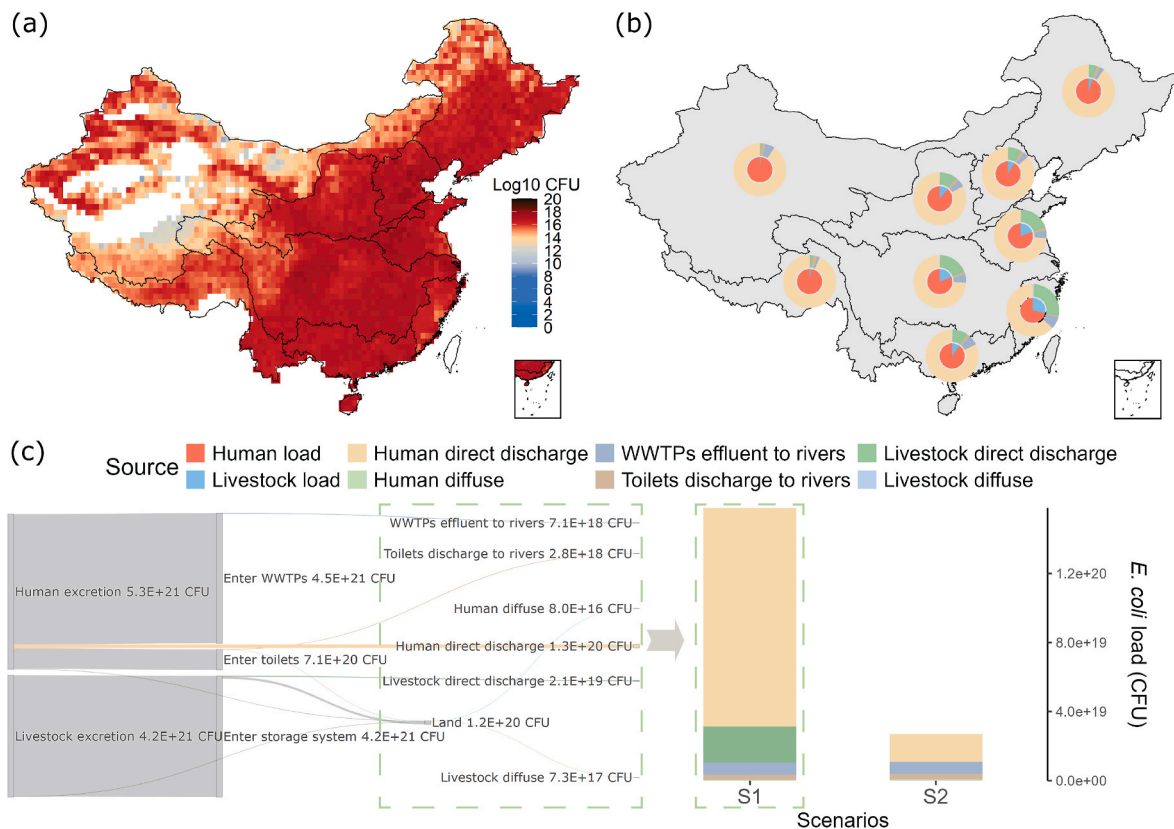


**Fig. 6.** (a): annual *E. coli* load to rivers in S1; (b): *E. coli* source attribution in nine river basins in S1; (c): Sankey diagram of the flow of *E. coli* load from source to rivers in S1 (the width of each flow is proportional to the relative amount of *E. coli* load being transported) and overall *E. coli* source attribution (S1 and S2), the nodes in the last column in the Sankey diagram and the bar plot of S1, highlighted in green rectangles, present identical source attribution information for S1. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

## 3.4. Source attribution

The simulated annual *E. coli* load to rivers in S1 is $1.6 \times 10^{20}$ CFU and $2.7 \times 10^{19}$ CFU in S2. Comparing *E. coli* load to rivers of S1 (Fig. 6a) with the human and livestock population (Fig. S8 and Fig. S9), *E. coli* load hotspots exhibits a similar east-west disparity, which follows the Hu Huanyong Line, with denser human and livestock populations concentrated in eastern China. Despite having lower *E. coli* loads in S2 (Fig. S10), the spatial distribution of hotspots is the same as in S1. Hotspots of *E. coli* load also align with the *E. coli* concentration hotspots in rivers. Monthly variation in *E. coli* load is minimal (Fig. S6).

The flow of *E. coli* load from source to rivers in S1 and source attribution of *E. coli* load to rivers for the whole of China is shown in Fig. 6c. Importantly, *E. coli* directly discharged from human and livestock account for 80.2 % and 13.1 % of the total *E. coli* load in rivers in S1, respectively, contrasting sharply with their shares of total *E. coli* excretion, 1.3 % (human direct discharge) and 0.2 % (livestock direct discharge). Subsequently, emissions from WWTPs emerge as the next major contributor (4.5 %), followed by emissions from toilets (1.8 %), with diffuse emissions from livestock (0.3 %) and humans (0.1 %) being the least significant. In comparison, the breakdown of source contributions in S2 is: human direct discharge contributes 59.1 % (consisting of waste from open defecation in urban areas and other unimproved sanitation), WWTPs 26.3 %, toilets 12.2 %, livestock diffuse sources 2.1 %, and human diffuse sources 0.3 %. Fig. 6b presents the distribution of *E. coli* loads from various sources in S1 in the nine river basins, with human sources emerging as the principal contributor in all basins.

Among livestock in S1, pigs contribute the most with 43.5 % (Fig. S11), followed by broilers at 24.9 % and laying hens at 18.4 %. The remaining livestock each account for less than 10 %. Regarding different breeding systems, *E. coli* load from smallholder farms account for 65.2 %, followed by industrial farms (22.4 %) and medium farms contributed the least (12.5 %) in S1. The shares of livestock and breeding systems in S1 and S2 are almost identical.

## 3.5. Sensitivity analysis

The sensitivity analysis indicates that the model is most responsive to the changes in human and livestock waste directly discharged into rivers (Table S18). Specifically, the introduction of 'flush to elsewhere' in S1 (1.2 % of the human waste directly discharged) results in a dramatic six-fold increase in both the mean and median modelled concentrations compared with S2 (baseline scenario for sensitivity analysis). For livestock, direct discharge of treated manure leads to 1.5 times higher concentrations than in the baseline scenario, while untreated manure discharge elevates the mean concentration to 18.1 times above the baseline. The sensitivity analysis also reveals a pronounced response to the human excretion rate of *E. coli*. A one-order-of-magnitude increase in the human *E. coli* excretion rate leads to a 9.9 times rise in the mean modelled concentrations. Moreover, the model demonstrates sensitivity to variations in the fraction of WWTPs employing diverse technologies. Upgrading primary treatment to higher levels of treatment can yield around 20 % reduction of *E. coli* concentrations. During the *E. coli* routing processes in the rivers, fluctuations in hydrological parameters such as discharge, river length and river residence time also exhibit a considerable impact. The model is not sensitive to resuspension of *E. coli* from the sediment.

## 4. Discussion

This study presents the first large-scale model to simulate *E. coli* concentrations in rivers and systematically assess the microbial water quality across China's river systems on a national scale for the first time. The model allows us to identify hotspots of high *E. coli* concentrations and the relative importance of sources. The modelling results can offer valuable insights to protect public health and environmental

sustainability. Meanwhile, sustainable management of microbial water quality in rivers support the achievement of the broader goals of Sustainable Development Goal 6 (SDG 6) on clean water and sanitation for all (Rose et al., 2023).

## 4.1. Model performance and uncertainties

Model comparison to observations indicates that the model outcomes overall represent the dynamics of *E. coli* concentrations in rivers in China. The RMSE of the model results is slightly higher compared with other large-scale models simulating *E. coli*/FC concentrations: 0.44–0.70 (Reder et al., 2015), but still demonstrates a reasonable level of accuracy. Discrepancies between the modelled and the observed concentrations are likely due to uncertainties to which the model is particularly sensitive.

The model is highly sensitive to the direct discharge of human and livestock waste (Table S18), significantly impacting *E. coli* loads and concentrations. As the data on direct discharge in China is scarce, to adopt a precautionary stance, it is assumed in this study that human waste categorized as 'flushed to elsewhere', open defecation (in urban areas) and other unimproved sanitation is directly discharged, although it may also be deposited on land, where the associated *E. coli* undergo decay. For livestock, it has been reported earlier that the direct discharge of livestock manure is the main source of pollutants, such as P and N (Strokal et al., 2016). Prior studies used varied and inconsistent information – with percentages of direct discharge of livestock manure varying between 0 and 75 % (see Supplementary Material S9). Nevertheless, policies by the State Council of the People's Republic of China aim for a 95 % or higher rate of manure treatment facility implementation in industrial farms by 2020, with the goal of achieving an over 75 % reuse rate of livestock manure (General Office of the State Council (GOSC), 2017; General Office of the State Council (GOSC), 2019), but the outcome of the policy implementation is unknown. Therefore, it is uncertain whether direct discharge of human faeces and livestock manure was still occurring in 2020, and to which extent. To minimize the uncertainty, the current study replaces the wastewater treatment rates from JMP data with those from the statistical yearbook of China (Ministry of Housing and Urban-Rural Construction of the People's Republic of China (MHURC), 2021). In the meantime, fractions of the rest of the sanitation systems, including 'flush to elsewhere' are normalized accordingly (described in Supplementary Material S3). Given that the statistical yearbooks are known to be the most reliable data source in China, leading us to believe that we have obtained the most accurate fractions of human direct discharge possible. Moreover, to further understand the uncertainty, two scenarios based on the reasonable range of direct discharge fractions were developed to explore the impact of these input data on the model outcome. As demonstrated in this study, scenario development serves as an effective methodology for handling uncertainty and gaining a holistic understanding regarding the response of the model to the variations in critical input data. Nonetheless, to achieve an optimal model performance and to address this potentially highly relevant source of faecal pollution, open, accurate data and further research on direct discharge is urgently needed.

Furthermore, the removal efficiency of the disinfection step in the WWTPs might be underestimated. Sun et al. (2023) found that in WWTPs in rural areas in China, a 6-log reduction can be reached when UV is combined with chlorination. Comparatively, a lower removal efficiency summarized from literature – a 1.5-log reduction for disinfection - is applied in the model for both urban and rural areas, while in urban areas, the treatment facilities are much more advanced (Ministry of Housing and Urban-Rural Construction of the People's Republic of China (MHURC), 2021).

Resuspension processes are not yet included in large-scale models simulating *E. coli* and FC (Jones et al., 2023; Reder et al., 2015; United Nations Environment Programme (UNEP), 2016), underscoring the need to assess their impacts. Results show that the model is not notably

sensitive to *E. coli* resuspension from sediment: neither low nor high sediment concentrations meaningfully increase overall *E. coli* levels. However, this does not imply that resuspension has no effect. Because resuspension typically occurs in short bursts following mechanical or hydrodynamic disturbances (Pachepsky & Shelton, 2011) a monthly time-step model that uses averaged river discharge masks short-term variability. Consequently, resuspension's overall contribution is not evident in most areas. An increase in *E. coli* concentration is observed only in regions where human and livestock inputs are minimal. In contrast, in areas already subject to high *E. coli* loading, the effect of resuspension is negligible. It is therefore recommended that future model setups incorporate resuspension at finer temporal resolutions, supported by more data on *E. coli* concentrations in sediments.

For monthly variation in *E. coli* concentrations, while observed data displayed multiple patterns, the model primarily predicted a single trend: elevated concentrations in cold seasons and reduced levels in warm seasons. While the monthly *E. coli* load to rivers is almost identical across the year, river discharge, being higher in the warm season, dominantly influences the *E. coli* concentration. Several processes not included in the model due to data availability, such as leakage of wastewater and sewage overflow, are expected to show seasonality. Leakage of wastewater due to low-quality pipes and improper construction is likely to happen in warm seasons in China (Xu et al., 2019; Zhao et al., 2018). Similarly, strong precipitation, which is more probable in warm seasons, can increase the volume of influent flowing into WWTPs, exceed the treatment capacity of the WWTPs and lead to sewage overflow (Li et al., 2023a; Xu et al., 2019). However, the model is not currently able to incorporate such short-term processes. Further research and data are needed to incorporate these processes to bring more accuracy to the model and better representation of the seasonal trend of *E. coli* load and concentrations.

Uncertainties also exist in the limited measured data used for model validation, which can impact the assessment of model accuracy. For example, observations of *E. coli* concentrations are strongly affected by rainfall events (Schilling et al., 2009; Xiao et al., 2013), or the samples might have been taken in surface waters that are not representative of the total modelled cell. In addition, in this study, many FC observations are used rather than *E. coli* data to validate the model. Given the variation in survival rates of FC and *E. coli* in rivers due to strain characteristics and environmental conditions (Anderson et al., 2005; Pachepsky & Shelton, 2011), FC may not fully represent *E. coli* behaviour. However, these FC data were the only ones available. Despite their limitations as proxies for *E. coli,* these data still allow for an initial understanding of the model's performance.

### 4.2. Microbial water quality

This study provides the first overview of microbial water quality in rivers across China. The model reveals that 45.6 %–78.1 % of rivers exhibit poor microbial water quality. However, microbial water quality has not received sufficient attention, as current water quality assessments in China do not consider microbial water quality (Ministry of Ecology and Environment of People's Republic of China (MEE), 2021; Ministry of Environmental Protection (MEP), 2011). This highlights a significant oversight of microbial water quality. Nevertheless, there are uncertainties and limitations in using *E. coli* as FIB to assess water quality. Studies have shown that its correlation with pathogens is inconsistent, varying by season and location (Pachepsky et al., 2016). Additionally, *E. coli* and pathogens exhibit different environmental growth and persistence patterns (Pachepsky et al., 2016). Despite these imperfection, *E. coli* remains the best available tool for monitoring microbial water quality (Pachepsky et al., 2018). Further research is needed to enhance our understanding of microbial contamination and its health implications in rivers in China. Utilizing modelled *E. coli* concentrations, quantitative microbial risk assessments can be conducted as a next step to evaluate health risks, inform policy

development, and improve water quality to safeguard public health.

### 4.3. Source attribution and priorities for mitigating microbial water quality

Considering the poor microbial water quality in rivers in China, actions need to be taken to reduce *E. coli* concentrations to protect human health. Identifying opportunities to mitigate *E. coli* emissions and concentrations based on the relative importance of sources is critical for developing effective management strategies.

Human sources are the predominant sources of *E. coli* in the rivers. This finding is consistent with other work modelling microorganism concentration in rivers (Jones et al., 2023; Vermeulen et al., 2019). The primary source of *E. coli* in rivers is the direct discharge of waste originating from humans and livestock, which highly impacts *E. coli* concentrations. The level of treatment of human waste within urban areas of China has reached 96.4 % as shown in Table S21. However, in rural regions, the collection and treatment rates average only 56.6 %. Besides, the share of WWTPs practising primary treatment is still high in some provinces like Xinjiang and Tibet. According to the sensitivity analysis, transitioning from secondary to tertiary treatment has limited improvement in reducing *E. coli* concentrations. Thus, enhancing human waste collection rates, expanding WWTPs coverage, and upgrading wastewater treatment levels, with a particular focus on phasing out primary treatment WWTPs, constitute effective measures for reducing *E. coli* concentrations. For livestock, the most important measure is to eradicate the direct discharge of manure to rivers, especially in smallholder farms.

Besides, mitigation strategies should be tailored to specific locations, prioritizing the primary pollution sources within the local environment to ensure efficient and effective management. Moreover, hotspots of *E. coli* also tend to harbour other pollutants including nitrogen, phosphorus, antibiotics and plastics in China (Li et al., 2024; Li et al., 2023b; Wang et al., 2018). Hence, a synergistic approach to pollutant management should be developed.

### 5. Conclusion

This is the first study simulating spatially explicit *E. coli* concentrations in rivers across China. The modelling approaches developed, such as the removal of microorganisms in WWTPs and onsite sanitation systems, the storage and treatment of livestock manure, and the seasonality of livestock manure application to land, contribute to a more advanced and accurate process-based modelling of pollutants in river systems. Moreover, the usage of scenarios capturing critical but uncertain input data represents an effective methodology for handling uncertainty.

According to the GloWPa-Ecoli C1 model.

- 45.6 %–78.1 % of rivers in China exhibit poor microbial water quality. *E. coli* concentrations in rivers in China in 2020 range from $10^{-1.2}$ to $10^{6.3}$ CFU/L in two scenarios. Hotspots of high *E. coli* concentrations are Haihe, Huaihe, Pearl, middle and downstream areas of Yellow and Yangtze River Basins, central and south of Songliao and northwest of Continental River Basins.
- Direct discharge of faecal waste from humans and livestock to rivers is the predominant *E. coli* source, followed by discharge from WWTPs. *E. coli* load from human sources dominate over livestock sources.
- *E. coli* concentrations in rivers in China can be reduced by enhancing human waste collection rates, expanding WWTPs coverage, phasing out primary treatment WWTPs and eradicating the direct discharge of manure to rivers, especially in smallholder farms.

The current modelling approaches can easily be adapted to evaluate diverse microorganisms and be applied across global regions. Such pathogens for instance include pathogenic *E. coli* strains, such as *E. coli*

O157:H7, or antibiotic-resistant bacteria. The insights gained from the study can inform policymakers and regulators in creating effective environmental policies and regulations to protect microbial water quality and safeguard public health.

## CRediT authorship contribution statement

**Songtao Mei:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Kai Wang:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Heike Schmitt:** Writing – review & editing, Supervision, Resources, Methodology. **Nynke Hofstra:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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

During the preparation of this work the authors used chat-GPT in order to polish the English writing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Declaration of competing interest

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

## Acknowledgement

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2025.126184.

## Data availability

Data will be made available on request.

## References

Anderson, K.L., Whitlock, J.E., Harwood, V.J., 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl. Environ. Microbiol. 71 (6), 3041–3048. https://doi.org/10.1128/aem.71.6.3041-3048.2005.

Blaustein, R.A., Pachepsky, Y., Hill, R.L., Shelton, D.R., Whelan, G., 2013. Escherichia coli survival in waters: temperature dependence. Water Res. 47 (2), 569–578. https://doi.org/10.1016/j.watres.2012.10.027.

Blaustein, R.A., Pachepsky, Y.A., Hill, R.L., Shelton, D.R., 2015. Solid manure as a source of fecal indicator microorganisms: release under simulated rainfall. Environ. Sci. Technol. 49 (13), 7860–7869. https://doi.org/10.1021/acs.est.5b01095.

California Water Boards, 2011. Runoff coefficient (C) fact sheet. https://www.waterboards.ca.gov/water_issues/programs/swamp/docs/cwt/guidance/513.pdf (accessed by January 2024).

Cho, K.H., Cha, S.M., Kang, J.-H., Lee, S.W., Park, Y., Kim, J.-W., et al., 2010. Meteorological effects on the levels of fecal indicator bacteria in an urban stream: a modeling approach. Water Res. 44 (7), 2189–2202. https://doi.org/10.1016/j.watres.2009.12.051.

Cho, K.H., Pachepsky, Y.A., Oliver, D.M., Muirhead, R.W., Park, Y., Quilliam, R.S., et al., 2016. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: State of the science and future opportunities. Water Res. 100, 38–56. https://doi.org/10.1016/j.watres.2016.04.064.

Coffey, R., Cummins, E., Bhreathnach, N., Flaherty, V.O., Cormican, M., 2010. Development of a pathogen transport model for Irish catchments using SWAT. Agric. Water Manag. 97 (1), 101–111. https://doi.org/10.1016/j.agwat.2009.08.017.

Coffey, R., Dorai-Raj, S., O'Flaherty, V., Cormican, M., Cummins, E., 2013. Modeling of pathogen indicator organisms in a small-scale agricultural catchment using SWAT. Hum. Ecol. Risk Assess. 19 (1), 232–253. https://doi.org/10.1080/10807039.2012.701983.

Collins, R., Rutherford, K., 2004. Modelling bacterial water quality in streams draining pastoral land. Water Res. 38 (3), 700–712. https://doi.org/10.1016/j.watres.2003.10.045.

Dean, K., Mitchell, J., 2022. Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters. Water Res. 211, 118051. https://doi.org/10.1016/j.watres.2022.118051.

Farnleitner, A.H., Ryzinska-Paier, G., Reischer, G.H., Burtscher, M.M., Knetsch, S., Kirschner, A.K., et al., 2010. Escherichia coli and enterococci are sensitive and reliable indicators for human, livestock and wildlife faecal pollution in alpine mountainous water resources. J. Appl. Microbiol. 109 (5), 1599–1608. https://doi.org/10.1111/j.1365-2672.2010.04788.x.

Garcia-Armisen, T., Prats, J., Servais, P., 2007. Comparison of culturable fecal coliforms and Escherichia coli enumeration in freshwaters. Can. J. Microbiol. 53 (6), 798–801. https://doi.org/10.1139/w07-033.%m.17668040.

General Office of the State Council (GOSC), 2019. Notice of the general office of the state Council on issuing the pilot work plan for the construction of "Waste-Free cities" (in Chinese). https://www.mee.gov.cn/zcwj/gwywj/201901/t20190123_690456.shtml (accessed by January 2024).

General Office of the State Council (GOSC), 2017. Opinions of the general office of the state Council on accelerating the resource utilization of livestock and poultry breeding wastes (in Chinese). https://www.gov.cn/zhengce/content/2017-06/12/content_5201790.htm (accessed by January 2024).

Global Water Pathogens Project (GWPP), 2020a. Pathogen Flow & Mapping Tool. https://tools.waterpathogens.org/maps (accessed by January 2024).

Global Water Pathogens Project (GWPP), 2020b. *World-country-level data* [dataset]. https://data.waterpathogens.org/en/dataset/world-countries/resource/d6d68118-1e4b-461c-97e3-75be4ac1e63f.

Han, S., Tian, F., Gao, L., 2020. Current status and recent trend of irrigation water use in China. Irrig. Drain. 69 (S2), 25–35. https://doi.org/10.1002/ird.2441.

Iqbal, M.S., Hofstra, N., 2018. Modeling Escherichia coli fate and transport in the kabul river basin using SWAT. Hum. Ecol. Risk Assess. 25 (5), 1279–1297. https://doi.org/10.1080/10807039.2018.1487276.

Islam, M.A., Amin, M.B., Roy, S., Asaduzzaman, M., Islam, M.R., Navab-Daneshmand, T., et al., 2019. Fecal colonization with multidrug-resistant E. coli among healthy infants in rural Bangladesh. Front. Microbiol. 10, 640. https://doi.org/10.3389/fmicb.2019.00640.

Islam, M.M.M., Sokolova, E., Hofstra, N., 2018. Modelling of river faecal indicator bacteria dynamics as a basis for faecal contamination reduction. J. Hydrol. 563, 1000–1008. https://doi.org/10.1016/j.jhydrol.2018.06.077.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. J. Appl. Microbiol. 123 (3), 570–581.

Jones, E.R., Bierkens, M.F.P., Wanders, N., Sutanudjaja, E.H., van Beek, L.P.H., van Vliet, M.T.H., 2023. DynQual v1.0: a high-resolution global surface water quality model. Geosci. Model Dev. (GMD) 16 (15), 4481–4500. https://doi.org/10.5194/gmd-16-4481-2023.

Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic escherichia coli. Nat. Rev. Microbiol. 2 (2), 123–140.

Li, R., Filippelli, G., Wang, L., 2023a. Precipitation and discharge changes drive increases in Escherichia coli concentrations in an urban stream. Sci. Total Environ. 886, 163892. https://doi.org/10.1016/j.scitotenv.2023.163892.

Li, S., Zhu, Y., Zhong, G., Huang, Y., Jones, K.C., 2024. Comprehensive assessment of environmental emissions, fate, and risks of veterinary antibiotics in China: an environmental fate modeling approach. Environ. Sci. Technol. 58 (12), 5534–5547. https://doi.org/10.1021/acs.est.4c00993.

Li, Y., Zhang, Q., Baartman, J., van Wijnen, J., Beriot, N., Kroeze, C., et al., 2023b. The plastic age: river pollution in China from crop production and urbanization. Environ. Sci. Technol. 57 (32), 12019–12032. https://doi.org/10.1021/acs.est.3c03374.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. Atmos. 99 (D7), 14415–14428. https://doi.org/10.1029/94JD00483.

Masters, N., Christie, M., Stratton, H., Katouli, M., 2015. Viability and stability of Escherichia coli and enterococci populations in fecal samples upon freezing. Can. J. Microbiol. 61 (7), 495–501. https://doi.org/10.1139/cjm-2015-0020.%M.26053765.

Mbanga, J., Abia, A.L.K., Amoako, D.G., Essack, S.Y., 2020. Quantitative microbial risk assessment for waterborne pathogens in a wastewater treatment plant and its receiving surface water body. BMC Microbiol. 20 (1), 346. https://doi.org/10.1186/s12866-020-02036-7.

Ministry of Ecology and Environment of People's Republic of China (MEE), 2021. China ecological environment quality Bulletin 2020 (in Chinese). https://www.mee.gov.cn/hjzl/sthjzk/zghjzkgb/202105/P020210526572756184785.pdf (accessed by January 2024).

Ministry of Environmental Protection (MEP), 2011. Surface water environment quality assessment methods (in Chinese). https://www.mee.gov.cn/gkml/hbb/bgt/201104/W020110401583735386081.pdf (accessed by January 2024).

Ministry of Environmental Protection (MEP), 2002. Environmental quality standard for surface water (GB3838-2002) (in Chinese). https://english.mee.gov.cn/standards_reports/standards/water_environment/quality_standard/200710/W020061027509896672057.pdf (accessed by January 2024).

Ministry of Housing and Urban-Rural Construction of the People's Republic of China (MHURC), 2021. Urban and rural construction statistical yearbook 2021 (in Chinese). https://www.mohurd.gov.cn/gongkai/fdzdgknr/sjfb/tjxx/jstjnj/index.html (accessed by January 2024).

Ministry of Housing and Urban-Rural Construction of the People's Republic of China (MHURC), 2016. Code for design of outdoor wastewater engineering (GB50014-2006) (in Chinese). https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefileib/201607/20160713_228080.html (accessed by January 2024).

Muirhead, R.W., Monaghan, R.M., 2012. A two reservoir model to predict Escherichia coli losses to water from pastures grazed by dairy cows. Environ. Int. 40, 8–14. https://doi.org/10.1016/j.envint.2011.11.009.

Murray, C.J., Ikuta, L., Sharara, K.S., Swetschinski, F., L Robles Aguilar, G., Gray, A., et al., 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399, 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0, 10325.

Musaazi, I.G., 2020. Predicting Pathogen Fate and Transport throughout the Sanitation Service Chain: Knowledge to Practice with the Global Water Pathogen Project. Master's thesis. San Diego State University. https://digitallibrary.sdsu.edu/islandora/object/sdsu%3A109763 (accessed by September 2022).

Okaali, D.A., Kroeze, C., Medema, G., Burek, P., Murphy, H., Tumwebaze, I.K., et al., 2021. Modelling rotavirus concentrations in rivers: assessing Uganda's present and future microbial water quality. Water Res. 204, 117615. https://doi.org/10.1016/j.watres.2021.117615.

Pachepsky, Y., Shelton, D., Dorner, S., Whelan, G., 2016. Can E. coli or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters? Crit. Rev. Microbiol. 42 (3), 384–393. https://doi.org/10.3109/1040841X.2014.954524.

Pachepsky, Y.A., Allende, A., Boithias, L., Cho, K., Jamieson, R., Hofstra, N., et al., 2018. Microbial water quality: monitoring and modeling. J. Environ. Qual. 47 (5), 931–938. https://doi.org/10.2134/jeq2018.07.0277.

Pachepsky, Y.A., Shelton, D.R., 2011. Escherichia coli and fecal coliforms in freshwater and estuarine sediments. Crit. Rev. Environ. Sci. Technol. 41 (12), 1067–1110. https://doi.org/10.1080/10643380903392718.

Pandey, P.K., Soupir, M.L., Rehmann, C.R., 2012. A model for predicting resuspension of Escherichia coli from streambed sediments. Water Res. 46 (1), 115–126. https://doi.org/10.1016/j.watres.2011.10.019.

Reder, K., Flörke, M., Alcamo, J., 2015. Modeling historical fecal coliform loadings to large European rivers and resulting in-stream concentrations. Environ. Model. Software 63, 251–263. https://doi.org/10.1016/j.envsoft.2014.10.001.

Rose, C., Parker, A., Jefferson, B., Cartmell, E., 2015. The characterization of feces and urine: a review of the literature to inform advanced treatment technology. Crit. Rev. Environ. Sci. Technol. 45 (17), 1827–1879. https://doi.org/10.1080/10643389.2014.1000761.

Rose, J.B., Hofstra, N., Hollmann, E., Katsivelis, P., Medema, G.J., Murphy, H.M., et al., 2023. Global microbial water quality data and predictive analytics: key to health and meeting SDG 6. PLOS Water 2 (8), e0000166. https://doi.org/10.1371/journal.pwat.0000166.

Schilling, K.E., Zhang, Y.-K., Hill, D.R., Jones, C.S., Wolter, C.F., 2009. Temporal variations of Escherichia coli concentrations in a large Midwestern river. J. Hydrol. 365, 79–85.

Strokal, M., Kroeze, C., Wang, M., Bai, Z., Ma, L., 2016. The MARINA model (model to assess river inputs of nutrients to seAs): model description and results for China. Sci. Total Environ. 562, 869–888.

Sun, J., Zhang, R., Xing, L., Wu, Q., Huang, Y., Lou, Y., et al., 2023. Occurrence and removal of conventional pollutants, estrogenicities, and fecal coliform in village sewage treatment plants along the Yangtze River, China. Environ. Sci. Pollut. Control Ser. 30 (7), 18014–18025. https://doi.org/10.1007/s11356-022-23467-x.

United Nations Environment Programme (UNEP), 2016. A snapshot of the world's water quality: towards a global assessment. https://globewq.info/assets/policy_brief_unep_wwqa_web.pdf (accessed by July 2023).

Vermeulen, L.C., Benders, J., Medema, G., Hofstra, N., 2017. Global cryptosporidium loads from livestock manure. Environ. Sci. Technol. 51 (15), 8663–8671. https://doi.org/10.1021/acs.est.7b00452.

Vermeulen, L.C., de Kraker, J., Hofstra, N., Kroeze, C., Medema, G., 2015. Modelling the impact of sanitation, population growth and urbanization on human emissions of Cryptosporidium to surface waters—a case study for Bangladesh and India. Environ. Res. Lett. 10 (9), 094017. https://doi.org/10.1088/1748-9326/10/9/094017.

Vermeulen, L.C., van Hengel, M., Kroeze, C., Medema, G., Spanier, J.E., van Vliet, M.T., et al., 2019. Cryptosporidium concentrations in rivers worldwide. Water Res. 149, 202–214.

Wang, M., Ma, L., Strokal, M., Ma, W., Liu, X., Kroeze, C., 2018. Hotspots for nitrogen and phosphorus losses from food production in China: a county-scale analysis. Environ. Sci. Technol. 52 (10), 5782–5791. https://doi.org/10.1021/acs.est.7b06138.

WorldPop, 2020. The spatial distribution of population density in 2020 based on country total adjusted to match the corresponding UNPD estimate. https://dx.doi.org/10.5258/SOTON/WP00675 (accessed by January 2024).

World Health Organization and United Nations Children's Fund (WHO/UNICEF), 2023. Joint monitoring programme for water supply, sanitation and hygiene: estimates on the use of water, sanitation and hygiene in China. https://washdata.org/data/downloads#CHN (accessed by January 2024).

World Health Organization and United Nations Children's Fund (WHO/UNICEF), 2018. Core questions on water, sanitation and hygiene for household surveys: 2018 update. https://washdata.org/sites/default/files/documents/reports/2019-03/JMP-2018-core-questions-for-household-surveys.pdf (accessed by January 2024).

World Health Organization (WHO), 2024. WHO bacterial priority pathogens list, 2024: bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. https://www.who.int/publications/i/item/9789240093461 (accessed by September 2024).

World Health Organization (WHO), 2022. Guidelines for drinking-water quality: fourth edition incorporating the first and second addenda. https://www.who.int/publications/i/item/9789240045064 (accessed by August 2023).

Xiao, G., Wang, Z., Chen, J.a., Qiu, Z., Li, Y., Qi, J., Liu, W., Shu, W., 2013. Occurrence and infection risk of waterborne pathogens in Wanzhou watershed of the Three Gorges Reservoir, China. J. Environ. Sci. 25, 1913–1924.

Xu, X., Liu, J., Zhang, S., Li, R., Yan, C., Wu, S., 2018. China Multi-Period Land Use Remote Sensing Monitoring Dataset (CNLUCC). https://doi.org/10.12078/2018070201 (in Chinese).

Xu, Z., Xu, J., Yin, H., Jin, W., Li, H., He, Z., 2019. Urban river pollution control in developing countries. Nat. Sustain. 2 (3), 158–160. https://doi.org/10.1038/s41893-019-0249-7.

Zhao, M.M., Chen, Y.-p., Xue, L.-g., Fan, T.T., Emaneghemi, B., 2018. Greater health risk in wet season than in dry season in the Yellow River of the Lanzhou region. Sci. Total Environ. 644, 873–883. https://doi.org/10.1016/j.scitotenv.2018.07.006.

Zhu, Y., Jiao, X., Meng, W., Yu, X., Cheng, H., Shen, G., et al., 2023. Drinking water in rural China: water sources, treatment, and boiling energy. Environ. Sci. Technol. 57 (16), 6465–6473. https://doi.org/10.1021/acs.est.2c09344.