UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Assessing the risk of E. coli contamination from manure application in Chinese farmland by integrating machine learning and phydrus

Fengxian Chen, Bin Zhou, Liqiong Yang, Jie Zhuang, Xijuan Chen

Angaben zur Veröffentlichung / Publication details:

Chen, Fengxian, Bin Zhou, Liqiong Yang, Jie Zhuang, and Xijuan Chen. 2024. "Assessing the risk of E. coli contamination from manure application in Chinese farmland by integrating machine learning and phydrus." *Environmental Pollution* 356: 124345. https://doi.org/10.1016/j.envpol.2024.124345.

Nutzungsbedingungen / Terms of use:

CC BY-NC-ND 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
CC-BY-NC-ND 4.0: Creative Commons: Namensnennung - Nicht kommerziell - Keine Bearbeitung
Weitere Informationen finden Sie unter: / For more information see:
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de

Assessing the risk of *E. coli* contamination from manure application in Chinese farmland by integrating machine learning and Phydrus☆

Fengxian Chen [a], Bin Zhou [b], Liqiong Yang [a], Jie Zhuang [c], Xijuan Chen [d,*]

[a] *Key Laboratory of Pollution Ecology and Environmental Engineering, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China*
[b] *Chair of model-based environmental exposure science, Faculty of Medicine, University of Augsburg, Augsburg 86159, Germany*
[c] *Department of Biosystems Engineering and Soil Science, Institute for a Secure and Sustainable Environment, The University of Tennessee, Knoxville, TN 37996, United States*
[d] *Sino-Spain Joint Laboratory for Agricultural Environment Emerging Contaminants of Zhejiang Province, College of Environmental and Resource Sciences, Zhejiang Agriculture and Forestry University, Hangzhou 311300, China*

A B S T R A C T

This study aims to present a comprehensive study on the risks associated with the residual presence and transport of *Escherichia coli* (*E. coli*) in soil following the application of livestock manure in Chinese farmlands by integrating machine learning algorithms with mechanism-based models (Phydrus). We initially review 28 published papers to gather data on *E. coli*'s die-off and attachment characteristics in soil. Machine learning models, including deep learning and gradient boosting machine, are employed to predict key parameters such as the die-off rate of *E. coli* and first-order attachment coefficient in soil. Then, Phydrus was used to simulate *E. coli* transport and survival in 23692 subregions in China. The model considered regional differences in *E. coli* residual risk and transport, influenced by soil properties, soil depths, precipitation, seasonal variations, and regional disparities. The findings indicate higher residual risks in regions such as the Northeast China, Eastern Qinghai-Tibet Plateau, and pronounced transport risks in the fringe of the Sichuan Basin fringe, the Loess Plateau, the North China Plain, the Northeast Plain, the Shigatse Basin, and the Shangri-La region. The study also demonstrates a significant reduction in both residual and transport risks one month after manure application, highlighting the importance of timing manure application and implementing region-specific standards. This research contributes to the broader understanding of pathogen behavior in agricultural soils and offers practical guidelines for managing the risks associated with manure use. This study's comprehensive method offers a potentially valuable tool for evaluating microbial contaminants in agricultural soils across the globe.

1. Introduction

China generates approximately 550 million tons of livestock manure annually, of which about one-third is suitable for composting and use in agriculture (Jia et al., 2018; Bao et al., 2019). Although composting effectively eliminates most pathogens, small quantities may persist and can be introduced into the soil during the fertilization (Alegbeleye and Sant'Ana, 2020; Black et al., 2021; Lepesteur, 2022). The pathogenic microorganisms found in manure predominantly include bacteria (e.g., *Escherichia coli, Salmonella, Leptospira,* and *Yersinia*), protozoa (e.g., *Cryptosporidium* and *Giardia*), and viruses (e.g., *Rotavirus*) (Abdugheni et al., 2023; Matthews, 2023). Notably, *E. coli* is commonly used as an indicator for fecal contamination due to its prevalence in manure (Khan and Gupta, 2020). While most strains of *E. coli* are benign, pathogenic variants do exist, such as enterohemorrhagic, enteroinvasive, and enterotoxin-producing *E. coli* (Gyles, 1992; Williams et al., 2010). Instances of public health issues caused by pathogenic *E. coli* strains were documented in many literatures (Hunter, 2003; Majowicz et al., 2014; Rani et al., 2021). The residual presence of pathogenic *E. coli* in surface soil poses a direct threat of contamination to fruits and vegetables and is easily carried by runoff into rivers and lakes (Balali et al., 2020; Willis et al., 2020; Morgado et al., 2022). Moreover, the presence of pathogenic *E. coli* in deep soil may transport deeper and threaten groundwater and drinking water (Khan and Gupta, 2020; Chique et al., 2021).

Extensive research has been conducted to evaluate the accumulation and transport risks of *E. coli*, using methods such as transport

☆ This paper has been recommended for acceptance by Dr. Sarah Harmon.
* Corresponding author.
E-mail address: chenxj@iae.ac.cn (X. Chen).

experiments, incubation experiments, field experiments, and mathematical modeling (Foppen and Schijven, 2006; Tufenkji, 2007; Bradford et al., 2014; Engström et al., 2015; Jang et al., 2017). These approaches have provided insights into the *E. coli* behaviors under various soil conditions and environmental factors, including soil texture, organic matter content, ionic strength, pH, moisture, temperature, water flow, soil heterogeneity, etc. (Bradford et al., 2013; Zhong et al., 2017). Common mechanism-based models are usually based on the traditional advection-diffusion equation (ADE), such as attachment/detachment model and filtration theory (Harvey and Garabedian, 1991; Johnson et al., 1995; Bradford et al., 2014). Although these studies have elucidated the mechanisms influencing *E. coli* survival and transport, the parameters derived from experiments and modeling (e.g., die-off rate, attachment coefficient, and detachment coefficient) are often region-specific and lack broad applicability (Balkhair et al., 2017; Wen et al., 2017; Baker et al., 2020; Guo et al., 2020).

Machine learning models offer the potential to broaden the application of the key parameters influencing the residual and transport of *E coli*, including first-order attachment coefficient, spatial removal rate, adsorption efficiency, and die-off rate of *E. coli* in soil (Park et al., 2016; Pang et al., 2020; Chen et al., 2023; Gomez-Flores et al., 2023). Common machine learning models include deep learning, random forest, gradient boosting machine, generalized linear model, and so on (Mahesh, 2020). However, these machine learning algorithms and predicted parameters are difficult to use for the management of manure on farmlands, primarily because they fail to consider factors such as region-specific variations in soil layer, temperature, and rainfall (Petersen and Hubbart, 2020; Phan-Thien et al., 2020; Hachad et al., 2022). Moreover, although predicting key parameters using current experimental and monitoring data is possible, the volume of available data remains inadequate for comprehensive modeling of the pathway from manure application to pathogen fate, owing to the extensive data demands of machine learning techniques (Tahmasebi et al., 2020; Zhong et al., 2021). Overall, machine learning models can enhance the generalizability of key parameters; while mechanism models, grounded in principles like the advection-diffusion equation, provide crucial insights into the fundamental processes influencing *E. coli* transport and survival.

This study aims to synergize machine learning and mechanism-based models to create an integrated modeling framework, enhancing understanding and management of the risks associated with *E. coli* residual and transport following manure application. The sub-objectives include: 1) predicting the attachment and die-off parameters of *E. coli* in soil; and 2) evaluating the behavior of *E. coli* in the surface soils of Chinese farmlands, along with the risk of its transport to deeper soil layers.

2. Materials and methods

2.1. Theory

The advection-dispersion equation describing the transport and survival of *E. coli* under one-dimensional uniform water flow can be expressed as:

$$\frac{\partial \theta C}{\partial t} + \rho_b \frac{\partial S}{\partial t} = \frac{\partial}{\partial x}\left(\theta D \frac{\partial \mathbb{C}}{\partial x}\right) - \frac{\partial q \mathbb{C}}{\partial x} - \mu_w \theta C - \mu_s \rho_b S \quad (1)$$

where C is the bacterial concentration (cell L^{-3}) in the aqueous phase, t is time (T^{-1}), D is the dispersion coefficient ($L^2\ T^{-1}$), θ is the volumetric water content, ρ is the soil bulk density (M L^{-3}), q is the flow rate (L T^{-1}); x is the spatial coordinates (L); S is the solid-phase concentrations associated with adsorption site (M M^{-1}); μ_s and μ_w is the die-off rate of bacteria (T^{-1}) in solid and liquid phase. L, M, and T represent the dimensions of length, mass, and time. It is worth noting that we cannot distinguish whether the bacterial die-off occurred in the liquid or solid phase. Therefore, we only used one parameter μ to represent die-off in both liquid and solid phases.

The die-off of *E. coli* is obtained either from field experiments or indoor incubation experiments, and it is approximately described using first-order kinetic model (Crane and Moore, 1986):

$$C_1 = C_0 e^{-\mu t} \quad (2)$$

where C_0 is the initial *E. coli* concentration (M L^{-3}), C_1 is the *E. coli* concentration (M L^{-3}) after a period, t is time (T^{-1}), and μ is a die-off rate (T^{-1}).

The attachment/detachment model was used to describe mass transfer between the aqueous and solid phase as:

$$\rho_b \frac{\partial S}{\partial t} = \theta \psi_t k_a C - k_d \rho_b S \quad (3)$$

where k_a is the first-order attachment coefficient (T^{-1}); k_d is the first-order detachment coefficient (T^{-1}). Previous studies showed that the detachment of bacteria in soil was much smaller than attachment of bacteria (Firouzi et al., 2015; Gharabaghi et al., 2015; Balkhair, 2017). Therefore, the first-order detachment coefficient in this study was assumed as 0, equivalent to filtration theory. k_a can be obtained from literatures according to the method in Chen et al. (2023).

For a given site, the parameter C, t, x, q in Equation (1) (bacterial concentration (C), time (t), spatial coordinate (x), and flow rate (q)) can be set according to the manure application amount and precipitation conditions; the parameters of soil property (i.e., bulk density (BD), Van Genuchten model parameters (VG), saturated hydraulic conductivity (K_s)) can be set according to soil survey data of Chinese agricultural soils; the dispersion coefficient or dispersivity in Chinese agricultural soil was obtained from Chen et al. (2024). If K_a and μ can be predicted by machine learning algorithms, the residual and transport of *E. coli* can be simulated.

2.2. Data collection and analysis

A comprehensive literature review was conducted using the Web of Science database to gather relevant data on the survival (die-off, 111 datasets) and attachment characteristics (58 datasets) of *E. coli* in soil. This collected data was then utilized to develop machine learning models. For modeling *E. coli* die-off, seven key factors were identified and used as input variables based on their importance and frequency in previous studies. These factors include pH, soil organic matter (SOM) content, temperature, soil water content, and the proportions of sand, silt, and clay in the soil. In terms of *E. coli* attachment, five principal factors were selected as input variables, guided by their prevalence in existing studies. These factors were transport scale, pore water velocity, SOM content, saturated hydraulic conductivity (Ks), and ionic strength of pore water (IS). The collected publications are listed in Table S1. The process and methodology of integrating these models to simulate the transport and survival of *E. coli* in Chinese agricultural soils are depicted in Fig. 1.

2.3. Machine learning algorithms

We employed the H2O AutoML, an automated machine learning framework developed by LeDell and Poirier (2020), which consolidates a range of prevalent machine learning algorithms including Deep Learning, Distributed Random Forest (DRF), Generalized Linear Model (GLM), Gradient Boosting Machine (GBM), Extremely Randomized Trees (XRT), and Extreme Gradient Boosting (XGBoost). GBM and XGBoost are categorized as gradient boosting algorithms, while DRF and XRT are based on random forest methodologies. Deep Learning utilizes artificial neural networks, and GLM represents a flexible generalization of ordinary linear regression (Mahesh, 2020). H2O AutoML automates the process of training these various algorithms (LeDell and Poirier, 2020).

To ascertain the robustness and reliability of our model's

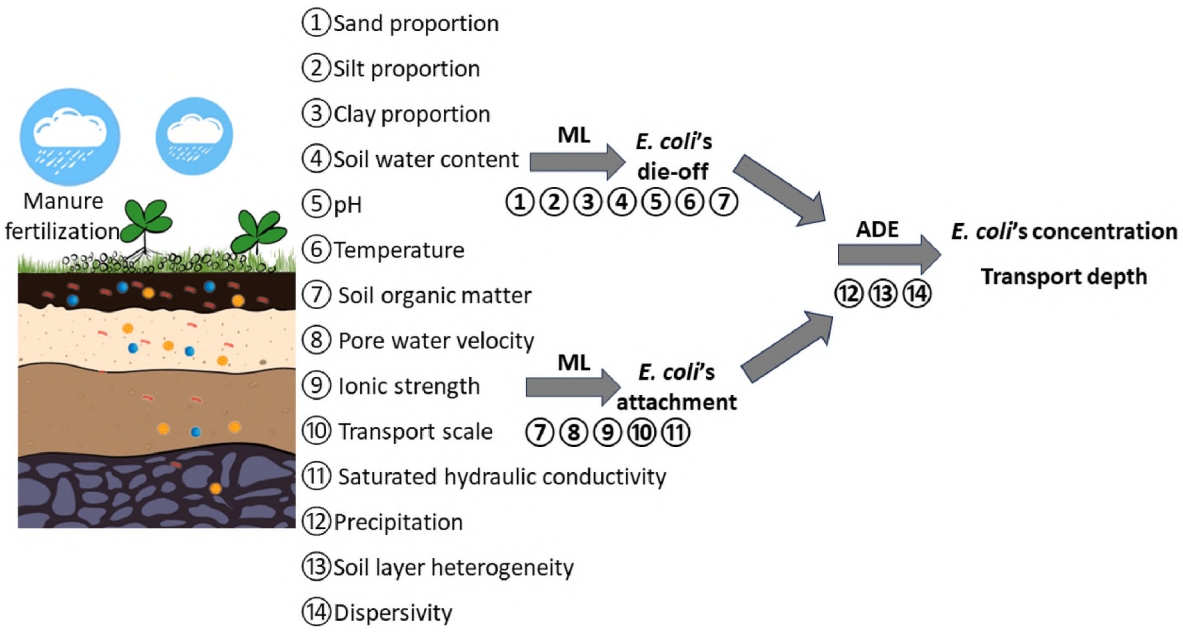


Fig. 1. The flowchart of the integrated modeling for transport and survive of *E. coli* in Chinese agricultural soils (ML: machine learning; ADE: Advection-dispersion equation).

performance, we executed a 5-fold cross-validation. The data was randomly divided into five distinct subsets. In each cross-validation cycle, the models were trained on four subsets (80% of the data) and evaluated on the remaining one (20% of the data). This process was repeated until each subset had been used once as the test set. The efficacy of the models was determined using metrics like root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), root mean squared logarithmic error (RMLSE), mean residual deviance (MRD), the slope of linear regression, and the coefficient of determination (R^2) in linear regression, as detailed in Table S2 and Table S3.

Within H2O AutoML, twenty different machine learning algorithms were simultaneously examined, and their performances were ranked on a leaderboard. For clarity, we presented only the results from the top-performing model, which was also employed for predicting the target variables. Variable importance scores range from 0 to 100%, indicating the relative impact of each input variable on the target variable.

2.4. Modeling with the mechanism-based model

In this study, we employed Phydrus (Collenteur et al., 2020), a Python adaptation of the Hydrus-1D (Simunek et al., 2005) to simulate the transport and survival of *E. coli* in soil environments. To effectively account for the soil layer heterogeneity, we divided the soil profile, extending from the surface down to 200 cm, into six distinct layers. These layers are delineated as follows: surface to 5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm. The maps used in this study, which detailed various Chinese geographical features including soil properties, precipitation, and temperature, offered a fine spatial resolution, approximately 20 km by 20 km (totally 23692 subregions). Detailed information on the data sources (Nachtergaele et al., 2009; Liu et al., 2020; Meng et al., 2021; Lembrechts et al., 2022; Liu et al., 2022; Chen et al., 2024) and the parameters used for the modeling process are outlined in Table 1.

In simulating water flow using Phydrus, the initial and boundary conditions were set as follows: 1) Upper boundary condition-established as an atmospheric boundary, allowing for the exchange of moisture through precipitation input; rainfall occurred weekly, with each event lasting for 24 h; This periodic rainfall simulation mimics the natural occurrence of weekly precipitation. 2) Lower boundary condition-fee drainage. 3) Initial Condition-at the start of the simulation, the initial moisture state of the soil was specified based on an assumed saturated state. Through these settings, the Phydrus model can simulate the dynamics of moisture in soil, taking into account the effects of rainfall. Rainfall occurs weekly, with each event lasting 24 h. The volume of precipitation for each event is determined based on China's monthly precipitation data (Peng et al., 2019). The lower boundary condition was defined as free drainage. The initial condition of *E. coli* concentration in soil was set to zero. The upper boundary condition of *E. coli* concentration was specified as a flux concentration, whereas the lower boundary was characterized by a zero gradient.

We assumed that the concentration of *E. coli* in manure at 10^5 Colony Forming Units (CFU) per gram of dry manure according to literature data (Blaustein et al., 2015). It was hypothesized that the release of

Table 1
The data sources for the modeling in machine leaning models and Phydrus.

Data	Description	Source
IS	Ionic strength of pore water	Dolling and Ritchie (1985); Nachtergaele et al. (2009)
Clay, Silt, Sand, SOM, pH, BD	Soil properties	Liu et al. (2020); Liu et al. (2022)
Ks, VGMP	Calculation from soil properties	Schaap et al. (2001)
Precipitation	Monthly precipitation	Peng et al. (2019)
SWC	Soil volumetric water content	Meng et al. (2021)
λ	Dispersivity	Chen et al. (2024)
Temperature	Soil temperature	Lembrechts et al. (2022)
Velocity	Equivalent to Ks under saturated condition	
Time	Summer: April to September Winter: October to March	
Scale	5, 10, 15, 30, 60, 100 cm	
K_a	Prediction by machine learning models	
μ **in Summer**	Prediction by machine learning models	
μ **in Winter**	Prediction by machine learning models	

E. coli from the manure coincided with the first instance of rainfall following the manure's application. Manure application was scheduled for two distinct seasons: in spring, before sowing at the end of March, and in autumn, post-harvest at the end of September. We defined the period from April to September as the Summer season and from October to March as the Winter season. The rate of manure application was standardized at 30 tons per hectare (Araji et al., 2001; Moshatati et al., 2019).

3. Results and discussion

3.1. Die-off of E. coli

Fig. 2(a) presents a box plot illustrating the distribution of the collected dataset for the die-off of *E. coli*. This dataset encompasses a wide range of input variables. Specifically, the content of clay, silt, and sand varied from 0.7 to 60%, 0.2–76.4%, and 3.7–99.1%, respectively. Temperature measurements spanned from 4 to 37 °C, while pH levels were observed between 3.92 and 8.74. Soil organic matter (SOM) showed variation from 0.99 to 110 g kg^{-1}, and soil volumetric water content (SWC) ranged from 0.0255 to 0.46. The target variable, which is the die-off rate (μ) of *E. coli*, was found to vary from 0.018 to 6.908 day^{-1}. This comprehensive dataset effectively represents the diverse conditions encountered in various field experiments concerning the die-off of *E. coli*.

Pearson correlation coefficient (r) measured the extend of a linear correlation between pairwise variables (Fig. 2(b)). The correlation coefficient between the die-off rate (μ) and pH was 0.49, signifying a moderate negative correlation. For μ and the other six input variables, the r values fluctuated between −0.17 and 0.17, indicating weak correlations. The Pearson correlations were medium to weak, indicating that none of the seven input variables independently offer a strong prediction of the die-off rate.

The correlation between the predicted and observed values for the test datasets based on the best-performing model - deep learning algorithm is shown in Fig. 2(c). The linear fitting of this correlation yielded a slope of 0.85, and the coefficient of determination (R^2) for the linear regression was 0.68. These values indicate a strong agreement between the model's predictions and the actual data.

Additionally, the importance of various variables used in the machine learning model indicates how "important" the variable is to the model (Fig. 2(d)). Among these variables, pH showed the highest importance in the predictions of μ. Numerous studies have investigated the impact of pH on the survival of *E. coli* in soil (Zhang et al., 2013; Liang et al., 2019; Xing et al., 2019). Consistent with the negative correlation in Fig. 2(b), it has been demonstrated that lower pH levels are associated with increased die-off rates of *E. coli*. SOM was second most important variable. However, research on the influence of SOM on *E. coli* survival has yielded varied findings. Although some studies suggest that higher levels of SOM tend to favor the survival of *E. coli* (Ma et al., 2011; Zhang et al., 2013), other research indicates that SOM levels do not significantly affect *E. coli*'s survival (Unc and Goss, 2003; Williams et al., 2015). A possible explanation could be that the survival of *E. coli* is influenced not only by the overall quantity of SOM but also by its immediate availability (Franz et al., 2008a,b). Temperature ranks as the third most significant variable in predicting the die-off rate of *E. coli*. Previous literature has consistently indicated that cooler temperatures are more conducive to the persistence of *E. coli* in environmental settings (Park et al., 2016; Underthun et al., 2018). In addition, *E. coli* survival rates were found to be greater in environments with a higher concentration of sand and lower levels of silt and clay (Guy and Visser, 1979; Cools et al., 2001). This trend could be attributed to the relatively greater availability of oxygen and reduced competition from indigenous microorganisms. Furthermore, elevated SWC has been linked to higher *E. coli* die-off rates, potentially due to the lower oxygen availability and intensified competition from native microorganisms (Lang and Smith, 2007; Rothrock et al., 2012).

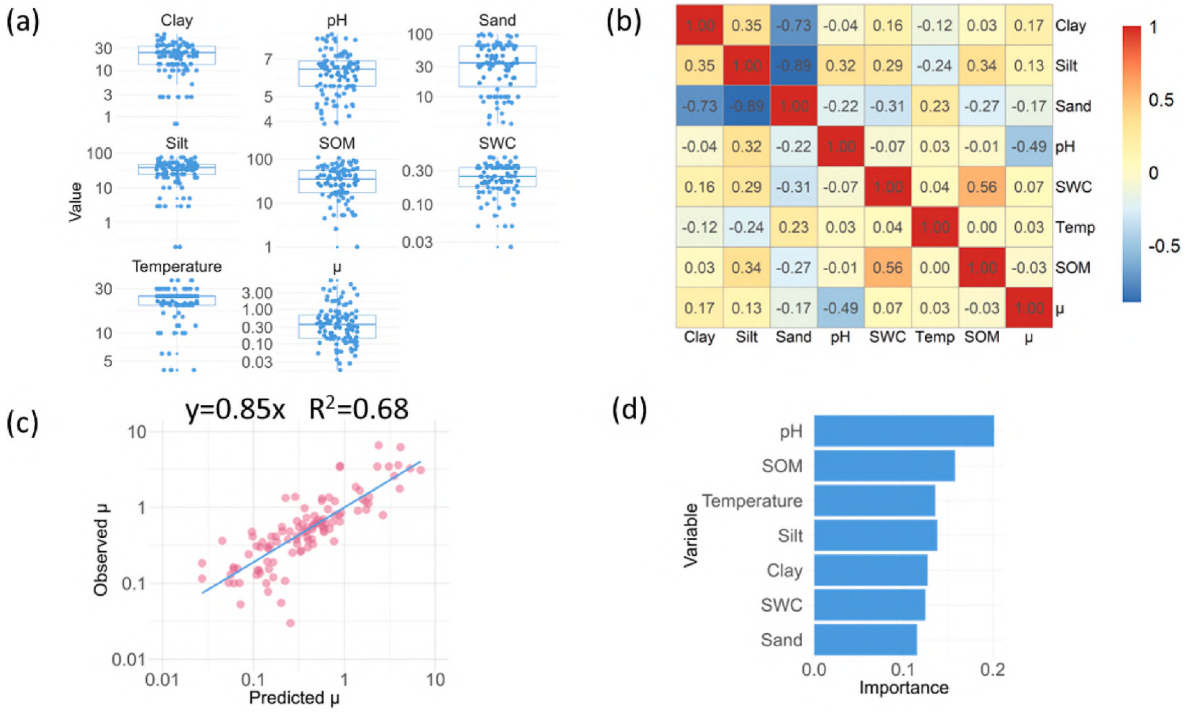


Fig. 2. (a) Collected datasets: clay, sand, and silt content (%),pH, soil organic matter (SOM, g kg^{-1}) content, soil water content (SWC, cm^3 cm^{-3}), temperature (°C) and die-off rate of *E. coli* (μ, day^{-1}); (b) Pearson correlation coefficient (r) between pairwise variables; (c) The correlation between predicted values and observed values for the test datasets; (d) The importance of variables.

3.2. Attachment of E. coli

Fig. 3(a) shows the box plot of the collected dataset for the attachment of *E. coli*. For the input variables, IS ranged from 0.0055 to 30 mM; the pore water velocity ranged from 2.26 to 936 cm day^{-1}; the transport scale ranged from 5 to 50 cm; the saturated hydraulic conductivity (Ks) varied from 1.81 to 413 cm day^{-1}; the SOM ranged from 1.03 to 60.9 g kg^{-1}. Regarding the target variable, the first-order attachment coefficient (K_a) ranged from 0.32 to 102 day^{-1}. Overall, the collected datasets covered a variety of soil types and water flow conditions.

The Pearson correlation coefficient (r) of K_a with other five input variables ranged from −0.16 to 0.71 (Fig. 3(b)). The Pearson correlations imply that pore water velocity had a strong positive correlation with K_a; Ks had a moderate positive correlation with K_a; the other three input variables had weak correlation with K_a.

The correlation between predicted values and observed values for the test datasets was plotted in Fig. 3(c), which is based on the best-performing model - GBM algorithm. The slope of linear fitting was 0.91, and the R^2 of linear regression was 0.81, indicating the model's predictions are very close to the actual data.

Regarding the importance of variables, pore water velocity showed the highest importance in the predictions of K_a (Fig. 3(d)), consistent with our previous study (Chen et al., 2023). The Pearson correlation coefficient indicates a positive correlation between pore water velocity and the K_a, suggesting that as pore water velocity increases, so does K_a. IS showed the second importance. It is widely accepted that the influence of IS on bacterial attachment aligns with the DLVO theory (Verwey, 1947; Derjaguin and Landau, 1993). According to this theory, an increase in IS compresses the electric double layer, potentially leading to enhanced bacterial adsorption. Nonetheless, higher IS might also have limited or negatively impact when coupling with factors such as SOM, soil pH and the presence of microbial extracellular polymers (Kim et al., 2009; Jimenez-Sanchez et al., 2015; Jin et al., 2021). Consequently, while IS is a significant input variable, its impact on bacterial adsorption is not unequivocal. This notion is further corroborated by the weak Pearson correlation coefficient between IS and K_a, as illustrated in Fig. 3 (b). SOM was the third most significant variable, and the effect of SOM on K_a was consistent with previous studies (Johnson and Logan, 1996; Jimenez-Sanchez et al., 2015). Saturated hydraulic conductivity (Ks), as a comprehensive indicator of soil texture, was the fourth important variable. Fig. 3(b) demonstrates that in natural soil, there is typically a strong positive correlation between Ks and pore water velocity. Consequently, the influence of Ks on K_a appears to be similar to the effect of pore water velocity. Transport scale was the fifth important variable and showed week negative correlation with K_a, consistent with the result of a recent study (Oudega et al., 2021).

3.3. Die-off rate and attachment coefficient of E coli in Chinese soil

The die-off rate of *E. coli* across various soil depths during different seasons shows obvious difference (Fig. 4(a)-(d)). In addition to the data presented in Fig. 4, additional data at other soil depth can be found in the supplementary materials. Notably, the die-off rate in the southeastern region and subtropical areas of Yunnan exceeds that in other parts of China. This pattern is likely influenced predominantly by soil pH (Chen et al., 2019; Wang et al., 2019). The impact of soil depth on *E. coli* die-off rate, as shown in Fig. 4(a) and (b) compared to Fig. 4(c) and (d), appears negligible. Seasonal variations, however, do affect die-off rate: Fig. 4(a) and (c) suggest lower die-off rate in winter compared to in summer, as seen in Fig. 4(b) and (d), aligning with previously noted temperature influences on die-off rate (Park et al., 2016; Underthun et al., 2018).

The first-order attachment coefficient of *E. coli* at varying soil depths shows obvious difference as well (Fig. 4(e) and (f)). The eastern section of Northeast China, the Tarim and Junggar Basins in Xinjiang, and the eastern Tibetan Plateau exhibit higher attachment coefficient. These higher attachment coefficients are likely due to a combination of factors including pore water velocity, ionic strength, soil organic matter, soil texture, and other factors. When comparing different soil depths, the attachment coefficient diminishes in deeper soil, indicating easier

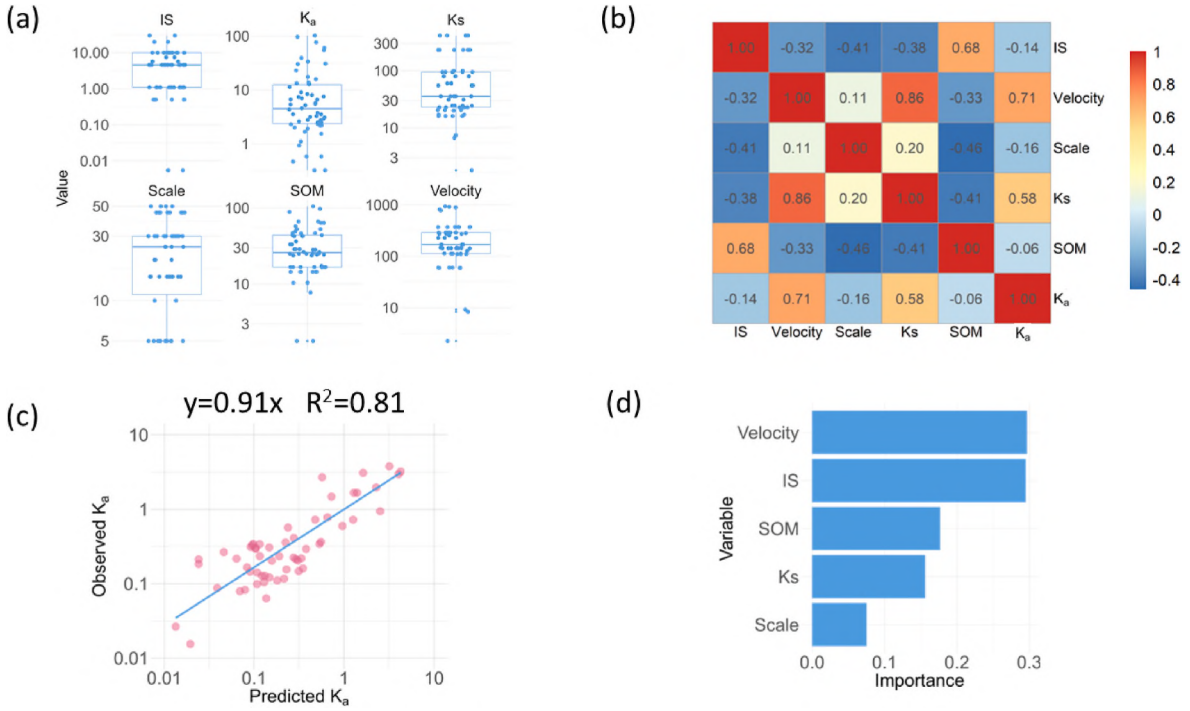


Fig. 3. (a) Collected datasets: ionic strength (IS, mM), saturated hydraulic conductivity (Ks, cm day^{-1}), transport scale (cm), soil organic matter (SOM, g kg^{-1}) content, pore water velocity (cm day^{-1}) and first-order attachment coefficient of *E. coli* (K_a, day^{-1}); (b) Pearson correlation coefficient (r) between pairwise variables; (c) The correlation between predicted values and observed values for the test datasets; (d) The importance of variables.

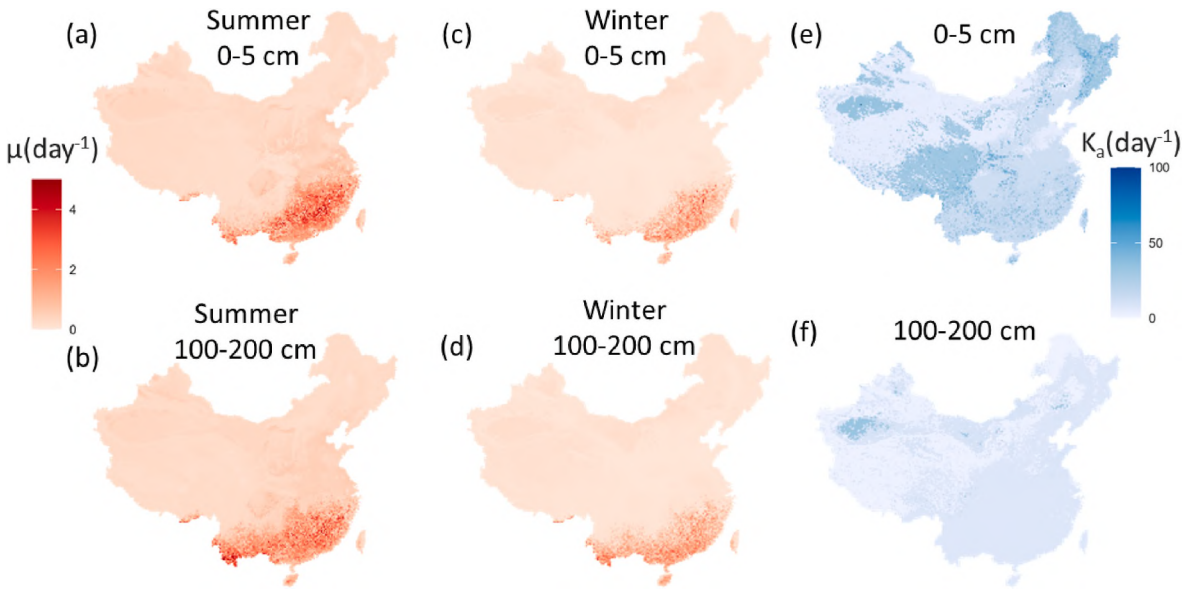


Fig. 4. Predicted die-off rate of *E. coli* (μ, day^{-1}) in summer (April to September) at soil depth of (a) 0–5 cm, (b) 100–200 cm; predicted die-off rate of *E. coli* (μ, day^{-1}) in winter (October to March) at soil depth of (c) 0–5 cm, (d) 100–200 cm; predicted first-order attachment coefficient of *E. coli* (K_a, day^{-1}) at soil depth of (e) 0–5 cm, (f) 100–200 cm (maps for other depths are in the Supplementary Material Figs. S1, S2, and S3).

bacterial transport, consistent with a recent study (Chen et al., 2021).

3.4. Residual of E. coli concentrations in topsoil

The concentration variations of *E. coli* in the topsoil layer (0–5 cm) in April and October demonstrates the Residual of *E. coli* (Fig. 5). In general, the highest concentrations are found in the Northeast, Eastern Qinghai-Tibet Plateau, followed by North China, with the Southeast region exhibiting the lowest concentrations. One week after manure application, *E. coli* concentrations exceeded 1000 CFU g^{-1} in most areas, excluding the Southeast. After two weeks, concentrations fell below 1000 CFU g^{-1} in most regions nationwide, except for some parts of the Northeast. By the four-week mark, concentrations in most areas declined to below 100 CFU g^{-1}. The concentration 100 CFU g^{-1} was usually as detection limit of *E.coli* in soil (Franz et al., 2008a,b; Zhang et al., 2013). Previous studies have confirmed a rapid decrease in *E. coli* concentrations in soil from Eastern China (Zhang et al., 2013; Wang et al., 2014). When comparing the concentrations of *E. coli* in summer and winter, it is observed that winter concentrations are marginally higher than those in summer. This difference can be attributed to two main reasons: higher die-off rates of *E. coli* during summer and the increased rainfall in summer compared to winter (Williams et al., 2005; Sharma et al., 2019). The most pronounced seasonal difference is observed at the junction between the Qinghai-Tibet Plateau and the Tarim Basin, where the risk of *E. coli* residue in winter significantly surpasses that in summer. This phenomenon could be due to the lower die-off rate of *E. coli* in this area

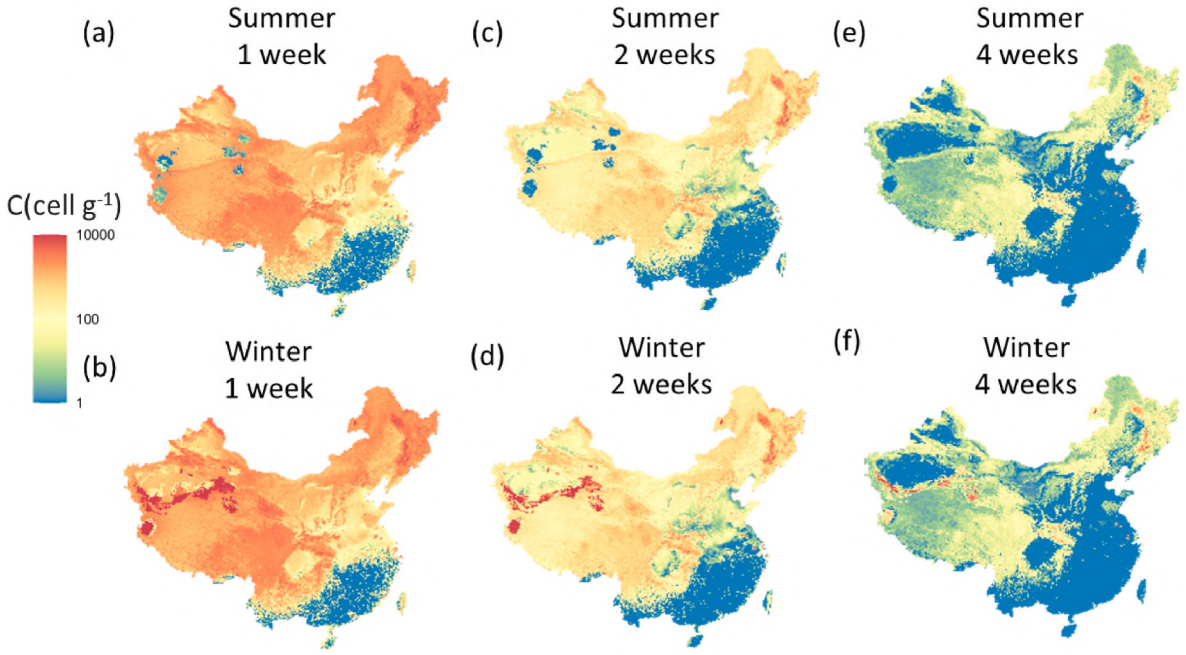


Fig. 5. *E. coli* concentrations (cell g^{-1}) of topsoil (0–5 cm) (a) one week, (c) two weeks, and (e) four weeks after manure application in summer (April); *E. coli* concentrations (cell g^{-1}) of topsoil (0–5 cm) (b) one week, (d) two weeks, and (f) four weeks after manure application in winter (October).

during winter, coupled with minimal precipitation, which results in a lack of the diluting effect brought about normally by rainfall (Kraay et al., 2020).

3.5. Transport risk of E. coli

In this study, the transport depth of *E. coli* is defined as the maximum depth where its concentration exceeds 10 CFU g^{-1}, representing a 99.99% reduction from the initial concentration of 10^5 CFU g^{-1}. Fig. 6 illustrates the transport depth of *E. coli* at various intervals following manure application in both summer and winter. Overall, the transport risk of *E. coli* is higher in summer compared to winter. Obviously, this phenomenon is related to more rainfall in summer (Brennan et al., 2010). In summer, the highest transport risks are observed at the fringe of the Sichuan Basin, Eastern Qinghai-Tibet Plateau, the Loess Plateau, the Shigatse Basin, and the Shangri-La region. In winter, the transport risks are most pronounced in the Loess Plateau, the fringe of the Sichuan Basin, the North China Plain, the Northeast Plain, the Shigatse Basin, and the Shangri-La region. The southeast shows minimal transport risk. A comparison of Figs. 5 and 6 reveals that most regions with greater *E. coli* transport risks often coincide with those having higher residual concentrations of *E. coli* in the topsoil. Despite the Southeast experiencing higher rainfall than other regions, the transport depth there remains limited due to the lower *E. coli* concentrations in the topsoil. Furthermore, some areas like the North China Plain exhibit a moderate residual concentration of *E. coli* in the topsoil. However, due to the low attachment coefficient, the transport risk in these regions is also heightened. McLeod et al. (2019) performed both laboratory-based column experiments and extended leaching experiments under field conditions. Their work focused on assessing the risk of *E. coli* leaching through soil in flat to rolling terrains in New Zealand. This assessment was based on various characteristics of bacterial transport, such as bacterial removal rates and microbial breakthrough curves. Our study not only considered bacterial attachment characteristics but also demonstrated that the survival of *E. coli* significantly influences the transport risk.

A week after manure application, the transport depth in high-risk areas generally ranged between 10 and 20 cm, whereas in the southeastern region, it remained below 5 cm. After two weeks, the transport risk diminishes across all areas, with only some regions exhibiting a transport depth exceeding 10 cm. By the four-week mark, most areas show a transport depth of less than 5 cm. The decline in *E. coli* concentrations in topsoil over time (Fig. 5) corresponds with a reduced risk of its transport (Fig. 6). This phenomenon further reinforces the significant impact of *E. coli* residual concentration on the risk of transport.

Research from the United States emphasized the significant roles of livestock density and waterfowl transport in influencing *E. coli* concentrations in water systems, which resonates with the biotic factors influencing our observed transport risks (Hansen et al., 2020). Additionally, the use of advanced predictive models like ANFIS in another study to enhance the accuracy of *E. coli* predictions highlights the potential of integrating similar methodologies in our research (Abimbola et al., 2020). These comparisons suggest that effective livestock management and predictive modeling are crucial for developing strategies to mitigate *E. coli* risks in both agricultural and environmental settings.

3.6. Suggestions, limitations, and future studies

Based on the residual concentration and transport depth of *E. coli* in Chinese soil, the following recommendations are proposed: 1) develop a manure fertilization strategy: specifically, it is advisable to apply manure during periods of low rainfall, ensuring a minimum interval of one month between the application of manure and the onset of the rainy season; 2) adopt region-specific manure residue standards: implement stricter microbial residue regulations in high-risk areas to better control microbial contamination.

Additionally, the conclusions drawn from this study on E. coli can inform the management of other soil-borne pathogens. The methodologies and findings, especially regarding the influence of soil properties, environmental conditions, and temporal factors on pathogen behavior, can guide targeted strategies to mitigate the risks associated with other pathogens. For instance, understanding the effects of soil texture and moisture on *E. coli* survival can help predict similar risks for pathogens like *Salmonella* or *Listeria*, which may behave differently under the same environmental conditions but still follow similar mechanistic pathways

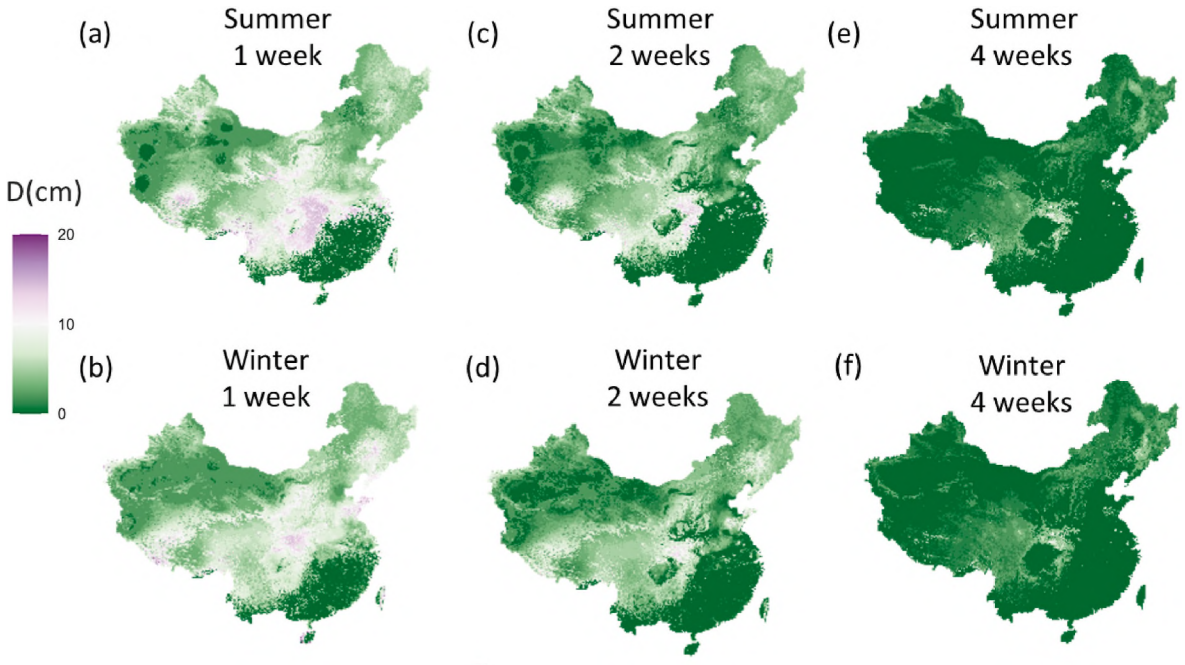


Fig. 6. *E. coli* transport depth (cm) (a) one week, (c) two weeks, and (e) four weeks after manure application in summer (April); *E. coli* transport depth (cm) (b) one week, (d) two weeks, and (f) four weeks after manure application in winter (October).

(Strawn et al., 2013).

This study is primarily limited by two aspects. One is that the omission of preferential flow and large pores in our modeling. Prior research indicates that large pores substantially influence pathogen migration in field soils (Wang et al., 2013; Chen et al., 2021). Future studies could address this limitation by multiplying a soil type-dependent coefficient to describe preferential flow (Gerke, 2006; Allaire et al., 2009; Luo and Schmidt, 2010). The second aspect is the lack of consideration for the detachment process of *E. coli*. This was mainly due to the challenge of extracting detachment parameters from existing literature. To overcome this, future research should focus on conducting experiments to gather large number of parameters related to *E. coli* detachment in various soil types.

4. Conclusions

By integrating the strengths of machine learning algorithms and mechanism-based models, we have developed a comprehensive framework that enhances the predictability and understanding of *E. coli* residual and transport in soil environments.

Our integrated approach highlighted how different soil properties, soil depths, seasonal variations, and regional differences can influence the residual and transport of *E. coli* in Chinese soil. Spatial analysis revealed that regions such as the Northeast China and Eastern Qinghai-Tibet Plateau exhibited higher *E. coli* residual risk. The fringe of the Sichuan Basin, Loess Plateau, the North China Plain, the Northeast Plain, the Shigatse Basin, and the Shangri-La region showed relative high transport risk. One month after applying organic fertilizer, both residual and transport risk were significantly reduced. Overall, this study offers practical guidelines for managing the risks associated with manure application in farmland, contributing to the sustainability and safety of agricultural practices.

Funding

This research was financially supported by the National Natural Science Foundation of China (Grant number: 52379051).

CRediT authorship contribution statement

Fengxian Chen: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Bin Zhou:** Writing – review & editing, Methodology. **Liqiong Yang:** Writing – review & editing. **Jie Zhuang:** Writing – review & editing, Supervision. **Xijuan Chen:** Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgement

Acknowledgement for the data support from “Soil SubCenter, National Earth System Science Data Center, National Science & Technology Infrastructure of China. (http://soil.geodata.cn)".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2024.124345.

References

Abdugheni, R., Li, L., Yang, Z.N., Huang, Y., Fang, B.Z., Shurigin, V., Mohamad, O.A.A., Liu, Y.H., Li, W.J., 2023. Microbial risks caused by livestock excrement: current research status and prospects. Microorganisms 11 (8), 1897.

Abimbola, O.P., Mittelstet, A.R., Messer, T.L., Berry, E.D., Bartelt-Hunt, S.L., Hansen, S. P., 2020. Predicting Escherichia coli loads in cascading dams with machine learning: An integration of hydrometeorology, animal density and grazing pattern. Science of the total environment 722, 137894.

Alegbeleye, O.O., Sant'Ana, A.S., 2020. Manure-borne pathogens as an important source of water contamination: an update on the dynamics of pathogen survival/transport as well as practical risk mitigation strategies. Int. J. Hyg Environ. Health 227, 113524.

Allaire, S.E., Roulier, S., Cessna, A.J., 2009. Quantifying preferential flow in soils: a review of different techniques. J. Hydrol. 378 (1–2), 179–204.

Araji, A.A., Abdo, Z.O., Joyce, P., 2001. Efficient use of animal manure on cropland–economic analysis. Bioresour. Technol. 79 (2), 179–191.

Baker, C.A., Lee, S., De, J., Jeong, K.C., Schneider, K.R., 2020. Survival of Escherichia coli O157 in autoclaved and natural sandy soil mesocosms. PLoS One 15 (6), e0234562.

Balali, G.I., Yar, D.D., Afua Dela, V.G., Adjei-Kusi, P., 2020. Microbial contamination, an increasing threat to the consumption of fresh fruits and vegetables in today's world. International journal of microbiology 2020.

Balkhair, K.S., 2017. Modeling fecal bacteria transport and retention in agricultural and urban soils under saturated and unsaturated flow conditions. Water Res. 110, 313–320.

Bao, W., Yang, Y., Fu, T., Xie, G.H., 2019. Estimation of livestock excrement and its biogas production potential in China. J. Clean. Prod. 229, 1158–1166.

Black, Z., Balta, I., Black, L., Naughton, P.J., Dooley, J.S., Corcionivoschi, N., 2021. The fate of foodborne pathogens in manure treated soil. Front. Microbiol. 12, 781357.

Blaustein, R.A., Pachepsky, Y.A., Shelton, D.R., Hill, R.L., 2015. Release and removal of microorganisms from land-deposited animal waste and animal manures: a review of data and models. J. Environ. Qual. 44 (5), 1338–1354.

Bradford, S.A., Morales, V.L., Zhang, W., Harvey, R.W., Packman, A.I., Mohanram, A., Welty, C., 2013. Transport and fate of microbial pathogens in agricultural settings. Crit. Rev. Environ. Sci. Technol. 43 (8), 775–893.

Bradford, S.A., Wang, Y., Kim, H., Torkzaban, S., Šimůnek, J., 2014. Modeling microorganism transport and survival in the subsurface. J. Environ. Qual. 43 (2), 421–440.

Brennan, F.P., O'Flaherty, V., Kramers, G., Grant, J., Richards, K.G., 2010. Long-term persistence and leaching of Escherichia coli in temperate maritime soils. Appl. Environ. Microbiol. 76 (5), 1449–1455.

Chen, S., Liang, Z., Webster, R., Zhang, G., Zhou, Y., Teng, H., Hu, B., Arrouays, D., Shi, Z., 2019. A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution. Sci. Total Environ. 655, 273–283.

Chen, J., Yang, L., Chen, X., Ripp, S., Radosevich, M., Zhuang, J., 2021. Bacterial mobility facilitated by soil depth and intact structure. Soil Tillage Res. 209, 104911.

Chen, F., Zhou, B., Yang, L., Chen, X., Zhuang, J., 2023. Predicting bacterial transport through saturated porous media using an automated machine learning model. Front. Microbiol. 14, 1152059.

Chen, F., Zhou, B., Yang, L., Zhuang, J., Chen, X., 2024. Annual atrazine residue estimation in Chinese agricultural soils by integrated modeling of machine learning and mechanism-based models. J. Hazard Mater., 134539

Chique, C., Hynds, P., Burke, L.P., Morris, D., Ryan, M.P., O'Dwyer, J., 2021. Contamination of domestic groundwater systems by verotoxigenic Escherichia coli (VTEC), 2003–2019: a global scoping review. Water Res. 188, 116496.

Cools, D., Merckx, R., Vlassak, K., Verhaegen, J., 2001. Survival of E. coli and Enterococcus spp. derived from pig slurry in soils of different texture. Appl. Soil Ecol. 17 (1), 53–62.

Collenteur, R., Vremec, M., Brunetti, G., 2020, May. Interfacing FORTAN Code with Python: an example for the Hydrus-1D model. In: EGU General Assembly Conference Abstracts, 15377.

Crane, B.S., Moore, J.A., 1986. Modeling enteric bacterial die-off: a review. Water Air Soil Pollut. 27, 411–439.

Derjaguin, B., Landau, L., 1993. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. Prog. Surf. Sci. 43 (1–4), 30–59.

Dolling, P.J., Ritchie, G.S.P., 1985. Estimates of soil solution ionic strength and the determination of pH in West Australian soils. Soil Res. 23 (2), 309–314.

Engström, E., Thunvik, R., Kulabako, R., Balfors, B., 2015. Water transport, retention, and survival of Escherichia coli in unsaturated porous media: a comprehensive review of processes, models, and factors. Crit. Rev. Environ. Sci. Technol. 45 (1), 1–100.

Firouzi, A.F., Homaee, M., Klumpp, E., Kasteel, R., Tappe, W., 2015. Bacteria transport and retention in intact calcareous soil columns under saturated flow conditions. J. Hydrol. Hydromechanics 63 (2), 102.

Foppen, J.W.A., Schijven, J.F., 2006. Evaluation of data from the literature on the transport and survival of Escherichia coli and thermotolerant coliforms in aquifers under saturated conditions. Water Res. 40 (3), 401–426.

Franz, E., Semenov, A.V., Termorshuizen, A.J., De Vos, O.J., Bokhorst, J.G., Van Bruggen, A.H., 2008a. Manure-amended soil characteristics affecting the survival of E. coli O157: H7 in 36 Dutch soils. Environ. Microbiol. 10 (2), 313–327.

Franz, E., Semenov, A.V., Termorshuizen, A.J., De Vos, O.J., Bokhorst, J.G., Van Bruggen, A.H., 2008b. Manure-amended soil characteristics affecting the survival of E. coli O157: H7 in 36 Dutch soils. Environ. Microbiol. 10 (2), 313–327.

Gerke, H.H., 2006. Preferential flow descriptions for structured soils. J. Plant Nutr. Soil Sci. 169 (3), 382–400.
Gharabaghi, B., Safadoust, A., Mahboubi, A.A., Mosaddeghi, M.R., Unc, A., Ahrens, B., Sayyad, G., 2015. Temperature effect on the transport of bromide and E. coli NAR in saturated soils. J. Hydrol. 522, 418–427.
Gomez-Flores, A., Bradford, S.A., Cai, L., Urík, M., Kim, H., 2023. Prediction of attachment efficiency using machine learning on a comprehensive database and its validation. Water Res. 229, 119429.
Guo, X., Hu, H., Meng, H., Liu, L., Xu, X., Zhao, T., 2020. Vertical distribution and affecting factors of Escherichia coli over a 0–400 cm soil profile irrigated with sewage effluents in northern China. Ecotoxicol. Environ. Saf. 205, 111357.
Guy, E.M., Visser, T.A., 1979. Adsorption and survival patterns of Escherichia coli and Streptococcus bovis in six New Zealand soils. N. Z. J. Agric. Res. 22 (2), 341–348.
Gyles, C.L., 1992. Escherichia coli cytotoxins and enterotoxins. Can. J. Microbiol. 38 (7), 734–746.
Hachad, M., Yarahmadi, H., Duy, S.V., Sauvé, S., Prévost, M., Dorner, S., 2022. Occurrence and partitioning behavior of E. coli and wastewater micropollutants following rainfall events. Resources, Environment and Sustainability 9, 100067.
Hansen, S., Messer, T., Mittelstet, A., Berry, E.D., Bartelt-Hunt, S., Abimbola, O., 2020. Escherichia coli concentrations in waters of a reservoir system impacted by cattle and migratory waterfowl. Sci. total environ. 705, 135607.
Harvey, R.W., Garabedian, S.P., 1991. Use of colloid filtration theory in modeling movement of bacteria through a contaminated sandy aquifer. Environ. Sci. Technol. 25 (1), 178–185.
Hunter, P.R., 2003. Drinking water and diarrhoeal disease due to Escherichia coli. J. Water Health 1 (2), 65–72.
Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M.N., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. J. Appl. Microbiol. 123 (3), 570–581.
Jia, W., Qin, W., Zhang, Q., Wang, X., Ma, Y., Chen, Q., 2018. Evaluation of crop residues and manure production and their geographical distribution in China. J. Clean. Prod. 188, 954–965.
Jimenez-Sanchez, C., Wick, L.Y., Cantos, M., Ortega-Calvo, J.J., 2015. Impact of dissolved organic matter on bacterial tactic motility, attachment, and transport. Environ. Sci. Technol. 49 (7), 4498–4505.
Jin, C., Zhao, L., Zhao, W., Wang, L., Zhu, S., Xiao, Z., Mo, Y., Zhang, M., Shu, L., Qiu, R., 2021. Transport and retention of free-living amoeba spores in porous media: effects of operational parameters and extracellular polymeric substances. Environ. Sci. Technol. 55 (13), 8709–8720.
Johnson, W.P., Logan, B.E., 1996. Enhanced transport of bacteria in porous media by sediment-phase and aqueous-phase natural organic matter. Water Res. 30 (4), 923–931.
Johnson, W.P., Blue, K.A., Logan, B.E., Arnold, R.G., 1995. Modeling bacterial detachment during transport through porous media as a residence-time-dependent process. Water Resour. Res. 31 (11), 2649–2658.
Khan, F.M., Gupta, R., 2020. Escherichia coli (E. coli) as an indicator of fecal contamination in groundwater: a review. Sustainable Development of Water and Environment: Proceedings of the ICSDWE2020 225–235.
Kim, H.N., Bradford, S.A., Walker, S.L., 2009. Escherichia coli O157: H7 transport in saturated porous media: role of solution chemistry and surface macromolecules. Environ. Sci. Technol. 43 (12), 4340–4347.
Kraay, A.N., Man, O., Levy, M.C., Levy, K., Ionides, E., Eisenberg, J.N., 2020. Understanding the impact of rainfall on diarrhea: testing the concentration-dilution hypothesis using a systematic review and meta-analysis. Environ. Health Perspect. 128 (12), 126001.
Lang, N.L., Smith, S.R., 2007. Influence of soil type, moisture content and biosolids application on the fate of Escherichia coli in agricultural soil under controlled laboratory conditions. J. Appl. Microbiol. 103 (6), 2122–2131.
LeDell, E., Poirier, S., 2020. H2o automl: scalable automatic machine learning. In: Proceedings of the AutoML Workshop at ICML, vol. 2020. ICML.
Lembrechts, J.J., van den Hoogen, J., Aalto, J., Ashcroft, M.B., De Frenne, P., Kemppinen, J., Kopecký, M., Luoto, M., Maclean, I.M., Crowther, T.W., Bailey, J.J., 2022. Global maps of soil temperature. Global Change Biol. 28 (9), 3110–3144.
Lepesteur, M., 2022. Human and livestock pathogens and their control during composting. Crit. Rev. Environ. Sci. Technol. 52 (10), 1639–1683.
Liang, C., Yao, Z., Du, S., Hong, M., Wang, K., Zhang, D., 2019. Sediment pH, not the bacterial diversity, determines Escherichia coli O157: H7 survival in estuarine sediments. Environ. Pollut. 252, 1078–1086.
Liu, F., Zhang, G.L., Song, X., Li, D., Zhao, Y., Yang, J., Wu, H., Yang, F., 2020. High-resolution and three-dimensional mapping of soil texture of China. Geoderma 361, 114061.
Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J.L., Song, X., Shi, Z., Zhu, A.X., Zhang, G.L., 2022. Mapping high resolution national soil information grids of China. Sci. Bull. 67 (3), 328–340.
Luo, L., Lin, H., Schmidt, J., 2010. Quantitative relationships between soil macropore characteristics and preferential flow and transport. Soil Sci. Soc. Am. J. 74 (6), 1929–1937.
Ma, J., Ibekwe, A.M., Yi, X., Wang, H., Yamazaki, A., Crowley, D.E., Yang, C.H., 2011. Persistence of Escherichia coli O157: H7 and its mutants in soils. PLoS One 6 (8), e23191.
Mahesh, B., 2020. Machine learning algorithms-a review. Int. J. Sci. Res. 9 (1), 381–386 *[Internet]*.
Majowicz, S.E., Scallan, E., Jones-Bitton, A., Sargeant, J.M., Stapleton, J., Angulo, F.J., Yeung, D.H., Kirk, M.D., 2014. Global incidence of human Shiga toxin–producing Escherichia coli infections and deaths: a systematic review and knowledge synthesis. Foodborne pathogens and disease 11 (6), 447–455.
Matthews, K.R., 2023. Manure management. In: The Produce Contamination Problem. Academic Press, pp. 47–66.
McLeod, M., Aislabie, J., McGill, A., Rhodes, P., Hunter, D.W.F., Thornburrow, D., 2019. Mapping the potential risk of Escherichia coli leaching through soils of the Waikato River catchment, New Zealand. Soil Res. 57 (2), 132–148.
Meng, X., Mao, K., Meng, F., Shi, J., Zeng, J., Shen, X., Cui, Y., Jiang, L., Guo, Z., 2021. A fine-resolution soil moisture dataset for China in 2002–2018. Earth Syst. Sci. Data 13 (7), 3239–3261.
Morgado, M.E., Hudson, C.L., Chattopadhyay, S., Ta, K., East, C., Purser, N., Allard, S., Ferrier, M.D., Sapkota, A.R., Sharma, M., Goldstein, R.R., 2022. The effect of a first flush rainwater harvesting and subsurface irrigation system on E. coli and pathogen concentrations in irrigation water, soil, and produce. Sci. Total Environ. 843, 156976.
Moshatati, A., Khodaei Joghan, A., Siadat, S.A., Mousavi, S.H., Rezaei, M., 2019. The effect of cattle manure and zeolite on bread wheat yield under drought stress condition. Environmental Stresses in Crop Sciences 12 (4), 1179–1188.
Nachtergaele, F., Velthuizen, H.V., Verelst, L., Wiberg, D., 2009. Harmonized World Soil Database (Hwsd). Food and Agriculture Organization of the United Nations, Rome.
Oudega, T.J., Lindner, G., Derx, J., Farnleitner, A.H., Sommer, R., Blaschke, A.P., Stevenson, M.E., 2021. Upscaling transport of bacillus subtilis endospores and coliphage phiX174 in heterogeneous porous media from the column to the field scale. Environ. Sci. Technol. 55 (16), 11060–11069.
Pang, H., Mokhtari, A., Chen, Y., Oryang, D., Ingram, D.T., Sharma, M., Millner, P.D., Van Doren, J.M., 2020. A predictive model for survival of Escherichia coli O157: H7 and generic E. coli in soil amended with untreated animal manure. Risk Anal. 40 (7), 1367–1382.
Park, Y., Pachepsky, Y., Shelton, D., Jeong, J., Whelan, G., 2016. Survival of manure-borne Escherichia coli and fecal coliforms in soil: temperature dependence as affected by site-specific factors. J. Environ. Qual. 45 (3), 949–957.
Peng, S., Ding, Y., Liu, W., Li, Z., 2019. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. Earth Syst. Sci. Data 11 (4), 1931–1946.
Petersen, F., Hubbart, J.A., 2020. Physical factors impacting the survival and occurrence of Escherichia coli in secondary habitats. Water 12 (6), 1796.
Phan-Thien, K., Metaferia, M.H., Bell, T.L., Bradbury, M.I., Sassi, H.P., de Ogtrop, F.F., Suslow, T.V., McConchie, R., 2020. Effect of soil type and temperature on survival of Salmonella enterica in poultry manure-amended soils. Lett. Appl. Microbiol. 71 (2), 210–217.
Rani, A., Ravindran, V.B., Surapaneni, A., Mantri, N., Ball, A.S., 2021. Trends in point-of-care diagnosis for Escherichia coli O157: H7 in food and water. Int. J. Food Microbiol. 349, 109233.
Rothrock, M.J., Frantz, J.M., Burnett, S., 2012. Effect of volumetric water content and clover (Trifolium incarnatum) on the survival of Escherichia coli O157: H7 in a soil matrix. Curr. Microbiol. 65, 272–283.
Schaap, M.G., Leij, F.J., Van Genuchten, M.T., 2001. Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251 (3–4), 163–176.
Sharma, M., Millner, P.D., Hashem, F., Vinyard, B.T., East, C.L., Handy, E.T., White, K., Stonebraker, R., Cotton, C.P., 2019. Survival of Escherichia coli in manure-amended soils is affected by spatiotemporal, agricultural, and weather factors in the Mid-Atlantic United States. Appl. Environ. Microbiol. 85 (5), e02392, 18.
Simunek, J., Van Genuchten, Sejna, M., 2005. The HYDRUS-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. University of California-Riverside Research Reports 3, 1–240.
Strawn, L.K., Gröhn, Y.T., Warchocki, S., Worobo, R.W., Bihn, E.A., Wiedmann, M., 2013. Risk factors associated with Salmonella and Listeria monocytogenes contamination of produce fields. Appl. Environ. Microbiol. 79 (24), 7618–7627.
Tahmasebi, P., Kamrava, S., Bai, T., Sahimi, M., 2020. Machine learning in geo-and environmental sciences: from small to large scale. Adv. Water Resour. 142, 103619.
Tufenkji, N., 2007. Modeling microbial transport in porous media: traditional approaches and recent developments. Adv. Water Resour. 30 (6–7), 1455–1469.
Unc, A., Goss, M.J., 2003. Impact of manure properties on the survival of manure Escherichia coli in soils. In: Proceedings of 2nd Canadian Organic Residuals Recycling Conference, vol. 24. Penticton, p. 25.
Underthun, K., De, J., Gutierrez, A., Silverberg, R., Schneider, K.R., 2018. Survival of Salmonella and Escherichia coli in two different soil types at various moisture levels and temperatures. J. Food Protect. 81 (1), 150–157.
Verwey, E.J.W., 1947. Theory of the stability of lyophobic colloids. J. Phys. Chem. 51 (3), 631–636.
Wang, Y., Bradford, S.A., Šimůnek, J., 2013. Transport and fate of microorganisms in soils with preferential flow under different solution chemistry conditions. Water Resour. Res. 49 (5), 2424–2436.
Wang, H., Ibekwe, A.M., Ma, J., Wu, L., Lou, J., Wu, Z., Liu, R., Xu, J., Yates, S.R., 2014. A glimpse of Escherichia coli O157: H7 survival in soils from eastern China. Sci. Total Environ. 476, 49–56.
Wang, C.Y., Zhou, X., Guo, D., Zhao, J.H., Yan, L., Feng, G.Z., Gao, Q., Yu, H., Zhao, L.P., 2019. Soil pH is the primary factor driving the distribution and function of microorganisms in farmland soils in northeastern China. Ann. Microbiol. 69 (13), 1461–1473.
Wen, J., Li, J., Wang, Z., Li, Y., 2017. Modelling water flow and Escherichia coli transport in unsaturated soils under drip irrigation. Irrigat. Drain. 66 (5), 738–749.
Williams, A.P., Avery, L.M., Killham, K., Jones, D.L., 2005. Persistence of Escherichia coli O157 on farm surfaces under different environmental conditions. J. Appl. Microbiol. 98 (5), 1075–1083.

Williams, N.D., Torres, A.G., Lloyd, S.J., 2010. Evolution and epidemiology of diarrheagenic Escherichia coli. Pathogenic Escherichia coli in Latin America, pp. 8–24.
Williams, M.L., LeJeune, J.T., McSpadden Gardener, B., 2015. Soil conditions that can alter natural suppression of Escherichia coli O157: H7 in Ohio specialty crop soils. Appl. Environ. Microbiol. 81 (14), 4634–4641.
Willis, C., McLauchlin, J., Aird, H., Amar, C., Barker, C., Dallman, T., Elviss, N., Lai, S., Sadler-Reeves, L., 2020. Occurrence of Listeria and Escherichia coli in frozen fruit and vegetables collected from retail and catering premises in England 2018–2019. Int. J. Food Microbiol. 334, 108849.
Xing, J., Wang, H., Brookes, P.C., Salles, J.F., Xu, J., 2019. Soil pH and microbial diversity constrain the survival of E. coli in soil. Soil Biol. Biochem. 128, 139–149.
Zhang, T., Wang, H., Wu, L., Lou, J., Wu, J., Brookes, P.C., Xu, J., 2013. Survival of Escherichia coli O157: H7 in soils from jiangsu province, China. PLoS One 8 (12), e81178.
Zhong, H., Liu, G., Jiang, Y., Yang, J., Liu, Y., Yang, X., Liu, Z., Zeng, G., 2017. Transport of bacteria in porous media and its enhancement by surfactants for bioaugmentation: a review. Biotechnol. Adv. 35 (4), 490–504.
Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., 2021. Machine learning: new ideas and tools in environmental science and engineering. Environ. Sci. Technol. 55 (19), 12741–12754.