Journal of Cleaner Production 297 (2021) 126704

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A predictive model of the impact of urbanization on bacterial loads in watersheds

Chantima Piyapong [a], Nitcha Chamroensaksri [b], Sayam Aroonsrimorakot [c], Lawan Eyosawat [d], Surasak Khankhum [e], Sunirat Rattana [f], Nuchsupha Sunthamala [e], Panya Warapetcharayut [g], Emmanuel Paradis [h, *]

[a] Department of Biology, Faculty of Science, Burapha University, Chonburi, 20131, Thailand
[b] National Biobank of Thailand (NBT), National Science and Technology Development Agency (NSTDA), 144 Thailand Science Park, Phahonyothin Road, Khlong Nueng, Khlong Luang District, Pathum Thani, 12120, Thailand
[c] Program in Appropriate Technology and Innovation for Environmental Security, Faculty of Environment and Resource Studies, Mahidol University, Nakhon Pathom, 73170, Thailand
[d] Regional Environmental Office 13, Ministry of Natural Resources and Environment, Chonburi, 20000, Thailand
[e] Department of Biology, Faculty of Science, Mahasarakham University, Mahasarakham, 44150, Thailand
[f] Regional Environmental Office 7, Ministry of Natural Resources and Environment, Saraburi, 18120, Thailand
[g] Environmental Research and Training Center, Department of Environmental Quality Promotion, Ministry of Natural Resources and Environment, Technopolis, Khlong Luang District, Pathum Thani, 12120, Thailand
[h] ISEM, IRD, Univ. Montpellier, CNRS, EPHE, Montpellier, France

ARTICLE INFO

Article history:
Received 6 May 2019
Received in revised form
2 March 2021
Accepted 9 March 2021
Available online 14 March 2021

Handling editor: Yutao Wang

Keywords:
Coliforms
Microbiological water quality
Mixed-effect models
Rivers
Thailand

ABSTRACT

Bacterial concentration is one of the most important aspects of water quality. Many regions in the world are affected by increasing urbanization and a potential increase in bacterial concentrations in waters. We used long-term data from 68 stations in eight watersheds in Eastern Thailand to quantify the temporal and geographical variation in total and fecal coliform bacteria. Descriptive statistics showed considerable seasonal, inter-annual, and geographical variation. In order to quantify this multi-level variation, we built a predictive model of bacterial loads. Using fixed- and mixed-effects regression models, we built a model including the effects of urbanization and other significant variables. The best model, fitted by restricted maximum likelihood, included the effects of season, year, urbanization as fixed effects, and of watershed and station as nested, random effects. Temporal variation was related to seasonal and annual variations. Spatial variation had a very significant impact on the bacterial concentrations. Urbanization was an important factor controlling concentrations of bacteria in rivers: we found that the proportion of urban area around a station had a statistically significant effect on log-transformed total coliform bacterial concentration with a slope equal to 1.3 (SE = 0.3), and on log-transformed fecal coliform bacterial concentration with a slope equal to 1.4 (SE = 0.3). Our model predicts that bacterial concentrations would be multiplied by 20 if land is transformed from non-urban to fully urban.
© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The growth of urban areas is a major environmental issue around the world. Conversion of lands into either agricultural lands or urban areas have different consequences with the former case being generally associated with increased soil erosion and increased chemical inputs in water (particularly nitrogen), whereas increased human population density is generally associated to increased waste disposal from urban areas which affect directly bacterial concentrations (Cabral and Marques, 2006; Diwan et al., 2018; Shukla et al., 2018). These issues affect water quality in rivers and waterbodies at different scales and depending on ecological factors and the management of waste (Causse et al., 2015; Jeong et al., 2019). Furthermore, land use change is a dynamic process taking place at different scales of time and space

* Corresponding author.
E-mail addresses: chantimap@go.buu.ac.th (C. Piyapong), nitcha.cha@nstda.or.th (N. Chamroensaksri), sayamthai88@gmail.com (S. Aroonsrimorakot), lawan_e@hotmail.com (L. Eyosawat), surasak.kk@msu.ac.th (S. Khankhum), osunirat@gmail.com (S. Rattana), nuchsupha.s@msu.ac.th (N. Sunthamala), panya41@hotmail.com (P. Warapetcharayut), Emmanuel.Paradis@ird.fr (E. Paradis).

https://doi.org/10.1016/j.jclepro.2021.126704
0959-6526/© 2021 Elsevier Ltd. All rights reserved.

(Song et al., 2018), so we may expect the relationship between land use changes and water quality to be complex.

One of the most important components of water quality is the quantity of bacteria in rivers, and particularly coliforms, as some of them are produced by human waste (Abraham, 2011; Rochelle-Newall et al., 2015). A crucial challenge is to predict the changes in concentrations of coliforms in relation to land use (Lyautey et al., 2010; St Laurent and Mazumder, 2012; Rochelle-Newall et al., 2016). These predictions are critical for environmental sustainability, but also to be able to take preventive decisions to avoid negative impacts on human and ecosystem health.

Bacterial contamination is one of the most important issues related to water pollution. The scientific literature on this subject points to two main sources of coliforms in surface waters: agriculture (Lyautey et al., 2010; St Laurent and Mazumder, 2012) and urbanization (DiDonato et al., 2009). The considerable ongoing changes in land use and occupation in Southeast Asia result in challenges in the management of environmental resources. Such challenges may be local (Soytong and Perera, 2017) or regional (Taylor, 2010; Paradis, 2018, 2020). Thailand is a country characterized by heterogeneous socioeconomic development among its provinces, resulting in intense rural-urban migration (Shatkin, 2004; Soytong and Perera, 2017). Eastern Thailand is characterized by highly contrasted land use: it includes large tracks of tropical forest in its interior, extensive agricultural lands, and fast-growing urban areas on the coast. The past and current projects of economic development in this area, particularly on the sea coast, raise concerns about water quality and its evolution with respect to changes in land use (Water Environmental Partnership in Asia, 2018; Eastern Economic Corridor Office, 2019).

In the context of increased urbanization in Eastern Thailand, the risk of water contamination by coliforms could be increased because of the increasing local population densities in industrial areas (Soytong and Perera, 2014; Water Environmental Partnership in Asia, 2018). Surface waters in Thailand are classified into five categories according to different quantitative criteria; these categories are labelled as: (1) extra clean, (2) very clean, (3) fairly clean, (4) moderately clean, and (5) unclassified in the previous classes (http://www.secot.co.th/secot_ww/std-e.html; accessed 2021-01-14). Only the categories 1–3 are proper for human consumption. The main water quality problems in four major river basins of Thailand are high loads in coliform bacteria (36% of the problems; Simachaya et al., 2000). In the Eastern Region, fecal coliform contaminations represent 17% of the water quality problems compared to other parameters such as total solids, turbidity, biochemical oxygen demand, or dissolved oxygen. Fecal coliform bacteria in the Bang Pakong River Basin, a watershed situated in the North of the Eastern Thailand Region, exceeded surface water standard category 3, which is specified as equal to or less than 4000 bacteria per 100 mL (Simachaya, 2003). This previous study suggested that organic loading caused by human communities, industry, pig farms, and aquaculture were the sources of pollution in this basin.

In this paper, we present a predictive model of bacterial concentrations in rivers and waterways. We assessed the variability in total and fecal coliform bacteria throughout Eastern Thailand using data from a long-term monitoring program of water quality. Using fixed- and mixed-effects regression models, we built a model that includes the effects of urbanization and other significant variables. We used this model to predict changes in bacterial concentrations caused by urbanization taking other effects into account. We conclude with some perspectives on water quality management in the context of growing urban areas.

2. Materials and methods

2.1. Sampling

Sampling took place over the Eastern Thailand region in eight watersheds covering a variety of ecological situations (Fig. 1). In order to consider seasonal variation, sampling was conducted four times each year in February, May, August, and November which made possible to cover the range of weather regimes in Eastern Thailand: end of dry (cool) season, first peak of rainy season, start of second peak of rainy season, and start of dry season, respectively (http://www.climate.tmd.go.th/content/file/75). The data used in this paper cover a period from December 2007 until August 2016. For logistic reasons, sampling was done in December instead of November in 2007 and in 2008, and in April 2008 instead of May 2008. Additionally, several stations were added to the monitoring program in order to increase its geographical coverage. Four stations (NK01, NK02, PT01, and PT02) were discontinued after four sampling sessions because they could not be accessed continuously to sample water. A total of 68 stations were sampled between four and 36 times resulting in a total of 1999 samples. Five stations were sampled four or six times, while the 63 others were sampled at least 19 times. The geographical coordinates (longitude, latitude) of

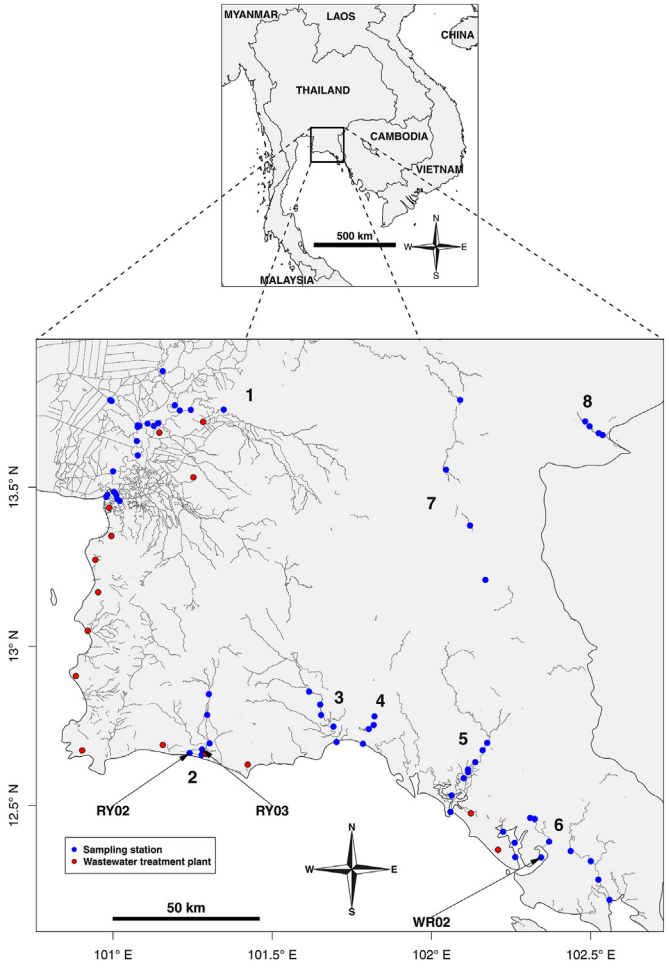


Fig. 1. Geographical distribution of sampling stations displayed with the main waterways in Eastern Thailand and the wastewater treatment plants. The bold numbers indicate the watersheds used in the analyses. The stations RY02, RY03, and WR02, which are used in Figs. 7 and 8, are shown with arrows. The locations of the wastewater treatment plants are from Dumrongthai (2019).

each station were recorded. The water samples were collected in sterile 250-mL plastic bottles and transported on ice to the laboratory of the Regional Environment Office of the Chonburi Province within 24 h.

2.2. Laboratory analyses

The enumerations of total coliform bacteria (TCB) and fecal coliform bacteria (FCB) were done by means of the multiple-tube fermentation (MTF) method according to the protocols of the American Public Health Association (APHA, 1992). For TCB counts, a series of five fermentation tubes of lauryl sulphate broth (LSB, Merck; also known as lauryl tryptose broth) were inoculated with appropriate volumes of 10-fold dilutions of water samples and incubated at 37 °C for 48 h. The number of five-tube series and the inoculation volumes were chosen depending on the quality and character of the water to be examined (American Public Health Association, 1992, p. 9–46). All gas-positive LSB tubes were subcultured to tubes of brilliant green lactose bile broth (BGLB, Merck) and incubated at 37 °C for 48 h. Gas-positive BGLB tubes were considered positive for the presence of TCB. Gas-positive LSB tubes were subjected to further analysis with *Escherichia coli* broth (ECB, Merck). The EC tubes were incubated at 44.5 °C for 24 h. Gas-positive ECB tubes were considered positive for the presence of FCB. The values of TCB and FCB (expressed in units of most probable number per 100 mL, or MPN/100 mL) were inferred using the following formula (American Public Health Association, 1992, p. 9–50):

$$\text{Bacterial concentration} = \frac{100 \times \text{number of positive tubes}}{\sqrt{\left(\begin{matrix}\text{mL sample in}\\ \text{negative tubes}\end{matrix}\right) \times \left(\begin{matrix}\text{mL sample in}\\ \text{all tubes}\end{matrix}\right)}}.$$

2.3. Land cover data

The land cover data were downloaded from the European Space Agency Climate Change Initiative (ESACCI, http://maps.elie.ucl.ac.be/CCI/viewer/download.php, accessed 2018-04-18). These remote sensing data provide land cover and land use from 1992 to 2015 with a resolution of 10" (ten arcseconds), which is equivalent to a distance of approximately 300 m. Land cover was classified into 38 categories ranging from "No data" to "Permanent snow and ice". The land cover data were extracted for Eastern Thailand (12.1–14° N, 100.8–102.7° E) and for the years 2007–2015. We calculated an index of urbanization (variable called 'Urbanization' below) for each station as the percentage of land use classified as "Urban areas" 1 km around the point defined by the coordinates of the station. This index was calculated for each station and each year from 2007 to 2015; its value varied between 0 (no urban area around the station) and 1 (only urban area around the station). We also summarized land use changes, evaluated pixel by pixel, during this period with a focus on urbanization by calculating the proportion of urban areas over Eastern Thailand and land conversion related to urbanization.

2.4. Statistical model-building

We modelled the variation in TCB and FCB with regression models using different approaches, starting from simple linear models and then building more complex models. We first used simple linear regressions to assess the form and the significance of the variation of these relationships. We then performed multiple regression models to test the relative statistical significance of several predictors ('Year', 'Season', 'Station', 'Urbanization') using *F*-tests, assessing the significance of each effect while taking others into account. Hereafter, variable names considered as effects in a model are written within single quotes. However, these models must be interpreted with caution because of confounding effects. For instance, it is difficult to test the effects of 'Station' and 'Urbanization' in the same model because these two effects are mostly confounded (i.e. both variables are geographically linked). Therefore, we used linear mixed-effects (LME) models using 'Watershed' and 'Station' as random-effect terms. This made it possible to test for the effect of 'Urbanization' while considering 'Watershed' and 'Station' as repeated, nested, correlated measures. Each individual effect was tested with a χ^2-based likelihood-ratio test where the null hypothesis is that the predictor has no effect on either TCB or FCB concentration while taking the effects of the other predictors in the model into account. Different LME models were compared with their Akaike information criteria (AIC; Akaike, 1973) after fitting the models by maximum likelihood (Pinheiro and Bates, 2000): the model with the smallest AIC was selected as the best one among those considered. TCB and FCB both had strongly skewed distributions so they were $\log_{10}$-transformed before analyses, as commonly done when analyzing such data (e.g. Lyautey et al., 2010). Since 'Urbanization' was a continuous variable, its effect in the model was linear; however, it could be that such an effect might not be linear, for instance, urbanization may have no effect if its index is less than a threshold. Nonlinear effects of 'Urbanization' were tested by transforming the urbanization index into two categories: a threshold was chosen and a new categorical variable was defined as "urban" for the stations with 'Urbanization' greater than or equal to the threshold, or as "non-urban" if 'Urbanization' was less than the threshold. Different values of threshold were assessed. Predictions for LME models were done with best linear unbiased predictors (BLUP) which include predictions for the random and the fixed effects (Robinson, 1991). The predictive power of a model was calculated as the ratio of its explained variance on the total variance calculated with:

$$R^2 = 1 - \frac{\widehat{\sigma}_r^2}{\widehat{\sigma}^2},$$

where $\widehat{\sigma}_r^2$ is the residual variance of the model, and $\widehat{\sigma}^2$ is the total variance.

All data analyses were done with R version 3.5.1 (R Core Team, 2019). The LME models were fitted and analyzed with the package `lme4` (Bates et al., 2015). The GIS data were handled with the packages `rgdal` (Bivand et al., 2018) and `raster` (Hijmans, 2018). A file with the commands used for these analyses is available as supplementary information.

3. Results

3.1. Summary of coliform bacterial concentrations

Very low or very large bacterial concentrations could not be quantified precisely in the laboratory with the MTF method because the required numbers of dilutions are not achievable in practice. The detection thresholds in our study were found to be 18 MPN/100 mL $\leq$ bacterial concentration $\leq$ 160,000 MPN/100 mL. However, presence of bacteria outside of this interval was found in 124 measures of TCB (6.2%) and 134 measures of FCB (6.7%). Out of the 1999 measures of TCB, 47 (2.4%) were below the lower-bound of the detection threshold (TCB concentration $<$ 18 MPN/100 mL),

and 77 (3.9%) were above the upper-bound limit (TCB concentration > 160,000 MPN/100 mL). For FCB, these numbers were 110 (5.5%) and 24 (1.2%), respectively. It is therefore possible to infer the presence of bacteria in very low or very large concentrations in these 258 measures. In order to include this information in the subsequent quantitative analyses, these values were replaced by values drawn randomly in the interval [1, 17] for those below 18 MPN/100 mL, or in the interval [160,000; 200,000] for those above 160,000 MPN/100 mL.

Table 1 gives a summary of the distributions of TCB and FCB. Bacterial concentration exceeded 10,000 MPN/100 mL in 599 (30%) of the measures of TCB and in 246 measures (12.3%) of those of FCB. A small number of measures exceeded 100,000 MPN/100 mL: 131 (6.6%) for TCB and 45 (2.2%) for FCB. The concentrations of FCB were greater than 4000 MPN/100 mL in 405 measures (20%). Both TCB and FCB displayed extensive spatial and temporal variation (Figs. 2 and 3, for TCB and FCB, respectively). All stations displayed strong seasonal variation in both TCB and FCB with the highest levels measured in summer. Some inter-annual variation were visible for some stations (e.g. BK13, CB02, RY01, TE01). The five stations with four or six measures (BTR01, NK01, NK02, PT01, and PT02) were dropped for the subsequent analyses which were thus done with the 63 stations with at least 19 measures out of the 36 sampling occasions (see Sampling subsection). The mean TCB for each station averaged over years and seasons ranged between 479 MPN/100 mL (Station WR01 located upstream of the River Chantaburi) and 115,576 MPN/100 mL (Station BPT01 near the seamouth of the River Bang Pakong). The mean FCB ranged between 48 MPN/100 mL (Station WR01) and 83,659 MPN/100 mL (Station BPT01). Three stations had a mean TCB larger than 100,000 MPN/100 mL: PR04 (upstream of the River Phang Rat on the border of the Provinces Rayong and Chantaburi), TE02 (upstream the River Bang Pakong), and BPT01 (see Table 1). The geographical distribution of the means for all stations showed considerable variation among basins (Fig. 4). Large mean TCB were observed throughout the study area, particularly downstream the rivers close to their seamouths.

3.2. Land cover use and changes between 2007 and 2015

The area considered for the land cover analysis covered 42,031 km^2 (including 10,323 km^2 covered by water) and showed little change in land use between 2007 and 2015 with only 287 km^2 (0.7%) of land that changed during this period. Out of these, 221 km^2 (77%) were converted into urban areas, mostly from cropland (204 km^2). The extent of urban areas changed from 283 km^2 in 2007, to 504 km^2 in 2015, an increase of 78%. The three main categories of land use in the study area in 2015 were: rainfed cropland (15,360 km^2, 48.4%, excluding water), herbaceous cover (5966 km^2, 18.8%), and evergreen, broadleaved forest (3429 km^2, 10.8%).

Table 1
Distribution summaries of total coliform bacteria (TCB) and fecal coliform bacteria (FCB).

	Number of measures greater than[a] (in MPN/100 mL)			Number of stations with mean greater than† (in MPN/100 mL)		
	10,000	50,000	100,000	10,000	50,000	100,000
TCB	599	248	131	38	7	3
FCB	246	84	45	14	2	0

[a] Out of 1999 measures from 68 stations. †Out of 63 stations.

3.3. Linear models

All effects were highly significant when considered alone in a regression model (Table 2). The variable 'Year' was considered either as a continuous variable or as a categorical one: in both cases the relationship was significant, but more variation was explained when 'Year' was a categorical variable, which indicates that TCB and FCB did not increase linearly over time. 'Year' was considered as a categorical variable in the subsequent analyses. 'Station' explained 47% of the variance in both TCB and FCB, whereas 'Watershed' explained 18%. On the other hand, 'Year' or 'Season' explained low quantities of variance. This shows that geographical variation is far more important than temporal variation.

A multiple regression model with the effects of 'Year', 'Season', and 'Station', as well as the interactions between 'Season' and 'Station' and between 'Season' and 'Year', was fitted. 'Watershed' and 'Urbanization' were not considered in this model because they are confounded with 'Station' (see subsection Statistical model-building). All effects as well as the interactions were highly significant (Table 3). This model explained 66% of the variation of TCB, and 64% of FCB.

3.4. Mixed-effects models

To further elaborate our models, we fitted an LME model considering 'Station' and 'Watershed' as random effects, with the former nested in the latter, and including 'Season', 'Year', the interaction between these two variables, and 'Urbanization', as main effects. All fixed effects of this model were strongly significant with either TCB or FCB as the response (Table 4). For TCB, the final model with all the above effects had AIC = 4494. Transforming the urbanization index as a categorical variable did not improve the fit: AIC = 4501 with a threshold of 0.5 (i.e. stations with less than 50% of urban areas around 1 km were considered as "non-urban"), or AIC = 4506 with a threshold of 0. The estimated variance associated with the random effect of 'Station' was σ^Station2=0.26 (SE = 0.52), and the variance associated with 'Watershed' was $\widehat{\sigma}^2_{\text{Watershed}} = 0.11$ (SE = 0.33), both smaller than the residual variance $\widehat{\sigma}^2 = 0.47$ (SE = 0.68).

For FCB, the final model had AIC = 4516. Transforming the urbanization index as a categorical variable did not improve the fit (AIC = 4527 with a threshold at 0.5, or AIC = 4535 with a threshold at 0). The estimated variances associated with the random effects were $\widehat{\sigma}^2_{\text{Station}} = 0.24$ (SE = 0.49) and $\widehat{\sigma}^2_{\text{Watershed}} = 0.11$ (SE = 0.33), again smaller than the residual variance $\widehat{\sigma}^2 = 0.47$ (SE = 0.69).

In order to use this final model for prediction, an artificial data set was built with all possible combinations of the five predictors ('Urbanizarion', 'Season', 'Year', 'Station', and 'Watershed'). The predicted value of the response (TCB or FCB) was calculated for each combination of the predictors. This made it possible to predict the effect of a given predictor while taking into account the effect of the others. Fig. 5 shows the predictions of TCB for each effect in the final model, taking the variation from other effects into account. Since 'Urbanization' is a continuous variable, the predicted values of TCB are on parallel straight lines with a slope equal to the estimated coefficient from the final model. These lines have different intercepts because they represent the variation in the other predictors. For the other predictors, which are all discrete, the variability from the others is represented with boxplots. This representation makes it possible to visualize the variability associated with the two random effects, 'Watershed' and 'Station', which are nested. In particular, some stations which are in the same watershed have different predicted values. Note that the predictions are

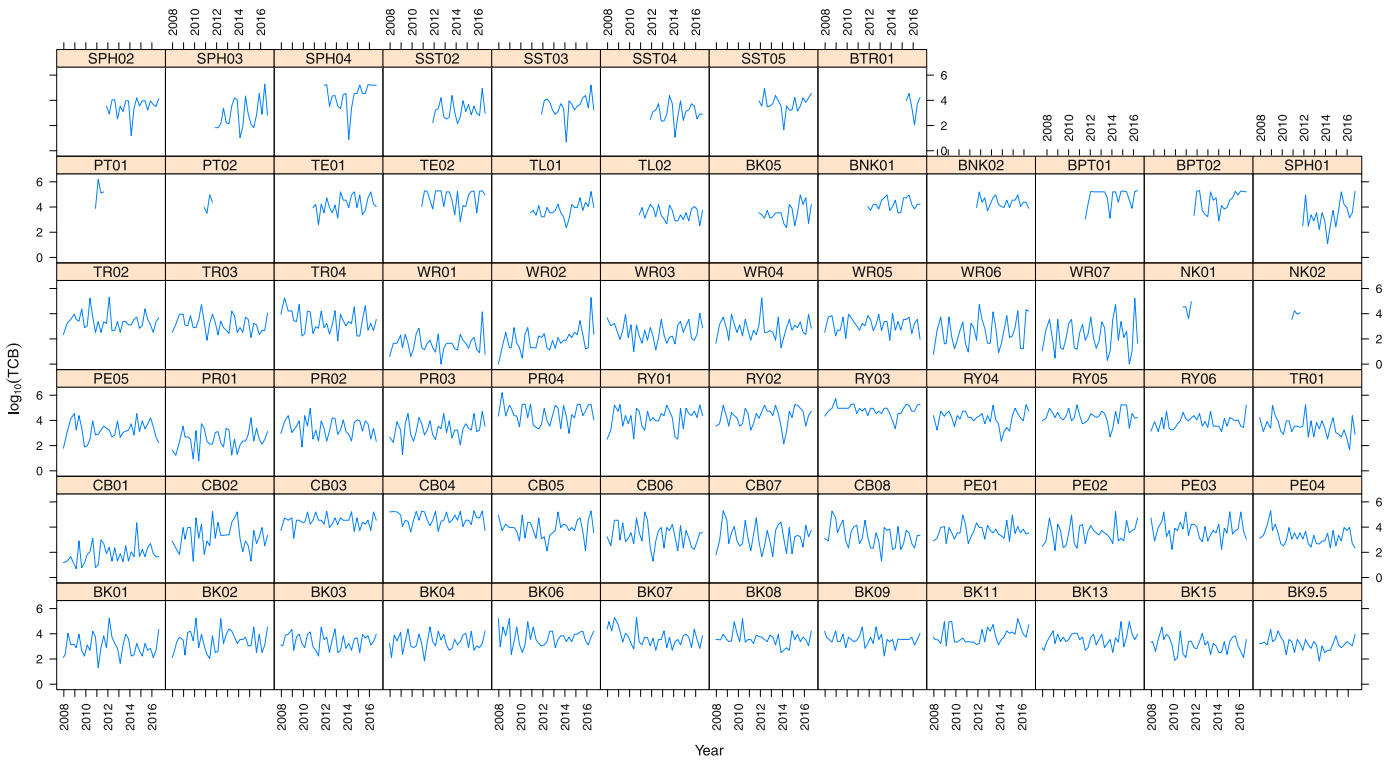


Fig. 2. Temporal variation in abundance of total coliform bacteria (TCB, in MPN/100 mL) in 68 stations in Eastern Thailand.

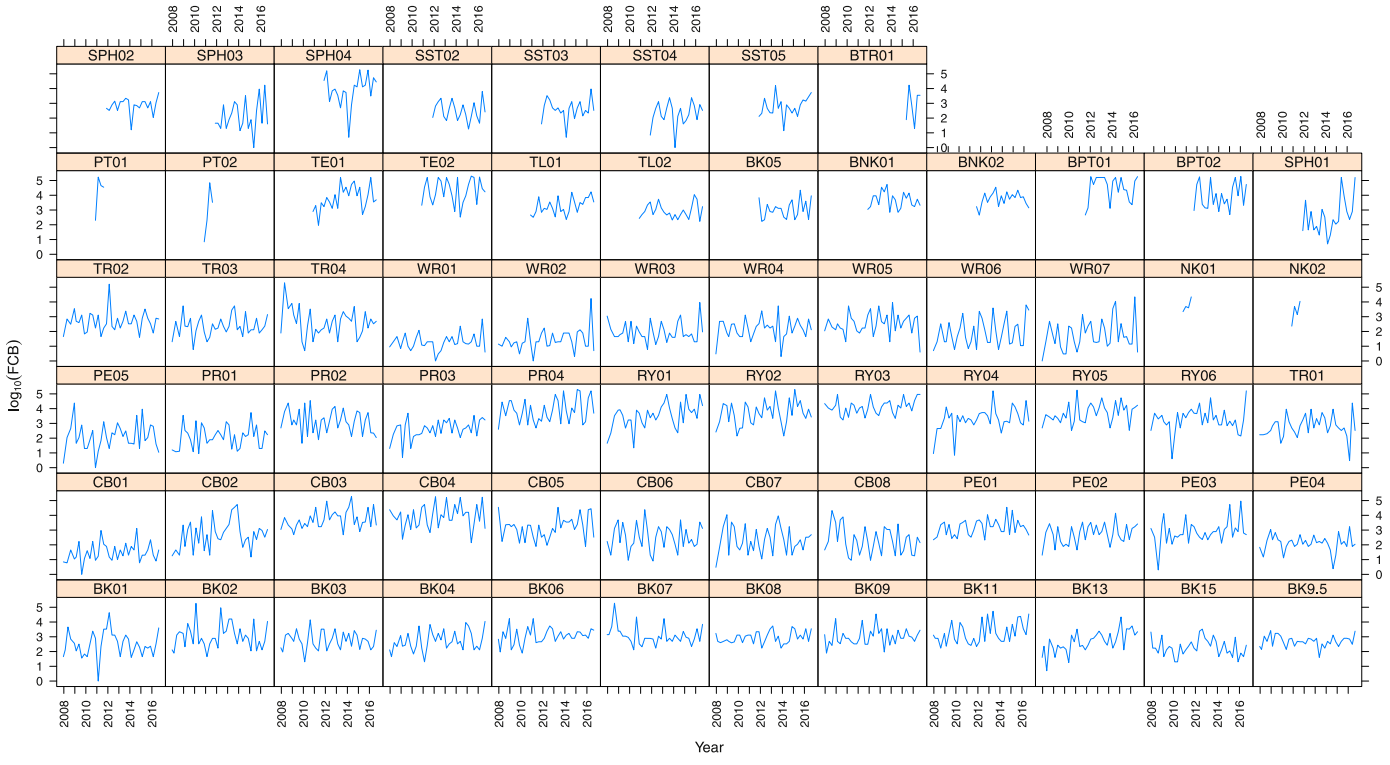


Fig. 3. Temporal variation in abundance of facal coliform bacteria (FCB, in MPN/100 mL) in 68 stations in Eastern Thailand.

made on the scale of the transformed response (i.e. $\log_{10}$). There is clearly considerable variability outside of each significant effect, meaning that a single effect explains only a portion of the global variability. Fig. 6 shows the predictions for FCB which are very similar to those for TCB. The plots of the predicted values against the observed ones show that the predictions were best in the range between 1000 MPN/100 mL and 10,000 MPN/100 mL (Fig. S3). Predictions tended to be overestimated for concentrations less than

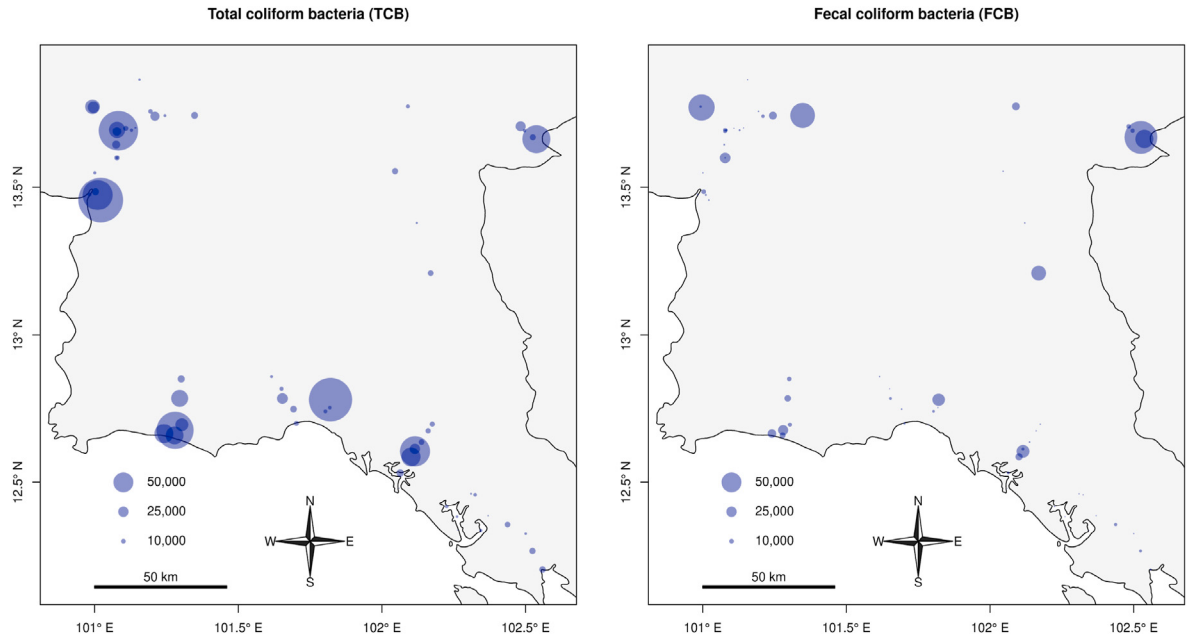


Fig. 4. Abundance of total and fecal coliform bacteria (in MPN/100 mL) in Eastern Thailand averaged over all sampling occasions for each station.

Table 2
Simple linear regressions of total coliform bacteria (TCB) and fecal coliform bacteria (FCB); cont.: continuous; cat.: categorical. *F*: value of the *F*-statistic testing the effect of the predictor, *df*: numbers of degrees of freedom of the test (d_1, d_2), *P*: *P*-value of the test, R^2: proportion of explained variance.

Predictor	TCB				FCB			
	F	*df*	*P*	R^2	*F*	*df*	*P*	R^2
Year (cont.)	5.28	1, 1997	0.021	0.00	46.04	1, 1997	< 0.001	0.02
Year (cat.)	5.47	9, 1989	< 0.001	0.02	9.41	9, 1989	< 0.001	0.04
Season	27.27	3, 1995	< 0.001	0.04	18.42	3, 1995	< 0.001	0.03
Station	25.79	67, 1931	< 0.001	0.47	25.80	67, 1931	< 0.001	0.47
Watershed	63.71	7, 1991	< 0.001	0.18	64.51	7, 1991	< 0.001	0.18
Urbanization	282.83	1, 1997	< 0.001	0.12	300.40	1, 1997	< 0.001	0.13

Table 3
Multiple linear regressions of total coliform bacteria (TCB) and fecal coliform bacteria (FCB). The *F*-tests are for the effect of each predictor taking the effects of the others into account. *F*: value of the *F*-statistic testing the effect of the predictor, *df*: numbers of degrees of freedom of the test (d_1, d_2), *P*: *P*-value of the test.

Predictor	TCB			FCB		
	F	*df*	*P*	*F*	*df*	*P*
Year (cat.)	9.70	8, 1695	< 0.001	14.30	8, 1685	< 0.001
Season[a]	46.47	3, 1919	< 0.001	29.94	3, 1919	< 0.001
Station	12.36	67, 1695	< 0.001	9.60	67, 1685	< 0.001
Season × Station	2.27	201, 1695	< 0.001	1.99	201, 1695	< 0.001
Season × Year	8.25	23, 1695	< 0.001	6.56	23, 1695	< 0.001

[a] Effect tested in a model without interactions.

Table 4
Linear mixed-effects model of total coliform bacteria (TCB) and fecal coliform bacteria (FCB). The χ^2-based tests are likelihood-ratio tests where the null hypothesis is that the predictor in the first column has no effect on either TCB or FCB as response while taking the effects of the others into account. χ^2: value of the χ^2-statistic testing the effect of the predictor, *df*: number of degrees of freedom of the test, *P*: *P*-value of the test.

Effect	TCB			FCB		
	χ^2	*df*	*P*	χ^2	*df*	*P*
Year (cat.)	78.02	9	< 0.001	89.58	9	< 0.001
Season	135.21	3	< 0.001	88.34	3	< 0.001
Urbanization	16.55	1	< 0.001	20.96	1	< 0.001
Season × Year	159.92	23	< 0.001	132.06	23	< 0.001

1000 MPN/100 mL, and underestimated for concentrations above 10,000 MPN/100 mL. The largest discrepancies were (on the log-scale) 3.6 and 2.8 for TCB and FCB, respectively.

The slope of the parallel lines associated with the effect of 'Urbanization' was estimated to be equal to 1.3 (SE = 0.3) for TCB and 1.4 (SE = 0.3) for FCB. Therefore, the process of urbanization in this region (increase of 'Urbanization' from 0 to 1) is predicted to result in a 20-times increase in TCB and 25-times increase in FCB (since the responses were $\log_{10}$-transformed).

To further assess the precision of these predictions, we focused on three stations with contrasting features: a station characterized by an urbanization index equal to one (RY03), a station with an urbanization index equal to zero but in the same watershed than the first one (RY02), and a station in an area with increasing urbanization during the study period (WR02). This makes it possible to formulate predictions without the additional effects of the predictors 'Watershed' and 'Station'. Furthermore, focusing on a specific station is likely to be more meaningful for management purposes. Fig. 7 shows the predicted values of TCB for each year and each season for these three stations. Since the effect of 'Urbanization' is linear in the final model, the predicted effect of this variable is to simply shift the temporal curves downwards (decreasing

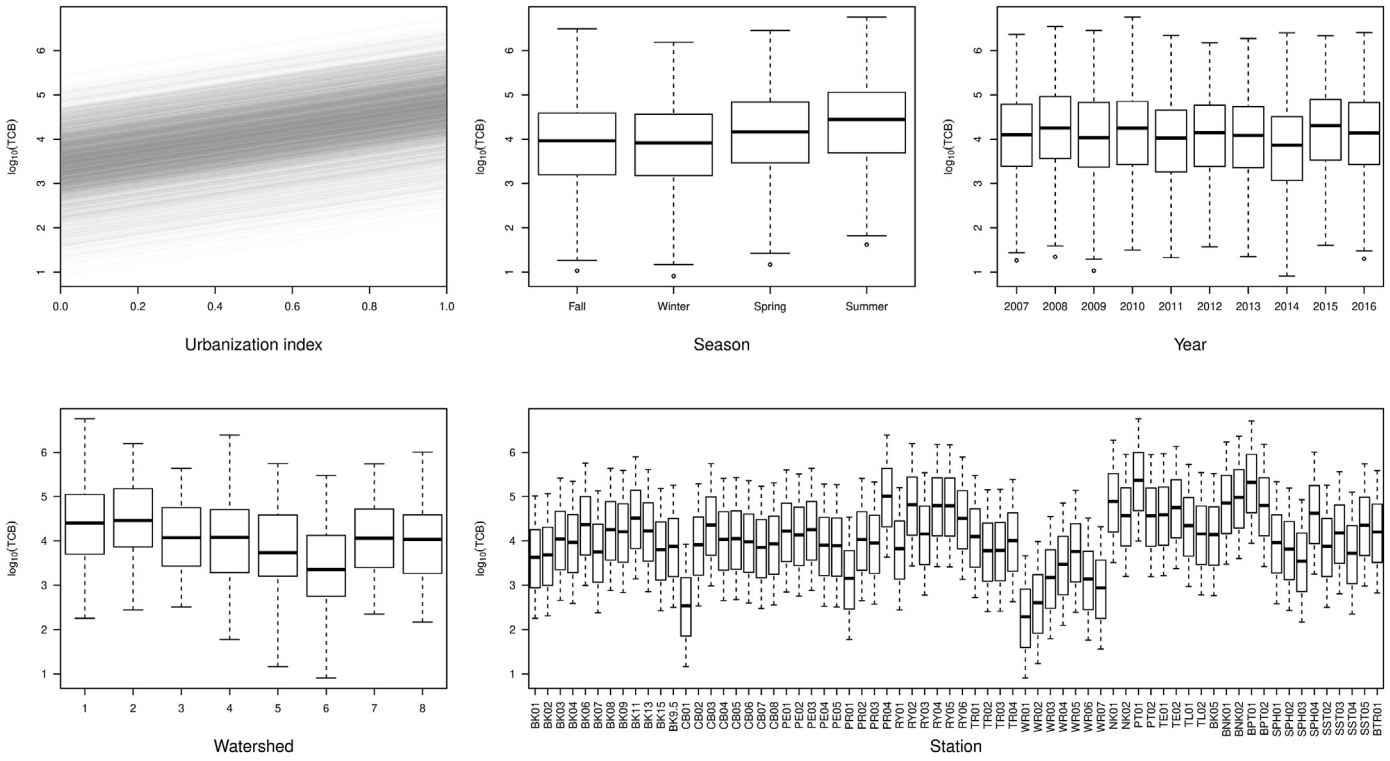


Fig. 5. Predictions of TCB (in MPN/100 mL) from the final mixed-effects model. Each panel shows the partial predictions for each effect included in the model; the displayed variation is due to the other effects.

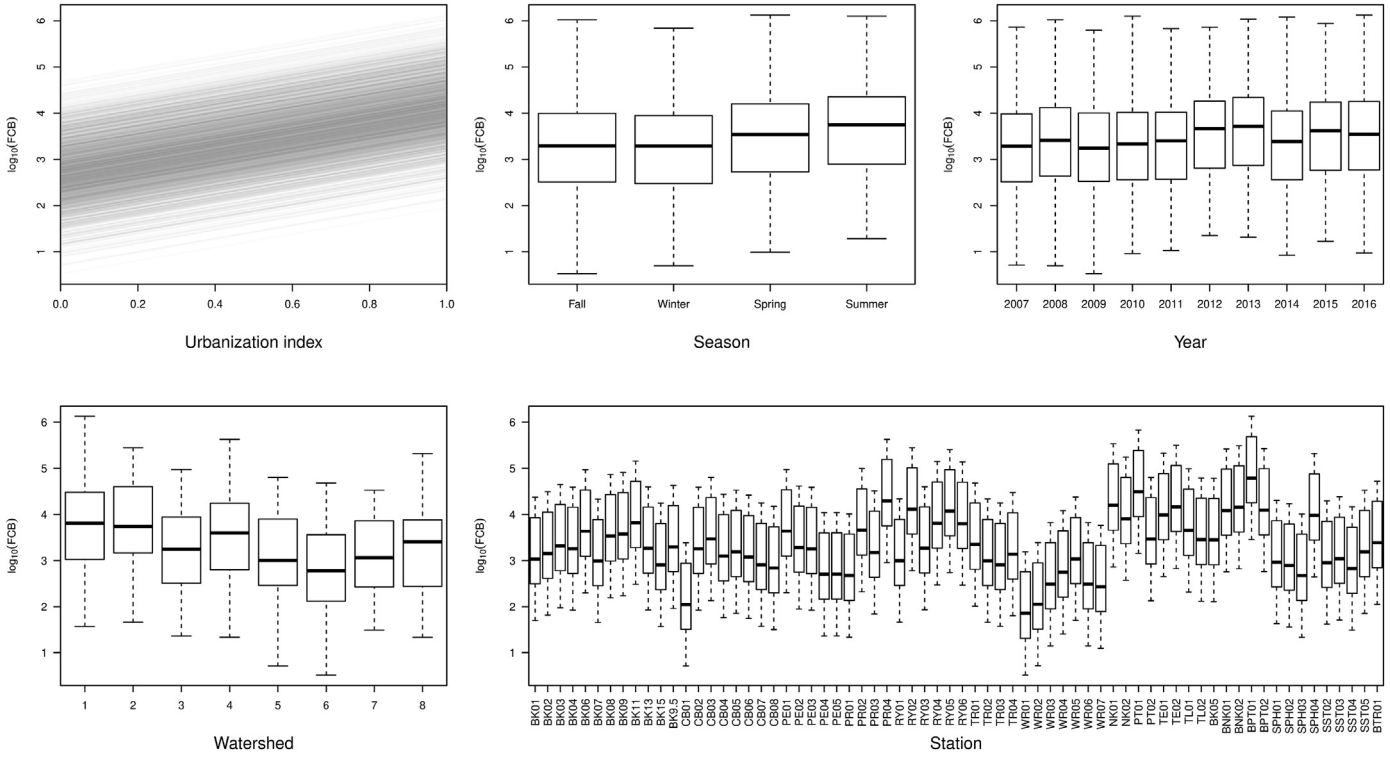


Fig. 6. Predictions of FCB (in MPN/100 mL) from the final mixed-effects model. Each panel shows the partial predictions for each effect included in the model; the displayed variation is due to the other effects.

'Urbanization') or upwards (increasing 'Urbanization'). For simplicity, only the predictions for Urbanization = 0 (continuous curve) and for Urbanization = 1 (dashed curve) are displayed. Clearly, the predictions for RY03 and for RY02 are very similar. In

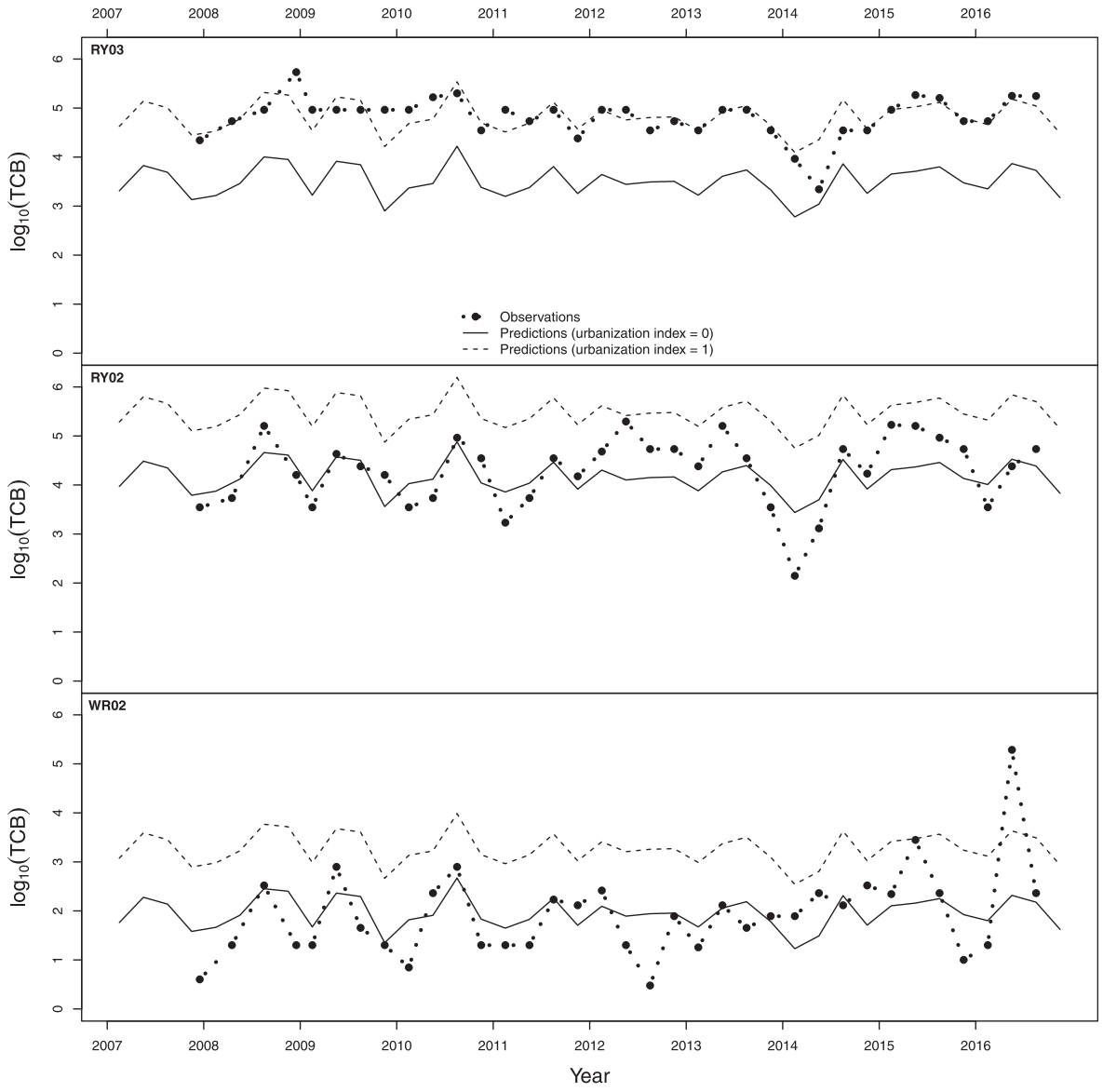


Fig. 7. Predicted values of TCB (in MPN/100 mL) for three stations.

the case of WR02, the urbanization will increase the values of TCB, although not reaching the values predicted for the two other stations. For the three stations, the observations are well predicted by the model. Fig. 8 shows the predictions of FCB.

4. Discussion

In this study of bacterial concentrations in waters of Eastern Thailand, we showed that total and fecal coliform bacteria showed considerable variation in the area studied both spatially and temporally. Several studies have shown similar extensive variation in bacteria in Portugal (Cabral and Marques, 2006), Canada (Lyautey et al., 2010), and India (Shukla et al., 2018). Temporal variation in our study was related to both seasonal and annual variation: the highest levels of TCB/FCB were measured in summer which could be explained by the highest temperatures over the year during this season. However, Vermeulen and Hofstra (2014) observed that concentrations of *Escherichia coli* in several major rivers in Netherlands and Belgium decreased with increasing temperatures because of bacterial die-off. On the other hand, summer in Thailand is also the season with high precipitation which could lead to increased transport of bacteria. We also observed considerable inter-annual variation in both TCB and FCB with no obvious trend. It is not clear what caused these variations, though a potential candidate explanation could be weather. This is currently under further investigation.

Spatial variation had a huge impact on the measured TCB and FCB. It is complicated to assesss this variation because of the nestedness of the effects (e.g. 'Station' and 'Watershed'). This nestedness implies that these effects must be included as random effects in a mixed model. The variance associated with the effect of 'Station' was estimated to be twice as large than the one associated with the effect of 'Watershed'. This was apparent also when plotting the predicted values associated with each effect (Figs. 5 and 6). Interestingly, urbanization appeared as a distinct effect from these two variables since the urbanization index was strongly significant even after including 'Station' and 'Watershed' as random effects. It is still unclear what could explain the important variation among stations and we are currently investigating other potential variables (e.g. distance to the shore, land use around the stations).

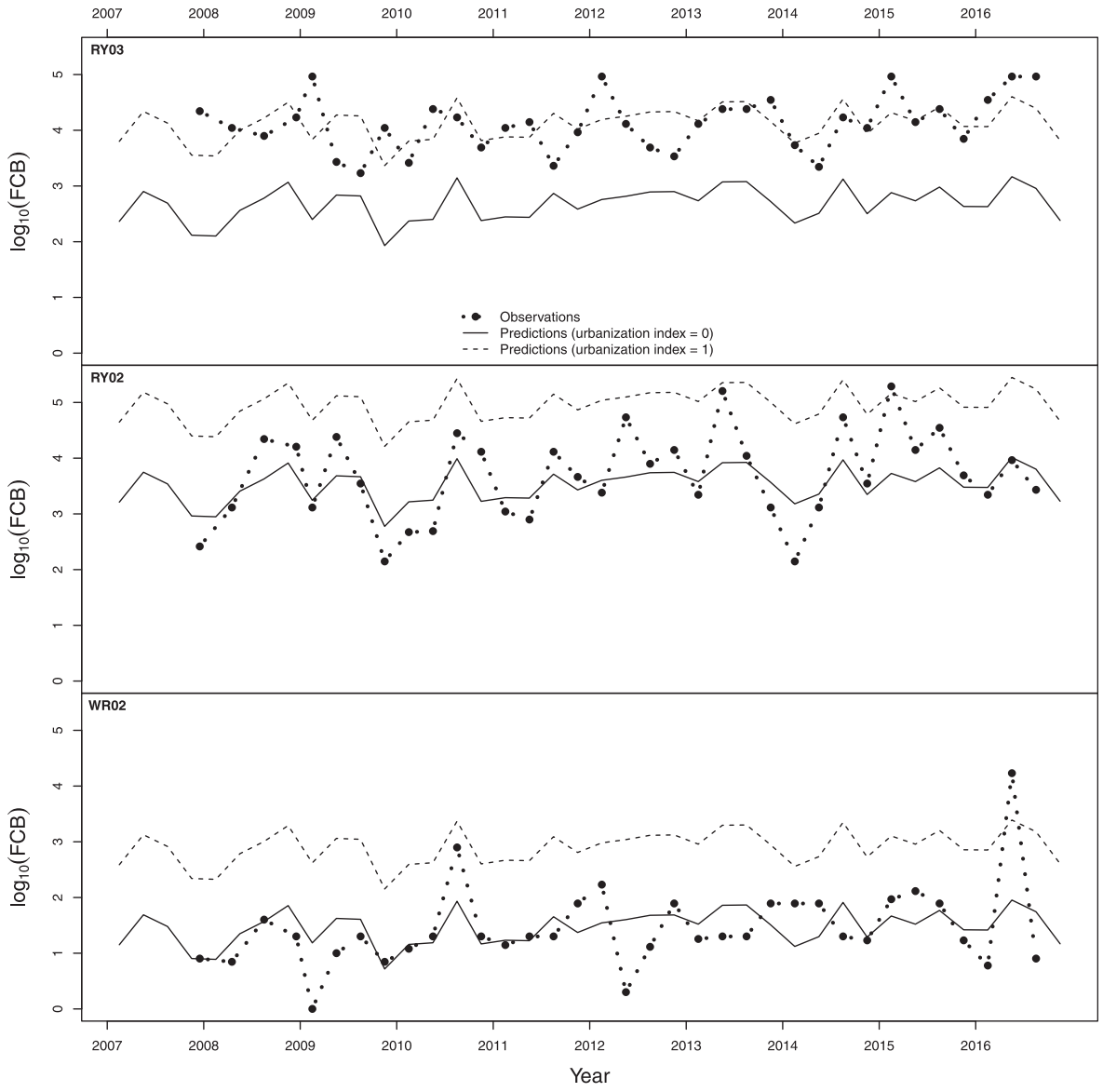


Fig. 8. Predicted values of FCB (in MPN/100 mL) for three stations.

It is difficult to compare our results with those from the literature because of possible (unreported) differences in the field and laboratory protocols, which could add some systematic or random variation in the data. Furthermore, the different methods may have different specificities towards bacterial species. On the other hand, the particularities of different studies (e.g. climate, landscape) may help to better understand the mechanisms controlling bacterial concentrations in rivers. Keeping in mind these caveats, previous studies reported maximum values of MPN equal to 2420 MPN/100 mL in Canada (Butler et al., 2014), 5172 MPN/100 mL in Ecuador (Rao et al., 2015), 8000 MPN/100 mL in Belgium (de Brauwere et al., 2011), and 10,000 MPN/100 mL in Netherlands (Vermeulen and Hofstra, 2014) or in a rural area of Laos (Causse et al., 2015). Lyautey et al. (2010) reported 2004 measures of *E. coli* concentration in Ontario (Canada) with annual means between 10 MPN/100 mL and 1000 MPN/100 mL (global median: 64 MPN/100 mL) and only three measures larger than 10,000 MPN/100 mL. Jeong et al. (2019) found that 67% of the *E. coli* in a watershed in Texas (U.S.A.) was contributed by wildflife. However, increase in wastewater discharges and urban lands due to population growth resulted in increased concentration of *E. coli* (Jeong et al., 2019). By contrast with these values from the literature, our results showed large values of MPNs, particularly for TCB, with 38 out of 63 stations (60%) showing a mean TCB above 10,000 MPN/100 mL. Most of the studies cited above were conducted in temperate countries or in rural tropical areas (Laos and Ecuador), which could explain these differences. Besides, our study covered several watersheds and a substantial geographic area and thus the data we analyzed were probably more heterogeneous than those in these studies. As an illustration of this point, our models predicted a mean TCB around 100 MPN/100 mL for some stations (CB01, WR01) and above 50,000 MPN/100 mL for others (e.g. BPT01). Nevertheless, it might be very useful to further assess the precision of the measures reported here by analyzing some samples with different methods. Because of its location, BPT01 receives impacts from several land use activities, mainly shrimp and fish aquaculture, compared to CB01 and WR01 where agriculture is the main activity.

Urban growth and its environmental impacts are very general issues in many countries. Several studies have shown that the growth of urban areas is not random and usually follows a dendritic

pattern known as urban sprawl (Pan et al., 2020; Xu et al., 2020). Studies in China, North America, and Europe showed that the impact of urbanization is very likely to have different impacts on different areas (Tang et al., 2005; Du and Huang, 2017). This impact is also very likely to differ with respect to sociological factors, for instance in relation to commuting behaviours which is typical of many large European and North American urban systems (Pan et al., 2019). The growth of urban areas is expected to increase the proportion of impervious surfaces as well as surface water runoff (Tang et al., 2005).

In the present study, we used a statistical approach based on linear mixed-effect models which was motivated by the geographical spread of the stations which covered several watersheds. The inclusion of random effects (here 'Station' and 'Watershed') makes it possible to fit complex models that would be impossible to fit with standard linear models. Additionally, linear models, even with random effects, make it possible to formulate predictions in a straightforward way which are easily interpretable.

An important result from our analyses is the importance of urbanization as a factor controlling concentrations of bacteria in rivers. Our models predict that TCB and FCB would be multiplied by 20 and 25, respectively, if land is transformed from non-urban to fully urban (i.e. urbanization index changed from zero to one). One motivation of monitoring water quality in this study area is its control in relation to the development of the Eastern Seabord (ESB) or the Eastern Economic Corridor (EEC). The transformation of the area started in 1982 and ESB has now the largest number of industrial estates in Thailand (i.e. 18 out of 38; Soytong and Perera, 2017). ESB, together with the Bangkok area, is a recipient of rural migrants from other regions of Thailand (Jampaklay et al., 2007). Special economic zones (SEZ) with extensive industrial and urban development have undeniable social and economic benefits, but their environmental sustainability is generally questionable (see review in Ślusarczyk and Grondys, 2018). Previous studies have focused on social conflict arising from industrialization in ESB (Soytong and Perera, 2017). Zhang and Tripathi (2018) found a relationship between atmospheric pollution by fine particulate matter and lung cancer in Eastern Thailand (actually a subset of the area considered in the present study). Therefore, health hazard is likely to be a general concern in ESB as in other SEZs. Other studies showed the importance of urban areas in the concentrations of bacteria in rivers (Wu and Jackson, 2016), lakes (Newton and McLellan, 2015), or creeks (DiDonato et al., 2009).

5. Conclusions

The present study shows that coliform bacterial concentrations in waters of Eastern Thailand varied considerably both spatially and temporally. Temporal variation was related to both seasonal and annual variation with the highest levels of bacterial concentrations measured in summer. These concentrations varied extensively among stations and among watersheds. The concentrations of fecal coliforms were greater than 4000 MPN/100 mL, the upper limit for consumable surface water, in 20% of the measures. The study also shows that urbanization is a very significant driver of bacterial concentrations in the rivers of Eastern Thailand and is predicted to cause a 20-fold increase of these concentrations. These results bring important information for land management. First, wastewater treatment in urban areas is a crucial element in improving water quality. Second, it is crucial to investigate the proximate, causal factors that affect TCB and FCB as a consequence of urbanization. Other components such as cleaner production, legal framework, institutional and financial arrangements, monitoring and enforcement, and cooperation with related agencies and local communities, should be taken into account as well. Most of all, an ecosystem approach is the key to development planning strategies as rivers themselves provide several basic ecosystem services.

CRediT authorship contribution statement

Chantima Piyapong: Writing – review & editing, Project administration, Funding acquisition. **Nitcha Chamroensaksri:** Conceptualization, Writing – original draft. **Sayam Aroonsrimorakot:** Supervision. **Lawan Eyosawat:** Conceptualization, Investigation. **Surasak Khankhum:** Conceptualization, Writing – review & editing. **Sunirat Rattana:** Conceptualization, Investigation, Writing – original draft. **Nuchsupha Sunthamala:** Conceptualization, Writing – review & editing. **Panya Warapetcharayut:** Supervision, Resources, Data curation. **Emmanuel Paradis:** Software, Formal analysis, Visualization, 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.

Acknowledgements

We are grateful to all the staff members who contributed to the data from the Regional Environmental Office 13 (Chonburi), Ministry of Natural Resources and Environment, including water sampling and laboratory analyses. We thank five anonymous reviewers for their constructive comments on previous versions of our paper. Many thanks to Colin Tosh for checking and correcting the English. Financial support was provided by a research grant from Burapha University through the National Research Council of Thailand (grant number: March 57, 2562). This is publication ISEM 2021-041 SUD.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2021.126704.

References

Abraham, W.R., 2011. Megacities as sources for pathogenic bacteria in rivers and their fate downstream. Int. J. Microbiol. 2011 https://doi.org/10.1155/2011/798292, 798292.

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle, in: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the Second International Symposium on Information Theory. Akadémia Kiado, Budapest, pp. 267–281.

American Public Health Association, 1992. Standard Methods for the Examination of Water and Wastewater, eighteenth ed. American Public Health Association, American Water Works Association, Water Environment Federation, Washington DC, USA.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Bivand, R., Keitt, T., Rowlingson, B., 2018. Rgdal: bindings for the 'geospatial' data abstraction library. URL: https://CRAN.R-project.org/package=rgdal. Rpackage version 1.3-6.

de Brauwere, A., de Brye, B., Servais, P., Passerat, J., Deleersnijder, E., 2011. Modelling *Escherichia coli* concentrations in the tidal Scheldt river and estuary. Water Res. 45, 2724–2738. https://doi.org/10.1016/j.watres.2011.02.003.

Butler, S., Webster, T., Redden, A., Rand, J., Crowell, N., Livingstone, W., 2014. Using remote sensing to identify changes in land use and sources of fecal bacteria to support a watershed transport model. Water 6, 1925. https://doi.org/10.3390/w6071925.

Cabral, J.P., Marques, C., 2006. Faecal coliform bacteria in Febros river (Northwest Portugal): temporal variation, correlation with water parameters, and species identification. Environ. Monit. Assess. 118, 21–36. https://doi.org/10.1007/s10661-006-0771-8.

Causse, J., Billen, G., Garnier, J., Henri-des Tureaux, T., Olasa, X., Thammahacksa, C., Latsachak, K.O., Soulileuth, B., Sengtaheuanghoung, O., Rochelle-Newall, E.,

Ribolzi, O., 2015. Field and modelling studies of Escherichia coli loads in tropical streams of montane agro-ecosystems. J. Hydro-Environ. Res. 9, 496–507. https://doi.org/10.1016/j.jher.2015.03.003.

DiDonato, G.T., Stewart, J.R., Sanger, D.M., Robinson, B.J., Thompson, B.C., Holland, A.F., Van Dolah, R.F., 2009. Effects of changing land use on the microbial water quality of tidal creeks. Mar. Pollut. Bull. 58, 97–106. https://doi.org/10.1016/j.marpolbul.2008.08.019.

Diwan, V., Hanna, N., Purohit, M., Chandran, S., Riggi, E., Parashar, V., Tamhankar, A.J., Lundborg, C.S., 2018. Seasonal variations in water-quality, antibiotic residues, resistant bacteria and antibiotic resistance genes of *Escherichia coli* isolates from water and sediments of the Kshipra River in Central India. Int. J. Environ. Res. Publ. Health 15, 1281. https://doi.org/10.3390/ijerph15061281.

Du, X.J., Huang, Z.H., 2017. Ecological and environmental effects of land use change in rapid urbanization: the case of Hangzhou, China. Ecol. Indicat. 81, 243–251. https://doi.org/10.1016/j.ecolind.2017.05.040.

Dumrongthai, P., 2019. Progress of water environment governance in Thailand. URL: http://wepa-db.net/3rd/en/meeting/20190222/pdf/D2_S3_Thailand.pdf. accessed: 2019-12-20.

Eastern Economic Corridor Office, 2019. Eastern economic corridor development. URL: http://www.eeco.or.th/en/background. accessed: 2019-02-20.

Hijmans, R.J., 2018. raster: geographic data analysis and modeling. URL: https://CRAN.R-project.org/package=raster.R package version 2.7-15.

Jampaklay, A., Korinek, K., Entwisle, B., 2007. Residential clustering among Nang Rong migrants in urban settings of Thailand. Asian Pac. Migrat. J. 16, 485–510. https://doi.org/10.1177/011719680701600403.

Jeong, J., Wagner, K., Flores, J.J., Cawthon, T., Her, Y., Osorio, J., Yen, H., 2019. Linking watershed modeling and bacterial source tracking to better assess *E. coli* sources. Sci. Total Environ. 648, 164–175. https://doi.org/10.1016/j.scitotenv.2018.08.097.

Lyautey, E., Lu, Z., Lapen, D.R., Wilkes, G., Scott, A., Berkers, T., Edge, T.A., Topp, E., 2010. Distribution and diversity of *Escherichia coli* populations in the south nation river drainage basin, eastern Ontario, Canada. Appl. Environ. Microbiol. 76, 1486–1496. https://doi.org/10.1128/AEM.02288-09.

Newton, R.J., McLellan, S.L., 2015. A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. Front. Microbiol. 6, 1028. https://doi.org/10.3389/fmicb.2015.01028.

Pan, H.Z., Page, J., Zhang, L., Chen, S., Cong, C., Destouni, G., Kalantari, Z., Deal, B., 2019. Using comparative socio-ecological modeling to support Climate Action Planning (CAP). J. Clean. Prod. 232, 30–42. https://doi.org/10.1016/j.jclepro.2019.05.274.

Pan, H.Z., Page, J., Zhang, L., Cong, C., Ferreira, C., Jonsson, E., Näsström, H., Destouni, G., Deal, B., Kalantari, Z., 2020. Understanding interactions between urban development policies and GHG emissions: a case study in Stockholm Region. Ambio 49, 1313–1327. https://doi.org/10.1007/s13280-019-01290-y.

Paradis, E., 2018. Nonlinear relationship between biodiversity and human population density: evidence from Southeast Asia. Biodivers. Conserv. 27, 2699–2712. https://doi.org/10.1007/s10531-018-1563-5.

Paradis, E., 2020. Modelling transition in land cover highlights forest losses and gains in Southeast Asia. Biodivers. Conserv. 29, 2539–2551. https://doi.org/10.1007/s10531-020-01987-7.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-effects Models in S and S-PLUS. Springer, New York.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org.

Rao, G., Eisenberg, J.N.S., Kleinbaum, D.G., Cevallos, W., Trueba, G., Levy, K., 2015. Spatial variability of *Escherichia coli* in rivers of northern coastal Ecuador. Water 7, 818–832. https://doi.org/10.3390/w7020818.

Robinson, G.K., 1991. The BLUP is a good thing: the estimation of random effects (with discussion). Stat. Sci. 6, 15–51.

Rochelle-Newall, E., Nguyen, T.M.H., Le, T.P.Q., Sengteheuanghoung, O., Ribolzi, O., 2015. A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions. Front. Microbiol. 6, 308. https://doi.org/10.3389/fmicb.2015.DD3D8.

Rochelle-Newall, E.J., Ribolzi, O., Viguier, M., Thammahacksa, C., Silvera, N., Latsachack, K., Dinh, R.P., Naporn, P., Sy, H.T., Soulileuth, B., Hmaimum, N., Sisouvanh, P., Robain, H., Janeau, J.L., Valentin, C., Boithias, L., Pierret, A., 2016. Effect of land use and hydrological processes on *Escherichia coli* concentrations in streams of tropical, humid headwater catchments. Sci. Rep. 6 https://doi.org/10.1038/srep32974, 32974.

Shatkin, G., 2004. Globalization and local leadership: growth, power and politics in Thailand's Eastern Seaboard. Int. J. Urban Reg. Res. 28, 11–26. https://doi.org/10.1111/j.0309-1317.2004.00500.x.

Shukla, A.K., Ojha, C.S.P., Mijic, A., Buytaert, W., Pathak, S., Garg, R.D., Shukla, S., 2018. Population growth, land use and land cover transformations, and water quality nexus in the Upper Ganga River basin. Hydrol. Earth Syst. Sci. 22, 4745–4770. https://doi.org/10.5194/hess-22-4745-2018.

Simachaya, W., 2003. A decade of water quality monitoring in Thailand's four major rivers: the results and the implications for management, in: proceedings of the Sixth International Conference on the Environmental Management of Enclosed Coastal Seas (Bangkok, Thailand, November 18–21, 2003), pp. 1–11. URL: http://infofile.pcd.go.th/water/10yr4majoremecs2003.pdf.

Simachaya, W., Watanamahart, P., Kaewkrajang, V., Yenpiem, A., 2000. Water quality situation in the chao phraya delta, in: the chao phraya delta: historical development, dynamics and challenges of Thailand's rice bowl, pp. 199–218. URL: http://www.documentation.ird.fr/hor/fdi:010024645.

Ślusarczyk, B., Grondys, K., 2018. The concept of sustainable development in the functioning of municipalities belonging to special economic zones in Poland. Sustainability 10, 2169. https://doi.org/10.3390/su10072169.

Song, X.P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. Nature 560, 639–643. https://doi.org/10.1038/s41586-018-0411-9.

Soytong, P., Perera, R., 2014. Use of GIS tools for environmental conflict resolution at map Ta phut industrial zone in Thailand. Sustainability 6, 2435–2458. https://doi.org/10.3390/su6052435.

Soytong, P., Perera, R., 2017. Spatial analysis of the environmental conflict between state, society and industry at the Map Ta Phut-Rayong conurbation in Thailand. Environ. Dev. Sustain. 19, 839–862. https://doi.org/10.1007/s10668-016-9768-9.

St Laurent, J., Mazumder, A., 2012. The influence of land-use composition on fecal contamination of riverine source water in southern British Columbia. Water Resour. Res. 48 https://doi.org/10.1029/2012WR012455. W00M03.

Tang, Z., Engel, B.A., Pijanowski, B.C., Lim, K.J., 2005. Forecasting land use change and its environmental impact at a watershed scale. J. Environ. Manag. 76, 35–45. https://doi.org/10.1016/j.jenvman.2005.01.006.

Taylor, D., 2010. Biomass burning, humans and climate change in Southeast Asia. Biodivers. Conserv. 19, 1025–1042. https://doi.org/10.1007/s10531-009-9756-6.

Vermeulen, L.C., Hofstra, N., 2014. Influence of climate variables on the concentration of *Escherichia coli* in the rhine, meuse, and drentse aa during 1985–2010. Reg. Environ. Change 14, 307–319. https://doi.org/10.1007/s10113-013-0492-9.

Water Environmental Partnership in Asia, 2018. Outlook on water environmental management in Asia 2018. URL: http://wepa-db.net/en/publication/2018_outlook/index.html. accessed: 2019-02-21.

Wu, J.Y., Jackson, L., 2016. Association of land use and its change with beach closure in the United States, 2004-2013. Sci. Total Environ. 571, 67–76. https://doi.org/10.1016/j.scitotenv.2016.07.116.

Xu, C., Haase, D., Su, M., Wang, Y., Pauleit, S., 2020. Assessment of landscape changes under different urban dynamics based on a multiple-scenario modeling approach. Environ. Plan. B: Urban Anal. City Sci. https://doi.org/10.1177/2399808320910161.

Zhang, H.R., Tripathi, N.K., 2018. Geospatial hot spot analysis of lung cancer patients correlated to fine particulate matter ($PM^{2.5}$) and industrial wind in Eastern Thailand. J. Clean. Prod. 170, 407–424. https://doi.org/10.1016/j.jclepro.2017.09.185.