Addressing and reducing parameter uncertainty in quantitative microbial risk assessment by incorporating external information via Bayesian hierarchical modeling

Seis, Wolfgang; Rouault, Pascale; Medema, Gertjan

DOI
10.1016/j.watres.2020.116202

Publication date
2020

Document Version
Final published version

Published in
Water Research

Water Research 185 (2020) 116202

Contents lists available at ScienceDirect

Water Research

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

Addressing and reducing parameter uncertainty in quantitative microbial risk assessment by incorporating external information via Bayesian hierarchical modeling

Wolfgang Seis[a,b,*], Pascale Rouault[a], Gertjan Medema[b]

[a] Kompetenzzentrum Wasser Berlin gGmbH, Cicerostraße 24, 10709, Berlin, Germany
[b] Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 13 November 2019
Revised 17 July 2020
Accepted 18 July 2020
Available online 19 July 2020

Keywords:
Bayesian hierarchical modeling
Meta-analysis
Quantitative microbial risk assessment
Norovirus
Performance target
Combining information

ABSTRACT

Probabilistic quantitative microbial risk assessment (QMRA) studies define model inputs as random variables and use Monte-Carlo simulation to generate distributions of potential risk outcomes. If local information on important QMRA model inputs is missing, it is widely accepted to justify assumptions about these model inputs by using external literature information. A question, which remains unexplored, is the extent to which previously published external information should influence local estimates in cases of nonexistent, scarce, and moderate local data. This question can be addressed by employing Bayesian hierarchical modeling (BHM). Thus, we study the effects and potential benefits of BHM on risk and performance target calculations at three wastewater treatment plants (WWTP) in comparison to alternative statistical modeling approaches (separate modeling, no-pooling, complete pooling). The treated wastewater from the WWTPs is used for restricted irrigation, potable reuse, or influences recreational waters, respectively. We quantify the extent to which external data affects local risk estimations in each case depending on the statistical modeling approach applied. Modeling approaches are compared by calculating the pointwise expected log-predictive density for each model. As reference pathogens and example data, we use locally collected Norovirus genogroup II data with varying sample sizes ($n = 4$, $n = 7$, $n = 27$), and complement local information with external information from 44 other WWTPs ($n = 307$). Results indicate that BHM shows the highest predictive accuracy and improves estimates by reducing parameter uncertainty when data are scarce. In such situations, it may affect risk and performance target calculations by orders of magnitude in comparison to using local data alone. Furthermore, it allows making generalizable inferences about new WWTPs, while providing the necessary flexibility to adjust for different levels of information contained in the local data. Applying this flexible technique more widely may contribute to improving methods and the evidence base for decision-making in future QMRA studies.

1. Introduction

Quantitative microbial risk assessment (QMRA) was developed as a common framework to assess risks caused by pathogenic microorganisms. QMRA aims at supporting decision-making related to the microbial safety of water systems. The method supports decision-making by quantifying health risks in local settings on a system level (Gonzales-Gustavson et al., 2019) as well as by deriving general statements on health risks and pathogen reduction targets as done e.g. for recreational waters and water reuse systems (Boehm et al., 2018; Soller et al., 2017).

Probabilistic QMRA studies define model inputs as random variables and use Monte-Carlo simulation to generate distributions of potential risk outcomes (WHO, 2016). However, in QMRA studies, local data may be limited (McBride et al., 2013) or completely absent (Mok et al., 2014; Soller et al., 2017). In such situations, distribution parameters for these random variables cannot be estimated precisely. Appropriate parameter estimates, however, are crucial for the robustness of simulation results and thus affect risk-based decision-making. In QMRA studies, deriving assumptions about point estimates for these parameters from literature is a widely accepted approach when local data is completely absent. Systematic scientific reviews and meta-analyses on important model inputs for QMRA are valuable pieces of information in these situations. For example, a recently published meta-analysis of the occurrence of norovirus (NoV) in wastewater by

* Corresponding author.
E-mail address: wolfgang.seis@kompetenz-wasser.de (W. Seis).

https://doi.org/10.1016/j.watres.2020.116202
0043-1354/

Table 1
Overview of locally available data.

Wastewater treatment plant	Health context	N sample size	Laboratory	*Sampling period*	*Sample preparation*	WWTP ID
Germany	*Wastewater is used for restricted agricultural irrigation*	7	*University of Barcelona*	*October-December 2014*	*Skimmed milk flocculation, as in Calgua et al. (2013)*	100
Spain	*Indirect potable reuse via managed aquifer recharge*	4	*University of Barcelona*	*February, April and June 2016*	*Skimmed milk flocculation as in Calgua et al. (2013)*	101
Germany	*Wastewater impacts recreational water under rain weather conditions*	27	*German Federal environment agency*	*November 2015-October 2018*	*Glass wool filtration as in (Wyn-Jones et al., 2011)*	102

Eftim et al. (2017) was used to inform QMRA studies on the risk of illness due to exposure to recreational waters contaminated with aged sewage (Boehm et al., 2018), and to assess the suitability of different wastewater treatment schemes to be used for direct potable reuse (Soller et al., 2017). However, the extent to which such external information influences local estimates against the background of nonexistent, scarce, and moderate local data is rarely explicitly addressed. This question can be addressed by employing Bayesian hierarchical modeling (BHM). While this technique has been applied e.g. to derive a generalized dose-repose relation for Adenovirus (Teunis et al., 2016), the evaluation of microbial recovery rates (Petterson et al., 2009), and the evaluation of treatment performances of ultrafiltration (Carvajal et al., 2017), the relation between locally available data on pathogen concentrations in water that is supported by information from external sources has not been addressed in the water-related QMRA literature so far.

Therefore, the objective of the present study is to investigate the effects of incorporating external pathogen data via BHM to support estimates of local pathogen concentrations in source waters and its effect on resulting health risks calculations. Results are compared to more commonly used estimation approaches (cf. Section 2.2).

2. Methods

In the present study, we investigate the effect of how different statistical approaches to including external information from the scientific literature may affect local risk and log-reduction calculations. We use Norovirus genogroup II (NoVII) data collected at three local municipal wastewater treatment plants (WWTP) as reference pathogens because NoV is among the most relevant causes of infection risk and dominated risk calculations in cases where multiple pathogens have been assessed (Boehm et al., 2018; Soller et al., 2017). While there are comparable estimation techniques available from classical statistics, we use Bayesian estimation methods based on Markov Chain Monte Carlo (MCMC) for all statistical analyses. At all locations, we estimate the required log-reductions of water treatment based on WWTP inflow concentrations for achieving predefined health targets (cf. 2.5).

2.1. Local data and external information

Local data: Local data were made available from three different wastewater treatment plants located in Europe. At all locations, there is a direct health relevance because the treated wastewater is either reused for irrigation or drinking water supply or directly influences nearby bathing waters (Table 1). The amount of local data varies between 4 data points at the WWTP for indirect potable reuse and 27 data points at the WWTP, which impacts recreational waters. Due to the different local sample sizes, large differences in parameter uncertainty can be expected. Samples were analyzed by laboratories in Germany or Spain using qPCR (Table 1).

2.2. External information

As external information, we use a systematic review and meta-analysis on NoV WWTP influent concentrations by Eftim et al. (2017). The authors analyzed the collected NoV data for geographical (Asia, Europe, North America, Oceania) as well as seasonal differences. Data were provided either directly from authors, collected from tables, or digitalized from graphs. For estimating the mean (μ) and the standard deviation (σ) of the population distribution of NoV observations of genogroups I and II, the authors grouped the data by continent and season and applied a bootstrapping approach on the grouped data. Based on this evaluation the authors reported a statistically significant difference between data from Europe and North America. For Europe, the authors collected 305 observations. We followed the procedure of Eftim et al. (2017), collecting NoVII data for European WWTPs either from reported tables or digitalizing them using the same software (GetData Graph Digitalizer). As our study focuses on estimates on the level of local WWTPs, we additionally grouped the data by WWTP (see SI). This is a major difference between our approach and the one published by Eftim et al. (2017), who stayed at a continent-level for data evaluation.

2.3. Statistical approaches and assumptions for local estimates

In statistical inference, observations, i.e. data, are used to estimate the unknown parameters of a statistical model; in the simplest form the parameters of an assumed probability distribution. Therefore, the degree to which external information influences local estimates inherently involves specific statistical assumptions about the relationship between these parameters at a local and general level. In previous meta-analyses and QMRA studies, various methods have been used to derive generalized estimates of the parameters of interest, including complete pooling (Boehm et al., 2018; Eftim et al., 2017), no-pooling (Boehm et al., 2019) and hierarchical modeling (Pouillot et al., 2015). If local, site-specific data are available, generally separate modeling is applied (Gonzales-Gustavson et al., 2019; McBride et al., 2013). To evaluate the influence of different modeling assumptions, we used the modeling approaches described below. We use the indices “i” for individual observations, “j” for individual WWTPs, and “NoV” for cases in which data are pooled across WWTPs, respectively. Concentrations are assumed to be lognormally distributed, which is a common assumption (Amoueyan et al., 2019; Pouillot et al., 2015; Soller et al., 2018) especially in situations were data a too scarce to decide between other potential candidate distributions (Boehm et al., 2018).

Separate models: Each $WWTP_j$ is regarded as being completely independent and results from one location do not influence the ex-

pectation of the other. Results are neither generalizable nor transferable.

Model:

$$\lg(NoV_i) \sim N(\mu_{j[i]}, \sigma^2_{j[i]})\ \mathrm{i} = 1....\mathrm{n}, \mathrm{j} = 1...\mathrm{J} \quad (1)$$

Complete pooling: The WWTPs and the applied analytical procedures are assumed to be completely comparable and observations from different WWTPs are assumed to come from a common population. They are assumed to be exchangeable with a common mean and variance.

$$\lg(NoV_i) \sim N(\mu_{NoV}, \sigma^2_{NoV})\ i = 1....\mathrm{n} \quad (2)$$

No pooling: For every WWTP j a separate mean μ_j is fitted, but the residual standard deviation is constant across WWTPs.

Model:

$$\lg(NoV_i) \sim N(\mu_{j[i]}, \sigma^2_{NoV})\ i = 1....\mathrm{n} \quad (3)$$

Partial pooling μ via classical hierarchical modeling: Within the classical Bayesian hierarchical (multilevel) framework, a separate mean μ_j is estimated for every WWTP j. The main additional assumption is that the local means μ_j come from a common distribution, which in turn is described by a normal distribution. The parameters of this common distribution η and τ^2 are referred to as hyperparameters. This population distribution can be regarded as a prior distribution for the local means μ_j. It is estimated from the data and does not depend on subjective choices of the analyst. The variance τ^2 represents the estimated between-WWTP-variability and can be used to estimate expectations of new WWTPs. By estimating all parameters on a data-level and WWTP-level simultaneously, information is shared across studies, which is referred to as "partial pooling". The within-WWTP-variability, i.e. the data variance σ^2_{NoV} is considered constant across groups.

Model:

$$\begin{aligned} &\lg(NoV_i) \sim N(\mu_{j[i]}, \sigma^2_{NoV})\ i = 1....\mathrm{n} \\ &\mu_j \sim N(\eta, \tau^2)\ \ j = 1, ...J \end{aligned} \quad (4)$$

Partial pooling μ and σ^2 via extended hierarchical modeling:

The classical hierarchical modeling approach can easily be extended by letting the individual within WWTP variances σ_j vary by WWTP. This leads to a further extension of the model by an additional distribution (estimated on a log-scale since variances are restricted to be positive) and additional hyperparameters (α, γ) which represent the average population variance and the between-WWTP-variability of variances.

Model:

$$\begin{aligned} &\lg(NoV_i) \sim N(\mu_{j[i]}, \sigma^2_{j[i]})\ \ i = 1,....\mathrm{n} \\ &\mu_j \sim N(\eta, \tau^2)\ \ j = 1, ...J \\ &\sigma^2_j \sim lognormal(\alpha, \gamma^2) j = 1, ...J \end{aligned} \quad (5)$$

Separate point estimates

To investigate the effect of parameter uncertainty in general, we additionally included the point estimates estimated with maximum likelihood derived from the local data, without accounting for parameter uncertainty.

Treatment of censored data

In the local data, no data points below the LOD were measured. However, in the literature, a small number of observations were below the LOD (see SI). For Bayesian inference, left-censored data were handled by replacing the normal likelihood of these observations with the complementary cumulative normal distribution function. A similar approach has been applied by Pouillot et al. (2015).

Model fitting

We use Bayesian estimation methods based on Markov Chain Monte Carlo (MCMC) for all statistical analyses. To run MCMC simulations, we use the programming languages *"R", "Stan"* (StanDevelopmentTeam, 2017) via the interface between *"R"* and *"Stan", "brms" (Bürkner, 2017).* "Stan" implements Hamiltonian Monte Carlo for MCMC. We used the same weakly informative prior distributions for the location and scale parameters (see SI). Due to the large number of WWTP with small sample sizes a small step size (< 0.99) was chosen. For every model, we ran four independent Markov Chains, with 20,000 iterations and a warm-up phase of 10,000. Thereby, 40,000 independent posterior samples were created for further evaluation. We chose this high number of samples as the posterior samples are not only used for estimation but also for subsequent risk simulation (cf. Section 2.5). Convergence of the Markov Chains were checked by inspecting whether the trace-plots of the four chains were well-mixed and by checking whether the Gelman-Rubin diagnostic statistic ($\hat{R}$) has converged to 1.

2.3. Calculating quantities of interest

As quantities of interest, we consider the marginal distributions of μ and σ of NoVII concentrations and the posterior predictive distribution (PPD), which represents the uncertainty about new observations, given the observed data and included external information. The average inflow concentration is of particular interest to determine the required log-credits at a local setting, as these are calculated from the difference between the average inflow and outflow concentrations. The PPD is used for subsequent forward Monte Carlo risk simulation.

2.4. Model comparison and generalized estimates for a new (unknown) WWTP

To compare the effects of the different statistical approaches on estimates for new, i.e. unknown WWTPs, the marginal distributions of the mean as well as the posterior predictive distribution for new data are simulated from the estimates for η, α, γ and τ for the hierarchical approaches and μ_{Nov} and σ_{NoV} for the complete pooling approach. For the "separate modeling" and "no pooling" approaches such estimations are not possible as these approaches assume independence of WWTPs. To assess the predictive performance of the different modeling approaches we calculate the expected pointwise leave-one-out *(loo)* log predictive density (elpd) using pareto-smooth importance sampling according to Vehtari et al. (2017). The higher a model's elpd, the more accurate the model's predictions are to be expected. Consequently, the model with the highest elpd is preferred. The loo-elpd values are calculated by applying the *"loo"* function from the *brms-package* to the fitted model objects. Models are refitted if k-values are > 0.7. Model comparison is conducted by applying the function *loo-compare,* which calculates the difference in the elpd between models.

2.5. Implications on risk calculation and performance targets

To estimate the effect of different statistical estimation approaches on risk simulation and thus performance target derivation, we run the following simulation:

- Simulation of the PPD of NoV influent concentrations for each local WWTP and each estimation approach (40,000 draws)
- Application of assumed log-reductions from source water to the point of exposure (Table 2)
- Application of use-specific exposure scenarios in terms of the volume ingested per event and number of exposure events per year (Table 2)
- Risk simulation and comparison of results to existing risk-based health targets (Table 2)

Table 2
Overview of site-specific exposure scenarios, health targets, and risk endpoints.

Application	Potable reuse	Restricted irrigation	Recreational water	Source
Endpoint risk assessment	Infection pppy	Disability-adjusted life years (DALYs) pppy	Single exposure risk of illness	Dutch drinking water regulation,
Health target	10^{-4} pppy	10^{-4}- 10^{-6} pppy	3%	(WHO, 2006)
N	365	50	1	(US EPA, 2012; NRMMC-EPHC-AHMC, 2006)
Volume ingested per exposure event	1L	1mL	100mL	(NRMMC-EPHC-AHMC, 2006) (Boehm et al., 2018)
P (illness|infection)		0.7	0.7	(WHO, 2011)
DALYs per case of disease		1.3×10^{-3}		(Mara and Sleigh, 2010)~
Log reduction	12–16	5–10	5–9	Assumption

- Repetition of the procedure to account for stochastic Monte-Carlo uncertainty

To isolate the effect of different statistical estimation approaches, we assume that the remaining model inputs are perfectly known without variation (Table 2). No specific treatment train is assumed but log-reductions include both technical barriers, and natural die-off from the WWTP inflow to the point of exposure. From the predicted inflow concentrations (PPD), the population distribution of the expected doses per exposure event is calculated by:

$$d = 10^{c_{influent} - \sum_{T=1}^{n} LRV_T} * V_{ingested} \quad (6)$$

where d is the dose per exposure event, $c_{influent}$ is the PPD of NoVII in lg GC/L, LRV is the log-removal value for treatment step T, n is the number of treatment steps, and $V_{ingested}$ is the ingested volume in L. From the population distribution of d, we calculate the population distribution of risk outcomes per exposure event, using the dose-response relation published by Teunis et al. (2008) for disaggregated viruses (Eq. (7)) (see SI). Note, the dose-response relationship of norovirus is subject to ongoing discussion and alternative relationships have been proposed (Messner et al., 2014; Schmidt, 2015; Van Abel et al., 2017).

$$P_{\text{inf|exposure}} = 1 - {}_1F_1(a,\ a+b,\ -d) \quad (7)$$

Subsequently, we estimate the sampling distribution of the average annual infection risk (P_{annual}) by randomly ("rand" in Eq. (8)) resampling N (number of exposure events) times from the population distribution of event probabilities for 10,000 times. To account for uncertainties caused by stochastic Monte Carlo sampling we repeat the full process, starting from sampling from the PPD for 100 times for the irrigation and swimming scenario, and 1000 times for the drinking water scenario. This leads to 100 and 1000 distributions of the average annual risk for each assumed LRV consisting of 10,000 calculations. For every distribution, the proportion of samples below the health targets is calculated (cf. Section 3.2).

$$P_{annual} = 1 - \prod_{i=1}^{N} \left(1 - rand(P_{\text{inf|exposure[i]}}\right) \quad (8)$$

3. Results

Fig. 1 shows all available data, which originate from 44 different WWTPs across Europe. The empirical geometric mean of the different WWTP varies between approximately 4.1 and 7.6 lg GC/L, while the variation of individual observations within a single WWTP varies by up to 6 orders of magnitude (WWTP 44). The three local WWTPs (100, 101, 102) have empirical means of 5.8 lg GC/L, 4.95 lg GC/L, and 5.98 lg GC/L, respectively. These values differ from the mean calculated from all data (5.42 lg GC/L) by approximately 0.5 log in both directions. The number of data points per WWTP varies between 2 and 49. With 27 data points, the local WWTP 102 is one of the largest data sets on a WWTP level, while with 4 data points WWTP 101 only provides little local information.

3.1. Marginal parameter distributions and PPD for new observations

Figs. 2 and 3 summarize the estimates and corresponding uncertainty of the parameters μ and σ for the local WWTPs. Fig. 4 shows the resulting PPD for new observations. Since in Bayesian inference probability is used to express uncertainty, all parameters have probability distribution themselves, referred to as marginal distributions.

Fig. 2 shows the marginal distributions of the mean NoVII concentrations. The estimates of the different approaches differ both regarding the location of the mean (μ) and the scale of the distribution, which reflects the uncertainty of the estimated mean.

Regarding the location, the "complete pooling" approach delivers the same estimate for all WWTPs (μ = 5.4 lg GC/L). The scenarios "separate point estimate" (dashed line), "separate modeling", and "no pooling" estimate identical means for the individual WWTPs, which correspond to the empirical means of the local data, as the three approaches estimate mean concentrations, separately. It can be observed, that the individual local point estimates differ from the "complete pooling" estimate by approximately 0.5 log units for WWTPs 101 and 102, while the difference is smaller (approximately 0.35 log units) for WWTP 100. This difference shows that by applying a "complete pooling" approach, which assumes that all data come from a common WWTP, the 305 data points from literature outweigh the local data even with a sample size of $N = 27$.

The two hierarchical modeling approaches, in turn, lead to point estimates (location) which lie between the "complete pooling" and the "separate estimates". The exact location of the estimate depends on the amount of local information, i.e. data per WWTP, and the variance ratio of the local data variances $\sigma^2_{j[i]}$ and the between WWTP variance τ^2. For WWTP 102 the estimated modes from hierarchical modeling are virtually identical to the separate estimates. For WWTPs 101 and 102 with 4 and 7 local data points, estimates are closer to the "separate" estimates than to the "complete pooling" estimate. This indicates that for the case of NoVII concentrations, already a small number of local data points may suffice to provide valuable additional information, against the background of the large variation of reported concentrations in the scientific literature.

Differences between approaches become more obvious when the corresponding uncertainty intervals are examined. Their widths

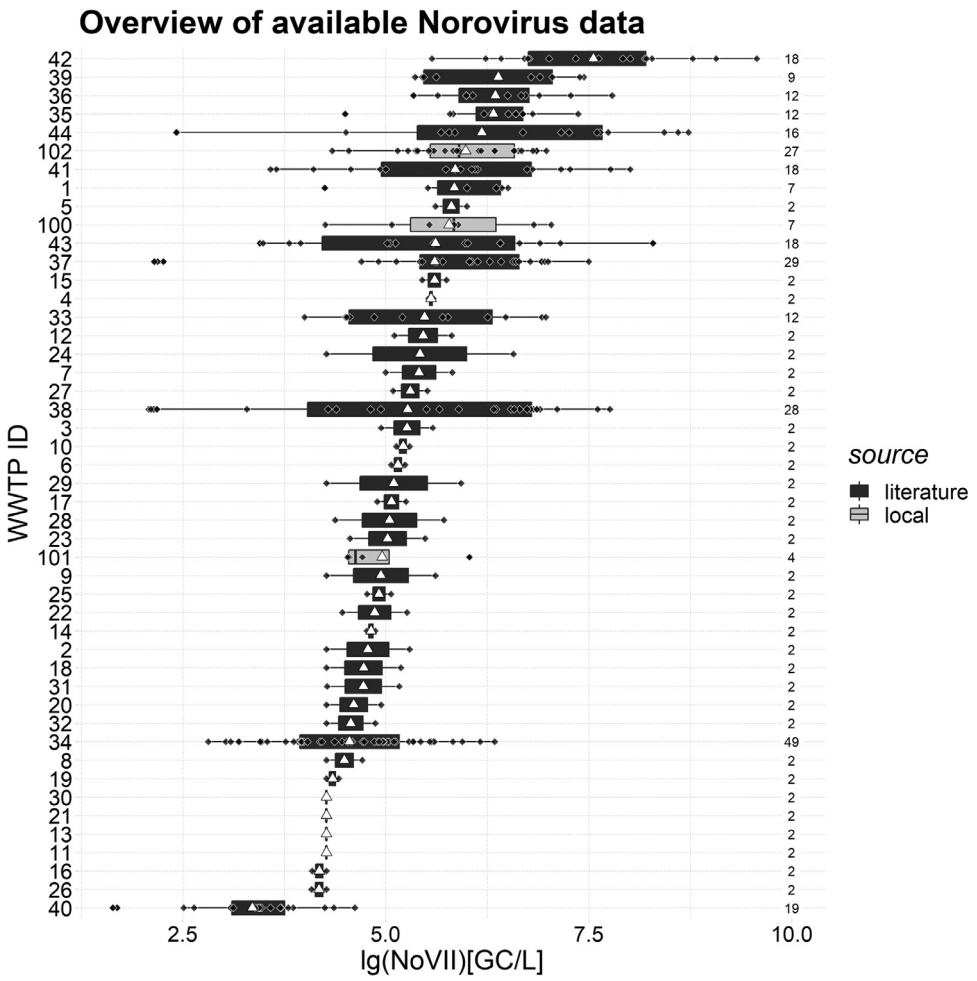

Fig. 1. Overview of local data and external data collected from literature. The numbers on the left represent the WWTP ID. The numbers on the right refer to the sample size at each wastewater treatment plant. White triangles represent the empirical means, black diamonds the individual observations.

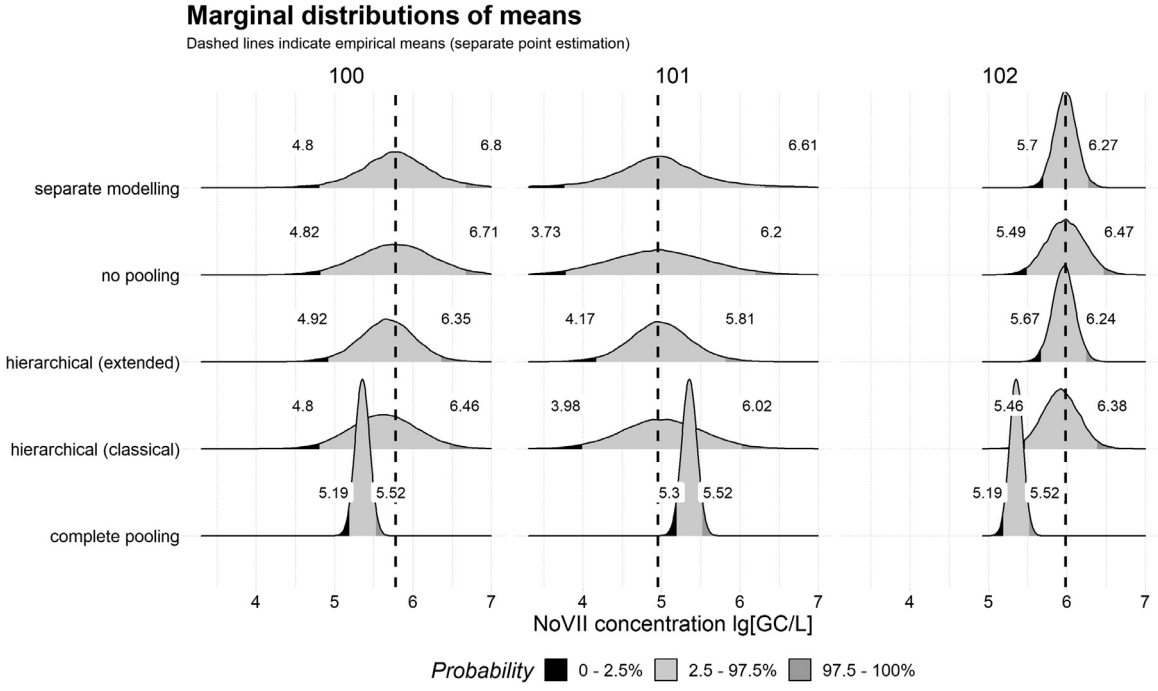

Fig. 2. Overview of calculated marginal distributions for the posterior mean NoVII concentrations using different estimation approaches. Numbers indicate the 95% uncertainty range.

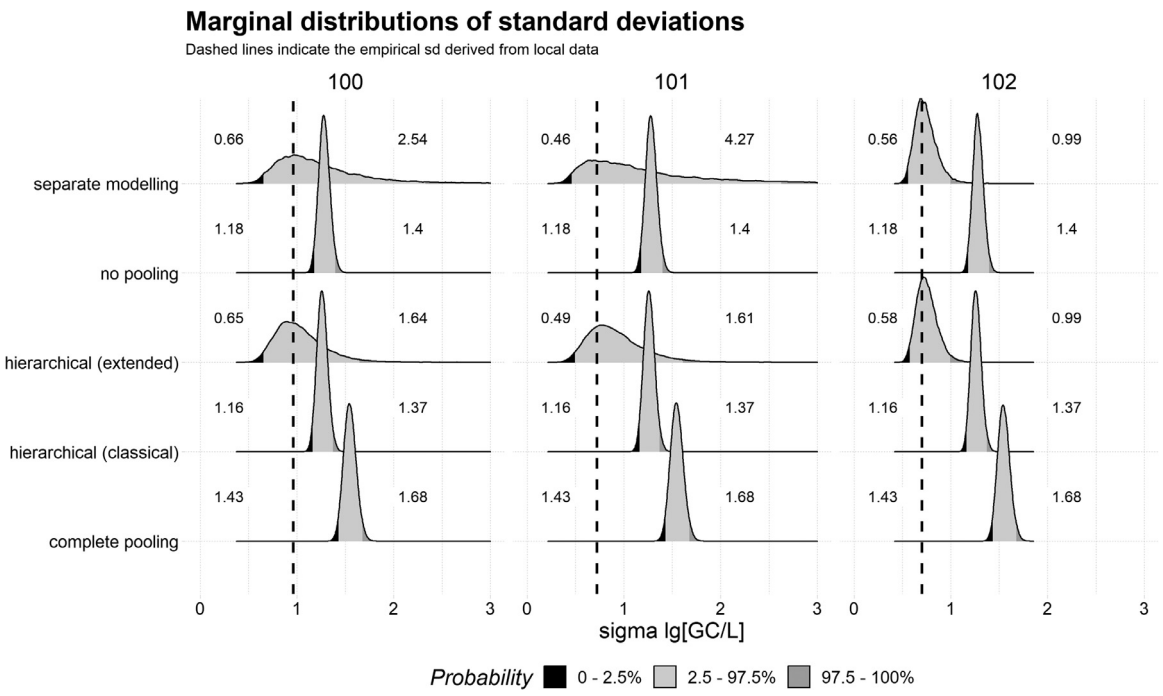

Fig. 3. Overview of the marginal distributions of the individual standard deviations using different modeling approaches. Numbers indicate the 95% uncertainty range.

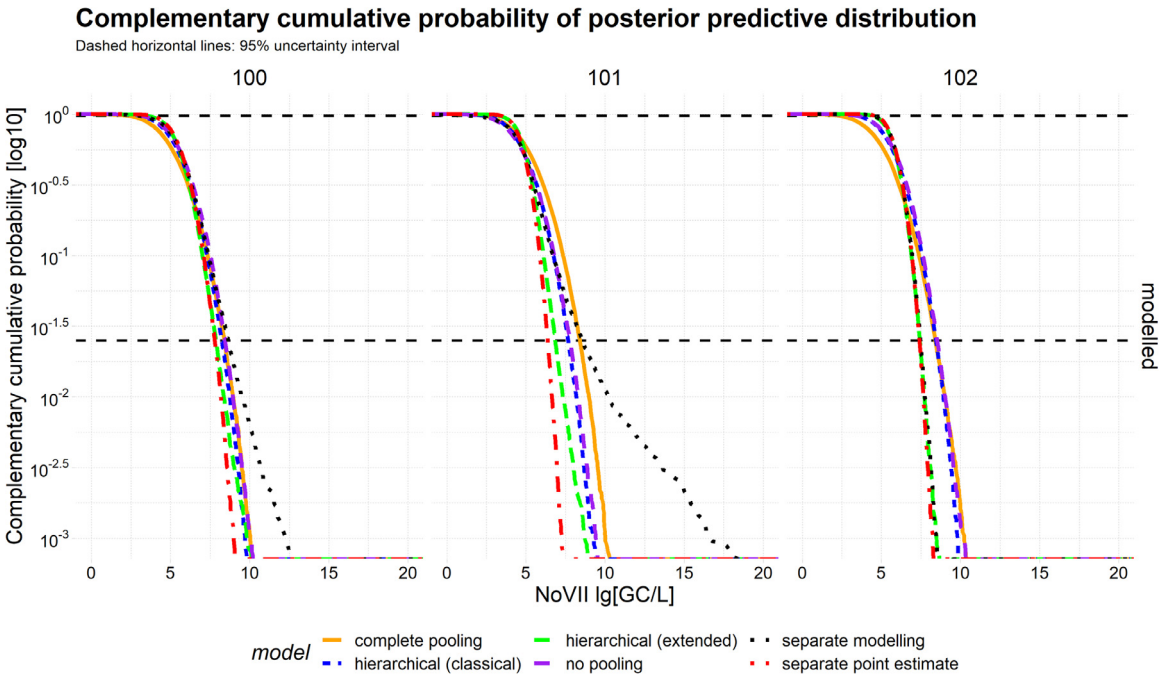

Fig. 4. Posterior predictive distribution plotted as the complementary cumulative probability density function. Horizontal dashed lines indicate the upper limit of the 95% uncertainty interval.

depend on the estimates of the standard deviation (Fig. 3) and the sample size at each location. By definition, "separate point estimation" does not account for parameter uncertainty at all (dashed line). Thus, no distinction is made whether the estimate derives from a WWTP with only 4 data points or one with 27. The "complete pooling" approach leads to the most precise estimate, as the standard error is calculated assuming that all 345 data points come from a common WWTP. For "no pooling" and "classical hierarchical" modeling the commonly estimated residual standard deviation is reduced in comparison to the "complete pooling" approach. However, since local standard errors are derived from local sample sizes the resulting uncertainty intervals become wider compared to the "complete pooling" approach.

Fig. 3 shows the marginal distribution of the residual standard deviation of the different estimation approaches. As for the marginal distributions of the means, the standard deviations of the different approaches differ in their location and the width of the uncertainty intervals. Regarding the location, "complete pooling" leads to the highest residual standard deviation, followed by "no pooling" and the "classical hierarchical" model. While these three approaches differ in the location (mode of the distribution), the uncertainty intervals of these three approaches are similar, be-

cause all three approaches estimate the residual standard deviation from all data points as they assume a constant σ across WWTPs. In contrast, “separate modeling”, “separate point estimation” and the “extended hierarchical” model show residual standard deviations, which are considerably lower (mode of the distribution) than the estimates from the named three approaches, as they estimate a separate σ for each WWTP along with the estimation of μ. However, these estimation approaches differ mainly regarding the width of the uncertainty intervals. “Separate point estimation” ignores parameter uncertainty completely (dashed line), while for “separate modeling” the width of the uncertainty interval is only influenced by the locally available data. For the “extended hierarchical” model, uncertainty intervals are the result of the information contained in the local data and the information provided by the meta-analysis from the remaining 46 WWTP. For WWTP 102, there is little difference between the estimate inferred by “separate modeling” and the “extended hierarchical model”. For WWTPs 100 and 101 differences are more visible. While the modes of the distributions correspond to each other, “separate modeling” with 7 and 4 data points leads to marginal posterior distributions with heavy tails, that still include values of up to 4.2 in the 95% uncertainty intervals. For extended hierarchical modeling, the upper confidence limits are reduced to values of approximately 1.6. Thereby, “extended hierarchical modeling” substantially reduces parameter uncertainty and thus predictive uncertainty on a local level in comparison to “separate modeling” (Fig. 4).

Fig. 4 shows the full PPD plotted as the complementary cumulative density function, thus the y-axis shows the probability of a single new observation exceeding the corresponding value on the x-axis. This visualization emphasizes the behavior of the distribution at its tails (Smeets et al., 2008), which is important for understanding the results of risk simulation (cf. Section 3.2). The PPD, moreover, incorporates complete parameter uncertainty regarding μ and σ, except from “separate point estimation” for which this is ignored by definition. As expected, "separate point estimation" leads to the narrowest PPD for all scenarios. For “separate modeling”, the resulting PPD corresponds to “separate point estimation” in the case of WWTP 102 with 27 data points. For WWTPs 101 and 102 with 7 and 4 data points, on the other hand, the inclusion of parameter uncertainty leads to completely different behavior, especially at the upper tails of the distribution, which reflects the large uncertainty of both μ and σ, caused by small sample size.

3.2. Effects on risk and performance target calculations

Figs. 5 –7 illustrate the calculated risk against the set health targets and assumed log-reductions. For the swimming scenario (Fig. 5) the “complete pooling” scenario leads to the lowest required log-reduction if the median risk (white line) is used for decision making since the average derived from all data lies below the average derived from the locally collected ones. However, if the uncertainty is taken into account, meaning that the complete 95% uncertainty interval (upper grey boundary) has to fall below the threshold level, then “separate modeling” and the “extended hierarchical” model lead to a required log-reduction of 7.4. This lies about 0.7 log below the reduction derived from “complete pooling” (8.1 log). “Separate modeling” and the “extended hierarchical” model have in common that both, the mean and the standard deviation are estimated individually for each WWTP. In contrast, the “no pooling” and “classical hierarchical” models estimate a common residual standard deviation after accounting for differences in the WWTP-specific means. Since the standard deviation of the locally collected data is lower than the one derived from all data, the estimation approaches which estimate the standard deviation separately lead to narrower uncertainty intervals.

The drinking water and irrigation scenario differ from the swimming scenario since risk calculations are based on multiple exposure events per year in contrast to the single exposure risk for swimming. Thus, a direct comparison between prediction intervals and risk calculations is less intuitive.

For the drinking water scenario, the lowest required log-reduction is derived from “separate point estimation” ($<$12 log). The required log-reduction derived from the remaining approaches depends on whether the median or the full uncertainty interval is used for decision-making. For the median, the “extended hierarchical” modeling approach leads to a required log-reduction of 12.8 log, followed by 13.2 - 13.3 log for the “classical hierarchical” and “no-pooling” approaches, 14 log for “complete pooling”, and $>$16 log for the “separate modeling” approach. If decisions are based on the complete uncertainty, differences are reduced to 0.5 - 1 log between approximately 14 log for the “no-pooling” and the two hierarchical models in comparison to 14.8 log for “complete pooling”, and “separate modeling” ($>$ 16). This order generally corresponds to the width of the posterior predictive distribution shown in Fig. 4. A striking feature is that even a reduction by 15 log does not suffice to reduce to the median risk below the applied threshold risk of infection of 10^{-4} per person per year for the “separate modeling” approach. Moreover, the increase of the assumed log-reduction reduces the median calculated risk quite linearly for all estimation approaches, except from “separate modeling” (see Fig. A2 in the SI). For “separate modeling” only the uncertainty interval increases but the calculated median risk stays between 10^{-1} and 1 per person per year (Fig. 6). While this behavior seems counterintuitive at first, it can be explained by investigating Fig. 4 in combination with the applied dose-response relationship and the corresponding exposure scenarios (see SI). From Fig. 4 it can be deduced that the probability of simulating a single observation, which is larger than 10^{15} GC/L is as low as $10^{-2.75}$ (0.00178) for the separate modeling approach. With an assumed log-reduction of 15, these probabilities correspond to the event of ingesting at least 1 virus in a single event. While these probabilities seem very low it has to be considered that for the annual average risk 365 samples are taken, which increases the probability of ingesting at least 1 virus particle per year to 0.48. As the probability of infection by ingesting a single virus particle is already 0.27 and rapidly increases to 0.5 with increasing dose, the calculated median risk stays between 10^{-1} and 1, even for a log-reduction of 15. The large difference between the result from “separate point estimation” (12 log) and “separate modeling” ($>$15 log) underlines the high importance of explicitly addressing parameter uncertainty in microbial risk simulation. A further interesting observation is that the Monte Carlo error is largest for the drinking water scenario. This is not related to the small sample size but to the narrower confidence intervals of the annual risk distribution caused by 365 samples per year in comparison to 50 and 1, respectively.

Risk calculations of the median annual risk for the irrigation scenario show similar behavior as the drinking water scenario. Uncertainty intervals are generally wider (Fig. A3 in SI) due to the lower number of exposure events per year, which lead to a less step curve in Fig. 7 and a smaller Monte Carlo error. If the median is used for decision-making differences between approaches lie between 0.5 and 1 log, with “separate point estimation” leading to the lowest required log-reduction, followed by the two hierarchical approaches, “complete pooling”, “no pooling”, and finally “separate modeling”. However, if the 95% uncertainty boundary is used, the difference between “separate modeling” and the “extended hierarchical” modeling approach increases to up to 2.8 log. The extended hierarchical model lies in turn 0.8 log below the “complete pooling” approach. This underlines the benefits of applying a flexible approach like BHM which integrates both literature and local information.

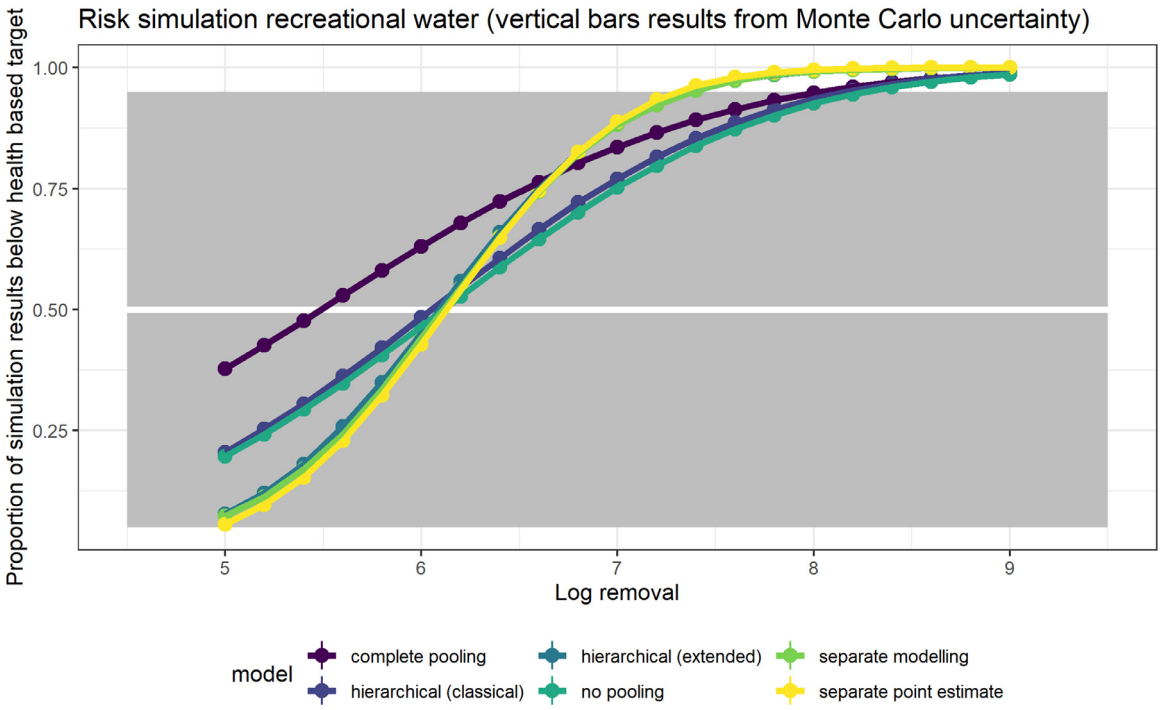

Fig. 5. Proportion of illness risk calculation below the recreational water target of 3% per swimming event. The white center line indicates whether the median of risk distribution is below the health-target. The upper bound of the grey area indicates whether the 95% uncertainty interval falls below the health target. Vertical bars represent the 95% uncertainty of the 100 repetitions. Interactive plots are provided in the SI.

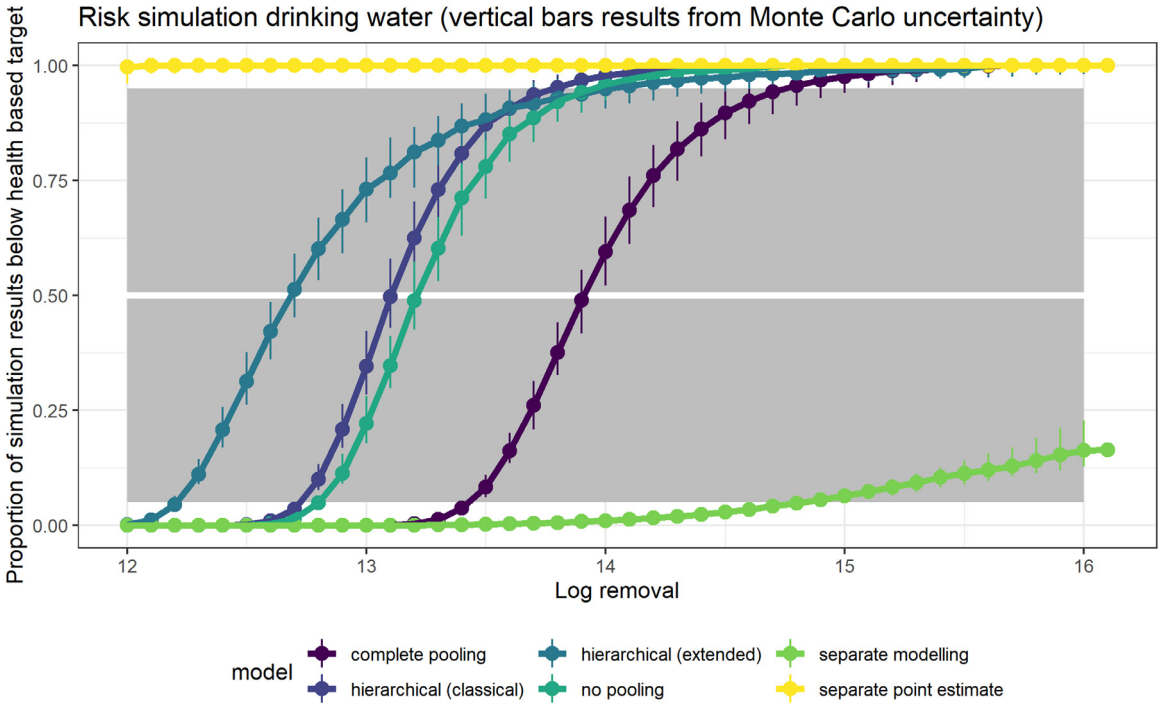

Fig. 6. Proportion of infection risk calculations below drinking water target of 10^{-4} pppy. The white center line indicates whether the median of risk distribution is below the health target. The upper bound of the gray area indicates whether the 95% uncertainty interval falls below the health target. Vertical bars represent the 95% uncertainty of the 1000 repetitions.Interactive plots are provided in the SI.

Interestingly, the location of the median risk derived from "separate point estimation", and the two hierarchical modeling approaches fall below the median risk of the "complete pooling" approach, even though their inferred means (cf. Fig. 2) are above the mean derived from "complete pooling". This can be explained by the larger standard deviation of the "complete pooling" approach in combination with the effect of multiple annual exposures (50 exposure events per year), i.e. resampling, as already described for the drinking water scenario.

3.3. Model comparison and generalized estimates for new WWTPs

Fig. 8 shows the generalized estimates for new data (prediction interval) and the uncertainty about the mean of a new, un-

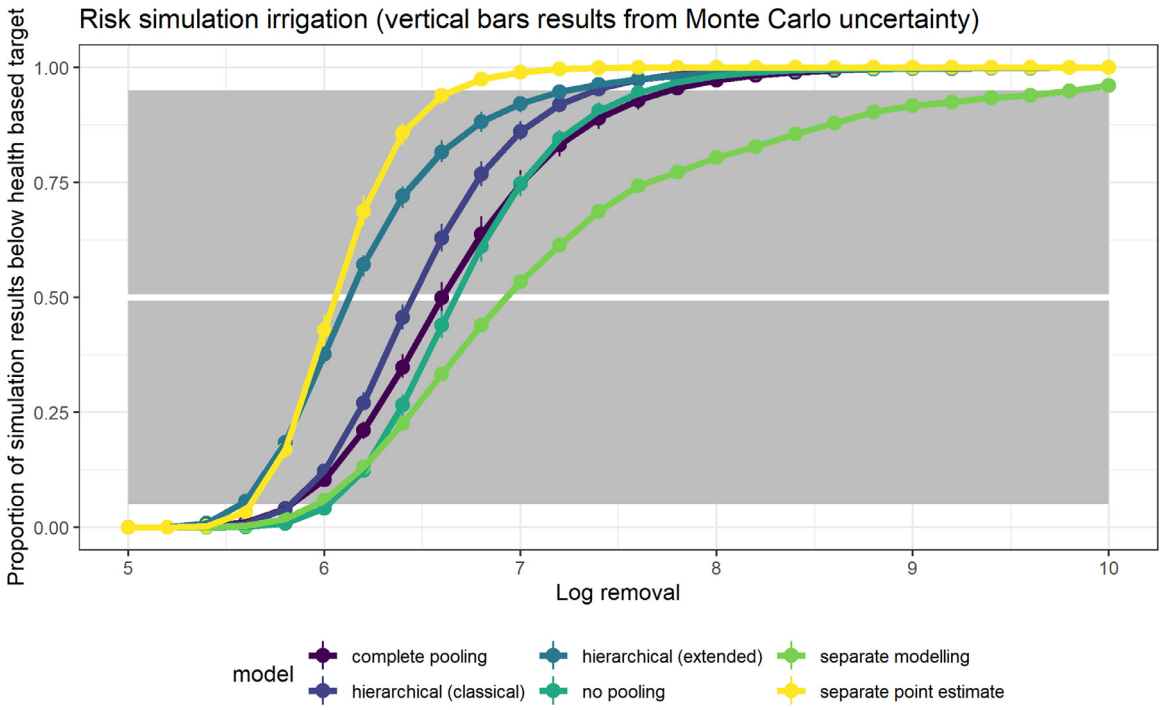

Fig. 7. Proportion of risk calculation below irrigation target of 10^{-4} DALYs pppy. The white center line indicates whether the median of risk distribution is below the health target. The upper bound of the grey area indicates whether the 95% uncertainty interval falls below the health target. Vertical bars represent the 95% uncertainty of the 100 repetitions.Interactive plots are provided in the SI.

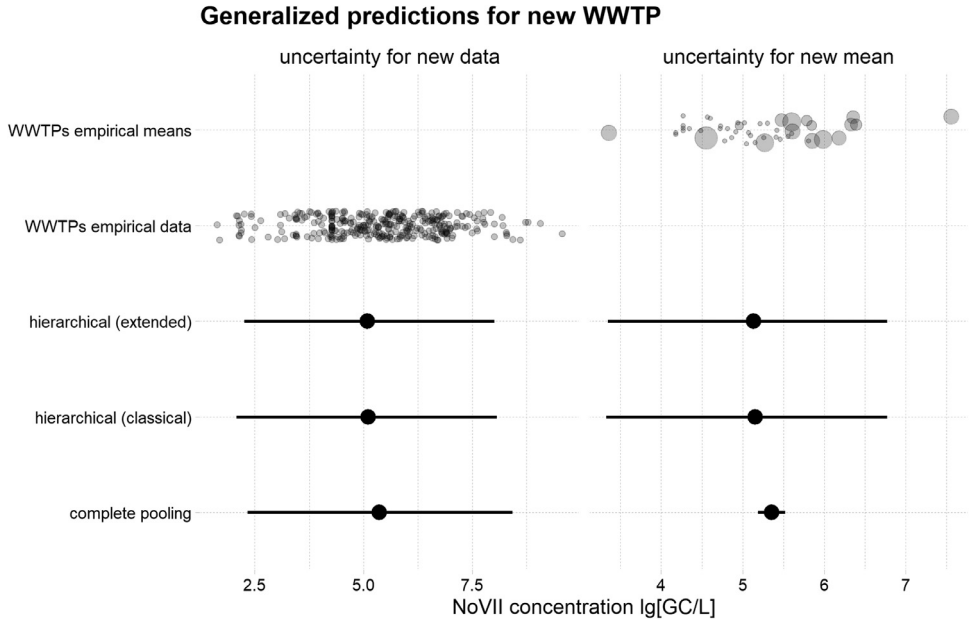

Fig. 8. Comparison of the posterior predictive distribution and the prediction of the mean for a new WWTP inferred from different estimation approaches. Empirical data and empirical means are shown in rows 1 and 2 (values have been "jittered" to reduce over-plotting). The point size of the empirical means is proportional to the sample size of at each WWTP.

observed WWTP, simulated from the estimates for η and τ for the hierarchical approaches and μ_{Nov} and σ_{NoV} for the complete pooling approach. An interesting observation is the width of the uncertainty intervals regarding the prediction of a new mean NoVII concentration at an unknown WWTP. As the "complete pooling" approach assumes that all data come from a comparable WWTP, the between-WWTP variability is not specifically addressed. The location of the mean and the standard error are estimated from all data. This leads to an elevated mean value and narrow confidence intervals in comparison to the hierarchical modeling approaches. The narrow confidence interval, however, is not consistent with the high between-WWTP-variability present in the empirical data.

In contrast, the 95% uncertainty intervals derived from hierarchical modeling cover 96% (45/47) of the observed empirical means. Thus, the two approaches provide a more uncertain, but also more realistic estimate of the existing between WWTP-variability. The calculated expected log-predictive density also sup-

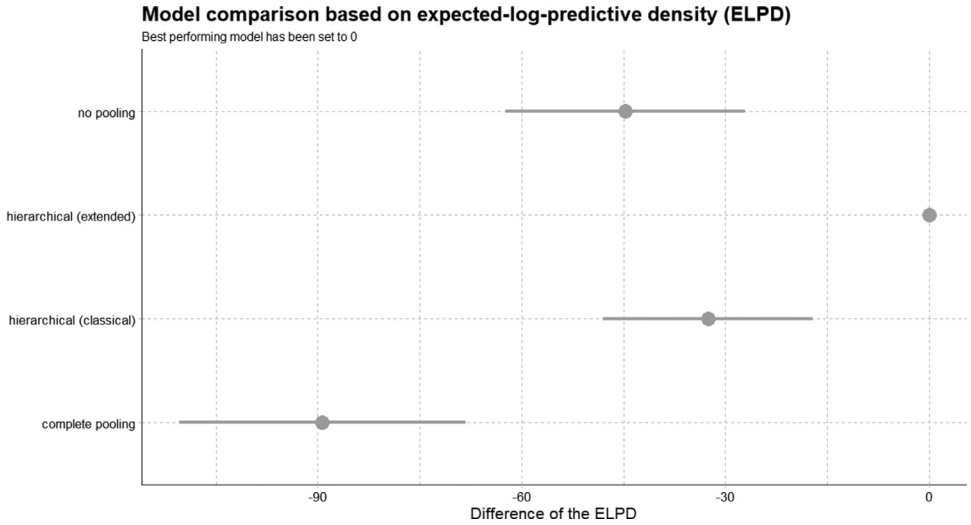

Fig. 9. Model comparison based on the expected log predictive density. The plot shows the differences in the calculated ELPD with the best model (extended hierarchical) set to zero.

ports the advantage of the two hierarchical modeling approaches over the "complete pooling" approach (Fig. 9). Results show that in the presented case the extended hierarchical model is expected to produce the most accurate predictions of NoVII concentrations. The "classical hierarchical" model and the "no-pooling" approach lead to similar results, whereas the "complete pooling" approach is expected to deliver the worst predictive accuracy in comparison to the other estimation approaches.

4. Discussion

In the present study, we investigated the effect of including external information from the scientific literature on local estimates using BHM based on the case of NoVII concentrations in the influent of municipal WWTPs. To the best of our knowledge, such an approach to quantitatively combine local and external information to reduce parameter uncertainty on a local level has not been conducted so far in the QMRA literature. In the literature two major meta-analyses were identified that address concentrations of NoVII (Eftim et al., 2017; Pouillot et al., 2015) in WWTPs, using different statistical approaches (bootstrapping, hierarchical modeling). However, in both studies, the focus was put on the derivation of generalizable estimates of NoVII concentrations instead of investigating the effect on a local scale.

Our results showed that the benefit of incorporating external information is largest in situations where local data are sparse ($N = 4$, $N = 7$). In such situations, parameter uncertainty would be so large that no reliable estimate of concentrations and log-reductions would be possible, without the use of external information. Our results also showed that even a small sample size in combination with external information might contribute valuable information to reduce the estimate of the required log-reduction to reasonable values. The fact that even small sample sizes affect estimates in hierarchical modeling, is an indication that the between-WWTP-variability is high and only little pooling occurs. Reasons for the high between-WWTP-variability may include differences in population size, season, climate, incidence in the population, and analytical methods and protocols. This shows that NoV concentrations from a single WWTP should not directly be used to inform assumptions about NoV concentration at other locations. However, since the degree of pooling that occurs in Bayesian hierarchical modeling depends on the between-WWTP-variability, this fact is accounted for and only little information is shared between WWTPs. Thus, employing this technique reduces the risk of not accounting for these variations by imposing too strong assumptions or pooling over interesting features in the dataset.

While a data set of four data points at WWTP 101, for which the effect was largest, is extremely small, small sample sizes at single locations are not uncommon in the QMRA literature. For example, McBride et al. (2013) conducted a discharge-based QMRA study and measured pathogens in stormwater from catchments with different characteristics. Within each catchment type, the sample size per site varied between 8 samples at 8 sites for dry season urban runoff (i.e. 1 sample per site) to 7 samples at 1 site for "forested open space stormwater". For the remaining catchment types, sample sizes varied between 3 and 4 samples per site. Amoueyan et al. (2019) used NoVII surface water concentration published by Lodder and de Roda Husman (2005) for a QMRA-based comparison of direct potable reuse and unintentional *de facto* reuse. In the latter publication, 8 NoVII observations are reported, which were collected at two rivers in the Netherlands (4 observations at each river). Thus, the situation of relying only on a couple of observations can be regarded as a realistic scenario, for which hierarchical modeling might be a suitable and transparent solution for addressing and potentially reducing parameter uncertainty.

Moreover, this information might also be of high practical relevance for applied risk-based decision-making, because operators may be confronted with the decision about whether to take additional efforts to collect local data or to plan potentially necessary risk reduction measures based on the available body of knowledge provided by the scientific literature or guideline documents. Our results show that even small sample sizes might be beneficial in combination with external information and might lead to lower required log-reduction, i.e. less resource depletion and cost, in comparison to e.g. a complete pooling approach. However, it is important to underline that small sample sizes are only beneficial if they are combined with external information, otherwise, parameter uncertainty is too high.

Our results also show that Bayesian hierarchical modeling allows for making generalizable estimates about new "locations" (in our case WWTPs) which explicitly account for the between location variability of model parameters. As a result, the predictions of a new mean μ becomes much wider than if a "complete pooling" approach is applied. Our prediction for a unknown mean influent concentration at an unknown WWTP covers a range be-

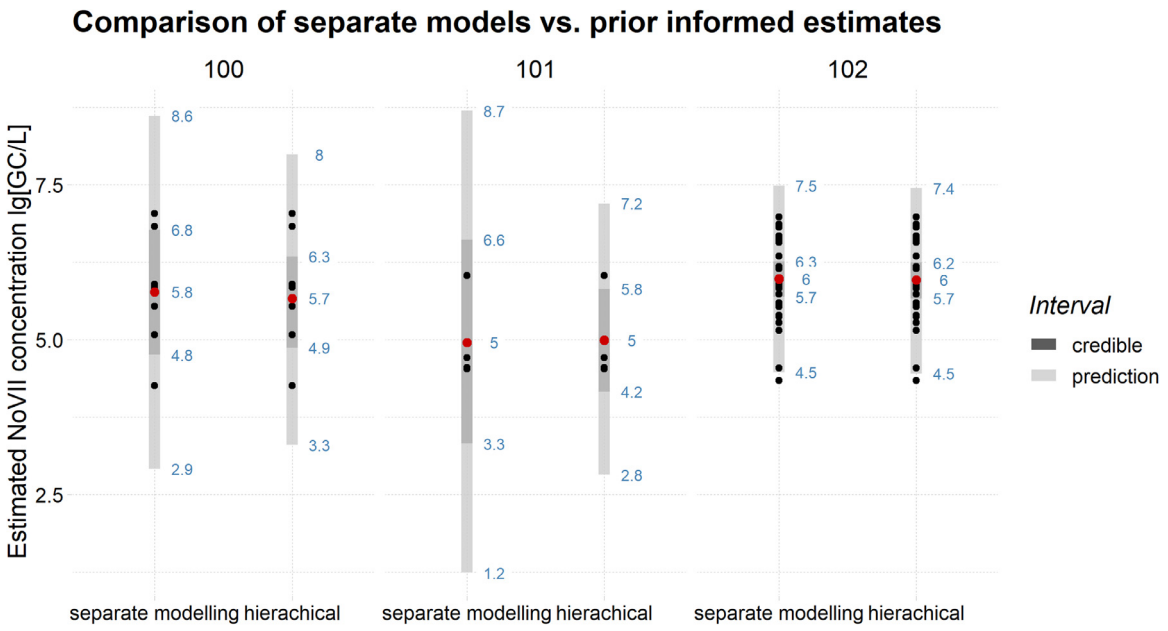

Fig. 10. Estimates and 95% uncertainty intervals for the mean (credible) and new data (prediction) derived from first fitting the extended hierarchical model only to the literature data and subsequently updating the deduced informative priors for the mean (μ) and sd (σ). Priors: $\mu = N(\eta = 5.04, \tau = 0.92)$, σ = lognormal($\alpha = 0, \gamma = 0.43$).

tween 3.5 and 6.5. Interestingly, if this range, which is derived only from European WWTP, is compared to the estimates reported by Eftim et al., 2017 for differences between geographic regions, all of the estimates reported by Eftim (North America, Europe, Asia, New Zealand) fall inside the derived prediction interval of the present study. Therefore, the statistically significant difference between NoV concentrations in North America in comparison to the ones in Asia and Europe reported in Eftim et al. (2017) has to be regarded with caution, because it might only be an artifact of the statistical procedure which does not account for the between-WWTP-variability. More generally, the question about appropriate methods for selecting one statistical model over the other for summarizing information via meta-analysis deserves more attention. In our study, we used the pointwise expected-log-predictive-density as a quantitative indication which statistical model might be the most accurate. Such or similar approaches, like the application of information criteria, might be reasonable ways to make decisions about which modeling approach to apply more transparent.

One additional advantage we see when using Bayesian hierarchical modeling is, that in contrast to complete pooling the generalized estimates from μ and σ represented by the estimates for η, τ, α, and γ in Eq. (5), may function as a prior distribution for subsequent studies. That means, that knowledge can be transferred by reporting only the parameterization of the priors for μ and σ, which can function as a starting point for subsequent updates. This feature might be potentially interesting because collecting raw data by contacting authors or digitalizing them from graphs, is very time-consuming. Especially in comparison to methods like bootstrapping which always requires the full raw data set, communicating a generalized prior for subsequent analysis might be a promising way to ease the use of literature information by subsequent studies. For illustration, we fitted the extended hierarchical model to the literature data only and subsequently updated the deduced informative prior using the local data Fig. 10. Parameter values differ slightly from the complete analysis as the data from the other two local WWTP are not included in the prior. It can be seen that as for the complete analysis parameter uncertainty can be reduced substantially in low data situations (100, 101) while estimates are not affected if the local information is strong (102). Moreover, while our study used the simple example of fitting a mean and standard deviation of a lognormal distribution, BHM can readily be extended to the estimation of other model parameters, like e.g. the slopes in a regression model.

5. Conclusion

- Bayesian hierarchical modeling allows for reducing local parameter uncertainty in comparison to separate modeling while being flexible enough to let local data dominate local estimates in contrast to e.g. “complete pooling”.
- External information provided e.g. by published meta-analyses helps to reduce local parameter estimates and should be considered to be included quantitatively into QMRA studies even when local data is available.
- In cases of limited sample sizes, parameter uncertainty may be too large to make reasonable inferences based on local data alone. Therefore, parameter uncertainty should be considered explicitly in future QMRA studies.
- In combination with external information, already small local data sets may help to reduce the required log-performance in local settings and justifies additional measurements before planning suitable risk reduction measures.
- Model comparison and selection methods, like e.g. information criteria or the expected log-predictive density, should be considered in future meta-analysis to decide between candidate statistical models.

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.

Acknowledgments

We thank the authors who provided the necessary data for the meta-analysis, especially John Flannery. Local Norovirus data were collected in the research projects FLUSSHYGIENE, funded by the

German Federal Ministry for Education and Research (BMBF) under sponsorship number 02WRS1278A) and DEMOWARE, funded by the European Union's 7th Framework Program for research, technological development and demonstration, theme ENV.2013.WATER INNO&DEMO-1 under grant agreement no 619040.

References

Amoueyan, E., Ahmad, S., Eisenberg, J.N.S., Gerrity, D., 2019. Equivalency of indirect and direct potable reuse paradigms based on a quantitative microbial risk assessment framework. Microb. Risk Anal. 12, 60–75.

Boehm, A.B., Graham, K.E., Jennings, W.C., 2018. Can We Swim Yet? Systematic Review, Meta-Analysis, and Risk Assessment of Aging Sewage in Surface Waters. Environmental Science & Technology.

Boehm, A.B., Silverman, A.I., Schriewer, A., Goodwin, K., 2019. Systematic review and meta-analysis of decay rates of waterborne mammalian viruses and coliphages in surface waters. Water Res. 164, 114898.

Bürkner, P.-C., 2017. brms: an R package for Bayesian multilevel models using stan. J. Stat. Softw. 1 (Issue 1) (2017).

Calgua, B., Rodriguez-Manzano, J., Hundesa, A., Suñen, E., Calvo, M., Bofill-Mas, S., Girones, R., 2013. New methods for the concentration of viruses from urban sewage using quantitative PCR. J. Virol. Methods 187 (2), 215–221.

Carvajal, G., Branch, A., Sisson, S.A., Roser, D.J., van den Akker, B., Monis, P., Reeve, P., Keegan, A., Regel, R., Khan, S.J., 2017. Virus removal by ultrafiltration: understanding long-term performance change by application of Bayesian analysis. Water Res. 122, 269–279.

Eftim, S.E., Hong, T., Soller, J., Boehm, A., Warren, I., Ichida, A., Nappier, S.P., 2017. Occurrence of norovirus in raw sewage – a systematic literature review and meta-analysis. Water Res. 111 (Supplement C), 366–374.

Gonzales-Gustavson, E., Rusiñol, M., Medema, G., Calvo, M., Girones, R., 2019. Quantitative risk assessment of norovirus and adenovirus for the use of reclaimed water to irrigate lettuce in Catalonia. Water Res. 153, 91–99.

Lodder, W.J., de Roda Husman, A.M., 2005. Presence of noroviruses and other enteric viruses in sewage and surface waters in the Netherlands. Appl. Environ. Microbiol. 71 (3), 1453.

Mara, D., Sleigh, A., 2010. Estimation of norovirus infection risks to consumers of wastewater-irrigated food crops eaten raw. J. Water Health 8 (1), 39–43.

McBride, G.B., Stott, R., Miller, W., Bambic, D., Wuertz, S., 2013. Discharge-based QMRA for estimation of public health risks from exposure to stormwater-borne pathogens in recreational waters in the United States. Water Res. 47 (14), 5282–5297.

Messner, M.J., Berger, P., Nappier, S.P., 2014. Fractional poisson–a simple dose-response model for human norovirus. Risk Anal. 34 (10), 1820–1829.

Mok, H.F., Barker, S.F., Hamilton, A.J., 2014. A probabilistic quantitative microbial risk assessment model of norovirus disease burden from wastewater irrigation of vegetables in Shepparton, Australia. Water Res. 54, 347–362.

NRMMC-EPHC-AHMC, 2006. National Guidelines for Water Recycling: managing health and environmental risks. Environment Protection and Heritage Council, the Natural Resource Management Ministerial Council and the Australian Health Ministers' Conference.

Petterson, S.R., Dumoutier, N., Loret, J.F., Ashbolt, N.J., 2009. Quantitative Bayesian predictions of source water concentration for QMRA from presence/absence data for E. coli O157:H7. Water Sci. Technol. 59 (11), 2245–2252.

Pouillot, R., Van Doren, J.M., Woods, J., Plante, D., Smith, M., Goblick, G., Roberts, C., Locas, A., Hajen, W., Stobo, J., White, J., Holtzman, J., Buenaventura, E., Burkhardt, W., 3rd, Catford, A., Edwards, R., DePaola, A., Calci, K.R., 2015. Meta-analysis of the reduction of norovirus and male-specific coliphage concentrations in wastewater treatment plants. Appl. Environ. Microbiol. 81 (14), 4669–4681.

Schmidt, P.J., 2015. Norovirus dose-response: are currently available data informative enough to determine how susceptible humans are to infection from a single virus? Risk Anal. 35 (7), 1364–1383.

Smeets, P.W.M.H., Dullemont, Y.J., Van Gelder, P.H.A.J.M., Van Dijk, J.C., Medema, G.J., 2008. Improved methods for modeling drinking water treatment in quantitative microbial risk assessment; a case study of Campylobacter reduction by filtration and ozonation. J. Water Health 6 (3), 301–314.

Soller, J.A., Eftim, S.E., Nappier, S.P., 2018. Direct potable reuse microbial risk assessment methodology: sensitivity analysis and application to State log credit allocations. Water Res. 128, 286–292.

Soller, J.A., Eftim, S.E., Warren, I., Nappier, S.P., 2017. Evaluation of microbiological risks associated with direct potable reuse. Microb. Risk Anal. 5, 3–14.

StanDevelopmentTeam 2017 Stan modeling language users guide and reference manual, Version 2.17.0. http://mc-stan.org.

Teunis, P., Schijven, J., Rutjes, S., 2016. A generalized dose-response relationship for adenovirus infection and illness by exposure pathway. Epidemiol. Infect. 144 (16), 3461–3473.

Teunis, P.F.M., Moe, C.L., Liu, P., Miller, S.E., Lindesmith, L., Baric, R.S., Le Pendu, J., Calderon, R.L., 2008. Norwalk virus: how infectious is it? J. Med. Virol. 80 (8), 1468–1476.

Van Abel, N., Schoen, M.E., Kissel, J.C., Meschke, J.S., 2017. Comparison of risk predicted by multiple norovirus dose-response models and implications for quantitative microbial risk assessment. Risk Anal. 37 (2), 245–264.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27 (5), 1413–1432.

US-EPA, 2012. Recreational Water Quality Criteria; OFFICE OF WATER 820-F-12-058.

WHO, 2006. Guidelines for the Safe Use of Wastewater, Excreta and Greywater, 2. World Health Organisation, Geneva, Switzerland, p. 222.

WHO, 2011. World Health Organization. (2011): Guidelines for drinking-water quality, 4th ed. World Health Organization https://apps.who.int/iris/handle/10665/44584.

WHO, 2016. Quantitative Microbial Risk Assessment: Application for Water Safety Management. Work Health Organisation, Geneva, Switzerland.

Wyn-Jones, A.P., Carducci, A., Cook, N., D'Agostino, M., Divizia, M., Fleischer, J., Gantzer, C., Gawler, A., Girones, R., Höller, C., de Roda Husman, A.M., Kay, D., Kozyra, I., López-Pila, J., Muscillo, M., José Nascimento, M.S., Papageorgiou, G., Rutjes, S., Sellwood, J., Szewzyk, R., Wyer, M., 2011. Surveillance of adenoviruses and noroviruses in European recreational waters. Water Res. 45 (3), 1025–1038.