Annals of Work Exposures and Health, 2026, 70, wxag019
https://doi.org/10.1093/annweh/wxag019
Published: 31 March 2026
Original article

The Chartered Society for Worker Health Protection
OXFORD

Advancing occupational exposure models: insights from a case study

Remy Franken*, Ilse Ottenbros, Calvin Ge, Wouter Fransman, and Tanja Krone

Risk Analysis for Prevention, Innovation and Development (RAPID), TNO, Princetonlaan 6, Utrecht 3584 CB, The Netherlands
*Corresponding author: Email: remy.franken@tno.nl.

Abstract

Currently used models in occupational exposure assessment are mostly mechanistic-based and their performance varies. The usage of more advanced modeling approaches such as machine learning (ML) or hybrid modeling approaches such as Bayesian Networks could improve accuracy of exposure predictions. Within this study, a case study was conducted where 5 different ML or hybrid modeling approaches were used to redevelop the "asbestos removal exposure assessment tool," an existing mechanistic exposure model, and their performance was compared. Multiple Linear Regression, Random Forest, Gradient Boosting Machines, Bayesian Network, and Neural Network models were developed using the same dataset and model determinants used to develop the mechanistic model. Random Forest and Gradient Boosting Machines performed best with regards to accuracy, followed by Bayesian Network, Multiple Linear Regression, the original model, and Neural Network. Such models show a promise for the development of more accurate models, but their limitations need to be considered before they can be implemented in a regulatory context. For these models, a trade-off exists where transparency in decision-making is traded against accuracy. Hybrid models such as Bayesian Networks might be a solution for this trade-off as expert knowledge and data-driven approaches are combined in a transparent model.

Keywords exposure assessment, exposure model, machine learning, Bayesian Network, asbestos exposure

What's important about this paper?

Advances in computational techniques, such as machining learning, may improve occupational exposure modeling, which has typically employed mechanistic or statistical models to date. This case study demonstrates that machining learning techniques were promising for making exposure models more accurate in their predictions. Applicability of such techniques within the field of occupational exposure modeling is discussed.

Introduction

Under the European REACH regulation (EC 1907/2006), manufacturers and importers of chemical substances must conduct quantitative occupational exposure assessments for relevant exposure scenarios. These exposure assessments can be conducted using exposure measurements, or by the use of exposure modeling tools. Regulatory guidance (ECHA R14, EN 689) explicitly supports the use of modeling, provided that the appropriate

Received: May 12, 2025. **Accepted:** February 16, 2026

and validated models are used. The significance of exposure modeling in workplace exposure assessment and management to ensure a reliable and efficient assessment is widely recognized (Fantke et al. 2022; Jones 2022; Schlüter et al. 2022). However, with a growing number of these models available the critical question is their reliability and validity (Fransman 2017; Spinazze et al. 2019; Schlüter and Tischer 2020), as regulators and users depend on these models to make informed decisions about workers' exposure. Previously, Tischer et al. (2017) described 3 key aspects of model evaluation: the conceptual evaluation (theoretical soundness of the model), external validation (alignment with measurement data or other models), and the operational analysis (between-user reliability).

In recent years, several studies have explored the evaluation of exposure models with a strong focus on external validation (Landberg et al. 2017; Schlüter and Tischer 2020) and between-user reliability (Schinkel et al. 2014; Lamb et al. 2017). The performance of commonly used REACH models such as ECETOC TRA, Stoffenmanager, and EMKG-EXPO has shown considerable variability, either overestimating exposure severely, resulting in unnecessary burden, or underestimating exposure posing a potential exposure risk to workers (Schlüter and Tischer 2020).

As exposure modeling likely takes on an increasingly prominent role in regulatory decision-making, the field faces new challenges. In response to these challenges, the International Society of Exposure Science (ISES) Europe has outlined a strategy for improving exposure models over the next decade (Schlüter et al. 2022). This strategy emphasizes the need for model refinement, the development of new methodologies in order to extend the applicability domains of these models, and targeted research for understudied areas in occupational exposure modeling.

The application of advanced statistical models in predictive chemical risk assessment has been increasing recently, eg for in application to real-time exposure studies (Patton et al. 2022), Bayesian based models for risk assessment in textile industry (Ak et al 2022) or the development of Trexmo+, which used ML algorithms to combine the output of ART, Stoffenmanager, and ECETOC TRA (Savic et al. 2020). Furthermore, the EFSA Roadmap for action on artificial intelligence (AI) in risk assessment and the Roadmap on aggregate exposure specifically mention the need for enhanced accuracy and reliability of exposure models, possible with the use of trustworthy AI deployment (PwC EU Services & Intellera Consulting 2022; Lamon et al. 2024).

AI applications such as machine learning (ML) algorithms could be one of the solutions for these challenges, being able to handle large number of features and/or covariates and identify complex and nonlinear patterns. Specifically, techniques such as data-driven ML models or hybrid modeling (combining ML models with expert knowledge) present a promising opportunity, particularly in reducing the under- and overestimations observed in the currently available models by capturing complex interactions not captured by traditional models. In this study, we explore how various types of ML algorithms could improve model predictions, and (briefly) their implications for regulatory risk assessment, which may contribute to the advancement of occupational exposure modeling. To this end, we use 5 different modeling techniques to create an exposure assessment model and compare them to each other and the current standard, the AREAT model (Franken et al. 2021).

Methods

Case study: asbestos removal exposure assessment tool

Following the conclusions of various validation studies (Schlüter and Tischer 2020) and the scientific support for the need for model improvement, combined with the rise of AI techniques such as ML or hybrid modeling, a case study was developed where an existing mechanistic model was redeveloped with the use of different hybrid modeling and ML techniques. For this case study, the "asbestos removal exposure assessment tool" (AREAT) was selected due to the availability of a large measurement dataset with complete contextual information (Franken et al. 2021).

AREAT was developed to estimate the asbestos exposure risk based on context variables such as the type of material used, the activities done on the material and control measures engaged to reduce the asbestos exposure. The goal of AREAT is to provide an estimate of the level of asbestos workers are exposed to, based on a mechanistic model created with expert knowledge and calibrated with observed measurements.

The mechanistic model, as detailed in Franken et al. (2021), estimates a dimensionless score based on equations (1) and (2):

$$C_{\mathrm{nf}} = (E_{\mathrm{nf}} * H_{\mathrm{nf}} * Lc_{\mathrm{nf1}} * Lc_{\mathrm{nf2}}) * D_{\mathrm{nf}} \quad (1)$$

$$C_{\mathrm{ff}} = (E_{\mathrm{ff}} * H_{\mathrm{ff}} * Lc_{\mathrm{ff1}} * Lc_{\mathrm{ff2}}) * D_{\mathrm{ff}} \quad (2)$$

Where E is the substance emission potential, defined by the product type (the matrix in which the asbestos is incorporated), the sum of the concentrations of chrysotile and amphibole fibers present in the material and the moisture content, H represents the activity emission potential, based on the type of activity involved to remove the asbestos product, Lc_1 and Lc_2 are local control measures in place, and D is the dispersion based on the effect of the room size and the ventilation rate. C (where the subscripts nf and ff denote the near-field and far-field) are aggregated following equation (3):

$$C_t = \frac{1}{t_{\text{total}}} \sum_{\text{tasks}} (t_{\text{exposure}} \times (C_{\text{nf}} + C_{\text{ff}})) + t_{\text{nonexposure}} \times 0 \quad (3)$$

Here, C_t is the dimensionless score for the total exposure during the time t. The model was calibrated using the available measurements in a linear regression model which translates the dimensionless model scores to estimated fiber concentrations in #/m^3 (equation (4)).

$$\ln(\gamma_{ij}) = \beta_{0j} + \beta_1 * \ln(C_t) + \delta_j + \varepsilon_{ij} \quad (4)$$

In equation (4), γ_{ij} represents the estimated asbestos fiber concentrations, C_t is the log-transformed dimensionless score from equations (1) to (3). β_{0j} is the intercept and β_1 is the coefficient of the AREAT score. δ_j and ε_{ij} represent the model error. The calibration of the model lead to a $\beta_0 = 11.41$, δ_j and ε_{ij} of 2.75 and 1.7, respectively. The slope (β_1) was forced on 1 as was the chosen approach for the AREAT model, leading to proportional model scores compared to actual exposure values.

Dataset

In this study, we will use the dataset used to calibrate equation (4) for AREAT to create 5 models using the modeling approaches detailed below. This training dataset is curated from different sources and contains information on 370 personal asbestos fiber exposure measurements from asbestos abatement tasks, taken in the Netherlands over the past decades (Franken et al. 2021). All asbestos fiber analyses were done following SEM/EDXA in accordance with ISO 14966 and reflect fibers with lengths greater than 5 μm and widths extending from the calibrated lower limit of visibility of 0.2 μm up to 3 μm (ISO 2019). Contextual information stored in the database contained detailed information such as the type of asbestos removed, the friability of the asbestos materials, chrysotile and amphibole concentrations, detailed information on the removal process, local control measures applied, and information with regards to the size and ventilation of the workroom. Inclusion criteria for data in this study were (i) core information was documented (eg sampling method, concentrations, uncertainty), (ii) all model determinants could be reliably assessed based on the contextual information, and (iii) relevant site information, unique containment IDs, and relevant information with regards to the laboratory, abatement company, and report were available. More detailed information can be found in the original paper (Franken et al. 2021).

To test the models and their prediction of the asbestos exposure based on the given context variables, we used a validation dataset. This validation dataset contains 270 personal asbestos exposure measurements from asbestos abatement activities, described using the same (context) information and variables as the first dataset (Franken et al. 2023). This dataset was used in order to validate the AREAT model by comparing the exposure as predicted by the model to the actual observed exposure in this. We used this set in the same vein in this study.

Five statistical modeling approaches

Five different statistical modeling methods were compared to the AREAT model: Multiple Linear Regression, Random Forest (RF; variant regression trees), Gradient Boosting Machines (GBM), Bayesian Networks (BN), and Neural Networks (NN). For comparability, the same properties and context information with regards to the asbestos removal processes as used in the AREAT model were used in the 5 statistical modeling methods. We will shortly discuss each model and its application to the data.

Multiple linear regression was selected as linear regression is one of the most used statistical models to estimate relationships between dependent and independent variables (Etemadi and Khashei 2021) and is a well-known and commonly used approach in model development. For this study, our linear regression model will use exposure as output following equation (5).

$$\ln(\gamma_j) = \boldsymbol{\beta_p} * \boldsymbol{I_{p,j}} + \boldsymbol{\beta_h} * \boldsymbol{I_{h,j}} + \boldsymbol{\beta_s} * \boldsymbol{I_{s,j}} + \boldsymbol{\beta_{LC}} * \boldsymbol{I_{LC,j}} + \boldsymbol{\beta_{ACH}} * \boldsymbol{I_{ACH,j}} + \boldsymbol{\beta_d} * \boldsymbol{I_{d,j}} + \boldsymbol{\beta_{ca}} * \boldsymbol{I_{ca,j}} + \boldsymbol{\beta_{cc}} * \boldsymbol{I_{cc,j}} + \epsilon \quad (5)$$

$\ln(\gamma)$ represents the measured fiber concentrations, β_p, β_h, β_s, β_{LC}, β_{ACH}, β_d, β_{ca} and β_{cc} are vectors that hold coefficients for the different levels in product type (p), task (h), state of the material (s), local control measures (LC), ventilation (ACH), room volume (d), and asbestos content of the material for both amphibole (ca) and chrysotile asbestos (cc), respectively, and I the indicator variable $I_{.,j}$ to select the correct level for measurement j ($j = 1,\ldots, N$). Depending on the data, multiple linear regression models can be prone to overfitting and cannot process missing data, which means that either an imputation method is needed or cases with missing data are not included in the model and cannot be predicted. The model was trained using a maximum likelihood algorithm to predict the exposure from the input variables using the caret package in R (Kuhn 2008).

RF (Breiman 2001) and GBM (Friedman 2001) were included for their accuracy, ease-of-use, and the capability of automatically handling interaction effects and nonlinear relationships (Zhang and Haghani 2015). These 2 models are known as “black box” models, which means that we put in all the information we have (in this case, the independent variables of the AREAT approach) and ask for a predictive model of the dependent variable (the asbestos fiber count). Both these models use a decision-tree-based approach with a myriad of trees. This means that for each model, thousands of decision trees are made that try to predict the outcome as best as they can based on the variables used. Each tree is made by starting with *m* randomly splits derived from the independent variables from the dataset, and selecting one of these *m* that can be used to split the dependent variable best in homogenous groups. Each of these groups or branches is then split again in the same fashion until a preset criterion is reached, such as that the groups become too small to split, or are so homogenous that splitting won’t improve them. A visual representation can be seen in Fig. 1a. After creating these trees, a prediction can be made by running the independent variables through all of the trees and taking the democratic consensus of the trees as the estimated outcome. The RF method is the original approach to this, the GBM model adds an optimization algorithm to the RF that uses the mistakes of the last tree to improve prediction on the current tree. The RF and GBM models were developed with the use of the caret package in R (Kuhn 2008).

NN (Riedmiller 1994) were included for its capability of capturing complex nonlinear relationships and their scalability to handle large datasets (Jiang et al. 2024). NN work through decomposition of relations in the data through several layers, as visualized in Fig. 1b. All nodes within each layer are connected to all nodes within the next layer, where the strength of these connections is adjusted to optimize the estimation of the dependent variable, in this case, the asbestos fibers, in the final layer or node. NN, like RF and GBM, are a black box approach which makes it hard to interpret the relations on which the model is build. NN are known for needing a large, high-quality dataset for its training, which is hard to obtain in exposure measurement. However, NN have been very popular, and this study is exploratory; therefore, the NN was built and compared with the other methods. The NN was estimated using the Neuralnet package in R (Fritsch et al. 2019).

BN were included in the analysis for their capability of handling missing data, handling uncertainty, and their transparency (Kitson et al. 2023). BN uses a directed a-cyclic graph to map variables and how they influence each other within the network as visualized in Figs. 1c and S1 for the actual BN derived, and as such does not necessarily have dependent and independent variables. When presented with a case, it will give a prediction for any variables not included in the case which allows estimation of the dependent variable and any missing variables based on the observed independent variables. BN uses probability theory as well as Bayes theorem to estimate a distribution for each node, thus incorporating uncertainty in each variable. The resulting network quantifies and visualizes the relations between the different variables. Before training a network on data, prior expectations and limitations can be set to allow for a quantification of expert knowledge beyond the mere selection of the used variables, as is the case in the other models. This also aids the algorithm to create a more stable model with less data. In this case, however, the model used uniform weak priors, with no information or expectation added before training. The BN was estimated with use of Bayes Server v10 (Bayes Server v10.10).

Model tuning

The RF, GBM, and NN models need certain tuning parameters to allow for optimal model estimation. To this end, the caret package (Kuhn 2008) was used to test several parameters. For the RF model, the *m* was tested for the values 5, 10, 20, and 30, resulting in an optimal *m* of 20. For the GBM, the number of trees (500, 1,000, 2,000), depth of interaction (5, 7, 9), shrinkage (0.075, 0.1, 0.15), and minimal observations per node (7, 10, 15) were tested, with optimal results in bold. For the NN, 3 layers (layer 1 = (20, 40), layer 2 = (0, 20, 40), layer 3 = (0, 20, 40)) were tested for the 46 different options resulting from the used variables, and subsequently a model with 2 hidden layers of 40 nodes was used.

Model evaluation

The 5 models were trained using the training data with 370 measurements. Then, all 5 new models and the AREAT model were used to estimate the exposure for the measurements from the validation data based on the context variables provided. As such, for each of these measurements, the measured exposure in the validation set is the golden standard to which the model estimation is compared. The correlation between the observed and the estimated exposure was tested with Pearson correlation coefficients (r), which indicates whether the estimation and observation follow the same trends in the data. The root mean squared error (RMSE) and the mean absolute error (MAE) of the difference between the observed and estimated scores indicate the average difference between the estimated and observed exposure in the validation set, with the RMSE laying more emphasis on extreme deviations than the MAE.

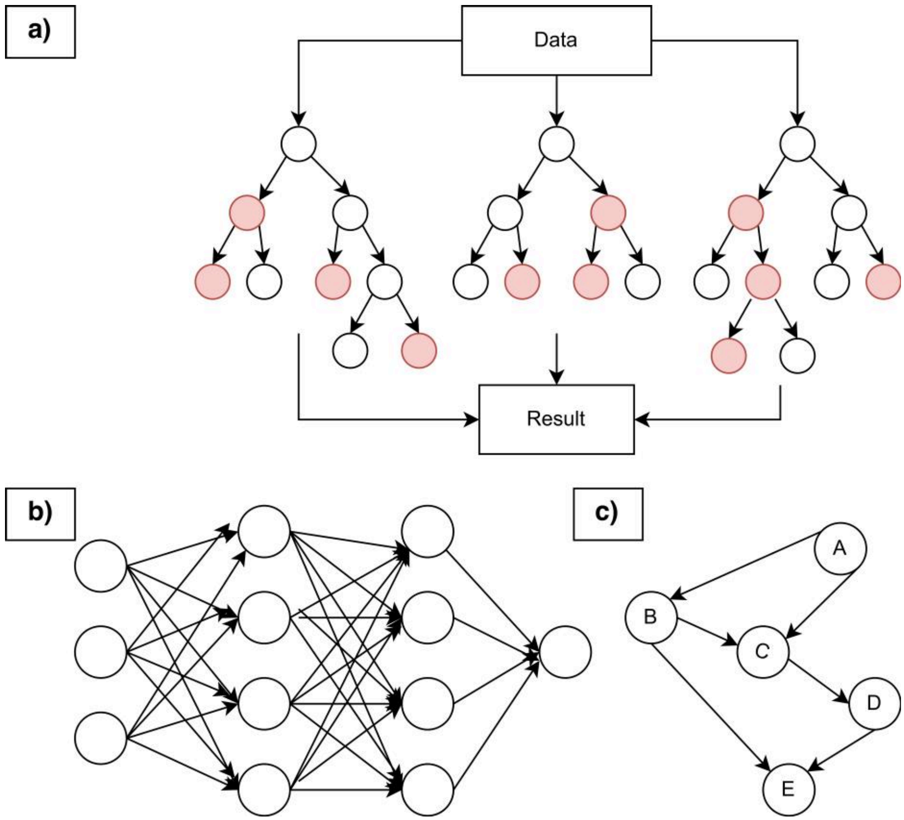

Figure 1 Visual representations of the (a) Random Forest, (b) Neural Network, and (c) Bayesian Network approach.

The model bias was calculated following equation (6), where ŷ is the natural log of the GM predicted exposure level by the model, y_i is the natural log of the measured exposure, and n_0 is the number of measurements in the validation dataset. The bias shows whether a model systematically over or underestimates the results by summing all errors and dividing by the number of measurements; a positive value indicates a tendency of the model to overestimate measured exposure, and negative value a tendency for underestimation of the exposure. The explained variance (R^2) indicates how much variation in the data is accounted for by the model. For the correlation and R^2, a value closer to 1 is better. For the other metrics, a lower value indicates a better fit. All analyses were conducted with the use of R statistical software (version 4.3.3).

$$\text{Bias} = \sum_{i=1}^{n0} \frac{(\hat{y}_i - y_i)}{n_0} \tag{6}$$

Results

For the model evaluation, both the observed and the estimated fiber count have been log transformed to obtain normally distributed results, and the fitmeasures are based on these results. Figure 2 shows the relation between the log-transformed model estimates and the log-transformed measured asbestos fiber concentrations for each developed model. When comparing the different models using the statistical evaluation metrics in Table 1, it can be observed that for all the statistical modeling approaches, the (Pearson) correlation between data and estimation and thus the explained variance expressed as the R^2 were improved compared to AREAT. RF and GBM performed best with regard to the RMSE on log-transformed data, while the BN showed the smallest absolute bias (closest to zero, smallest systematic over- or underestimation) and MAE. This indicates that the BN is generally speaking closest to the observed value, but when outliers are weighted more strongly, it is outperformed by the GBM and RF.

Figure 3 shows the log bias between the measured and predicted concentrations, with the line in each block indicating the median error. From the figure, a value of 0 means a perfect prediction, so models with a median close to 0 indicate good central accuracy. The variability of the boxplot indicates the consistency of the predictions. As

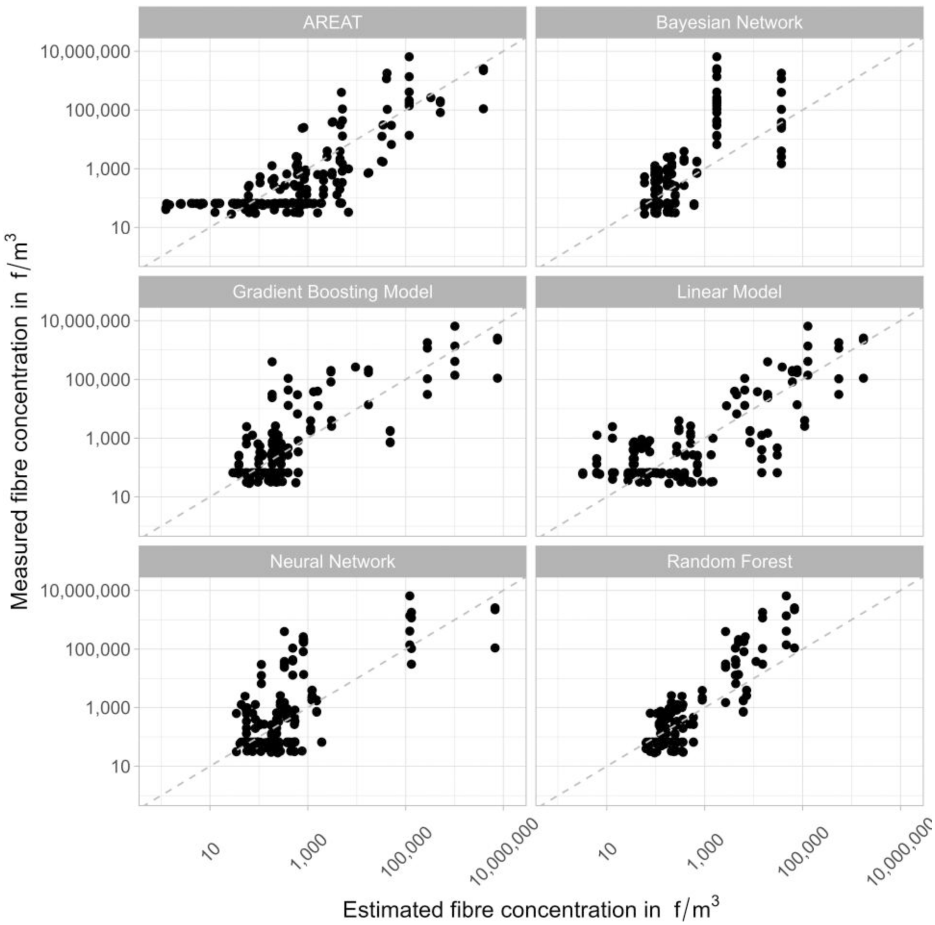

Figure 2 Estimated asbestos fiber concentrations compared to measure asbestos fiber concentrations for the different models.

Table 1 Validation statistics between all the different models.

Model	Pearson	RMSE	MAE	Bias	R^2
AREAT	0.71	2.13	1.80	0.65	0.51
Bayesian Network	0.78	1.71	1.1	−0.14	0.61
Gradient Boosting Machines	0.79	1.60	1.16	0.40	0.62
Multiple linear regression	0.74	2.04	1.6	−0.15	0.55
Neural Network	0.68	1.90	1.41	−0.14	0.46
Random Forest	0.89	1.43	1.15	0.27	0.79

MAE = mean absolute error; RMSE = root mean squared error.

can be seen, the AREAT and the linear model show a larger variability in error between estimations, even though the linear regression model has a small average error, or bias.

As several models do not provide a confidence interval, we cannot compare these across methods. However, in 95% of the estimations, the AREAT model a factor of 24 or less off from the observations. For the LM, RF, GBM, NN, and BN this factor is, respectively, 16.8, 14.5, 24.5, 76.3, and 50.1. Another indication of this variability is how often the estimation is less than a factor 2 off the observation. For the AREAT model, this is 79% of the time, for the LM, RF, GBM, NN, and BN this is, respectively, 67%, 84%, 84%, 85%, and 81%.

Discussion

The 5 different modeling approaches as presented in the case study illustrate possible advancements by ML and

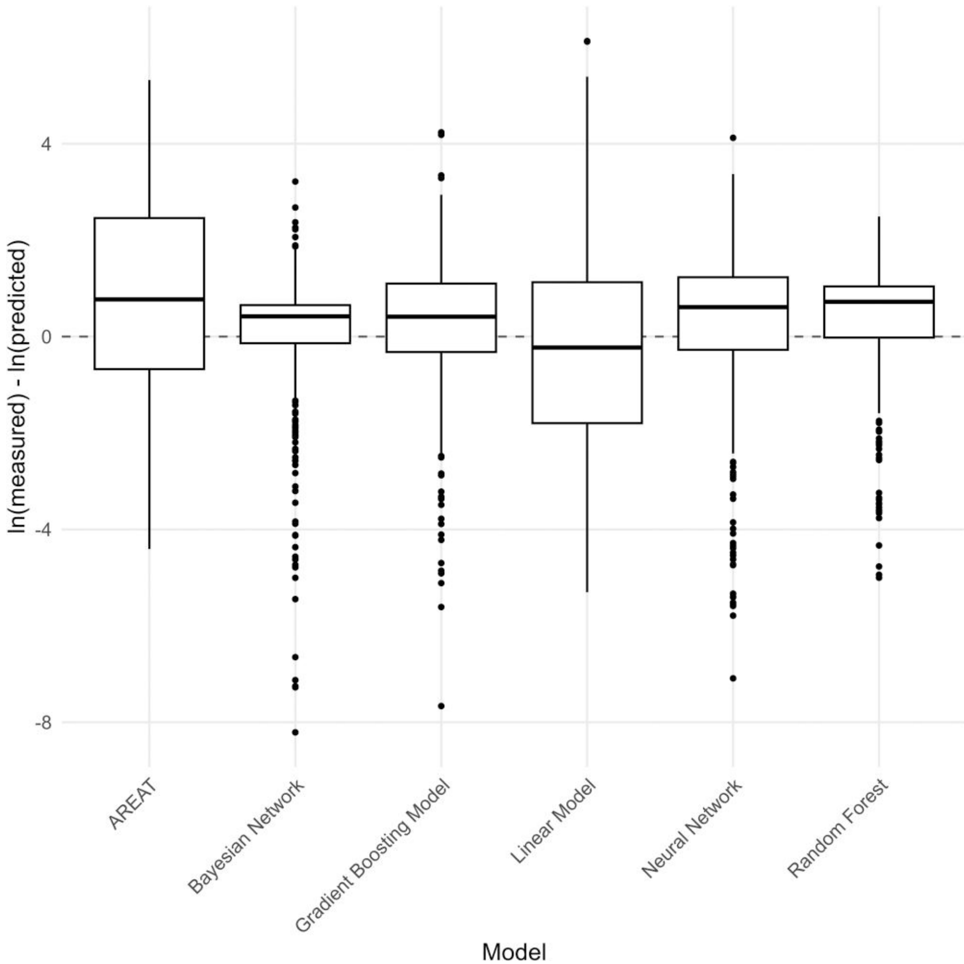

Figure 3 Variation in model residuals for each different model.

hybrid approaches to develop occupational exposure models. These 5 approaches were compared to the original model. Results showed that performance metrics were best for the RF, GBM, or the BN, and were better than the existing AREAT model. The usage of a NN showed an improved estimation to a lesser degree for the MAE and RMSE, but a lower (Pearson) correlation to the measured data or explanation of the variance (R^2). Multiple linear regression did lower the bias but had no impact on the other model evaluation parameters (correlation, RMSE, MEA, R^2).

The case study presented in this paper illustrates possible advancements in occupational exposure modeling through the application of advanced statistical models. The regulatory landscape for occupational exposure assessment such as REACH relies on tiered approaches that prioritize simplicity and conservatism, often at the expense of precision. Despite an increase in the use of advanced statistical models in various domains, the transparency of the methods and the reasoning behind the decisions and estimates of these complex models often remain unclear. This inability to explain the reasoning of a model hampers their future usage (Hagras 2018). This challenge arises from the trade-off between interpretability and predictive power, as demonstrated in our case study. Linear regression models are transparent and interpretable, especially for domain experts, but they often oversimplify complex exposure pathways. For example, in the current study, the multiple linear regression model associated higher air exchange rates (ACH) with increased asbestos fiber concentrations—an interpretation that does not align with general expectations. However, this may be partially explained by the nature of asbestos abatement activities: higher ACH values typically indicate that abatement occurs within enclosed containments, where elevated exposures are indeed expected. In contrast, lower ACH values are generally associated with open or semi-controlled environments, since enclosures are not required here, where exposures tend to remain below the Dutch occupational exposure limit (OEL) of 2,000 f/m^3. The coefficients of the linear regression model are presented in Table S1 in the supplementary materials.

More advanced ML approaches, such as RF, GBM, and NN, offer improved predictive accuracy as demonstrated in this case study. However, they do so at the cost of interpretability. Continuing with the ACH example, these models identified ACH as the most important predictor through feature importance metrics. Yet they offer no insight into the nature of the relationship between ACH and asbestos concentrations due to their black-box nature.

To address the trade-off between transparency and accuracy, hybrid modeling approaches such as BN offer a promising solution. These models allow for the incorporation of expert knowledge to define relationships (eg, how ACH influences asbestos exposure) while enabling data-driven refinement within those expert-defined constraints. Applying this to the ACH example, an expert could specify that high ACH values are typically associated with high exposures due to containment-based abatement, the ML component could then adjust the strength of this relationship based on observed data and describe the dependencies such that the relation between ACH and exposure is dependent on abatement location. This approach prevents models from making nonphysical or unrealistic predictions while still leveraging data-driven improvements (Heckerman 2008). This approach can also aid in fine-tuning the model, for example, to examine the predictions in the validation set where the model is performing badly and see if expert knowledge can improve this.

The graphical style of the networks further allows interpretation and visualization of the relations between predictors and predictions in a way comprehensive to the average researcher without a Bayesian background. Another practical point is the allowance for missing data. In BN, unknown inputs are taken as a distribution of their most probable values based on the other, known determinants. This allows the network to both use incomplete data in model creation and make predictions with missing input data, increasing the data available for both training and predicting. Additionally, model estimates may improve over time when techniques improve (such as ventilation or local control measures), which are easily updated within BN as new data can be added over time to improve parametrization without re-estimating the whole model.

In addition to the improvement of existing models, hybrid modeling may also be a useful tool to extend the applicability domains of models by aggregation of different exposure routes, which aligns with EFSA's roadmap for action for advancing aggregate exposure to chemicals in the EU (Lamon et al. 2024) as well as other European initiatives such as PARC. The generation of sub-networks within a single model allows for utilizing determinants of exposure relevant for both routes and keeping separate route specific exposure determinants as well.

Conclusion

The case study shows promise for the use of advanced statistical techniques that combine expert knowledge with data-driven methods to improve current- or develop new occupational exposure models, in such a way that the models are explainable while becoming more accurate and future proof.

Supplementary material

Supplementary material is available at *Annals of Work Exposures and Health* online.

Funding

No external funding was provided for this study.

Conflicts of interest

The authors declare no conflict of interest relating to the material presented in this article. Its contents, including any opinions and/or conclusions expressed, are solely those of the authors.

Data availability

The data underlying this article cannot be shared publicly as the data are owned by many different third party companies who shared their data with TNO for analysis.

References

Ak MF, Yucesan M, Gul M. 2022. Occupational health, safety and environmental risk assessment in textile production industry through a Bayesian BWM-VIKOR approach. Stoch Environ Res Risk Assess. 36: 629–642. https://doi.org/10.1007/s00477-021-02069-y.

Breiman L. 2001. Random forests. Mach Learn. 45:5–32. https://doi.org/10.1023/A:1010933404324

Etemadi S, Khashei M. 2021. Etemadie multiple linear regression. Measurement. 186:110080. https://doi.org/10.1016/j.measurement.2021.110080.

Fantke P et al. 2022. The European exposure science strategy 2020–2030. Environ Int. 170:107555. https://doi.org/10.1016/j.envint.2022.107555.

Franken R et al. 2021. The development and calibration of a mechanistic asbestos removal exposures assessment tool (AREAT). Ann Work Expo Health. 65: 789–804. https://doi.org/10.1093/annweh/wxaa112.
Franken R, Schinkel J, Spaan S. 2023. Validation and recalibration of the asbestos removal exposure assessment tool (AREAT). Ann Work Expo Health. 67: 650–662. https://doi.org/10.1093/annweh/wxad010.
Fransman W. 2017. How accurate and reliable are exposure models? Ann Work Expo Health. 61:907–910. https://doi.org/10.1093/annweh/wxx068.
Friedman JH. 2001. Greedy function approximation: a gradient boosting machine. Ann Stat. 29:1189–1232. https://doi.org/10.1214/aos/1013203451.
Fritsch S, Guenther F, Wright M. 2019. _neuralnet: Training of Neural Networks_. R package version 1.44.2. https://CRAN.R-project.org/package=neuralnet
Hagras H. 2018. Towards human-understandable explainable AI. Computer (Long Beach Calif). 51:28–36. https://doi.org/10.1109/MC.2018.3620965.
Heckerman D. 2008. A tutorial on learning with Bayesian networks. In: Holmes DE, Jain LC, editors. Innovations in Bayesian networks. Studies in computational intelligence. Vol. 156: Springer. p. 33–82. https://doi.org/10.1007/978-3-540-85066-3_3
ISO 14966. 2019. Ambient air - determination of numerical concentration of inorganic fibrous particles - scanning electron microscopy method. International Organization for Standardization.
Jiang C et al. 2024. Network properties determine neural network performance. Nat Commun. 15:5718. https://doi.org/10.1038/s41467-024-48069-8.
Jones MR. 2022. Models in occupational hygiene. Ann Work Expo Health. 66:417–418. https://doi.org/10.1093/annweh/wxac016.
Kitson NK et al. 2023. A survey of Bayesian network structure learning. Artif Intell Rev. 56:8721–8814. https://doi.org/10.1007/s10462-022-10351-w.
Kuhn M. 2008. Building predictive models in R using the caret package. J Stat Softw. 28:1–26. https://doi.org/10.18637/jss.v028.i05.
Lamb J et al. 2017. Between-user reliability of tier 1 exposure assessment tools used under REACH. Ann Work Expo Health. 61:939–953. https://doi.org/10.1093/annweh/wxx074.
Lamon L et al. 2024. Roadmap for action for advancing aggregate exposure to chemicals in the EU. EFSA Support Publ. 21:EN-8971. 364 pp. https://doi.org/10.2903/sp.efsa.2024.EN-8971.
Landberg HE, Axmon A, Westber H, Tinnerberg H. 2017. A study of the validity of two exposure assessment tools: Stoffenmanager and the advanced REACH tool. Ann Occup Expo Work. 61:575–588. https://doi.org/10.1093/annweh/wxx008.
Patton AN, Medvedovsky K, Zuidema C, Peters TM, Koehler K. 2022. Probabilistic machine learning with low-cost sensor networks for occupational exposure assessment and. Ann Work Expo Health. 66:580–590. https://doi.org/10.1093/annweh/wxab105
PwC EU Services & Intellera Consulting. 2022. Roadmap for action on artificial intelligence for evidence management in risk assessment. EFSA Support Publ. 19: EN-7339. https://doi.org/10.2903/sp.efsa.2022.EN-7339.
Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.
Riedmiller M. 1994. Rprop—description and implementation details. Technical Report. University of Karlsruhe.
Savic N, Lee EG, Gasic B, Vernez D. 2020. TREXMO plus: an advanced self-learning model for occupational exposure assessment. J Expo Sci Environ Epidemiol. 30: 554–566. https://doi.org/10.1038/s41370-020-0203-9
Schinkel J et al. 2014. Reliability of the advanced REACH tool (ART). Ann Occup Hyg. 58:450–468. https://doi.org/10.1093/annhyg/met081.
Schlüter U et al. 2022. Exposure modeling in Europe: how to pave the road for the future as part of the European Exposure Science Strategy 2020–2030. J Expo Sci Environ Epidemiol. 32:499–512. https://doi.org/10.1038/s41370-022-00455-4.
Schlüter U, Tischer M. 2020. Validity of tier 1 modeling tools and impacts on exposure assessments within REACH registrations—ETEAM projects, validation studies and consequences. Int J Environ Res Public Health. 17:4589. https://doi.org/10.3390/ijerph17124589.
Spinazzè A et al. 2019. How to obtain a reliable estimate of occupational exposure? Review and discussion of models' reliability. Int J Environ Res Public Health. 16:2764. https://doi.org/10.3390/ijerph16152764.
Tischer M, Lamb J, Hesse S, Van Tongeren M. 2017. Evaluation of tier one exposure assessment models (ETEAM): project overview and methods. Ann Work Expo Health. 61: 911–920. https://doi.org/10.1093/annweh/wxx066.
Zhang Y, Haghani A. 2015. A gradient boosting method to improve travel time prediction. Transp Res Part C Emerg Technol. 58:308–324. https://doi.org/10.1016/j.trc.2015.02.019.