

Accuracy, Conservatism and Parsimony of Three Vapour Intrusion Models Used in Sweden

Jeroen Provoost¹, Stephanie Nouwen² & Jan Bronders³

¹ Independent researcher, Finland

² Open University Netherlands (OUNL), Department of Science, Heerlen, Netherlands

³ Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium

Correspondence: Jeroen Provoost, Independent researcher, Finland. E-mail: jprovoost-research@yahoo.co.uk

Received: July 25, 2018

Accepted: August 4, 2018

Online Published: September 27, 2018

doi:10.5539/mas.v12n10p128

URL: <https://doi.org/10.5539/mas.v12n10p128>

Abstract

This study presents an evaluation of three screening-level models, namely the Dilution Factor (DF) model from 1996, the update version from 2005, as well as the Johnson and Ettinger model (JEM) from 1997, that are applied within the frameworks for contaminated land management (CLM) in Sweden. This evaluation applies, besides a deterministic approach (point estimate), a probabilistic assessment plus sensitivity analysis. The latter approach allows the models to be ranked according to conservatism, accuracy and parsimony by contrasting predicted and observed soil and indoor air concentrations for two contaminants (benzene and ethylbenzene), as to determine their suitability for application within CLM. The results and conclusions from this study suggest that the most accurately model for predicting the soil and indoor concentration is the JEM followed by the DF 2005 and 1996. Predictions of the soil air concentration are primarily driven by variation in physico-chemical parameters. The variation in indoor air concentration by physico-chemical and/or soil parameters for the DF models, while for the JEM soil parameters dominate. The deterministic analysis showed that default parameter values could be revised as to increase the conservatism, and be closer to the probabilistic 95-percentile predicted indoor air concentration. The DF 1996 model includes a limited number of parameters, and this analysis shows that a model with more parameters is more accurate, and less conservative. The DF 2005 seems to be the most parsimonious model as it is accurate, sufficiently conservative, and has 14 parameters, whereas the DF 1996 with 9 parameters is the most conservative and the JEM with 27 parameters the most accurate with an increased probability to produce false negative predictions. For the latter some of the dominant parameters cannot easily be measured on site.

Keywords: accuracy, conservatism, parsimony, vapour, intrusion, model, Sweden

1. Introduction

In Sweden contaminated soils are regulated by the Sweden Environmental Protection Agency (S-EPA), who issued guideline values in 1996 as part of their assessment framework for contaminated land management (CLM), which was expanded in 1998 to include petrol stations (S-EPA, 1996; 1998). The framework includes tiered risk assessment and triggers for remedial actions, allowing the owner of the site and S-EPA to either set priorities or determine the need for risk reduction.

The generic exposure model as described in S-EPA (1996) and resulting guideline values are applicable for typical Swedish land-use and are sufficiently conservative based on the precautionary principle. If generic guideline values do not apply (for example if site conditions, such as building type, differ) a site specific assessment should be conducted. For consistency reasons a model for the prediction of site specific concentrations and exposure (S-EPA, 2005a) was issued, and includes lists of typical parameter values (physical and chemical, bioconcentration factors, (eco)toxicological reference values), as well as background concentrations and exposure. The 2005 model includes more parameters in comparison to the 1996 model, which increases the parsimony (predict with the fewest possible assumptions). The S-EPA published a supplementary guidance to risk assessment of contaminated areas (S-EPA, 2005b; 2009) that advises on how to adapt the model to site specific conditions. Additional information on the approach in Sweden can be found in Carlon et al. (2007), and its progress on managing contaminated land in van Liedekerke et al. (2014).

The parameters needed to run the model from 1996 include measured concentrations in soil and groundwater,

resulting in phase distribution. The model does not include sorption and degradation during transport, which in turn is mainly driven by soil properties. The 1996 and adapted 1995 model are both used for the derivation of general and site specific guideline values. Model predictions can be made for different land-use types (for example industrial area, considered less sensitive, or residential area, considered sensitive) and routes (for example consumption of locally grown vegetables and groundwater, or inhalation of air contaminated by vapour intrusion). Exposure from the different routes are summed and compared to toxicological reference values, like the Tolerable Daily Intake (mg/kg.bw.day) for oral exposure, and Reference Concentration (mg/m³) for inhalation exposure. The reader is referred, for a description of the legal framework, scientific basis, technical approach, receptors taken into account and the derivation method for guideline values, to reports on contaminated land management in Sweden (S-EPA, 1996; 1998; 2005a; 2005b; 2009, Carlon et al., 2007).

Risk assessment of contaminated soils is associated with variation and implies the use of models within an assessment framework. Models are applied to predict a point estimate of for example the soil and indoor air concentration based on a set of default parameters, also called the deterministic approach. This framework should provide sufficient conservatism in protecting the general public health, as well as adequately discriminate between sites that need further action, which raises the issue of accuracy (Hers, 2004; Fitzgerald, 2009). The Swedish guidelines do not reflect a worst case scenario, but aims to protect most individuals ($\pm 95\%$) (Öberg, 2006; Sanders, Bergback & Öberg, 2006).

A major pathway of exposure is inhalation by humans of indoor air that is contaminated, resulting from nearby buried sub-surface volatile chemicals. Soil vapour can enter a building and, if contaminated, affect the indoor air quality negatively (Kaplan et al., 1993; Fugler & Adomait, 1997). Predictions the indoor air concentration in a building and the related human exposure is convoluted and is influenced by physico-chemical, environmental and building factors, which are subject to variation (McAlary et al., 2011; Provoost et al., 2010). As opposed to a deterministic approach, variation can be included in a probabilistic approach by computing probability distribution functions that propagate to a distribution of predicted air concentrations. The probabilistic approach, which includes a sensitivity analysis, gives insight into what parameters drive the predicted outcomes, and the effect of variation thereon.

The Swedish Dilution Factor model from 1996 (S-EPA, 1996) and the Johnson and Ettinger model (Johnson and Ettinger, 1991; 1997), both frequently used within the framework of CLM in Sweden, were assessed for their accuracy and conservatism by Provoost et al. (2008, 2010, 2013) and via a probabilistic approach by Provoost et al (2014). However, the peer-reviewed literature does not include research on the Swedish Dilution Factor model from 2005 (S-EPA, 2005a). The model was not investigated for accuracy, conservatism or parsimony for regulatory purposes.

Different models can be selected depending on the context it is used in, so regulators may want to use a model with a sufficient level of conservatism, while those that caused the contamination may desire predicted indoor air concentrations that are very accurate (McAlary and Johnson, 2009, Provoost et al. 2013). Thus, the purpose of this study is to rank the three vapour intrusion (VI) models for their accuracy, conservatism and parsimony, so allow users to make a better informed decision on what model to apply.

Chapter 2 describes the models used for this study and gives further information of the contaminated site that provides input for the observed air concentrations. Chapter 3 defines the terms accuracy, conservatism and parsimony in relation to the deterministic and probabilistic approach. Chapter 4 provide the results, and chapter 5 and 6 the conclusion and discussion, followed by further research needs.

2. Description of the Models and Site

2.1 Description of the Three Models

2.1.1 Swedish Dilution Factor Model from 1996

In 1996 the S-EPA issued a framework for risk management within contaminated land management. The guidance contains models and data to be used for the derivation of generic guideline values, when exceeded indicate a possible effect on human health and/or the environment (S-EPA, 1996). The guideline includes multiple routes of exposure to humans, and for volatile organic compounds (VOC) the vapour intrusion (VI) route dominates the exposure. The model for VI, hereafter called DF SE 1996, is a dilution factor (DF) between the source in the soil and the indoor air as shown in Figure 1.

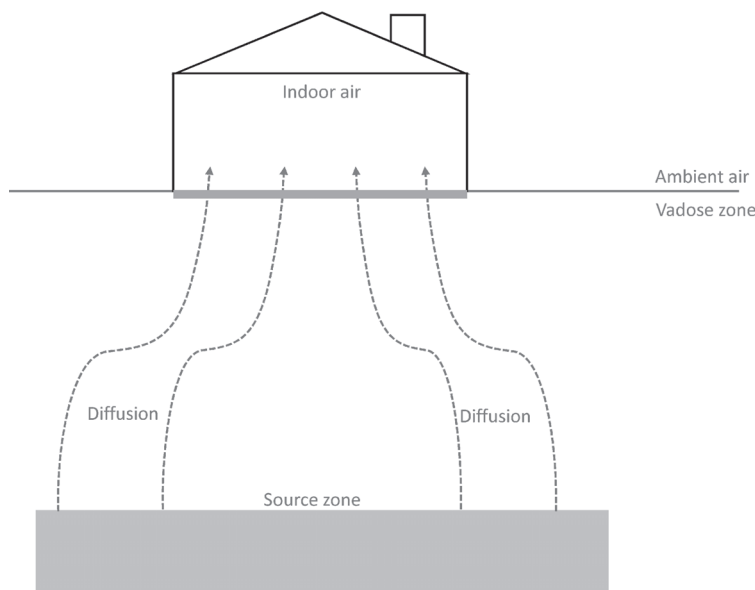


Figure 1. Conceptual model for the 1996 DF model from Sweden

The transport process is based on diffusion, and because of uncertainty the source concentration is assumed to be constant over time and thus conservative. Equilibrium fugacity of the VOC is assumed between the solid phase, pore water, and soil air. Equilibrium between the pore water and soil air is predicted by using the Henry's constant (vapour pressure and solubility). Soil and contaminant properties are used to predict the soil air concentration, from which the indoor air concentration is calculated by applying a DF. The below formulas indicate the default parameter value where applicable. A default parameter value is used for the volume and ventilation rate of the single compartment building.

The concentration in pore water (C_w) is calculated from the total concentration of the contaminant in the soil (C_s) by using the formula:

$$C_w = C_s \times \left[K_d + \frac{(\theta_w + \theta_a \times H)}{\rho_b} \right] \quad (1)$$

Where the K_d is the distribution coefficient soil-water (l/kg), θ_w the soil water content ($0.3 \text{ dm}^3 \text{ water/dm}^3 \text{ soil}$), θ_a the soil air content ($0.2 \text{ dm}^3 \text{ air/dm}^3 \text{ soil}$), H the Henry's constant (-) and the ρ_b the dry soil bulk density (1.5 kg/dm^3).

If the distribution coefficient soil-water (K_d) is not provided it is calculated from:

$$K_d = K_{oc} \times f_{oc} \quad (2)$$

Where K_d is the distribution coefficient soil-water (l/kg), K_{oc} is the distribution factor between water and organic carbon (l/kg) and f_{oc} the organic carbon content of the soil (0.02 -) by weight.

The soil pore air concentration is calculation via the formula:

$$C_a = H \times C_w \quad (3)$$

Where C_a is the vapour concentration in air (mg/dm^3), H Henry's constant (-) and C_w the concentration in pore water (mg/dm^3).

The concentration in the indoor air (C_{ia}) is given by:

$$C_{ia} = C_a \times DF_{ia} \quad (4)$$

Where C_{ia} is the concentration in the indoor air (mg/dm^3), C_a the vapour concentration in air (mg/dm^3) and the DF_{ia} the dilution factor for indoor air to soil air (-).

The DF_{ia} is derived from empirical data from MDEP (1994) and the default was set to 1:20 000. The DF_{ia} can be adapted to the floor type. The DF_{ia} for an open floor basement was put on 1:5 000 and for a concrete basement floor about 1:70 000.

2.1.2 Swedish Dilution Factor model from 2005

In 2005 the model for calculating soil screening values (SSV), hereafter called SE DF 2005, underwent a revision. The main difference is the DF for VI which is calculated instead of assumed to be a fixed factor of 1:20 000. The calculation of the DF_{ia} from equation 4 was adapted, and now includes dimensions of the building and temporal perspective (S-EPA, 2005a) and allows the model to be used for site specific assessments.

The model assumes that a zone of influence (C_d) is situated directly under the floor as indicated in Figure 2. The concentration in the building (C_{hus}) depends on the volume of the building (V_{hus}), surface floor (A), concentration in the C_d and the soil air intrusion in the building (L_a).

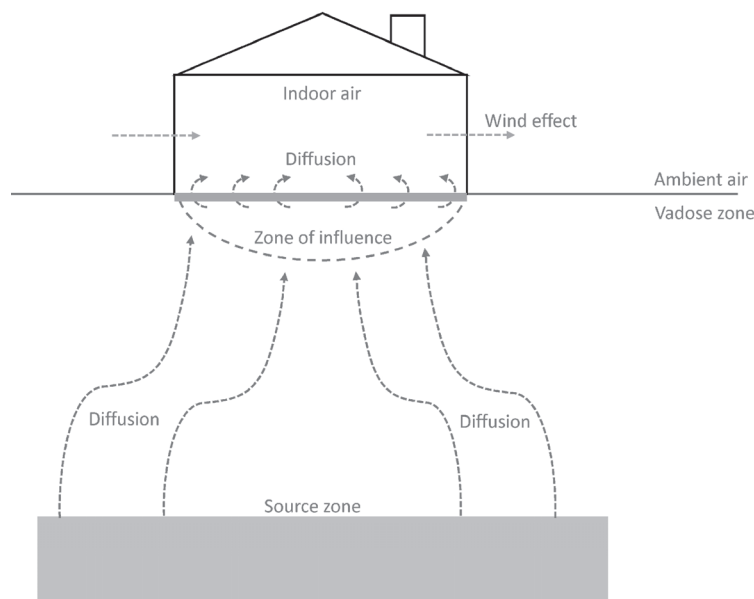


Figure 2. Conceptual model for the 2005 model from Sweden

The depth of the contaminant (Z) influences the flux of soil vapour (D) to C_d . D is a function of the porosity and water content of the soil and is related to the concentration of the contaminant in the soil pore air in the contaminated zone (C_a). The calculation of C_a uses as well equations 1-3, while the DF for indoor air to soil air (DF_{ia}) is given by:

$$DF_{ia} = \frac{L_a}{V_{hus} \times l} \times \frac{A \times D}{L_a \times Z + A \times D} \quad (5)$$

Where DF_{ia} is the dilution factor for indoor air to soil air (-), A the surface floor (100 m²), D soil vapour flux (m²/day), L_a the soil air vapour intrusion in the building (2.4 m³/day), V_{hus} the volume of the building (240 m³), l the indoor air exchange rate (12/day or 0.5/hour), and Z the depth of the contaminant (0.35 m).

The diffusion of vapour in the soil is estimated via:

$$D = D_0 \times \frac{\theta_a^{(10/3)}}{\epsilon^2} \quad (6)$$

Where D is the diffusion (m²/day), D_0 the diffusion in air (0.7 m²/day or 8.10⁻⁶ m²/s), θ_a air filled porosity (0.1 -), ϵ total soil porosity (0.4 -).

The DF as presented above does not take into account the reduction as a result of temporal depletion of the contaminant. Contaminants for which the VI is an important route of exposure often have a low K_d and a high Henry constant, resulting in a depletion of the contaminant closest to the zone of influence under the building. This might result in reduced temporal indoor air concentrations. To account for this effect the model as applied by the US EPA was incorporated (US EPA, 1996) and depends on the chemical specific physico-chemical parameters. The model calculates an average soil air concentration over 5 years and the formula is given below.

$$DF_{SSL} = \frac{A}{V_{hus} \times l} \times 2 \times \sqrt{\frac{D_a}{\pi \times T} \left[\frac{\theta_w + H \times \theta_a + K_d \times \rho_b}{H} \right]} \quad (7)$$

Where DF_{SSL} is the time specific DF (-), A the surface floor (100 m^2), V_{hus} the volume of the building (240 m^3), l the indoor air exchange rate (12/day or 0.5/hour), D_a the apparent diffusion (m^2/day), π (3.14), T time (1825 days), θ_w the soil water content ($0.3 \text{ dm}^3 \text{ water}/\text{dm}^3 \text{ soil}$), H the Henry's constant (-), θ_a the soil air content ($0.2 \text{ dm}^3 \text{ air}/\text{dm}^3 \text{ soil}$), K_d is the distribution coefficient soil-water (l/kg) and ρ_b the dry soil bulk density ($1.5 \text{ kg}/\text{dm}^3$).

The apparent diffusion (D_a) is calculated from:

$$D_a = \frac{D}{\frac{\rho_b \times K_d + \theta_w}{H} + \theta_a} \quad (8)$$

Where D_a is the apparent diffusion (m^2/day), D the diffusion (m^2/day), ρ_b the dry soil bulk density ($1.5 \text{ kg}/\text{dm}^3$), K_d is the distribution coefficient soil-water (l/kg), H Henry's constant (-), θ_w the soil water content ($0.3 \text{ dm}^3 \text{ water}/\text{dm}^3 \text{ soil}$) and θ_a the soil air content ($0.2 \text{ dm}^3 \text{ air}/\text{dm}^3 \text{ soil}$).

The overall DF ($DF_{\text{ia,inne}}$) is then calculated by combining the DF from equation 4 with 7:

$$DF_{\text{ia,inne}} = \frac{1}{\frac{1}{DF_{\text{ia}}} + \frac{1}{DF_{\text{SSL}}}} \quad (9)$$

Where $DF_{\text{ia,inne}}$ is the overall dilution factor, DF_{ia} the DF for indoor air to soil air, and DF_{SSL} the time specific DF (-).

The overall DF accounts for diffusion through an uncontaminated soil layer (DF_{ia}) and for dilution due to depletion of the soil layer closest to the zone of influence (DF_{SSL}). The indoor air concentration is then calculated by applying:

$$C_{\text{ia}} = C_a \times DF_{\text{ia,inne}} \quad (10)$$

Where C_{ia} is the concentration in the indoor air (mg/dm^3), C_a the vapour concentration in air (mg/dm^3) and the $DF_{\text{ia,inne}}$ the overall dilution factor for indoor air to soil air (-).

2.1.3 Johnson and Ettinger model from 1991

In 1998 the US EPA developed a model, hereafter called JEM, to predict the indoor air concentrations as a result of sub-surface contamination. The model is based on publications from Johnson and Ettinger (1991; 1997) and Johnson et al. (1999). The Johnson and Ettinger model, hereafter called JEM, includes both diffusion and convection as transport of vapour to the indoor air, as indicated in Figure 3.

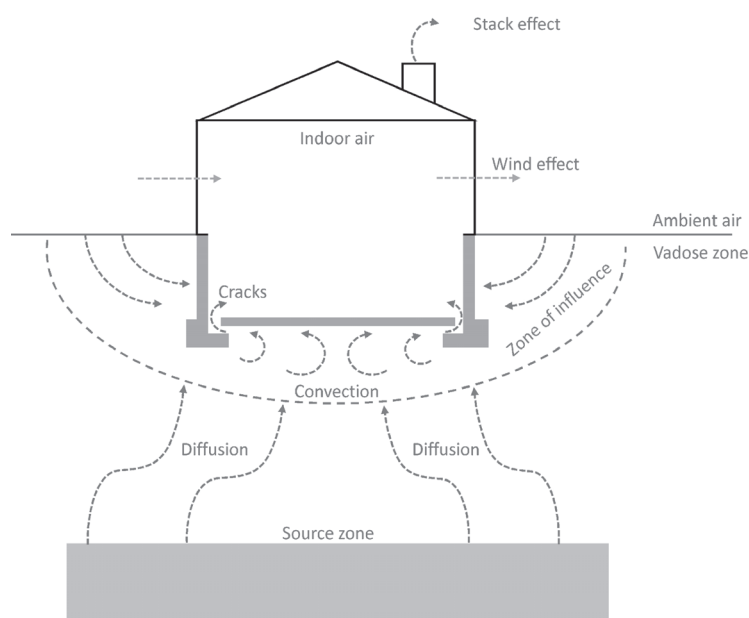


Figure 3. Conceptual model for the 1991 JEM

The model allows the user to include a set of site-specific data as well as parameters related to physico-chemical, soil and building properties. The general model is given below:

$$\alpha = \frac{\left[\left(\frac{D_T^{\text{eff}} \times A_B}{Q_{\text{building}} \times L_T} \right) \times \exp\left(\frac{Q_{\text{soil}} \times L_{\text{crack}}}{D_{\text{crack}} \times A_{\text{crack}}} \right) \right]}{\left[\exp\left(\frac{Q_{\text{soil}} \times L_{\text{crack}}}{D_{\text{crack}} \times A_{\text{crack}}} \right) + \left(\frac{D_T^{\text{eff}} \times A_B}{Q_{\text{building}} \times L_T} \right) + \left(\frac{D_T^{\text{eff}} \times A_B}{Q_{\text{soil}} \times L_T} \right) \times \left[\exp\left(\frac{Q_{\text{soil}} \times L_{\text{crack}}}{D_{\text{crack}} \times A_{\text{crack}}} \right) - 1 \right] \right]} \quad (11)$$

Where α is the dilution factor between soil air and indoor air (-), D_T^{eff} the total overall effective diffusion coefficient (cm^2/s), A_B the area of the enclosed space (cm^2), Q_{building} the ventilation rate (cm^3/s), L_T the source-building separation (cm), Q_{soil} the volume flow rate of soil gas into the enclosed space (cm^3/s), L_{crack} the enclosed space foundation thickness (cm), D_{crack} the effective diffusion coefficient through the cracks (cm^2/s), and A_{crack} the area of total cracks (cm^2) (US EPA, 2004).

2.2 Description of the Site

Few well-documented site, with sufficient temporal and spatial data to contrast predictions and observations, are available in the public domain. From those available a site situated in the city Vilvoorde (Belgium) was selected, as it was used for over 30 years by the coating industry to produce paints and varnishes, resulting in a contamination with volatile aromatic and chlorinated hydrocarbons.

At the source of the contamination the soil as well as the groundwater is contaminated. The migration of the plume shows a near-surface and deep groundwater contamination (Provoost et al., 2014). The conceptual site model (CSM) revealed that the aromatic hydrocarbons benzene (hereafter BE) and ethylbenzene (hereafter EB) were the contaminants of concern, given their concentrations ranges in the soil. Therefore the dominant route for exposure of humans is considered to be migration of soil vapour to the building (indoor air), resulting from sub-surface soil contamination. The soil was on average 1.5 m thick, with sandy-loam near the sub-surface and a loamy soil at groundwater level. The building floor was made out of concrete with an average thickness of 0.3 m (with cracks and gaps). The vadose zone contaminated covers around 4 000 m^2 , has an estimated volume of 50 000 m^3 , with an average benzene concentrations of 408 mg/kg soil and 491 mg/kg soil for ethylbenzene. The volume of contaminated groundwater is around 63 000 m^3 with an average concentration of 1 330 mg/l benzene and 580 g/l ethylbenzene. Samples were taken in the building overlying the vadose zone contamination. For more details on the site, CSM and observed concentrations the reader is referred to Provoost et al. (2002, 2014), and upon request the full study report Bronders et al. (2000) can be made available.

3. Methodology

3.1 Approaches and Visualization

3.1.1 Deterministic Approach

The deterministic approach requires the selection of a single value for each of the model parameters to arrive at a conservative predicted air concentration. Consequently, model parameters were adjusted to specific conditions on the site, such as the soil type, contaminant concentration and its depth below ground level, and volume of the building. This approach is usually applied by risk assessors within a CLM framework, and results in a point value for exposure.

However, model parameters have variation, so the deterministic approach obscures the propagation of either uncertainty or variability to the predicted air concentrations (Ragas et al., 2009). Consequently deterministically predicted concentrations do not consider the range or likelihood of the outcomes, so may underestimate the exposure risk (false negative prediction) (Krupnick et al., 2006). To address this shortcoming a probabilistic approach was applied.

3.1.2 Probabilistic Approach

A probabilistic approach is an assessment that produces a distribution for the predicted air concentration, generally by assigning a probability distribution function to represent variability or uncertainty in input parameters (EPA, 2001). Furthermore, predicting VOC exposure as a result of inhaling of indoor air is subject to variation, which has various sources. The first source is variation in model parameters (1), which in turn are either uncertain or variable. A parameter can be uncertain because insufficient data is obtained to derive a value. An example is the fraction organic carbon in the soil, which is not routinely measured during field investigations, but rather estimated from the soil organic matter. Parameters can also vary because they describe a population with different values, for example the thickness of the concrete floor in the building. Variability on the contrary cannot be reduced by gathering more data (Cullen & Frey, 1999; Finley & Paustenbach, 1994; McKone & Bogen, 1990). Other source are uncertainty related to the model itself (2), scenario uncertainty (in the application of the model) (3) and uncertainty due to simplification relevant for a given decision context (4). The probabilistic approach used in this

study address parameter variation only (source 1), because models are applied as made available with their specific scenarios (source 2 and 3) and simplification (source 4). This approach requires the user to consider uncertainty, variability and interdependencies of parameters and includes various probability distribution functions in the prediction (van der Sluijs et al., 2004), and allows a sensitivity analysis of the dominant parameters, effect of variation therein.

As indicated in chapter 2 three VI models were selected for this study, and were made available in a spreadsheet version, as to allow for the calculation of probabilistic distributions of the predicted air concentration. For the probabilistic approach a Monte Carlo simulation (Hammersley & Hanscomb, 1964; McKone & Ryan, 1989; EPA, 2001) was applied, by utilizing the add-in to Excel® named Crystal Ball® (Crystal Ball, 2000), for which each parameter distribution function is sampled 10 000 times. The sensitivity analysis, as part of the probabilistic approach, provides insight in how variation, because of variability and uncertainty in input parameters, propagate through the model into variation of the predicted air concentration (Hammersley & Hanscomb, 1964; McKone & Ryan, 1989; Provoost et al., 2014). The probability distribution function for model parameters were derived from the literature as well as from the site investigation report.

3.1.3 Visualisation of the Results

The graphical representation of the deterministic and probabilistic approach, including the sensitivity analysis, are visualized by box-and-whisker plot, bar charts and tables. A plot is provided for each combination of contaminant (benzene or ethylbenzene), model (DF SE 1996, 2005 or JEM) and medium (soil air or indoor air).

The plots display the minimum and maximum, 25, 75 and 95-percentile (x) and median (□) predicted soil or indoor air concentration. Each plots displays the predicted deterministic concentration (○) and, in the case of the indoor air concentration, for reason of comparison, the tolerable concentration in air (TCA) (◇). For both contaminants the observed soil and indoor air concentrations are shown for comparison with the predictions. The plots provide an insight in the spread (difference between the whiskers minimum and maximum concentration), and the mid-spread (the middle 50 % of the predicted concentrations), also called inter-quartile range. The location of the median value relative to the 25 and 75 quartiles value indicates the amount of asymmetry in the data, also called skewness (Provoost et al, 2014).

The results from the sensitivity analysis are displayed via tables indicating for each of the contaminants and model the contribution of dominant parameters.

A graph containing the outcome of the three statistical criteria (RMSE, ME or CMR, as defined below) for each combination of contaminant (BE and EB) and model (DF SE 1996, 2005 or JEM) provides a measure for the inter model accuracy for predicted indoor air concentrations.

3.2 Defining Terms

3.2.1 Accuracy

The accuracy of a model is defined as the difference between the predicted and observed (soil or indoor) air concentrations, and the closer the proximity the more accurate the model (Hers et al., 2002; Warmink et al., 2010; Provoost et al, 2010). Within this context the accuracy of a model is defined by applying three statistical criteria, namely the maximum relative error (ME), root mean squared error (RMSE) and coefficient of residual mass (CRM), as proposed by Loague and Green (1991). The lower the score, which is unitless, for a criterion, the more accurate a model predicts concentrations in close proximity to the observed concentrations, and are thus a measure for the precision. The three criteria are defined below, where the O means the observed concentration, P the predicted concentration, and n the number of pairs (observation and prediction).

Maximum relative Error (ME):

$$ME = \frac{\max_{i=1}^n [abs(O_i - P_i)]}{\bar{O}} \quad (12)$$

The ME provides the maximum difference between the O and P concentrations.

Root Mean Squared Error (RMSE):

$$RMSE = 100 \frac{\sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}}}{\bar{O}} \quad (13)$$

The RMSE provides a measure of the average difference between all O and P concentrations.

Coefficient of Residual Mass (CRM):

$$CRM = - \frac{\left[\sum_{i=1}^n O_i - \sum_{i=1}^n P_i \right]}{\sum_{i=1}^n O_i} \tag{14}$$

The CRM provides a measure whether the model over- or under-predicts in comparison to observations. If the CRM has a negative (–) value the predicted concentration by the model under-predicts in relation to the observed concentration and inversely.

3.2.2 Conservatism

The conservatism is defined as to what extent the predicted air concentration reflects variation as a result of variability or uncertainty in model parameters (Eklund & Burrows, 2009; Johnston & MacDonald, 2010; Labieniec et al., 1996), while maintaining a 95-percentile level of protection. Hereto, the 95-percentile predicted concentration of the probabilistic distribution is used as an adequate level of protection (US EPA, 2001; Ferguson, 1999). The conservatism is determined by comparing the deterministically predicted concentration with the 95-percentile value of the probabilistic distribution (Provoost et al. 2014).

3.2.3 Parsimony

A parsimonious model accomplishes a desired level of prediction with as few parameters as possible (Pitt & Myung, 2002), or as “simple as possible, as complex as necessary” (Bilitewski B. et al, 2013). In relation to the probabilistic approach, a parsimonious model obtains the simplest model possible, that explains as much of the variation in indoor air as possible (Coles, 2001).

For that reason, the number of model parameters was used as a measure for the parsimony of the models under investigation. Preference is given to a model which is parsimonious while still accurate and sufficiently conservative. The 1996 DF model and new 2005 VI model, both from Sweden, and the US-EPA 1991 JEM, includes respectively 9, 14 and 27 model parameters.

4. Results

4.1 Probabilistic Analysis

4.1.1 Soil Air

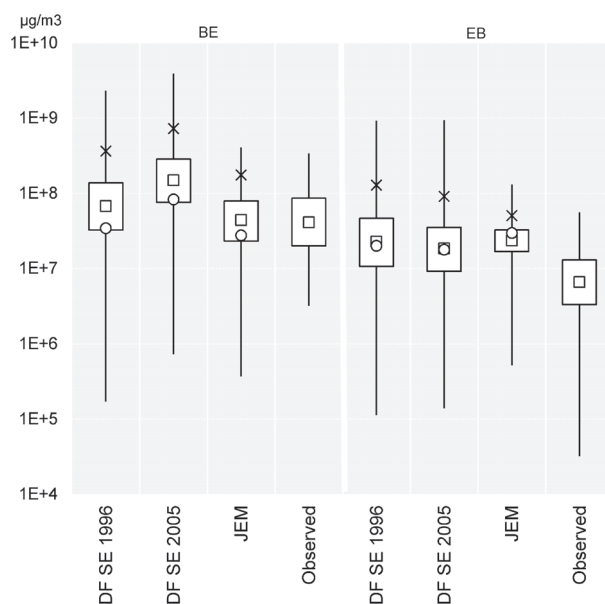


Figure 4. Box-and-whisker plots for predicted and observed soil air concentration by model and contaminant x 95-percentile, o deterministic value, □ median, BE benzene, EB ethylbenzene, 10 000 predictions, 35 observations.

The ranges for the predicted concentrations cover the observed ranges and reveal that the mid-spread of predictions and observations are within 1 order of magnitude (OOM). The median and deterministic predicted soil air concentrations are in the mid-range, therefore close proximity. Deterministic predicted concentrations tend to be slightly lower than the median predicted concentration, with the exception of the JEM for ethylbenzene.

The 95-percentile predicted concentrations are ± 1 OOM higher than the deterministic predicted concentration, which indicates a low conservatism (level of protection). The predicted and observed concentrations are not much skewed.

4.1.2 Indoor Air

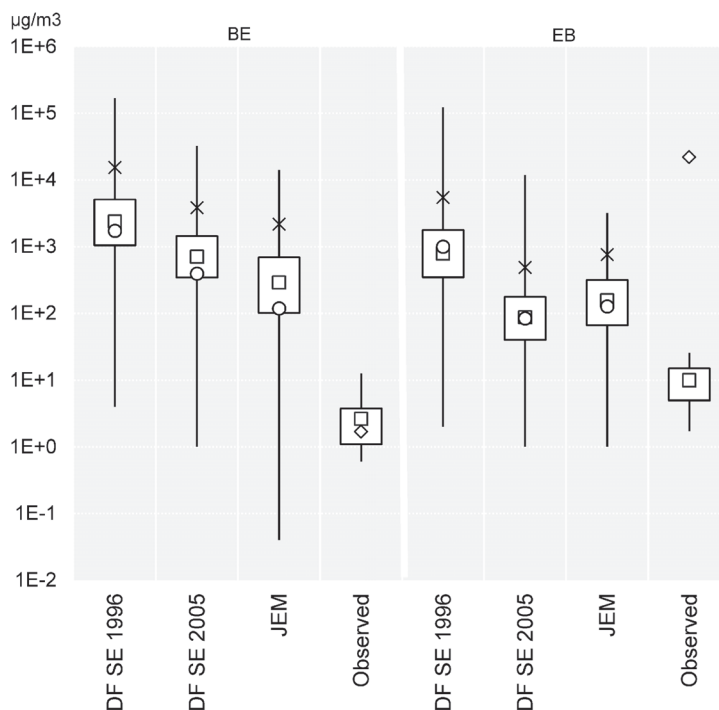


Figure 5. Box-and-whisker plots for predicted and observed indoor air concentration by model and contaminant x 95-percentile, o deterministic value, □ median, ◇ tolerable concentration in air (TCA), BE benzene, EB ethylbenzene, 10 000 predictions, 52 observations.

The ranges for the predicted concentration are higher when compared to the observed ranges, with the exception of the JEM for BE. The ranges and mid-spread of predictions and observations differ more than 1 OOM. The median and deterministic predicted indoor air concentrations are in the mid-spread, therefore close proximity. Deterministic predicted concentrations are in close proximity to the median predicted concentration. The 95-percentile predicted concentration is 1 to 2 OOM higher than the deterministic predicted concentration, which indicates a low conservatism (level of protection). The predicted and observed concentrations are slightly skewed.

4.2 Sensitivity Analysis

4.2.1 Soil air

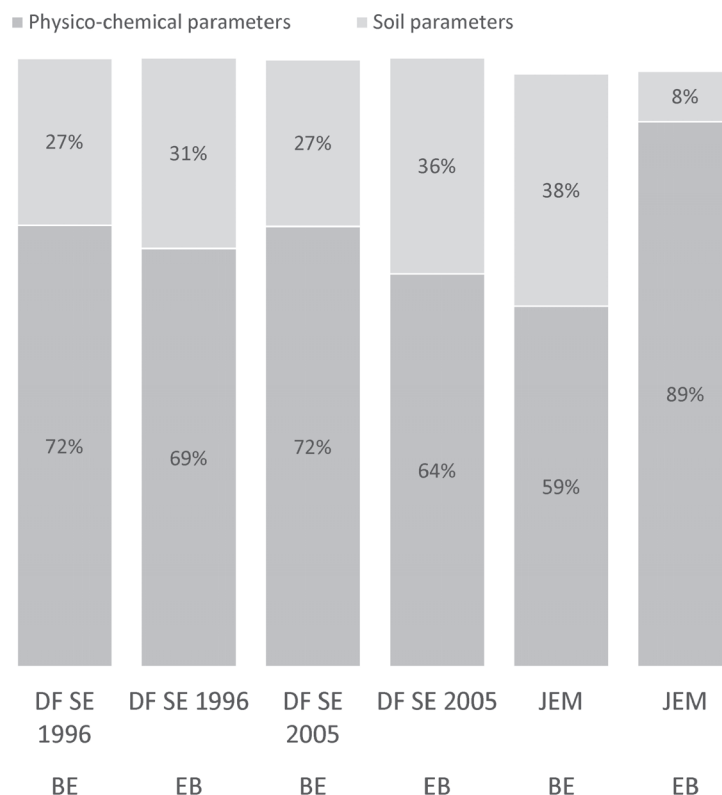


Figure 6. Contribution of groups of model parameters to the variation in predicted soil air concentration by model and contaminant

BE benzene, EB ethylbenzene.

The below table 1 provides more details for each of the model parameter groups.

Table 1. Contribution of model parameters to the variation in the predicted soil air concentration by model and contaminant

Contaminant	BE		EB		EB	
	DF SE 1996	DF SE 1996	DF SE 2005	DF SE 2005	JEM	JEM
Physico-chemical parameters	72.4%	68.6%	72.2%	64.5%	59.1%	89.4%
Henry constant (at ref. T)	1.6%	11.5%	2.2%	13.8%		53.6%
initial soil concentration	42.8%	37.9%	52.5%	47.2%	59.1%	14.8%
octanol-water partition coefficient			17.5%	3.5%		
organic carbon-water coefficient	28.0%	19.2%				
diffusivity in air						0.9%
diffusivity in water						1.0%
pure component water solubility						19.2%
Soil parameters	27.4%	31.4%	27.4%	35.5%	38.1%	8.3%
organic carbon fraction in soil	27.4%	31.4%	27.4%	35.5%	38.1%	
soil temperature						2.7%
soil total porosity						1.8%
soil vapour permeability						1.1%
soil water filled porosity						2.7%

Contaminant	BE	EB	BE	EB	BE	EB
	DF SE		DF SE		JEM	
Model	1996	DF SE 1996	2005	DF SE 2005	JEM	JEM
Total	99.9%	100.0%	99.6%	100.0%	97.3%	97.7%

Figure 6 provides an overview of the different groups of model parameters contributing to the variation. The physico-chemical parameters overall contribute most (59-89%) to the variation of the predicted benzene and ethylbenzene soil air concentration, while the soil parameters contribute less (8-38%).

Table 1 reveals that for the groups of physico-chemical parameters mainly the initial soil concentration drives the predictions, with the exception of the JEM for ethylbenzene, where the Henry constant dominates. For the group of soil parameter, the organic carbon fraction in soil contributes most tot the variation.

4.2.2 Indoor Air

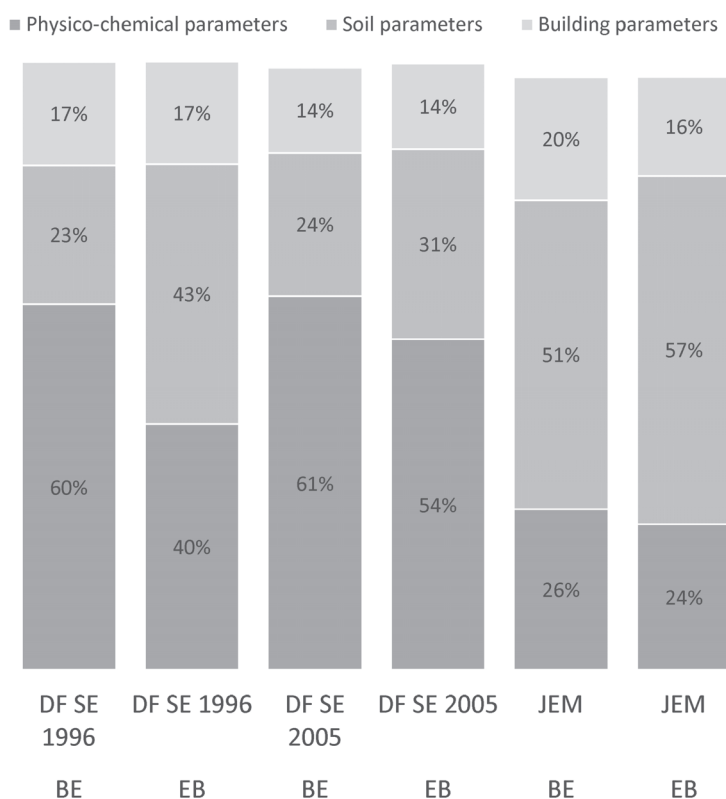


Figure 7. Contribution of groups of model parameters to the variation in predicted indoor air concentration by model and contaminant

BE benzene, EB ethylbenzene.

The below table 2 provides more details for each of the model parameter groups.

Table 2. Contribution of model parameters to the variation in the predicted indoor air concentration by model and contaminant

Contaminant	BE		EB		BE		EB	
	DF	SE	DF	SE	DF	SE	DF	SE
Model	1996		1996		2005		2005	
	JEM		JEM		JEM		JEM	
Physico-chemical parameters	60.1%	40.5%	61.4%	54.3%	26.5%	24.0%		
Henry constant (at ref. T)	1.4%	9.5%	1.8%	11.7%				14.6%
initial soil concentration	36.4%	31.0%	44.6%	39.6%	26.5%	4.2%		
octanol-water partition coefficient	22.3%		15.0%	3.1%				
pure component water solubility								5.1%
Soil parameters	22.8%	42.7%	23.6%	31.3%	50.7%	57.2%		
organic carbon coefficient		16.4%						
organic carbon fraction in soil	22.8%	26.2%	23.6%	31.3%	17.8%			
soil dry bulk density								1.3%
soil total porosity								1.3%
soil vapour permeability					31.1%	54.6%		
soil water filled porosity					1.8%			
Building parameters	17.0%	16.9%	14.0%	14.0%	20.3%	16.3%		
dilution factor from soil to indoor air	17.0%	16.9%						
indoor air exchange rate					1.3%	1.0%		
intrusion rate of pore air			10.7%	11.2%				
soil-building pressure differential					18.9%	13.6%		
ventilation rate for air in the building			3.3%	2.8%				
thickness of soil stratum								1.6%
Total	99.9%	100.0%	99.0%	99.6%	97.4%	97.5%		

Figure 7 provides an overview of the different groups of model parameters contributing to the variation in indoor air. The group of physico-chemical parameters contribute 40-61% to the variation of the predicted indoor air concentration, with the exception of the JEM where the contribution reduces to 24-36%. The contribution of the group of soil parameters vary for the DF SE models (1996, 2005) between 23-43%, while this increases to 51-57% for the JEM. For all three models the building parameters do not contribute much to the variation of the indoor air concentration (14-20%).

Table 2 reveals that for the group of physico-chemical parameters mainly the initial soil concentration drives the predictions, except for EB in the JEM, where the Henry constant dominates. For the group of soil parameter the organic carbon fraction in the soil drives the variation for both the DF SE (1996, 2005) models (22-31%), while for the JEM the soil vapour permeability contribute most to the variation (31-54%). The group of building parameters contribute the least (14-20%).

Table 3. Model parameters that contribute most to the variation in the predicted indoor air concentration by model

Parameter type	Parameter	DF SE 1996	DF SE 2005	JEM
Physico-chemical	Initial soil concentration	■	■	■
	Henry's constant			■
Soil	Organic carbon fraction in soil	■	■	
	Soil vapour permeability			■
Building	Dilution factor from soil to indoor air	■		
	Intrusion rate of pore air		■	
	Soil-building pressure differential			■

Table 3 provides per model an overview of the parameters that contribute most to the variation of the indoor air concentration. The JEM has different parameters driving the variation in predicted indoor air concentration, than the both DF SE models. The equations presented in Chapter 2 show that the three model have implemented similar fugacity algorithms for calculating the soil air concentration (equation 3), resulting in similar parameters driving the predictions, like for example the initial soil concentration or organic carbon fraction in the soil.

However the three models incorporated different mathematical algorithms (equation 4, 9 and 11) for predicting the indoor air concentration from the soil air concentration, so different building parameters dominate the variation (dilution factor from soil to indoor air vs. the intrusion rate of pore air vs. soil-building pressure differential).

4.3 Accuracy

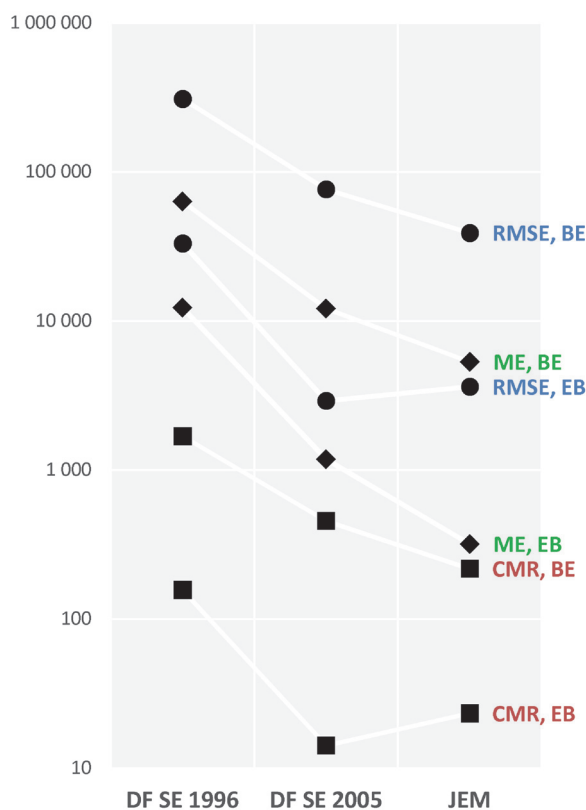


Figure 8. Statistical criteria by contaminant and model for the predicted indoor air concentration

BE: benzene, EB: ethylbenzene, ME: Maximum relative Error, RMSE: Root Mean Squared Error, CRM: Coefficient of Residual Mass.

Figure 8 shows the scores (unitless) of the three statistical criteria for both contaminants and three models and reveals a decreasing value, meaning increasing accuracy, from DF SE 1996 to DF SE 2005, and further for JEM. Thus, the latter can be considered to be the most accurate model. The statistical criteria produce systematically a higher outcome for BE than EB, which suggests that indoor air concentration for EB are more accurately predicted than for BE. The positive CRM indicate that the models have a tendency to over-predict indoor air concentrations in relation to the observed concentrations. The model DF SE 2005 performs better for the RMSE and CMR for EB, when compared to the other two models.

4.4 Parsimony

The number of parameters for the DF SE 1996, 2005 and JEM VI model for which a distribution function was derived (see chapter 3.2) are respectively 9, 14 and 27. The variation for predicting the indoor air concentration results in a four OOM spread, with the mid-spread mostly within one OOM (Figure 5). The predicted indoor air concentration for the DF SE 1996 and 2005 are mainly driven by physico-chemical parameters and less by soil or building parameters. The exception is the JEM where half of the variation is caused by soil parameters (Figure 7).

The most accurate level of prediction of the indoor air concentration, with as few parameters as possible, is obtained by the DF SE 2005 model. The dominate parameters that drive the variation in predicted indoor air concentration (Table 3) are measured routinely during site investigations, like the initial soil concentration and organic carbon fraction in the soil (derived from organic matter). The soil vapour permeability for the JEM model will require additional sampling and analysis. The dominant building parameters for JEM are not routinely measured during standard investigations and will incur additional site investigation (costs).

5. Conclusions and Discussion

A deterministic and probabilistic approach (includes a sensitivity analysis) was applied to provide insight in the accuracy, conservatism and parsimony of three VI models used in Sweden for human health risk assessment within the framework of CLM.

The model that most accurately predicts the soil air concentration is the JEM followed by the DF SE 2005 and 1996 (Figure 4). Predictions are predominantly driven by variation in physico-chemical model parameters (Figure 6). The most accurate model for predicting the indoor air concentration is the JEM, followed by DF SE 2005 and 1996 (Figure 5 and 8). The variation in indoor air is driven by physico-chemical and/or soil parameters, with the exception of JEM, where soil parameters dominate (Figure 7).

The deterministic predicted soil and indoor air concentrations were in the mid-spread of the probabilistic range, often close to the median, but below the 95-percentile concentration (Figure 4 – 7). Thus, default parameter values could be revised as to increase the conservatism, and be closer to the 95-percentile predicted indoor air concentration. This revision should take into account the dominant parameters for each model (Table 1 and 2, Figure 6 and 7).

Application of the three models on a site requires that parameters are adapted to the CSM. Table 3 indicates for each of the model what parameters drive the predictions, so site specific data can be gathered with the dominant parameters in mind. Some parameters cannot easily be measured on site, like the intrusion rate of pore air, so a value needs to be selected from the distribution function representing the CSM. Literature, handbook or manuals can be consulted for the ranges and representative default parameter settings.

The above findings are in agreement with Provoost et al. (2008; 2010a; 2013; 2014), where the JEM model was among the most accurate and driven by the same parameters, while the DF SE 1996 was amongst the most conservative models, being less accurate, however producing less false negative predictions. A false negative is a predicted air concentration that is lower than the observed concentration, resulting in a reduced need to take further action (site investigation or remediation), while this might be triggered based on observed concentrations. The minimum concentration of the whiskers in figure 4 and 5 are below the minimum of the observed concentrations, therefore suggest a review on the probability for false negative predictions.

The sensitivity analysis reveals the contribution of each of the parameter to the predicted concentrations, resulting in a better prioritisation of additional data gathering on a site or in literature. Compared to the deterministic point estimated, the probabilistic approach revealed instead which parameters dominate the predicted indoor air concentration, hence the human health risk estimate (Johnson & MacDonald, 2010). Nevertheless, the use of results from a probabilistic approach is challenging for regulatory decision makers, as there is no rule-of-thumb to decide on the amount of acceptable variation as a result of variability or uncertainty of input parameters (van der Sluijs et al., 2004; EPA, 1997; Morgan & Henrion, 1990; Saltelli et al., 2004; Burmaster & Anderson, 1994). A guide for making decisions under variation (uncertainty) and how to communicate this are proposed by van der

Sluijs et al. (2004) and EPA (2001).

The DF SE 1996 model is a screening level model, which shows a higher level of conservatism (lower accuracy). A limited number of model parameters are included, and this analysis shows that a model with more parameters is more accurate, and less conservative. A model can be selected depending on the users desired level of conservatism and/or accuracy. A more conservative, less accurate, model like the DF SE 1996, could be selected for deriving general soils screening values, while a more accurate model can be applied for site specific risk assessment (Provoost et al. 2013). The DF SE 2005 seems to be a parsimonious model as it includes sufficient parameters to be accurate, but is sufficiently conservative with 14 parameters, while the DF SE 1996 with 9 parameters is conservative and the JEM with 27 parameters accurate. For the latter some of the dominant parameters cannot easily be measured on site.

The sensitivity analysis revealed dominant parameters for each of the models, where some have variability and some uncertainty, Fisher et al. (2002) suggests to use two models as to account for variability and explains differences in predictions.

6. Further Research Needs

The probabilistic approach applied in this study was limited to variation in model parameters, which have either variability or uncertainty. Both are propagated into a single output distribution, unless a 2D simulation is performed in which both sources of variation are separated and their contribution calculated (Ragas et al., 2009). A 2D analysis shows how much of the variation is caused by parameters that have either variability or uncertainty.

A similar study is required for other well documented sites, as to provide a broader insight in the accuracy, conservatism and parsimony of the three VI models. Additional research can confirm if the same model parameters drive the predicted soil and indoor air concentrations.

Other contaminant, like chlorinated hydrocarbons can add to this broader view. Consequently extrapolating the results from this study to the application on other sites require caution.

References

- Bilitewski, B., Darbra, R. M. & Barceló, D. (2013). *Global risk-based management of chemical additives II: Risk-based assessment and management strategies*, page 101, Springer-Verlag Berlin, Heidelberg, ISBN 978-3-642-34572-2. <https://doi.org/10.1007/978-3-642-34572-2>
- Bronckers, J., Cornelis, C., Olivier, I., Patyn, J., Provoost, J., Wilczek D. & Smolders, R. (2000). *Descriptive soil research site Astral near Brussels, phase two, 2000/IMS/R/161*
- Burmester, D. E. & Anderson, P. D. (1994). Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Analysis*, 14(4), <https://doi.org/10.1111/j.1539-6924.1994.tb00265.x>
- Carlou, C., D'Alessandro, M., & Swartjes, F. A. (2007). *Derivation methods of soil screening values in Europe. A review of national procedures towards harmonisation*. pages 320. European Commission Joint Research Centre, Ispra Italy, EUR 22805 EN-2007. Retrieved from <https://publications.europa.eu/en/publication-detail/-/publication/206489ef-386d-4bcb-8b39-25b41d4d3c45/language-en>
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*, page 109, Springer-Verlag London, ISBN 978-1-85233-459-8. <https://doi.org/10.1007/978-1-4471-3675-0>
- Crystal Ball (2000). *User's manual 2000*, Decisioneering Inc.
- Cullen, A. C. & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment. A handbook for dealing with variability and uncertainty in models and inputs*. Pages 335, Plenum Press, New York, ISBN 978-0-306-45956-6. Retrieved from <https://www.springer.com/gp/book/9780306459566#aboutBook>
- Eklund, B. & Burrows, D. (2009). Prediction of indoor air quality from soil-gas data at industrial buildings. *Ground Water Monitoring and Remediation*, 29(1), 118-125, <https://doi.org/10.1111/j.1745-6592.2008.01220.x>
- EPA (1997). *Guiding principles for Monte Carlo analysis*. EPA - Risk Assessment Forum, EPA/630/R- 97/001. Retrieved from <http://www.epa.gov/raf/publications/pdfs/montecar.pdf>
- EPA (2001). *Risk assessment guidance for superfund: volume III – Part A, Process for conducting Probabilistic Risk Assessment*, Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, Report number: EPA 540-R-02-002. Retrieved from <https://www.epa.gov/risk/risk-assessment-guidance-superfund-rags-volume-iii-part>

- Ferguson, C. C. (1999). Assessing risks from contaminated sites: policy and practice in 16 European countries. *Land Contamination and Reclamation*, 7(2), 33-54, *Land Contamination and Reclamation*, 7(2), 33-54. https://www.commonforum.eu/Documents/DOC/Caracas/Ferguson_Paper_Policies.pdf
- Finley, B., & Paustenbach, D. (1994). The benefits of probabilistic exposure assessment: three case studies involving contaminated air, water and soil. *Risk Analysis*, 14(1), 53-73, <https://doi.org/10.1111/j.1539-6924.1994.tb00028.x>
- Fisher, B. E. A., Ireland, M. P., Boyland, D. T., & Critten, S. P. (2002). Why use one model? An approach for encompassing model uncertainty and improving best practice. *Environmental Modeling and Assessment*, 7, 291-299, <https://doi.org/10.1080/08920750701682007>
- Fitzgerald, J. (2009). One regulatory perspective on the vapor intrusion pathway. *Ground Water Monitoring and Remediation*, 29(1), 51-52. <https://doi.org/10.1111/j.1745-6592.2008.01213.x>
- Fugler, D., & Adomait, M. (1997). Indoor infiltration of volatile organic contaminants: measured soil gas entry rates and other research results from Canadian houses. *Journal of Soil Contamination*, 6(1), 9-13, <https://doi.org/10.1080/15320389709383542>
- Hammersley, J. S., & Handscomb, D.C. (1964). *Monte Carlo Methods*. John Wiley and Sons Inc., New York. <https://doi.org/10.1007/978-94-009-5819-7>
- Hers, I. (2004). *Field-based evaluation of processes and models for soil vapour intrusion into buildings*. PhD thesis, University of British Columbia, Department of Civil Engineering
- Hers, I., Zapf-Gilje, R., Evans, D., & Li, L. (2002). Comparison, validation, and use of models for predicting indoor air quality from soil and groundwater contamination. *Soil and Sediment Contamination*, 11(4), 491-527. <https://doi.org/10.1080/20025891107140>
- Johnson, P. C., & Ettinger, R. E. (1991). Heuristic model for predicting the intrusion rate of contaminant vapors into buildings. *Environmental Science and Technology*, 25(8), 1445-1452. <https://doi.org/10.1021/es00020a013>
- Johnson, P. C., & Ettinger, R. E. (1997). *Users guide for the Johnson and Ettinger (1991) model for subsurface vapor intrusion into buildings*. US EPA, Office of Emergency and Remedial Response, Link: <https://www.epa.gov/risk/users-guide-evaluating-subsurface-vapor-intrusion-buildings>
- Johnston, J. E., & MacDonald, G. J. (2010). Probabilistic approach to estimating indoor air concentrations of chlorinated volatile organic compounds from contaminated groundwater: a case study in San Antonio, Texas. *Environmental Science and Technology*, 45, 1007-1013. <https://doi.org/10.1021/es102099h>
- Johnson, P. C., Kemblowski, M. K., & Johnson, R. L. (1999). Assessing the significance of subsurface contaminant vapor migration to enclosed spaces, Site-specific alternative to generic estimates. *Journal of Soil Contamination*, 8(3), 389-421
- Loague, K. M., & Green, R. E. (1991). Statistical and graphical methods for evaluating solute transport models: overview and application. *Journal of Contaminant Hydrology*, 7, 51-73, [https://doi.org/10.1016/0169-7722\(91\)90038-3](https://doi.org/10.1016/0169-7722(91)90038-3)
- Kaplan, M. B., Brandt-Rauf, P., Axley, J. W., Shen, T. T., & Sewell, G. H. (1993). Residential releases of number 2 fuel oil: A contributor to indoor air pollution. *American Journal of Public Health*, 8(1), 84-88, <https://doi.org/10.2105/ajph.83.1.84>
- Krupnick, A., Morgenstern, R., Batz, M., Nelson, P., Burtraw, D., Shih, J. S., & McWilliams, M. (2006). *Not a sure thing: making regulatory choices under uncertainty*. U.S. Environmental Protection Agency. Retrieved from <http://rff.org/rff/Documents/RFF-Rpt-RegulatoryChoices.pdf>
- McAlary, T. A., & Johnson, P. C. (2009). GWMR focus issue on vapor intrusion. *Ground Water Monitoring and Remediation*, 29(1), 40-41. <https://doi.org/10.1111/j.1745-6592.2008.01212.x>
- McAlary, T., Provoost, J., Dawson, H., & Swartjes, F. (editor) (2011). *Book on Dealing with Contaminated Sites - From Theory towards Practical Application, chapter 10 – Vapour intrusion*. January 2011, 409-453, Springer Publishers, Netherlands, ISBN 978-90-481-9756-9, https://doi.org/10.1007/978-90-481-9757-6_10
- McKone, T. E., & Bogen, K. T. (1990). Uncertainty in health risk assessment – an integrated case study based on tetrachloroethylene in California groundwater. *Regulatory Toxicology and Pharmacology*, 15(1), 86-103, [https://doi.org/10.1016/0273-2300\(92\)90087-P](https://doi.org/10.1016/0273-2300(92)90087-P)

- McKone, T. E., & Ryan, P. B. (1989). Human exposures to chemicals through food chains: an uncertainty analysis. *Environmental Science and Technology*, 23, 1154-1163.
- MDEP. (1994). *Background documentation for the development of MCP numerical standards*. Massachusetts Department of Environmental Protection, USA, Link: <https://archive.org/details/backgrounddocume00mass>
- Morgan, M. G., & Henrion, M. (1990). *Uncertainty, A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press. <https://doi.org/10.1080/00137919308903109>
- Öberg, T. (2006). *Probabilistisk riskbedömning, Fas 1. Rapport till Hållbar Sanering*, report 5532, in Swedish, Swedish Environmental Protection Agency, Stockholm, Sweden. Retrieved from <https://www.naturvardsverket.se/Documents/publikationer/620-5532-1.pdf>
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6(10), 421-425. [https://doi.org/10.1016/S1364-6613\(02\)01964-2](https://doi.org/10.1016/S1364-6613(02)01964-2)
- Provoost, J., Cornelius, C., & Seuntjens, P. (2002). *Estimation of human indoor air exposure from soil contamination with benzene and ethylbenzene using probabilistic risk assessment*, Book Risk Analysis III, In Brebbia CA (Ed.), Wit Press, Southampton, United Kingdom. Retrieved from <http://www.witpress.com/978-1-85312-915-5.html>
- Provoost, J., Reijnders, L., Swartjes, F., Bronders, J., Seuntjens, P., & Lijzen, J. (2008). Accuracy of seven vapour intrusion algorithms for VOC in groundwater. *Journal of Soils and Sediments – Protection, Risk Assessment, and Remediation*, 9(1), 62-73. <https://doi.org/10.1007/s11368-008-0036-y>
- Provoost, J., Bosman, A., Reijnders, L., Bronders, J., Touchant, K., & Swartjes, F. (2010). Vapour intrusion from the vadose zone – seven algorithms compared. *Journal of Soils and Sediment*, 10(3), 473-483. <https://doi.org/10.1007/s11368-009-0127-4>
- Provoost, J., Reijnders, L., & Bronders, J. (2013). Accuracy and conservatism of vapour intrusion algorithms for contaminated land management. *Journal of Environment and Pollution*, 2(2), ISSN 1927-0917. <https://doi.org/10.5539/ep.v2n2p71>
- Provoost, J., Reijnders, L., Bronders, J., Van Keer, I., & Govaerts, S. (2014). Probabilistic risk assessment for six vapour intrusion algorithms. *Journal of Environment and Pollution*, 3(2), ISSN 1927-0917, <https://doi.org/10.5539/ep.v3n2p1>
- Ragas, A. M. J., Brouwer, F. P. E., Buchner, F. L., Hendriks, H. W. M., & Huijbregts, M. A. J. (2009). Separation of uncertainty and interindividual variability in human exposure modelling. *Journal of exposure science and environmental epidemiology*, 19, 201-212. <https://doi.org/10.1038/jes.2008.13>
- Sanders, P., Bergback, B., & Öberg, T. (2006). Uncertain numbers and uncertainty in the selection of input distributions – consequences for a probabilistic risk assessment of contaminated land. *Risk Analysis*, 26(5), 1363-1375, <https://doi.org/10.1111/j.1539-6924.2006.00808.x>
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons publishers. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/0470870958.fmatter/pdf>
- S-EPA. (1996). *Development of generic guideline values, Model and data used for generic guideline values for contaminated soils, in Sweden*. Swedish Environmental Protection Agency, Stockholm, Report 4639, ISBN 91-620-4639-x. Retrieved from <https://www.naturvardsverket.se/Om-Naturvardsverket/Publikationer/ISBN/4000/91-620-4639-x/>
- S-EPA. (1998). *Proposed guideline levels for contaminated petrol stations*. In Swedish, Swedish Environmental Protection Agency, Stockholm, Report 4889, ISBN 91-620-4889-9. Retrieved from <https://naturvardsverket.se/Om-Naturvardsverket/Publikationer/ISBN/4000/91-620-4889-9/>
- S-EPA. (2005a). *Model to calculate guideline values for soil*. Draft version 050704 for review, in Swedish. Swedish Environmental Protection Agency, Stockholm.
- S-EPA. (2005b). *Guidance to risk assessment of contaminated areas*. Draft version for review, in Swedish, Swedish Environmental Protection Agency, Stockholm.
- S-EPA. (2009). *Report describing the model for calculating generic guideline values for contaminated soil*. In Swedish, Swedish Environmental Protection Agency, Stockholm, ISBN: 978-91-620-5976-7. Retrieved from <https://www.naturvardsverket.se/Om-Naturvardsverket/Publikationer/ISBN/5900/978-91-620-5976-7/>

- US EPA. (1996). *Soil Screening Guidance - Technical Background document*. EPA/540/R-95/128, PB96-963502, US EPA, Washington. Retrieved from <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=100025LM.PDF>
- US EPA. (2001). *Risk Assessment Guidance, vol. III, Part A: Process for conducting probabilistic risk assessment (RAGS)*. Washington DC, Office of Emergency and Remedial Response. Retrieved from http://www.epa.gov/oswer/riskassessment/rags3adt/pdf/rags3adt_complete.pdf
- US EPA. (2004). *User's guide for evaluating subsurface vapor intrusion into buildings*. United States Environmental Protection Agency, Office of Emergency and Remedial Response, Washington. Retrieved from https://www.epa.gov/sites/production/files/2015-11/documents/2004_0222_3phase_users_guide.pdf
- van der Sluijs, J. P., Janssen, P. H. M., Petersen, A. C., Klopogge, P., Risbey, J. S., Tuinstra, W., & Ravetz, J. R. (2004). *RIVM/MNP Guidance for uncertainty assessment and communication: Tool catalogue for uncertainty assessment*. Copernicus Institute and RIVM, Utrecht and Bilthoven, report nr: NWS-E-2004-37. Retrieved from <http://www.nusap.net/downloads/toolcatalogue.pdf>
- van Liedekerke, M., Prokop, G., Rabl-Berger, S., Kibblewhite, M., & Louwagie, G. (2014). *Progress in the management of contaminated sites in Europe*. European Commission - Joint Research Centre, Institute for Environment and Sustainability, ENISBN 978-92-79-34846-4. <https://doi.org/10.2788/4658>
- Warmink, J. J., Jansen, J. A. E. B., Boij, M. J., & Krol, M. S. (2010). Identification and classification of uncertainties in the application of environmental models. *Environmental Modelling and Software*, 25, 1518-1527. <https://doi.org/10.1016/j.envsoft.2010.04.011>

Appendix A

Probabilistic distributions and deterministic values, site Astral, BE and EB

The tables provide an overview of the distributions used for the probabilistic (first) and deterministic (below) predictions. Each table include references to the sources of the values.

Table A1. Chemical related parameters benzene

Parameter	Abbr.	DF SE 1996	DF SE 2005	JEM
Henry coefficient	H	T (0.167;0.227;0.33) (-) 0.227 (-)	T (0.167;0.227;0.33) (-) 0.227 (-)	
Organic carbon – water partitioning coefficient	K _{oc}	LN (71.8;49.3) (l kg ⁻¹) 71.8 (l kg ⁻¹)		LN (71.8;49.3) (cm ³ g ⁻¹) 71.8 (cm ³ g ⁻¹)
Octanol-water partitioning coefficient	K _{ow}		N (149.27;64.23) (-) 149.27 (-)	
Diffusion in air	D _a		T (0.68;0.76;0.84) (m ² d ⁻¹) 0.7603 (m ² d ⁻¹)	T (7.87E-2;8.80E-2;9.72E-2) (cm ² s ⁻¹) 8.80E-2 (cm ² s ⁻¹) U (9.80E-06;1.11E-05) (cm ² s ⁻¹)
Diffusion in water	D _w			1.05E-05 (cm ² s ⁻¹)
Reference temperature	T _r			25 (°C)
Boiling point	T _b			353.24 (K)
Critical temperature	T _c			562.16 (K)
Henry coefficient at reference temperature T _r	H _r			LN (5.53E-03;5.16E-04) (atm m ³ mol ⁻¹) 5.53E-03 (atm m ³ mol ⁻¹)
Vaporisation enthalpy at boiling point	ΔH _{v,b}			7342 (cal mol ⁻¹)
Solubility	S			N (1767.6;407.1) (cm ³ g ⁻¹) 1770 (cm ³ g ⁻¹)
Soil concentration	C _s	T (0.14; 408.267; 3192.702) (mg kg ⁻¹) 408.267 (mg kg ⁻¹)	T (0.14; 408.267; 3192.702) (mg kg ⁻¹) 408.267 (mg kg ⁻¹)	T (140;408267;3192702) (μg kg ⁻¹) 408267 (μg kg ⁻¹)

N = normal, LN = lognormal, T = triangular, U = uniform

Parameter	Abbr.	DF SE 1996	DF SE 2005	JEM
References: US EPA (2004), Ettinger (2000), Hers et al. (2002, 2003), Johnson et al. (2002), Neelamegam et al. (2005), Otte et al. (2001), OVAM (2004), Provoost et al. (2002), Tillman & Weaver (2006), Waitz et al. (1996).				

Table A2. Chemical related parameters ethylbenzene

Parameter	Abbr.	DF SE 1996	DF SE 2005	JEM
Henry coefficient	H	T (0.0285;0.351;0.466) (-) 0.351 (-)	T (0.0285;0.351;0.466) (-) 0.351 (-)	
Organic carbon – water partitioning coefficient	K _{oc}	LN (242.7;127.6) (l kg ⁻¹) 243 (l kg ⁻¹)		LN (242.7;127.6) (cm ³ g ⁻¹) 243 (cm ³ g ⁻¹)
Octanol-water partitioning coefficient	K _{ow}		N (1416.34;268.34) 1416.34 (-)	
Diffusion in air	D _a		U (0.583;0.648) (m ² d ⁻¹) 0.6156 (m ² d ⁻¹)	U (6.75E-02;7.50E-02) (cm ² s ⁻¹) 7.13E-2 (cm ² s ⁻¹)
Diffusion in water	D _w			U (7.80E-06;9.00E-06) (cm ² s ⁻¹) 8.40E-6 (cm ² s ⁻¹)
Reference temperature	T _r			25 (°C)
Boiling point	T _b			409,34 (K)
Critical temperature	T _c			617,2 (K)
Henry coefficient at reference temperature Tr	H _r			LN (6.87E-03;2.37E-03) (atm m ³ mol ⁻¹) 6.87E-03 (atm m ³ mol ⁻¹)
Vaporisation enthalpy at boiling point	ΔH _{v,b}			8501 (cal mol ⁻¹)
Solubility	S			LN (177;52.3) (mg l ⁻¹) 177 (mg l ⁻¹)
Soil concentration	C _s	T (0.620;491.829;3233.256) (mg kg ⁻¹) 491.829 (mg kg ⁻¹)	T (0.620;491.829;3233.256) (mg kg ⁻¹) 491.829 (mg kg ⁻¹)	T (620;491829;3233256) (μg kg ⁻¹) 491829 (μg kg ⁻¹)

N = normal, LN = lognormal, T = triangular, U = uniform

References: US EPA (2004), Ettinger (2000), Hers et al. (2002), Hers et al. (2003), Johnson et al. (2002), Neelamegam, Dattatreya & Gopalakrishnan (2005), Otte, Lijzen, Swartjes & Versluijs (2001), OVAM (2004), Provoost, Cornelis & Seuntjens (2002), Tillman & Weaver (2006), Waitz, Feijer, Kreule & Swartjes (1996).

Table A3. Soil properties

Parameter	Abbr.	DF SE 1996	DF SE 2005	JEM
Total porosity	n			U (0.40;0.50) (-) 0.45 (-)
Water filled porosity	Θ _w	U (0.13;0.39) (dm ³ dm ⁻³) 0.26 (dm ³ dm ⁻³)	U (0.13;0.39) (dm ³ dm ⁻³) 0.26 (dm ³ dm ⁻³)	U (0.13;0.39) (cm ³ cm ⁻³) 0.26 (cm ³ cm ⁻³)
Air filled porosity	Θ _a	U (0.10;0.35) (dm ³ dm ⁻³) 0.225 (dm ³ dm ⁻³)	U (0.10;0.35) (dm ³ dm ⁻³) 0.225 (dm ³ dm ⁻³)	
Soil temperature	T _s			T (12.2;15.7;18.8) (°C) 15.7 (°C)
Dry soil bulk density	ρ _b	T (1.25;1.5;1.75) (kg dm ⁻³) 1.5 (kg dm ⁻³)	T (1.25;1.5;1.75) (kg dm ⁻³) 1.5 (kg dm ⁻³)	T (1.25;1.50;1.75) (g cm ⁻³) 1.5 (g cm ⁻³)
Air permeability soil	k _v			U (1.0E-12;1.0E-8) (cm ²) 5.0E-9 (cm ²)

Parameter	Abbr.	DF SE 1996	DF SE 2005	JEM
Organic carbon content soil	f_{oc}	T (0.0009;0.0346;0.1574) (-)	T (0.0009;0.0346;0.1574) (-)	T (0.0009;0.0346;0.1574) (-)
Depth of contaminant	L_T		T (0.8;1.15;1.53) (m)	T (80;115;153) (cm)
Thickness soil layer	h_A		1.15 (m)	115 (cm)
Time	Days		T (1460;1825;2190) (d)	T (80;115;153) (cm)
			1825 (d)	115 (cm)

N = normal, LN = lognormal, T = triangular, U = uniform

References: US EPA (2004), Ettinger (2000), Hers et. al. (2002), Hers et. al. (2003), Johnson et.al. (2002), Neelamegam, Dattatreya & Gopalakrishnan (2005), Otte,Lijzen, Swartjes & Versluijs (2001), OVAM (2004), Provoost, Cornelis & Seuntjens (2002), Tillman & Weaver (2006), Waitz, Feijer, Kreule & Swartjes (1996).

Table A4. Building characteristics (foundation and indoor air)

Parameter	Abbr.	DF SE 1996	DF SE 2005	JEM
Intrusion rate pore air			T (1.44;7.2;14.4) (m ³ d ⁻¹)	
			7.2 (m ³ d ⁻¹)	
Length space floor	L_B			22000 (cm)
Width space floor	W_B			12500 (cm)
Height space	H_B			900 (cm)
Surface floor			27500 (m ²)	
Volume indoor air			247500 (m ³)	
Ventilation rate building	ER		U (4.08;7.92) (d ⁻¹)	U (0.17;0.33) (h ⁻¹)
			6 (d ⁻¹)	0.25 (h ⁻¹)
Soil–building pressure difference	ΔP			T (0;40;200) (g cm ⁻¹ s ⁻²)
				40 (g cm ⁻¹ s ⁻²)
Enclosed space floor thickness	L_{crack}			T (25;35;50) (cm)
				35 (cm)
Depth foundation till surface	L_F			T (25;35;50) (cm)
				35 (cm)
Floor-wall seam crack width	w			T (0.05;0.1;1.0) (cm)
				0.1 (cm)
Dilution factor for indoor air to soil air	DFia	T (5000;20000;70000) (-)		
		20000 (-)		

N = normal, LN = lognormal, T = triangular, U = uniform

References: US EPA (2004), Ettinger (2000), Hers et. al. (2002), Hers et. al. (2003), Johnson et.al. (2002), Neelamegam, Dattatreya & Gopalakrishnan (2005), Otte,Lijzen, Swartjes & Versluijs (2001), OVAM (2004), Provoost, Cornelis & Seuntjens (2002), Tillman & Weaver (2006), Waitz, Feijer, Kreule & Swartjes (1996).

Declaration of conflicting interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received no financial support for their research and/or authorship of this article.

Author's consent and contribution

Each authors has made substantial contributions to the design of the study, analysis and interpretation of the data, and participated in drafting the manuscript, or revising it critically for intellectual content. All authors gave approval of the version to be submitted.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).