

The development of an empirical formula to calculate concentration and arrival time of a chemical contaminant at groundwater production wells in sandy aquifers

Master thesis
Tanneke Blok
3819566

Earth Surface and Water
Utrecht University

First supervisor:
Prof Jack Schijven
Utrecht University and
National Institute for Public Health and Environment (RIVM)

Second supervisor:
Dr Frank Swartjes
National Institute for Public Health and Environment (RIVM)

Date:
20th of April, 2016



Rijksinstituut voor Volksgezondheid
en Milieu
*Ministerie van Volksgezondheid,
Welzijn en Sport*



Universiteit Utrecht

Abstract

In groundwater protection zones two acts apply. Firstly the Dutch Soil Protection Act which regulates all soils and groundwater on the basis of Intervention Values. Secondly there is the Dutch Drinking Water Act which states that all drinking water must comply with the Drinking Water Quality standard. If a contaminant plume is present in the groundwater protection zone the plume itself must have a lower concentration than the Intervention Value. However, once the plume starts migrating, it must reach the well at a concentration lower than the Drinking Water Quality Standard. It is possible that the contaminant plume is under the Intervention Value but reaches the well at concentrations above the Drinking Water Quality Standard. Therefore the plume must still be mitigated or managed. The question remains, what is the allowable concentration of the contaminant at the plume?

Mostly contaminant concentrations are calculated with analytical solutions to the transport equation or with numerical models. However this study has set up four different empirical equations to calculate the concentration and arrival time of a plume at a drinking water production well. These equations are relatively easy in use compared to the traditional methods and are able to predict the concentration with a factor 6 error and the arrival time with an error of 41 days.

However these empirical equations are only applicable within the prescribed parameter space and therefore still have some limitations.

Acknowledgement

I would like to thank my supervisors Jack Schijven, Frank Swartjes and Piet Otte for their help and guidance during this study. Especially Jack Schijven who made time for weekly meetings to answer my questions and helped me keep track of my goals, deserves special mention. Frank Swartjes and Piet Otte were invaluable in understanding the legislation and the practical application of my empirical formulas.

Furthermore I would like to show my appreciation for Majid Hassanizadeh for helping me with the initial set-up and for making the model dimensionless and Amir Raouf for answering all PMWIN-related questions.

Monique van der Aa and Maarten Schipper added useful insights in their area of expertise. Monique van der Aa provided me with papers and reports on drinking water companies in the Netherlands while Maarten Schipper came with the amazing idea of using the Box-Behnken design to help reduce the number of model runs. Their input is much appreciated.

Without the help of any of these persons my study would not have been finished in time.

Table of Contents

1	Introduction.....	1
1.1	Background and problem specification.....	1
1.2	Objective	3
1.3	Approach	4
2	Method.....	6
2.1	PMWIN domain	6
2.2	Governing equations	7
2.3	Parameter space.....	10
2.4	Scenarios	15
2.5	Multivariate regression analysis.....	19
3	Results	21
3.1	PMWIN	21
3.2	Empirical formula	25
4	Discussion.....	40
4.1	Discussion results	40
4.2	Comparison	41
4.3	Restrictions.....	41
5	Conclusion and recommendations.....	42
	References.....	43
	Papers, books and reports.....	43
	Websites	44
	Programs	45
	Appendix.....	i
	A – List of phreatic aquifers in the Netherlands.....	ii
	B – Reproducing the MODFLOW and MT3DMS model.....	iii
	Reproducing MODFLOW	iv
	Reproducing MT3DMS	viii
	C – R-script.....	xiii

List of figures

Figure 1 Flowchart of the tiered approach.....	3
Figure 2 Model domain	6
Figure 3 Frequency distribution of the depth of the top of the well (m blow surface)	11
Figure 4 Frequency distribution of the size of the 25-year zone	12
Figure 5 Frequency distribution of well screen length.....	13
Figure 6 Frequency distribution of pumping rates in 2010.....	13
Figure 7 Frequency distribution of the K_D value of dissolved species.....	14
Figure 8 The breakthrough curves of scenarios 1 to 3 compared to the analytical solution.....	21
Figure 9 Breakthrough curves of test scenarios 4 to 9.....	22
Figure 10 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{max}})$	26
Figure 11 Sensitivity analysis for $\text{Log}(C^*_{\text{max}})$	27
Figure 12 Influence of X^* , Z_W^* , i , $\text{Log}(Q^*)$, $\text{Log}(m)$, R and Da on $\text{Log}(C^*_{\text{max}})$	28
Figure 13 Comparison of results and Bland-Altman plot for vt^*_{max}	29
Figure 14 Sensitivity analysis for vt^*_{max}	30
Figure 15 Influence of X^* , Z_W^* , i , $\text{Log}(Q^*)$, m , R and Da on vt^*_{max}	31
Figure 16 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{half}})$	32
Figure 17 Sensitivity analysis for $\text{Log}(C^*_{\text{half}})$	33
Figure 18 Influence of X^* , Z_W^* , i , $\text{Log}(Q^*)$, $\text{Log}(m)$, R and Da on $\text{Log}(C^*_{\text{half}})$	34
Figure 19 Comparison of results and Bland-Altman plot for $\text{vt}^*_{\text{half}}$	35
Figure 20 Sensitivity analysis for $\text{vt}^*_{\text{half}}$	36
Figure 21 Influence of X^* , Z_W^* , i , $\text{Log}(Q^*)$, m , R and Da on $\text{vt}^*_{\text{half}}$	37
Figure 22 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{max}})$	38
Figure 23 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{max}})$ with $R>1$	38

List of tables

Table 1 Conversion of stream function to dimensionless stream function	8
Table 2 Conversion of transport equation to dimensionless transport equation.....	9
Table 3 Conversion of domain to dimensionless values	10
Table 4 Values of the constants and ranges of the variables.....	14
Table 5 Dimensionless parameter values for the test scenarios	15
Table 6 Dimensionless parameter values for the Box-Behnken scenarios	16
Table 7 Dimensionless parameter values for the Central Composite Design scenarios.....	18
Table 8 Concentrations and arrival times from the test scenarios	21
Table 9 Concentrations and arrival times from the Box-Behnken scenarios	22
Table 10 Concentrations and arrival times from the Central Composite Design scenarios	24

1 Introduction

1.1 Background and problem specification

About two thirds of drinking water production in the Netherlands has groundwater as its source. Roughly 25% of these groundwater production wells are located near a source of a historical (older than 1987) chemical contamination and it is uncertain whether this contamination will enter drinking water wells (Swartjes et al., 2014). Thirty-one groundwater production wells already have contaminant concentrations above the Drinking Water Quality Standard and forty-two wells have concentrations between 75% and 100% of the Drinking Water Quality Standard (Wuijts et al., 2014). Drinking water companies need to know when a contaminant will enter their well, and at what concentration.

The Dutch Soil Protection Act (Wet bodembescherming), (Ministry of VROM and Ministry of Landbouw en Visserij, 1986) provides regulations for all soils and groundwater. According to the Dutch Soil Protection Act, soil and groundwater contamination must be mitigated or managed on the basis of Intervention Values, when the contamination threatens human health or the ecosystem. These Intervention Values indicate the concentration at which the functional properties of a soil are (potentially) gravely reduced and remediation must take place (Ministry of VROM, 2013). According to the Dutch Drinking Water Act (Drinkwaterwet), (Ministry of VROM, 2009), drinking water must meet the European Drinking Water Quality Standards (Council of the European Union, 1998). The Drinking Water Quality Standards are defined as the maximum allowable concentration of a contaminant in drinking water. Importantly, these Drinking Water Quality Standards are much stricter than the Intervention Values. While the area around a groundwater well may contain a contaminant as long as its concentration is under Intervention Values, the concentration in the well must be under the Drinking Water Quality Standard.

An organic contaminant that migrates towards a well may be diluted and broken down which reduces its concentration. If the concentration is reduced to below the Drinking Water Quality Standard, then the contaminant does not pose a threat to drinking water quality. However, if the contaminant concentration is not reduced to concentrations below the Drinking Water Quality Standard, the contaminated groundwater should to be cleaned up or contained or the drinking water production well must be closed.

For example, benzene has an Intervention Value of 30µg/L (Ministry of VROM, 2013) and a Drinking Water Quality Standard of 1µg/L (Ministry of IenM, 2011). If the concentration of benzene is above 30µg/L in groundwater, remediation must take place. However, if groundwater concentrations are below 30µg/L in the field but above 1µg/L at the well, then there is no obligation for remediation, although the water is unsuitable as drinking water. Benzene in groundwater should be reduced – taking into account dilution and reduction – to such a concentration that the benzene does not enter the well at concentrations above 1µg/L. The question remains, what could the permissible concentration at the site of contamination be?

The difference in criteria between the Soil Protection Act and the Drinking Water Act confuses the various stakeholders (municipalities, drinking water companies, etc.). To avoid confusion on how to act in the presence of chemical groundwater contamination an approach with four Tiers (0-3) was suggested by Otte et al. (2007) (figure 1).

Tier 0 is the initial characterization of chemical groundwater contamination. Tier 0 is aimed at identifying the presence of unacceptable risks. These are cases with non-aqueous phase liquids, the presence of vulnerable objects (groundwater production wells) and/or a contaminant plume with a concentration above the Intervention Value and a plume size of more than 6000m³ (Otte et al., 2007).

If one of such unacceptable risks is present then Tier 2 is started. If there is a possible risk or more information is required then Tier 1, a general risk assessment, is started.

Tier 1 takes the spreading of the plume, and thus an increase in plume size (volume) into account. If the plume size grows with more than 1000m³ in a year then the risk is considered to be unacceptable (Otte et al., 2007). This growth is determined using look-up tables with prescribed values for natural groundwater velocity, hydraulic conductivity and hydraulic gradient. If there is a risk which might be unacceptable or more information is required to form a conclusion, then a more site-specific assessment is started in Tier 2.

In Tier 2, site-specific data is used to assess the risk of a contaminant plume (Otte et al., 2007). Many analytical solutions for solute transport exist, however, they are not applicable to complex scenarios with pumping wells. Therefore numerical models are often used but they require in-depth knowledge of hydrology, solute transport processes and transport models. Instead a simple formula to calculate the concentration of a contaminant in the well and that can be used by people with a basic knowledge of the abovementioned subjects is proposed. This formula takes into account natural attenuation processes like dilution, dispersion, retardation and decay. If the formula predicts concentrations above the drinking water quality standard in the drinking water production well then tier 3 comes into play. Tier 3 consists of site-specific modelling and monitoring. The assessment of natural attenuation is also part of Tier 3. Site-specific modelling entails the use of numerical models like MODFLOW and MT3DMS.

In each Tier it is assessed whether the plume migrates towards vulnerable objects and when there is an unacceptable risk in one of the Tiers, remediation must be started.

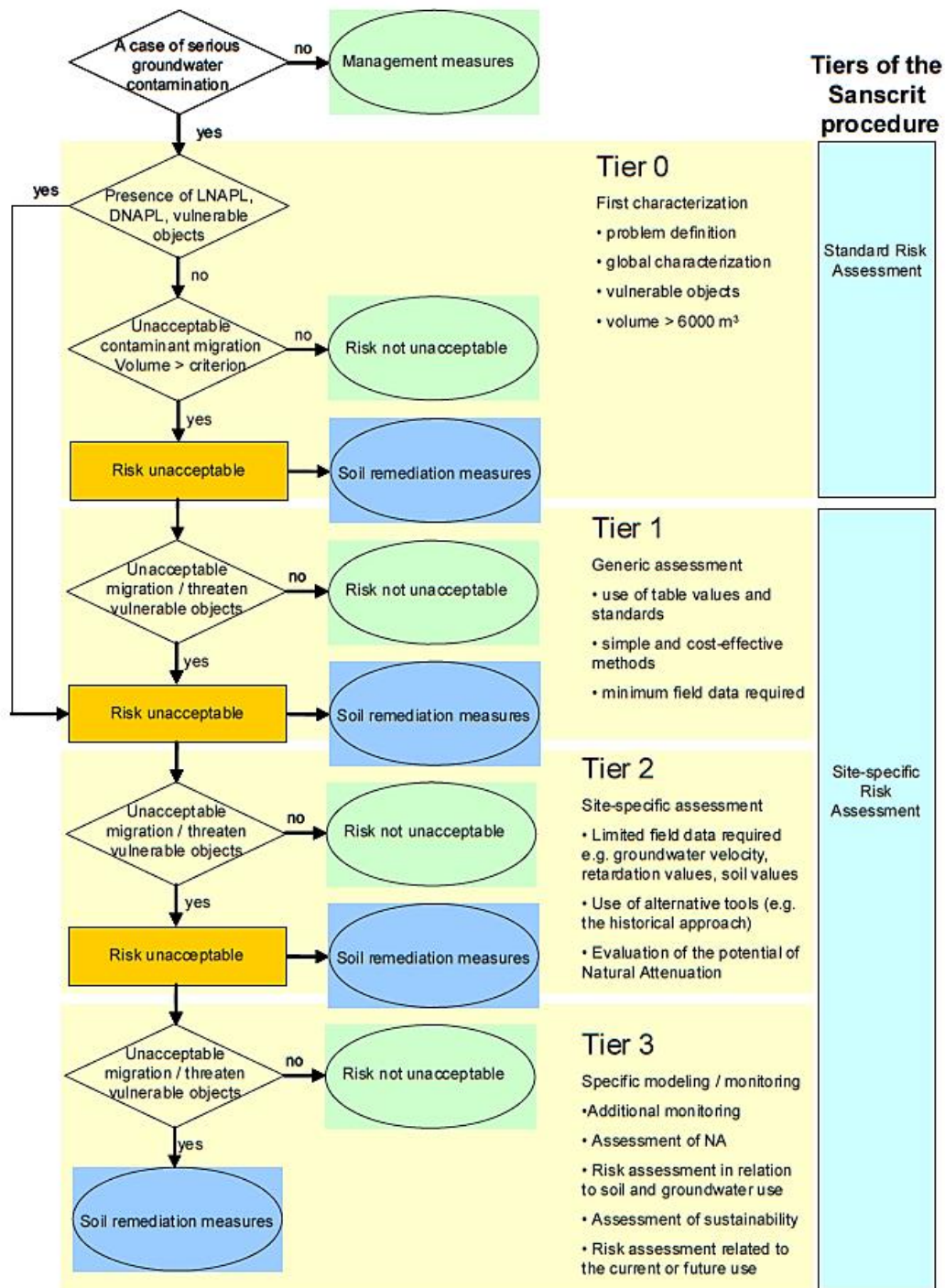


Figure 1 Flowchart of the tiered approach

This flowchart depicts the different tiers in the risk assessment of a contaminant plume. Taken from Otte et al., 2007.

1.2 Objective

Often chemical transport is calculated using analytical solutions for the transport equation or using numerical models. Analytical solutions often do not apply to field situations since the field situations are more complex than the analytical solutions allow. Numerical modelling is often used in field situations. Numerical modelling requires in-depth knowledge of the field situation but also about the physics involved in chemical transport. Since every contaminant differs, each contaminant needs a different set of parameter values. Municipalities often do not have this expert knowledge required for developing contaminant-specific numerical models.

This study aims to develop empirical formulas that are easy to use for a wide range of chemical contaminants. This means that the formulas do not need to be modified for different field situations and that anyone with a basic knowledge of chemical transport can apply these. The empirical formula will predict the concentration and arrival time of a chemical contaminant depending on seven model parameters, like the distance between the well and the plume, the well screen length, the hydraulic gradient, the pumping rate, anisotropy, retardation and decay.

1.3 Approach

In this study empirical formulas were set up by applying multivariate regression analysis on numerical model results. A wide range of different situations were modelled on the basis of seven parameters. The choices and assumptions made are further described in this chapter and in chapter 2.3 parameter space.

In order to develop the empirical formulas a model domain has been determined. The governing equations for ground water flow and contaminant transport were made dimensionless in order to reduce the number of parameters and to set up a dimensionless model domain. The parameter space was chosen in such a way that it covered the range of these (dimensional) parameters in the Netherlands. Within this parameter space various scenarios were set up to be modelled in PMWIN. PMWIN, or processing modflow for windows, is a 3D numerical model interface made by Simcore Software (©1991-2013). It is used to simulate groundwater flow (MODFLOW) and contaminant transport (MT3DMS) in complex scenarios and has been well investigated. For this study PMWIN version 8.0.40 was used. The PMWIN results were analysed using multivariate regression analysis which led to a set of empirical formulas. The set of empirical formulas were used to predict the peak, or maximum concentration at the well (C^*_{max}), the arrival time of the maximum concentration (t^*_{max}), half the maximum concentration (C^*_{half}) and the time half the maximum concentration was reached (t^*_{half}).

It is important to know the peak concentration because then it is easy to see if the Drinking Water Quality Standards will be exceeded. It is also important to know when this peak concentration will arrive since this gives an estimate of the time left to monitor the plume and to take action. The time half the maximum concentration is reached is the official arrival time. It is also useful to estimate when the contaminant will reach the well and whether the Drinking Water Quality Standard is exceeded.

To set up the domain some assumptions were made, which are discussed below.

List of assumptions:

1. Unconfined aquifer
2. Homogeneous anisotropic sandy soil
3. Fully water-saturated conditions
4. Groundwater flow in a straight line from the contaminant to the well
5. Only one well
6. Diffusion is negligible
7. Scale dependence of dispersion is not taken into account
8. Equilibrium sorption
9. Anaerobic decay
10. Only dissolved phase undergoes decay
11. There was no chain of decay

Unconfined aquifers are the most vulnerable aquifers since they lack a confining clay layer that slows the entry of contaminants into the aquifer. Therefore drinking water wells in unconfined aquifers are more vulnerable than the ones in confined aquifers (Wuijts et al., 2014). When the empirical

formulas show that a contaminant did not reach the well in an unconfined aquifer, it would not do so in a confined aquifer (1).

Aquifers used for extraction of drinking water are sandy aquifers of a heterogeneous and anisotropic nature. Anisotropy is the ratio between the horizontal and vertical hydraulic conductivity and therefore is a measure of relative ease of flow in the horizontal and vertical direction. Clay lenses were not included in the domain resulting in a horizontally homogeneous anisotropic aquifer (2). The influence of groundwater table fluctuations on the contaminant was not considered, fully saturated conditions were modelled (3).

Groundwater flow was assumed to be from the contaminant in a straight line towards the well, as a worst-case scenario (4). This means that contaminant transport is only towards the well and in the depth direction and thus all other directions can be ignored.

Next to natural groundwater flow there was radial flow due to the presence of the drinking water production well. While groundwater extraction was often done with multiple wells clustered in a larger area, the cumulative pumping rates of the wells were superimposed in one well in this study (5).

The transport of the contaminant is determined by advection, hydrodynamic dispersion, adsorption and decay.

- Advection is the movement of a solute with the bulk of the groundwater due to groundwater flow (Fetter, 1999). Density differences between the solute and water were neglected in this study.
- Hydrodynamic dispersion is the sum of diffusion and mechanical dispersion. In this study diffusion is assumed to be negligible compared to mechanical dispersion (6). Mechanical dispersion is caused by deviations of the average transport velocity. These deviations are caused by differences in pore size, travel length through pores and velocity across a pore. Larger pores have a lower velocity and the longer the pathway from A to B the lower the apparent velocity will be. Next to that, the velocity across a pore is highest in the middle and lowest at the edges. These differences in flow velocity cause a smearing of the plume front (Fetter, 1999). Dispersion is scale dependant, however this is not taken into account in this study (7)
- Adsorption is the removal and subsequent release of the solute through temporary sorption to mineral grains and organic carbon. Adsorption can be subdivided in equilibrium sorption in which an equilibrium value is reached and kinetic sorption in which the process is not fast enough to reach equilibrium (Fetter, 1999). In this case linear equilibrium adsorption was considered (8). This means that there is a linear relationship between the adsorbed and dissolved concentration. Adsorption causes a delay, or retardation, in the breakthrough curve.
- The decay rate depends on, amongst others, the availability of different chemicals like oxygen, iron and manganese. In this study only anaerobic decay was considered and the decay rate was constant over distance (9).

Both the adsorbed and the dissolved phase of a contaminant may undergo decay, however only solute decay was considered here since decay in the adsorbed phase is negligible (10). Many chemicals like benzene follow a chain of decay in which the daughter products are contaminants as well. In this study these decay chains, and consequently the daughter products, were not considered (11).

2 Method

2.1 PMWIN domain

The PMWIN domain was based on the size of the average groundwater protection zone (figure 2) and was a rectangle with length L , width W and height H (m). X was the horizontal distance (m) between the contaminant plume and the well, ΔZ is the vertical distance (m) between the plume and the well while Z_w gave the well screen length (m).

The natural groundwater flow was driven by a hydraulic gradient i . The hydraulic gradient was calculated with $i = \frac{h_u - h_d}{L}$ where h_u is the upstream hydraulic head and h_d is the downstream hydraulic head (m). The anisotropy, m , was the dimensionless ratio between k_x/k_z where k_x is the horizontal hydraulic conductivity and k_z is the vertical hydraulic conductivity (m/d). The porosity was given by n . The hydraulic gradient, the porosity and the hydraulic conductivity determine the natural groundwater velocity. Q was the pumping rate (m^3/d) and determined the radial velocity. The plume had a volume of V_c (m^3) and a concentration C ($\mu\text{g}/\text{L}$). ρ^b was the bulk density of the soil (kg/m^3) and K_D the distribution coefficient (m^3/kg). These two parameters, together with the porosity, determined the retardation R . α_L was the dispersivity (m) and μ the decay rate (d^{-1}). These different parameters will be explained further in the next paragraphs.

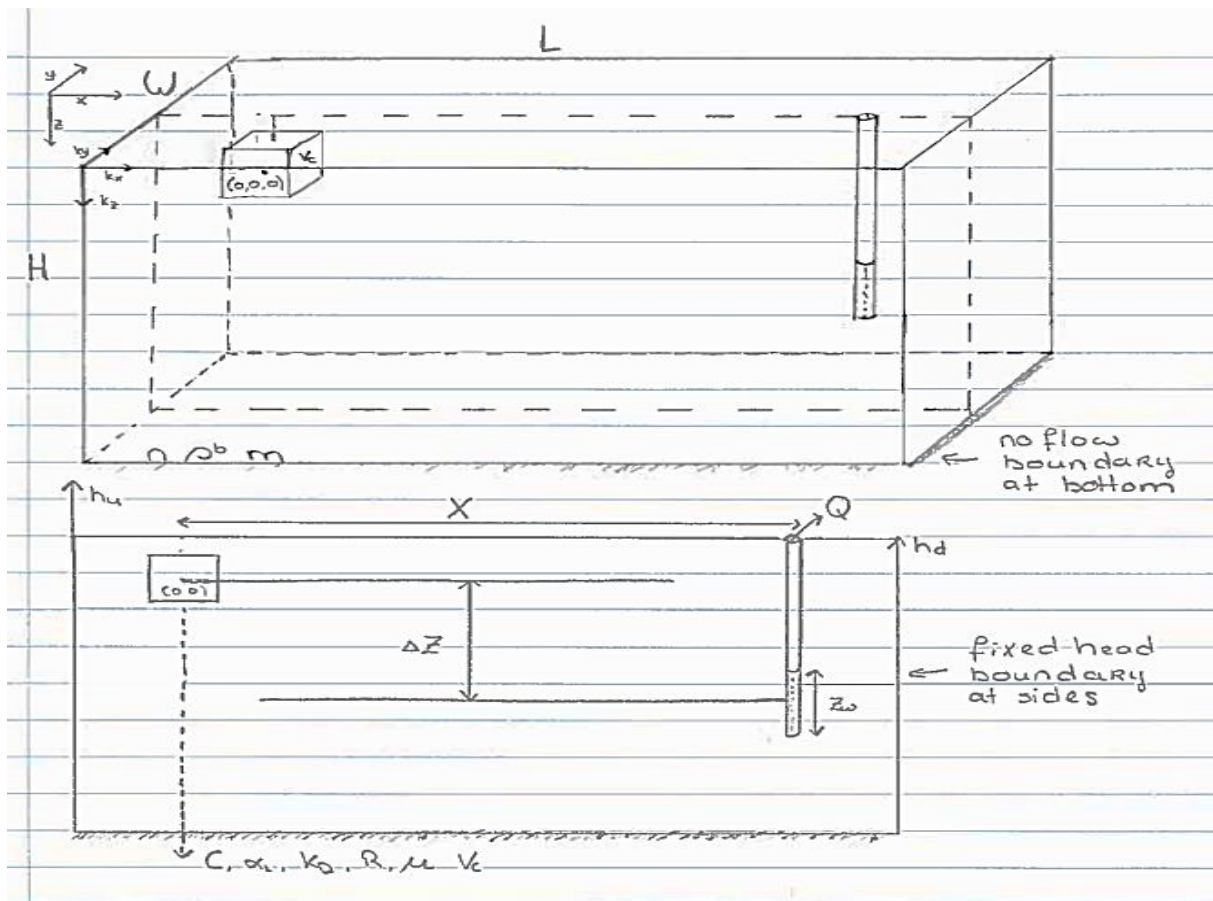


Figure 2 Model domain

The rectangular domain holds a contaminant plume close to the surface and a pumping well with a screen length of Z_w and a pumping rate Q . The domain has length L , width W and height H . There is a horizontal (X) and a vertical (ΔZ) distance between the plume and the well. The soil characteristics are the hydraulic conductivity (k), the porosity (n), the soil bulk density (ρ^b), the anisotropy (m), the upstream hydraulic head (h_u) and the downstream hydraulic head (h_d). The contaminant characteristics are the contaminant concentration (C), the dispersivity (α_L), the partitioning coefficient (K_D), the retardation factor (R), the decay rate (μ) and the volume of the contaminant plume (V_c). There is a no-flow boundary condition at the bottom of the domain and a constant head boundary at the left and right side of the domain.

The domain was modelled using PMWIN and was made dimensionless. By using a dimensionless model, less parameters variations were needed to reach the same results as a dimensional model. To determine the conversion factors for the domain first the stream function and the transport equation were made dimensionless. The stream function is used to calculate groundwater flow and the transport equation calculates the movement and the concentration of a contaminant.

The dimensionless domain was then implemented in PMWIN (Appendix B). The domain was divided into 10 layers, 10 rows and 65 columns (dimensional: 10m×100m×100m). PMWIN automatically implemented a no-flow boundary condition at the bottom of the domain to prevent seepage out of the domain. At the left and right side of the domain fixed head boundary conditions were applied to create a constant head situation and thus a constant hydraulic gradient.

The grid was refined to by a factor 10 in x and y direction around the contaminant plume and the well was placed in a grid cell at distance X and ΔZ with and a pumping rate Q. An observation point was placed in every well grid cell to observe the concentration in the well.

Advection was modelled using the finite difference method with upstream weighing and a Courant number of 0.1.

2.2 Governing equations

Transport in the PMWIN domain was calculated using the stream function (eq. 1, modified from Harbaugh et al., 2000). The stream function describes how a change in groundwater flow velocity in x, y and z direction and a source/sink term affects the change in storage over time. Since there was a steady state situation the change in storage was zero. In this case the sink term was the pumping rate per time.

$$\frac{\partial}{\partial x} \left(k_x \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left(k_y \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left(k_z \frac{\partial h}{\partial z} \right) - \frac{Q}{L^3} = \frac{\partial}{\partial x} (q_x) + \frac{\partial}{\partial y} (q_y) + \frac{\partial}{\partial z} (q_z) - \frac{Q}{L^3} = 0 \quad (\text{eq. 1})$$

With:

k is the hydraulic conductivity in x, y and z direction [L/T]

h is the hydraulic head [L]

Q is the pumping rate [L³/T]

L is the length of the domain [L³]

q is the Darcy velocity in x, y and z direction [L/T]. $q = nv$ where v is the porewater velocity [L/T] and

n is the porosity [-]

In order to make the stream function dimensionless x, y and z were divided by L and h was divided by h_d leading to equation 2. In order to look at the pore water velocity $q = nv$ was used to replace the Darcy velocity with the pore water velocity.

$$\frac{1}{L} \frac{\partial}{\partial x^*} \left(k_x \frac{h_d}{L} \frac{\partial h^*}{\partial x^*} \right) + \frac{1}{L} \frac{\partial}{\partial y^*} \left(k_y \frac{h_d}{L} \frac{\partial h^*}{\partial y^*} \right) + \frac{1}{L} \frac{\partial}{\partial z^*} \left(k_z \frac{h_d}{L} \frac{\partial h^*}{\partial z^*} \right) - \frac{Q}{L^3} = \frac{1}{L} \frac{\partial}{\partial x^*} (nv_x) + \frac{1}{L} \frac{\partial}{\partial y^*} (nv_y) + \frac{1}{L} \frac{\partial}{\partial z^*} (nv_z) - \frac{Q}{L^3} = 0 \quad (\text{eq. 2})$$

k_x , k_y and k_z were made dimensionless by dividing by k_x and the whole equation was multiplied by L leading to equation 3.

$$\frac{\partial}{\partial x^*} \left(k_x^* \frac{h_d}{L} \frac{\partial h^*}{\partial x^*} \right) + \frac{\partial}{\partial y^*} \left(k_y^* \frac{h_d}{L} \frac{\partial h^*}{\partial y^*} \right) + \frac{\partial}{\partial z^*} \left(k_z^* \frac{h_d}{L} \frac{\partial h^*}{\partial z^*} \right) - \frac{Q}{L^2 k_x} = \frac{\partial}{\partial x^*} \left(\frac{n}{k_x} v_x \right) + \frac{\partial}{\partial y^*} \left(\frac{n}{k_x} v_y \right) + \frac{\partial}{\partial z^*} \left(\frac{n}{k_x} v_z \right) - \frac{Q}{L^2 k_x} = 0 \quad (\text{eq. 3})$$

The equation was then multiplied by L/h_d to simplify it (eq. 4).

$$\begin{aligned} \frac{\partial}{\partial x^*} \left(k_x^* \frac{\partial h^*}{\partial x^*} \right) + \frac{\partial}{\partial y^*} \left(k_y^* \frac{\partial h^*}{\partial y^*} \right) + \frac{\partial}{\partial z^*} \left(k_z^* \frac{\partial h^*}{\partial z^*} \right) - \frac{Q}{L k_x h_d} = \\ \frac{\partial}{\partial x^*} \left(\frac{nL}{k_x h_d} v_x \right) + \frac{\partial}{\partial y^*} \left(\frac{nL}{k_x h_d} v_y \right) + \frac{\partial}{\partial z^*} \left(\frac{nL}{k_x h_d} v_z \right) - \frac{Q}{L k_x h_d} = 0 \end{aligned} \quad (\text{eq. 4})$$

According to Darcy's law $v_x = -\frac{k_x}{n} \frac{\partial h}{\partial x}$, $v_y = -\frac{k_y}{n} \frac{\partial h}{\partial y}$ and $v_z = -\frac{k_z}{n} \frac{\partial h}{\partial z}$. These Darcy velocities were made dimensionless following the abovementioned steps leading to $v_x = -\frac{k_x h_d}{n L} \frac{\partial h^*}{\partial x^*} = -v_{res} \frac{\partial h^*}{\partial x^*} = v_{res} v_x^*$, $v_y = -k_x \frac{h_d}{L} k_y^* \frac{\partial h^*}{\partial y^*} = v_{res} v_y^*$ and $v_z = -k_x \frac{h_d}{L} k_z^* \frac{\partial h^*}{\partial z^*} = v_{res} v_z^*$ with $v_{res} = k_x \frac{h_d}{nL}$.

This was implemented in equation 4 and $\frac{Q}{L h_d k_x}$ was substituted by Q^* leading to equation 5 which is the dimensionless stream function. All the conversions applied can be found in table 1.

$$\begin{aligned} \frac{\partial}{\partial x^*} \left(k_x^* \frac{\partial h^*}{\partial x^*} \right) + \frac{\partial}{\partial y^*} \left(k_y^* \frac{\partial h^*}{\partial y^*} \right) + \frac{\partial}{\partial z^*} \left(k_z^* \frac{\partial h^*}{\partial z^*} \right) - Q^* = \frac{\partial}{\partial x^*} (v_{res} v_x^*) + \frac{\partial}{\partial y^*} (v_{res} v_y^*) + \frac{\partial}{\partial z^*} (v_{res} v_z^*) - \\ Q^* = 0 \end{aligned} \quad (\text{eq. 5})$$

Table 1 Conversion of stream function to dimensionless stream function

Dimension	Dimensionless	Dimension	Dimensionless
x	$x^* = \frac{x}{L}$	h	$h^* = \frac{h}{h_d}$
y	$y^* = \frac{y}{L}$	Q/L^3	$Q^* = \frac{Q}{L k_x h_d}$
z	$z^* = \frac{z}{L}$	v_y	$v_y^* = \frac{v_y}{v_{res}}$
k_x	$k_x^* = \frac{k_x}{k_x} = 1$	v_z	$v_z^* = \frac{v_z}{v_{res}}$
k_y	$k_y^* = \frac{k_y}{k_x}$	$v_{res} = k_x \frac{h_d}{nL}$	
k_z	$k_z^* = \frac{k_z}{k_x}$	n	$n^* = \frac{n}{n} = 1$
v_x	$v_x^* = \frac{v_x}{v_{res}}$		

Contaminant transport not only depends on the stream function but also on the transport processes (eq. 6, modified from Zheng and Wang, 1999). The transport equation describes how concentration changes over time and distance under the influence of adsorption, hydrodynamic dispersion, advection and decay.

$$R \frac{\partial C}{\partial t} + \frac{\partial v_i C}{\partial x_i} = \frac{\partial}{\partial x_i} \left(D_{ij} \frac{\partial C}{\partial x_j} \right) - \mu C \quad (\text{eq. 6})$$

$$\text{Where } D_{ij} = (D_m + \alpha_T |v|) \delta_{ij} + \frac{(\alpha_L - \alpha_T) v_i v_j}{|v|}$$

With:

R is the retardation factor [-]. $R = 1 + \frac{\rho_b K_D}{n}$ where ρ_b is the bulk density [kg/m^3], K_D is the distribution coefficient [m^3/kg] and n is the porosity [-]

C is the concentration [M/L^3]

t is the time [T]

v_i is the pore water velocity [L/T] and expands to 3 terms $\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$

μ is the decay rate coefficient of the dissolved phase [T^{-1}]

D_{ij} is the dispersion tensor [L^2/T] and expands to 9 terms $\begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}$

D_m is the molecular diffusion [L^2/T]

α_T is the transverse dispersivity [L]

α_L is the longitudinal dispersivity [L]

δ_{ij} is the 3x3 unit matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$|v|$ is the length of the velocity vector [L/T] where $|v| = \sqrt{v_x^2 + v_y^2 + v_z^2}$

$R \frac{\partial C}{\partial t}$ describes linear equilibrium adsorption of the contaminant to soil and organic carbon particles. When $R > 1$ there was adsorption and transport of the contaminant was retarded. Hydrodynamic dispersion causes spreading of the contaminant plume, which is described by $\frac{\partial}{\partial x_i} \left(D_{ij} \frac{\partial C}{\partial x_j} \right)$. The larger the hydrodynamic dispersion the greater the concentration gradient. Advection is described with $-\frac{\partial v_i C}{\partial x_i}$ and includes the dependence of groundwater flow velocity. The larger the velocity the greater the concentration difference over distance. Finally, $-\mu C$ describes how the concentration drops with the decay rate coefficient. The larger the decay rate coefficient the larger the drop in concentration.

To make the transport equation dimensionless the conversions from the stream function were used for x, y, z and v . Next to that C was made dimensionless by dividing it by the initial concentration C_0 (table 2). Time was made dimensionless by dividing by the residence time $t_{res} = \frac{L}{v_{res}}$.

Applying these conversions yields equation 7.

$$R \frac{v_{res}}{L} \frac{\partial C^*}{\partial t^*} + \frac{v_{res}}{L} \frac{\partial v_i^* C^*}{\partial x_i^*} = \frac{1}{L} \frac{\partial}{\partial x_i^*} \left(\frac{D_{ij}}{L} \frac{\partial C^*}{\partial x_j^*} \right) - \mu C^* \quad (\text{eq. 7})$$

Table 2 Conversion of transport equation to dimensionless transport equation

Dimension	Dimensionless	Dimension	Dimensionless
C	$C^* = \frac{C}{C_0}$	n	$n^* = \frac{n}{n} = 1$
t	$t^* = \frac{t}{t_{res}} = \frac{v_{res} t}{L}$	x	$x^* = \frac{x}{L}$
v_x	$v_x^* = \frac{v_x}{v_{res}}$	y	$y^* = \frac{y}{L}$
v_y	$v_y^* = \frac{v_y}{v_{res}}$	z	$z^* = \frac{z}{L}$
v_z	$v_z^* = \frac{v_z}{v_{res}}$	μ	$Da = \mu^* = \frac{\mu L}{v_{res}}$
$v_{res} = k_x \frac{h_d}{nL}$			

By multiplying equation 7 with $\frac{L}{v_{res}}$ and incorporating L^{-1} in D_{ij} the transport equation was made dimensionless (eq. 8). The dimensionless decay rate is the dimensionless Damkohler number. This equation is similar in form to the original transport equation.

$$R \frac{\partial C^*}{\partial t^*} + \frac{\partial v_i^* C^*}{\partial x_i^*} = \frac{\partial}{\partial x_i^*} \left(D_{ij}^* \frac{\partial C^*}{\partial x_j^*} \right) - Da C^* \quad (\text{eq. 8})$$

$$\text{Where } D_{ij}^* = \left(\frac{D_m}{v_{res} L} + \alpha_T^* |v^*| \right) \delta_{ij} + \frac{(\alpha_L^* - \alpha_T^*) v_i^* v_j^*}{|v^*|}$$

The domain was then made dimensionless by using the abovementioned conversion. This was done by dividing all parameters with the unit length, except for the hydraulic heads, by distance L (so height H, width W, travel length X, vertical distance ΔZ and well screen length Z_w were divided by the domain length)(table 3). Bulk density ρ^b and distribution coefficient K_D were not made dimensionless since these parameters were only directly included in PMWIN and not in the stream function and only indirectly in the transport equation.

Table 3 Conversion of domain to dimensionless values

Dimension	Dimensionless	Dimension	Dimensionless
L	$L^* = \frac{L}{L} = 1$	ΔZ	$\Delta Z^* = \frac{Z}{L}$
W	$W^* = \frac{W}{L}$	Z_w	$Z_w^* = \frac{Z_w}{L}$
H	$H^* = \frac{H}{L}$	ρ^b	n/a
X	$X^* = \frac{X}{L}$	K_D	n/a
x	$x^* = \frac{x}{L}$	n	$n^* = \frac{n}{n} = 1$
y	$y^* = \frac{y}{L}$	Q/L^3	$Q^* = \frac{Q}{L k_x h_d}$
z	$z^* = \frac{z}{L}$	h	$h^* = \frac{h}{h_d}$
k_x	$k_x^* = \frac{k_x}{k_x} = 1$	V_c	$V_c^* = \frac{V_c}{L^3}$
k_y	$k_y^* = \frac{k_y}{k_x}$	C	$C^* = \frac{C}{c_0}$
k_z	$k_z^* = \frac{k_z}{k_x}$	μ	$Da = \mu^* = \frac{\mu L}{v_{res}}$
v_x	$v_x^* = \frac{v_x}{v_{res}}$	t	$t^* = \frac{t}{t_{res}} = \frac{v_{res} t}{L}$
v_y	$v_y^* = \frac{v_y}{v_{res}}$	$v_{res} = k_x \frac{h_d}{nL}$	
v_z	$v_z^* = \frac{v_z}{v_{res}}$		

2.3 Parameter space

In this model 13 parameters, namely the horizontal distance between the contaminant plume and the well (X), the domain length (L), the aquifer depth (H), the vertical distance between the contaminant and the well (ΔZ), the well screen length (Z_w), the hydraulic gradient (i), the pumping rate (Q), the anisotropy (m), the contaminant volume (V_c), the contaminant concentration (C), the retardation (R), the dispersivity (α_L) and the decay rate (μ) could be varied. To model each possible combination of these parameters, for five values, would lead to $5^{13} = 1,220,703,125$ scenarios, which is not feasible within the timeframe set for this study. Reduction of the amount of scenarios was achieved by reducing the number of variables. This was done by first making the model dimensionless and secondly by keeping some variables constant. The number of scenarios was further reduced by making a “smart” selection of points within a full factorial design. This however will be discussed in chapter 2.4 Scenarios.

Dimensionless model

By making the model dimensionless, the horizontal distance between the well and the plume (X), the aquifer depth (H) and the domain width (W) all became dependent on the domain length (L), meaning that a variation in L causes a variation in all these parameters. Therefore L was kept constant. An L of 6500m (table 4) was chosen to have a buffer zone of 250m before the plume and after the well at the greatest distance X (see parameter space - variables). This buffer zone was used to prevent interference from the constant head boundary conditions with the contaminant transport.

Contaminant transport does not take place in the y-direction, because groundwater flow was from the plume straight to the well and therefore the size of W did not matter as long as it was sufficiently wide for all the cases. A W of 1000m was chosen (table 4).

By making the model dimensionless the concentrations became relative to the initial concentration of the plume. This means that the initial concentration of the plume was divided by the initial concentration of the plume and thus the dimensionless starting concentration was constant at 1 (table 4).

This reduced the number of parameters from 13 to 10.

Constant parameters

Aquifer depths (H) range from 20 to 120 to 330m (rounded off from Schijven et al., 2010). However because the depth was <5% of the length of the domain the effect of depth will most likely be smaller than that of length. Therefore H was kept constant at a value of 120m (table 4).

As mentioned in the tiered approach a plume size of over 6000m^3 was considered an unacceptable risk (Otte et al., 2007), however smaller plume sizes may also form a possible risk and therefore a plume size of 1000m^3 was chosen (table 4).

A contaminant plume mostly originates from a surface source and therefore lies near to the surface. Thus the plume was located in the top layer in the PMWIN domain. The frequency distribution of the depth of the top of the well screen (figure 3) (Zwolsman, 2011) shows that most well screens have a depth of 50m below the surface, therefore the top of the well was set at 50m below surface. The vertical distance from the plume to the well (ΔZ) was kept as constant as possible but due to variations in the well screen length (Z_w), ΔZ was subsequently varied as a secondary variable (table 4).

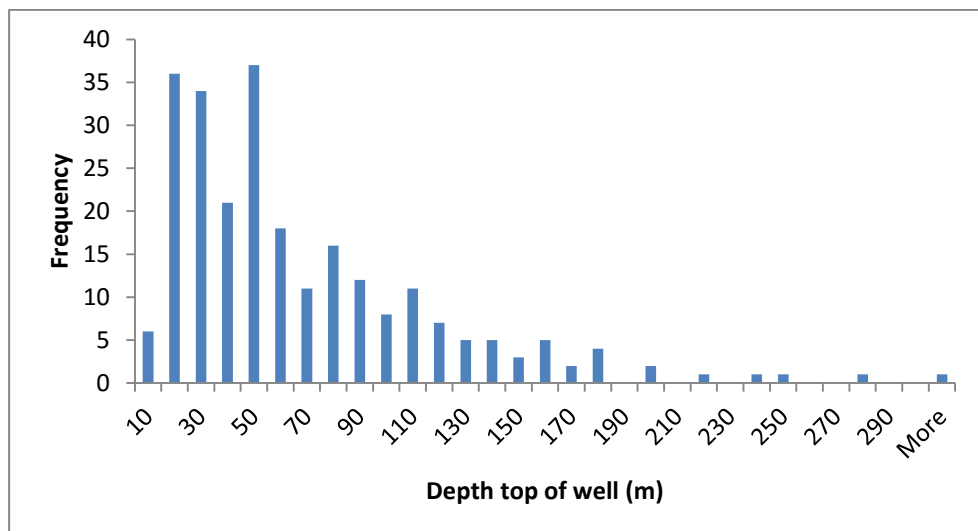


Figure 3 Frequency distribution of the depth of the top of the well (m) below surface

Lastly, the dispersivity (α_L) was kept constant at 130m as to have $Pe=50$ (table 4). The Peclet number (Pe) is the ratio between advective transport and dispersive transport (Fetter, 1999). With a $Pe=50$ there was more advective transport than dispersive transport.

Only 7 parameters with 5 values remain. This led to $5^7=78,125$ scenarios which was quite a reduction, but not enough.

Keeping values constant led to parameter reduction but it also forms a constraint on the empirical formula since it is only applicable within the parameter range provided. This means that the empirical formula is only applicable in situations where $H=120$, $V_c=1000\text{m}^3$, $\Delta Z=50\text{m}$ and $\alpha_L=130\text{m}$. Future studies may explore the possibilities in varying or changing the value of these constants.

Variables

The 25 year zone, or groundwater protection zone, is defined as a zone in which water particles travel 25 years to reach the well (Wuijts et al., 2007). This zone is defined by the act on environmental management (Wet Milieubeheer) and subdivided in environmental decrees by the Provincial Governments (Ministry of Volksgezondheid en Milieuhygiëne, 1979). This means that a tracer will travel 25 years to the well but a non-conservative contaminant will travel longer. In 25 years a contaminant can be (partly) broken down and thus it will reach the well at a lower concentration than a conservative contaminant. Therefore the biggest distance X was taken as the maximum length of the 25 year zone. The size of the 25 year zone was taken from so called 'gebiedsdossiers' (WMD-water, 2011, Hoogeveen and Witjes, 2012 and Lodder and Steinweg, 2013), the websites of the provinces (maps from Utrecht, Gelderland and Overijssel) and the environmental decree Noord-Brabant (provinciale milieu verordening Noord-Brabant, 2010). From these different distances a frequency distribution was made (figure 4) to determine the range of variation for the horizontal distance between the well and the plume (X). A few runs in PMWIN revealed that distances of X greater than 4500m led to arrival times of more than 50 years and were therefore of no interest for this study. X was varied from 100, 750, 1600, 2300 to 3050m (table 4).

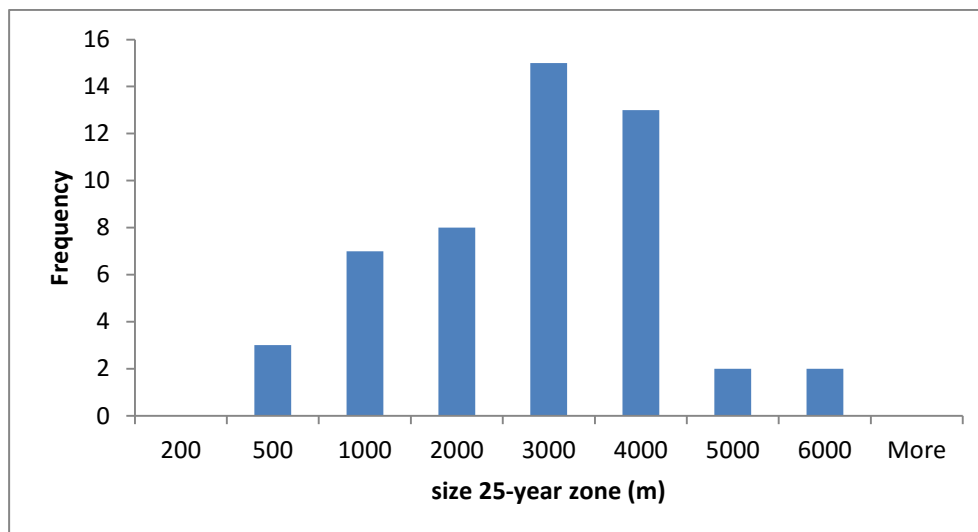


Figure 4 Frequency distribution of the size of the 25-year zone

Zwolsman (2011) gives the depth of the top and bottom of each drinking water well screen in the Netherlands. From this the well screen length (Z_w) was derived and a frequency distribution set up (figure 5) which led to a range of variation. The well screen cannot be larger than the height of the aquifer, therefore all values above 120m were disregarded. The well screen length was varied from 10, 30, 50, 70 to 90m (table 4).

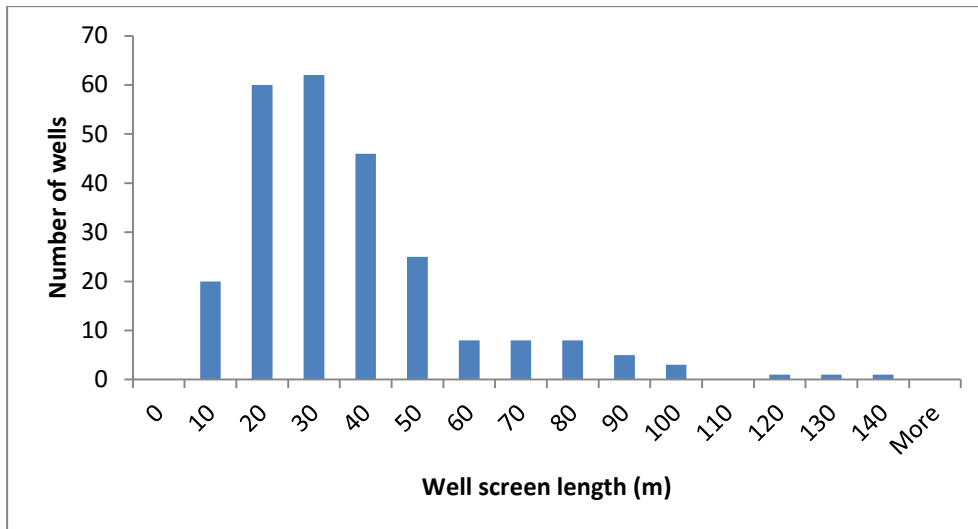


Figure 5 Frequency distribution of well screen length

In the domain there was both radial flow due to the pumping wells and natural flow due to a hydraulic gradient. In the Netherlands, this hydraulic gradient (i) is around 0.01m/m to 0.001m/m (van Rooijen, 1989). Davidson and Wilson (2011) found an i of 0.0033m/m and 0.005m/m in two aquifers. This falls within the range given by van Rooijen (1989). An i of 0.001m/m led to a very low flow velocity and therefore higher values were chosen. The hydraulic gradient was increased from 0.0055, 0.0067, 0.0078, 0.0089 to 0.01m/m (table 4). The gradient was calculated as $i = \frac{h_u - h_d}{L}$. The downstream hydraulic head (h_d) was kept constant at $H=120\text{m}$ in order to keep a fully saturated domain. $h_d=120\text{m}$, $L=6500\text{m}$ and i was also known. Then the required upstream hydraulic head (h_u) was calculated with $h_u = (i L) + h_d$. Therefore, h_u was varied from 155.75, 163.55, 170.7, 177.85 to 185m (table 4).

The frequency distribution of pumping rates in Dutch drinking water wells (figure 6) (Zwolsman, 2011) led to a range of pumping rates (Q). Q was increased from 200, 10250, 20100, 30350 to 40000m³/d (table 4).

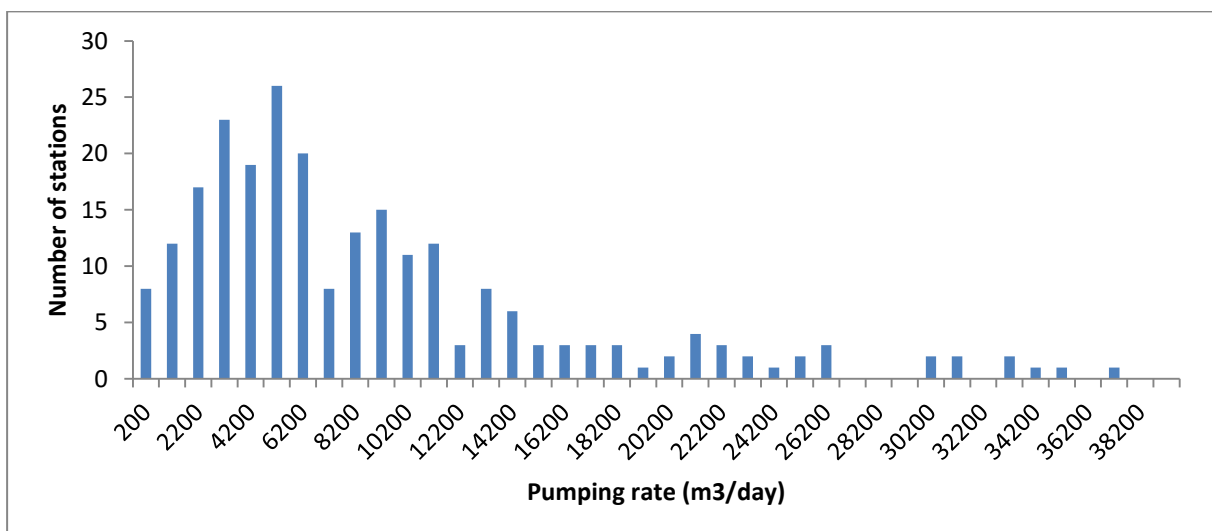


Figure 6 Frequency distribution of pumping rates in 2010

Another factor in groundwater flow velocity was the hydraulic conductivity (k) in x , y and z direction. Only k_z was varied, which means that in the dimensionless situation only the ratio between k_x and k_z was varied. This ratio is also known as the anisotropy (m). k_x has a range of 1-100m/d in sandy soils (Hendriks, 2010). By looking at the k_x value in the soil around 20 unconfined drinking water wells

spread over the Netherlands in Dinoloket it (TNO, 2015) became clear that most soils have a k_x value between 10-100m/d and some have a value between 1-100m/d. To obtain a variation in anisotropy ($m=k_x/k_z$) from 1, 13.25, 25.5, 37.75 to 50, k_z was varied from 20, 1.5, 0.78, 0.52 to 0.4m/d when $k_x=20$ m/d (table 4). This falls within the range given by Dinoloket.

The retardation (R) was varied by keeping the porosity (n) and the bulk density (ρ^b) constant and only changing the partitioning coefficient (K_D). The K_D values were obtained from a list of contaminants (Otte et al., 2001) taking only those that are soluble in water ($S > 1000$ mg/L) and looking at the occurrence of different K_D values in a histogram (figure 7). The K_D values were calculated from the soil organic carbon-water partitioning coefficient (K_{OC}) values using $K_D=K_{OC}\times f_{OC}$ where $f_{OC}=0.028$ (Medina et al., 1981).

A quick test run in PMWIN revealed that when R was larger than 10 the contamination will reach the well after 50 years. Therefore a maximal R value of 10 was chosen to keep the arrival time within 50 years. This led to the range of $R=1, 3.25, 5.5, 7.75$ and 10 (table 4). With $n=0.35$ (Hendriks, 2010 and Koorevaar et al., 1983) and $\rho^b=2000$ kg/m³ these R values were then translated back to the K_D values that were used in MODLOW. The K_D values were 0, 0.00039, 0.00079, 0.0012 and 0.0016m³/kg (table 4).

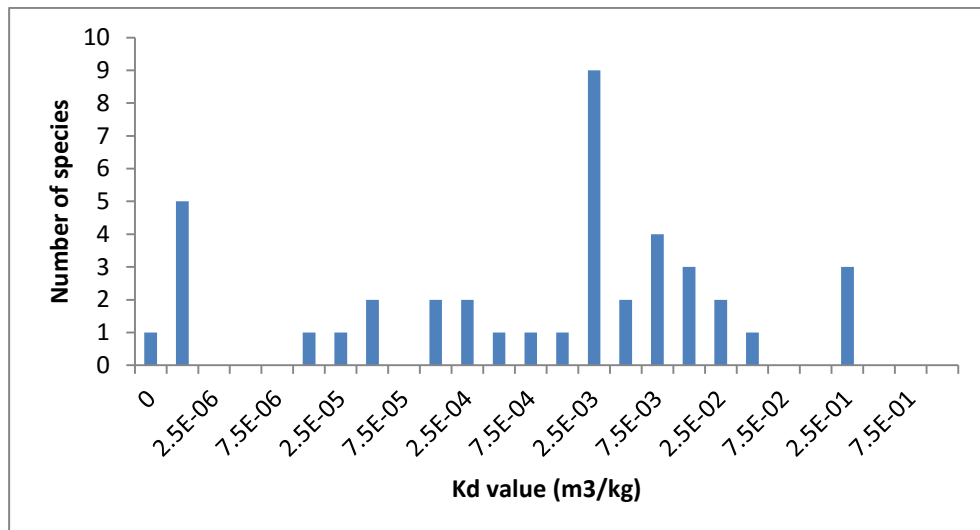


Figure 7 Frequency distribution of the K_D value of dissolved species

The decay rate (μ) was varied from 0, 0.0013, 0.0025 to 0.0038 to 0.005d⁻¹ (table 4) (as provided by the RIVM). Higher decay rates led to excessive concentration reductions and therefore were not taken into account.

All these parameters and constants were made dimensionless using the conversion from table 3. A full overview of the parameter ranges and constants both dimensional and dimensionless can be found in table 4. The mean values of each parameter are shown in bold and were used as central or reference values.

Table 4 Values of the constants and ranges of the variables
The bold printed values are the mean values of the parameters

Constants	Value	Dimensionless constants	Value
L	6500m	L*	1
W	1000m	W*	0.15
H	120m	H*	0.018
ΔZ	50m	ΔZ^*	0.0077
h_d	120m	h_d^*	1

k_y	2m/d	k_y^*	1
C	1	C^*	1
V_C	1000m ³	V_C^*	3.64×10 ⁻⁹
n	0.35	n^*	1
ρ^b	2000kg/m ³	n/a	n/a
α_L	130	α_L^*	0.02
α_{Ty}	13	α_{Ty}^*	0.002
α_{Tz}	13	α_{Tz}^*	0.002
Variable	Range	Dimensionless variables	Range
X	100 – 750 – 1600 – 2300 – 3050	X^*	0.015 – 0.11 – 0.24 – 0.35 – 0.47
Z_W	10 – 30 – 50 – 70 – 90m	Z_W^*	0.0015 – 0.0046 – 0.0077 – 0.011 – 0.014
i	0.0055 – 0.0067 – 0.0078 – 0.0089 – 0.01m/m		
h_u^o	155.75 – 163.55 – 170.7 – 177.85 – 185m	h_u^*	1.30 – 1.36 – 1.42 – 1.48 – 1.54
Q	200 – 10250 – 20100 – 30350 – 40000m ³ /d	Q^*	0.000013 – 0.00066 – 0.0013 – 0.00195 – 0.0026
k_x k_z	20 m/d 20 – 1.5 – 0.78 – 0.52 – 0.4m/d	k_x^* k_z^*	1 1 – 0.075 – 0.04 – 0.026 – 0.02
K_D'	0 – 0.00039 – 0.00079 – 0.0012 – 0.0016m ³ /kg	m	1 – 13.25 – 25.5 – 37.75 – 50
μ	0 – 0.0013 – 0.0025 – 0.0038 – 0.005d ⁻¹	R	1 – 3.25 – 5.5 – 7.75 – 10
		Da	0 – 7.75 – 15.5 – 23.25 – 31

2.4 Scenarios

Nine test scenarios (table 5) were run to check the model. Scenarios 1 to 3 did not have a pumping well and there was a continuous input of a tracer. The scenario outcomes were compared to an analytical solution supplied on the website of the University of Illinois (Valocchi et al., 2016) as a model check. This analytical solution is applicable in a 1D situation in a semi-infinite domain. To achieve the 1D situation the pumping well was removed and the plume covered the whole depth and width of the domain. The required input parameters were the dispersion coefficient (D), the pore water velocity (v), the input concentration (C_0) and the background value or initial concentration (which was zero). The results from PMWIN and the model from the University of Illinois should overlap as much as possible.

Scenarios 4 to 9 do had a fully penetrating pump, which led to a varying flow velocity in x and z direction, and can only be checked quantitatively. It was expected that the contaminant will reach the well at 100 m earlier than the well at 6000 m. An increase in i would lead to faster transport and thus earlier arrival. A higher pumping rate will increase the flow velocity but it also increases dilution.

Table 5 Dimensionless parameter values for the test scenarios

In which X^* is the horizontal distance from the plume to the well, Z_W^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number (the dimensionless decay rate).

Test scenario	X^*	Z_W^*	i	Q^*	m	R	Da
1	0.015	0	0.001	0	1	1	0
2	0.31	0	0.001	0	1	1	0
3	0.92	0	0.001	0	1	1	0
4	0.31	1	0.001	0.00032	1	1	0
5	0.015	1	0.001	0.00032	1	1	0

6	0.92	1	0.001	0.00032	1	1	0
7	0.31	1	0.01	0.00032	1	1	0
8	0.31	1	0.001	$1.3 \cdot 10^{-5}$	1	1	0
9	0.31	1	0.001	0.0026	1	1	0

A Box-Behnken (hereafter BB) design was applied (table 6) to reduce the number of scenarios further. This design was developed to decrease the number of scenarios (Cavazzuti, 2013) by combining two-level factorial designs and incomplete block designs (Box and Behnken, 1960 and Cavazzuti, 2013). By applying criteria of rotatability, the number of scenarios was greatly reduced while still remaining representative. The resulting design can be used to set up a quadratic polynomial (Box and Behnken, 1960).

This particular BB design with seven factors consists of boxes, in which four factors were kept constant, while the other three factors were varied with their minimum and maximum value until every possible combination of the three factor values were varied. Then, in the next box, three other factors were varied. This led to seven boxes in which three factors were varied. In the eighth box all factors were set to their mean value to form a centre point. Variations on the centre point were added to the eighth block as well. For these small variations each parameter was individually varied between the maximum and minimum value while the other parameters were kept at their mean value. This led to 72 scenarios in the BB design.

To increase the amount of data points a partial Central Composite design (hereafter CCD) was added (table 7). A CCD originally was a full factorial design for two factors with an added central point and star points (Cavazzuti, 2013). Since the BB already uses a minimum, mean and maximum these star- and central points could be eliminated. This means that only the two remaining values were used to form a CCD design leading to 64 scenarios. Using the CCD to increase data points led to a better result when using multivariate regression since a line was now plot using five points instead of three. In the end, the BB design and the CCD had 136 scenarios instead of $5^7=78,125$ scenarios.

Table 6 Dimensionless parameter values for the Box-Behnken scenarios

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number (the dimensionless decay rate). The bold values are varied between their minimum and maximum values in that particular box.

Block	BB Scenario	X^*	Z_w^*	i	Q^*	m	R	Da
1	1	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	1	1	15.5
	2	0.24	0.0077	0.0078	0.0026	1	1	15.5
	3	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	50	1	15.5
	4	0.24	0.0077	0.0078	0.0026	50	1	15.5
	5	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	1	10	15.5
	6	0.24	0.0077	0.0078	0.0026	1	10	15.5
	7	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	50	10	15.5
	8	0.24	0.0077	0.0078	0.0026	50	10	15.5
2	9	0.015	0.0077	0.0078	0.0013	25.5	1	0
	10	0.46	0.0077	0.0078	0.0013	25.5	1	0
	11	0.015	0.0077	0.0078	0.0013	25.5	10	0
	12	0.46	0.0077	0.0078	0.0013	25.5	10	0
	13	0.015	0.0077	0.0078	0.0013	25.5	1	31
	14	0.46	0.0077	0.0078	0.0013	25.5	1	31
	15	0.015	0.0077	0.0078	0.0013	25.5	10	31
	16	0.46	0.0077	0.0078	0.0013	25.5	10	31
3	17	0.24	0.0015	0.0078	0.0013	1	5.5	0
	18	0.24	0.014	0.0078	0.0013	1	5.5	0
	19	0.24	0.0015	0.0078	0.0013	50	5.5	0
	20	0.24	0.014	0.0078	0.0013	50	5.5	0
	21	0.24	0.0015	0.0078	0.0013	1	5.5	31

	22	0.24	0.014	0.0078	0.0013	1	5.5	31
	23	0.24	0.0015	0.0078	0.0013	50	5.5	31
	24	0.24	0.014	0.0078	0.0013	50	5.5	31
4	25	0.015	0.0015	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5
	26	0.46	0.0015	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5
	27	0.015	0.014	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5
	28	0.46	0.014	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5
	29	0.015	0.0015	0.0078	0.0026	25.5	5.5	15.5
	30	0.46	0.0015	0.0078	0.0026	25.5	5.5	15.5
	31	0.015	0.014	0.0078	0.0026	25.5	5.5	15.5
	32	0.46	0.014	0.0078	0.0026	25.5	5.5	15.5
5	33	0.24	0.0077	0.0055	$1.3 \cdot 10^{-5}$	25.5	5.5	0
	34	0.24	0.0077	0.01	$1.3 \cdot 10^{-5}$	25.5	5.5	0
	35	0.24	0.0077	0.0055	0.0026	25.5	5.5	0
	36	0.24	0.0077	0.01	0.0026	25.5	5.5	0
	37	0.24	0.0077	0.0055	$1.3 \cdot 10^{-5}$	25.5	5.5	31
	38	0.24	0.0077	0.01	$1.3 \cdot 10^{-5}$	25.5	5.5	31
	39	0.24	0.0077	0.0055	0.0026	25.5	5.5	31
	40	0.24	0.0077	0.01	0.0026	25.5	5.5	31
6	41	0.015	0.0077	0.0055	0.0013	1	5.5	15.5
	42	0.46	0.0077	0.0055	0.0013	1	5.5	15.5
	43	0.015	0.0077	0.01	0.0013	1	5.5	15.5
	44	0.46	0.0077	0.01	0.0013	1	5.5	15.5
	45	0.015	0.0077	0.0055	0.0013	50	5.5	15.5
	46	0.46	0.0077	0.0055	0.0013	50	5.5	15.5
	47	0.015	0.0077	0.01	0.0013	50	5.5	15.5
	48	0.46	0.0077	0.01	0.0013	50	5.5	15.5
7	49	0.24	0.0015	0.0055	0.0013	25.5	1	15.5
	50	0.24	0.014	0.0055	0.0013	25.5	1	15.5
	51	0.24	0.0015	0.01	0.0013	25.5	1	15.5
	52	0.24	0.014	0.01	0.0013	25.5	1	15.5
	53	0.24	0.0015	0.0055	0.0013	25.5	10	15.5
	54	0.24	0.014	0.0055	0.0013	25.5	10	15.5
	55	0.24	0.0015	0.01	0.0013	25.5	10	15.5
	56	0.24	0.014	0.01	0.0013	25.5	10	15.5
8	57	0.24	0.0077	0.0078	0.0013	25.5	5.5	15.5
	58	0.015	0.0077	0.0078	0.0013	25.5	5.5	15.5
	59	0.46	0.0077	0.0078	0.0013	25.5	5.5	15.5
	60	0.24	0.0015	0.0078	0.0013	25.5	5.5	15.5
	61	0.24	0.014	0.0078	0.0013	25.5	5.5	15.5
	62	0.24	0.0077	0.0055	0.0013	25.5	5.5	15.5
	63	0.24	0.0077	0.01	0.0013	25.5	5.5	15.5
	64	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5
	65	0.24	0.0077	0.0078	0.0026	25.5	5.5	15.5
	66	0.24	0.0077	0.0078	0.0013	1	5.5	15.5
	67	0.24	0.0077	0.0078	0.0013	50	5.5	15.5
	68	0.24	0.0077	0.0078	0.0013	25.5	1	15.5
	69	0.24	0.0077	0.0078	0.0013	25.5	10	15.5
	70	0.24	0.0077	0.0078	0.0013	25.5	5.5	0
	71	0.24	0.0077	0.0078	0.0013	25.5	5.5	31
	72	0.24	0.0077	0.0078	0.0013	25.5	1	0

Table 7 Dimensionless parameter values for the Central Composite Design scenarios

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number (the dimensionless decay rate).

CCD Scenario	X^*	Z_w^*	Q^*	i	m	R	Da
73	0.11	0.0046	0.00066	0.0067	13.25	3.25	23.25
74	0.11	0.0046	0.00066	0.0089	13.25	3.25	7.75
75	0.11	0.0046	0.00066	0.0067	37.75	3.25	7.75
76	0.11	0.0046	0.00066	0.0089	37.75	3.25	23.25
77	0.11	0.0046	0.00066	0.0067	13.25	7.75	7.75
78	0.11	0.0046	0.00066	0.0089	13.25	7.75	23.25
79	0.11	0.0046	0.00066	0.0067	37.75	7.75	23.25
80	0.11	0.0046	0.00066	0.0089	37.75	7.75	7.75
81	0.11	0.0046	0.00195	0.0067	13.25	3.25	7.75
82	0.11	0.0046	0.00195	0.0089	13.25	3.25	23.25
83	0.11	0.0046	0.00195	0.0067	37.75	3.25	23.25
84	0.11	0.0046	0.00195	0.0089	37.75	3.25	7.75
85	0.11	0.0046	0.00195	0.0067	13.25	7.75	23.25
86	0.11	0.0046	0.00195	0.0089	13.25	7.75	7.75
87	0.11	0.0046	0.00195	0.0067	37.75	7.75	7.75
88	0.11	0.0046	0.00195	0.0089	37.75	7.75	23.25
89	0.11	0.011	0.00066	0.0067	13.25	3.25	7.75
90	0.11	0.011	0.00066	0.0089	13.25	3.25	23.25
91	0.11	0.011	0.00066	0.0067	37.75	3.25	23.25
92	0.11	0.011	0.00066	0.0089	37.75	3.25	7.75
93	0.11	0.011	0.00066	0.0067	13.25	7.75	23.25
94	0.11	0.011	0.00066	0.0089	13.25	7.75	7.75
95	0.11	0.011	0.00066	0.0067	37.75	7.75	7.75
96	0.11	0.011	0.00066	0.0089	37.75	7.75	23.25
97	0.11	0.011	0.00195	0.0067	13.25	3.25	23.25
98	0.11	0.011	0.00195	0.0089	13.25	3.25	7.75
99	0.11	0.011	0.00195	0.0067	37.75	3.25	7.75
100	0.11	0.011	0.00195	0.0089	37.75	3.25	23.25
101	0.11	0.011	0.00195	0.0067	13.25	7.75	7.75
102	0.11	0.011	0.00195	0.0089	13.25	7.75	23.25
103	0.11	0.011	0.00195	0.0067	37.75	7.75	23.25
104	0.11	0.011	0.00195	0.0089	37.75	7.75	7.75
105	0.35	0.0046	0.00066	0.0067	13.25	3.25	7.75
106	0.35	0.0046	0.00066	0.0089	13.25	3.25	23.25
107	0.35	0.0046	0.00066	0.0067	37.75	3.25	23.25
108	0.35	0.0046	0.00066	0.0089	37.75	3.25	7.75
109	0.35	0.0046	0.00066	0.0067	13.25	7.75	23.25
110	0.35	0.0046	0.00066	0.0089	13.25	7.75	7.75
111	0.35	0.0046	0.00066	0.0067	37.75	7.75	7.75
112	0.35	0.0046	0.00066	0.0089	37.75	7.75	23.25
113	0.35	0.0046	0.00195	0.0067	13.25	3.25	23.25
114	0.35	0.0046	0.00195	0.0089	13.25	3.25	7.75
115	0.35	0.0046	0.00195	0.0067	37.75	3.25	7.75
116	0.35	0.0046	0.00195	0.0089	37.75	3.25	23.25
117	0.35	0.0046	0.00195	0.0067	13.25	7.75	7.75
118	0.35	0.0046	0.00195	0.0089	13.25	7.75	23.25
119	0.35	0.0046	0.00195	0.0067	37.75	7.75	23.25
120	0.35	0.0046	0.00195	0.0089	37.75	7.75	7.75
121	0.35	0.011	0.00066	0.0067	13.25	3.25	23.25
122	0.35	0.011	0.00066	0.0089	13.25	3.25	7.75

123	0.35	0.011	0.00066	0.0067	37.75	3.25	7.75
124	0.35	0.011	0.00066	0.0089	37.75	3.25	23.25
125	0.35	0.011	0.00066	0.0067	13.25	7.75	7.75
126	0.35	0.011	0.00066	0.0089	13.25	7.75	23.25
127	0.35	0.011	0.00066	0.0067	37.75	7.75	23.25
128	0.35	0.011	0.00066	0.0089	37.75	7.75	7.75
129	0.35	0.011	0.00195	0.0067	13.25	3.25	7.75
130	0.35	0.011	0.00195	0.0089	13.25	3.25	23.25
131	0.35	0.011	0.00195	0.0067	37.75	3.25	23.25
132	0.35	0.011	0.00195	0.0089	37.75	3.25	7.75
133	0.35	0.011	0.00195	0.0067	13.25	7.75	23.25
134	0.35	0.011	0.00195	0.0089	13.25	7.75	7.75
135	0.35	0.011	0.00195	0.0067	37.75	7.75	7.75
136	0.35	0.011	0.00195	0.0089	37.75	7.75	23.25

2.5 Multivariate regression analysis

To develop the empirical formula, the data obtained from PMWIN were analysed using modelling software R. The R Studio interface, version 0.99.491, was used (©2009-2015, RStudio Inc.). The complete R-script can be found in appendix C.

First, a data frame was created using the information from tables 6 and 7. Then a multivariate regression was fit through the different combinations of outputs and all the parameters. So the influence of the dimensionless horizontal distance between the well and the plume (X^*), the dimensionless well screen length (Z_w^*), the hydraulic gradient (i), the dimensionless pumping rate (Q^*), the anisotropy (m), the retardation (R), and the dimensionless decay rate (Da), the square of each of these parameters and their multiplications on C^*_{max} , C^*_{half} , t^*_{max} and t^*_{half} was analysed. This led to a formula in which all factors, their squares and their combinations were included. However, not all parameters in the formulas were significant. By applying AIC (Akaike Information Criterion), the significance of different parameters was tested (Akaike, 1974) to develop the least parsimonious formula. This way a different formula was developed for C^*_{max} , C^*_{half} , t^*_{max} and t^*_{half} . Because Q^* and m had a large range of variation, they were transformed on a natural logarithmic scale. For the same reason C^*_{max} , C^*_{half} , t^*_{max} and t^*_{half} were transformed on a natural logarithmic and square root scale. Because the lowest value for Da was zero, Da was transformed on a root scale instead of a natural logarithmic scale.

The resulting empirical formulas were tested by comparing their results with the PMWIN results and by making a Bland-Altman plot. A Bland-Altman plot is a plot of the difference between two methods (PMWIN points minus empirical formula points) against the average of two methods (PMWIN points plus empirical formula divided by 2) (Bland and Altman, 1986). Furthermore, a sensitivity analysis was carried out by varying the factors one by one within their range. C^*_{max} , C^*_{half} , t^*_{max} and t^*_{half} were the most sensitive to the parameter that lead to a large variation in C^*_{max} , C^*_{half} , t^*_{max} and t^*_{half} .

The empirical formulas are only applicable in situations that fit within the parameter space. Extrapolation is not advisable since the predictions are faulty.

The empirical formulas outcomes were also tested with a 3D analytical transport equation. However since this analytical solution did not take wells into account some adjustments had to be made. The empirical formulas were extrapolated to situations where the pumping rate was very small and there was no anisotropy.

The analytical solution was set up for an instantaneous point source and included advection, dispersion, and decay (eq. 9) (adjusted from Bear, 1988). Initially there was no contaminant present in the soil. At t=0 a mass M_0 was added at $x=0$, $y=0$ and $z=0$. In this case $M_0=v_x A C_0$.

$$C = \frac{v_x A C_0}{n R \sqrt{D_x D_y D_z} \left(\frac{4\pi t}{R}\right)^{\frac{3}{2}}} \exp\left(-\frac{\left(x - \frac{v_x t}{R}\right)^2}{\frac{4D_x t}{R}} - \frac{y^2}{\frac{4D_y t}{R}} - \frac{z^2}{\frac{4D_z t}{R}} - \frac{\mu t}{R}\right) \quad (\text{eq. 9})$$

Where $D_x \approx \alpha_L v_x$, $D_y \approx \alpha_L v_y$ and $D_z \approx \alpha_L v_z$ assuming that dispersion is much larger than diffusion.

Since there was no well there was only groundwater flow in the x-direction, and thus only transport in the x-direction, the y and z terms can be removed. Dispersion on the other hand does happen in y and z-direction.

In order to compare the empirical formula and the analytical solution the analytical solution was made dimensionless (eq.10) with the conversions from table 3.

$$C^* = \frac{\frac{v_x^* A}{L^2}}{n R \sqrt{D_x^* D_y^* D_z^*} \left(\frac{4\pi t^*}{R}\right)^{\frac{3}{2}}} \exp\left(-\frac{\left(X^* - \frac{v_x^* t^*}{R}\right)^2}{\frac{4D_x^* t^*}{R}} - \frac{Da t^*}{R}\right) \quad (\text{eq. 10})$$

Where $D_x = \alpha_L^* v_x^*$, $D_y = \alpha_T^* v_y^*$ and $D_z = \alpha_T^* v_z^*$

By using t^*_{\max} (from the extrapolated empirical formula) as input time, C^*_{\max} was calculated at a given distance X^* . Taking the natural logarithm of this result allowed comparison between the empirical formula and the analytical solution. The analytical solution would never fit the empirical equation due to the simplifications and extrapolation applied.

3 Results

3.1 PMWIN

Breakthrough curves for the different test scenarios are shown in figure 8. The maximum concentration and half the maximum concentration are shown in tables 8, 9 and 10. The time when the above-mentioned concentrations were reached also was noted.

The first three test scenarios show that it takes longer for the plume to reach 6000 m than it takes to reach 2000 and 100 m. After arriving the concentrations increase to a maximum concentration of 1. At 100 m the concentration rapidly increases to 1 while at a larger distance it takes more time.

Table 8 Concentrations and arrival times from the test scenarios

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor, Da is the Damkohler number (the dimensionless decay rate), C^*_{half} is half the maximum concentration, t^*_{half} is the arrival time, C^*_{max} is the maximum concentration and t^*_{max} is the time the maximum concentration is reached.

Test scenario	X^*	Z_w^*	i	Q^*	m	R	Da	C^*_{half}	t^*_{half}	C^*_{max}	t^*_{max}
1	0.015	0	0.001	0	1	1	0	0.50	0.23	1.00	1.00
2	0.31	0	0.001	0	1	1	0	0.50	5.67	1.00	16.42
3	0.92	0	0.001	0	1	1	0	0.48	17.49	0.98	28.06
4	0.31	1	0.001	0.00032	1	1	0	$3.29 \cdot 10^{-2}$	5.13	$6.58 \cdot 10^{-2}$	8.55
5	0.015	1	0.001	0.00032	1	1	0	$9.17 \cdot 10^{-2}$	0.35	$1.83 \cdot 10^{-1}$	4.42
6	0.92	1	0.001	0.00032	1	1	0	$1.47 \cdot 10^{-2}$	15.45	$2.94 \cdot 10^{-2}$	20.06
7	0.31	1	0.01	0.00032	1	1	0	$2.61 \cdot 10^{-2}$	0.65	$5.21 \cdot 10^{-2}$	2.06
8	0.31	1	0.001	$1.3 \cdot 10^{-5}$	1	1	0	$1.46 \cdot 10^{-2}$	5.50	$2.92 \cdot 10^{-2}$	8.85
9	0.31	1	0.001	0.0026	1	1	0	$3.56 \cdot 10^{-2}$	4.46	$7.12 \cdot 10^{-2}$	7.96

The analytical solution from the website from the University of Illinois (Valocchi et al., 2016) (figure 8) nearly coincides with the PMWIN results. Since the analytical solution required dimensional input the PMWIN outcomes were converted to dimensional outcomes. When the outcomes of both methods (roughly) coincide, it means that the dimensionless model outcome coincides with the dimensional analytical solution input and therefore the conversion is correct.

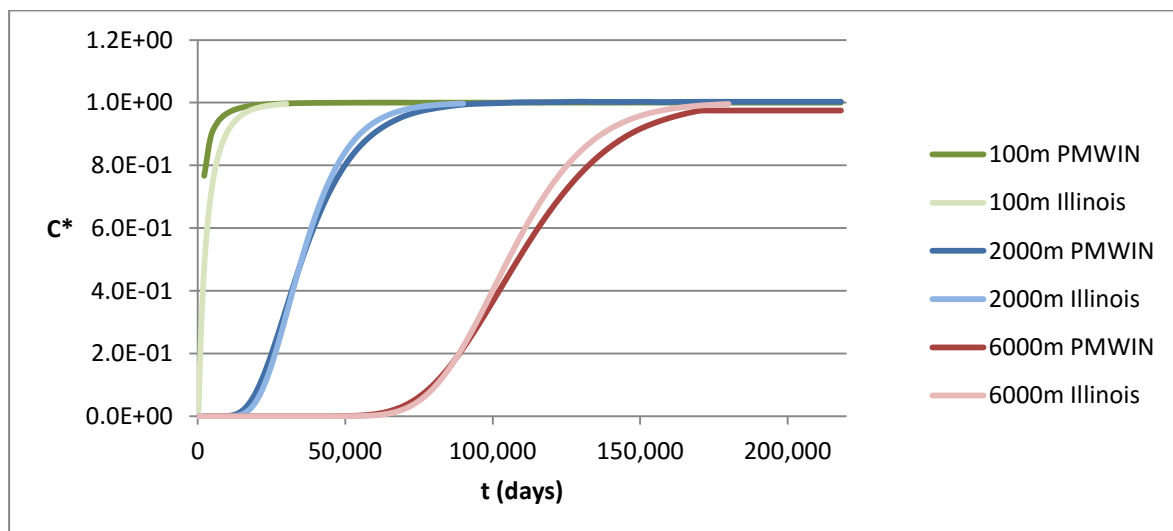


Figure 8 The breakthrough curves of scenarios 1 to 3 compared to the analytical solution

The lines denoted with PMWIN are the breakthrough curves modelled with PMWIN while the lines denoted with Illinois are the breakthrough curves calculated by the analytical solution of the University of Illinois (Valocchi et al., 2016).

In scenarios 4 to 9 the larger the horizontal distance between the plume and the well, the lower the maximum concentration was and the more elongated the plume got (figure 9, scenarios 4, 5 and 6). An increase in the hydraulic gradient led to an earlier arrival but to a small concentration drop as well (scenario 7). When the pumping rate was increased the plume arrived slightly earlier at the well and the maximum concentration was slightly higher, while a decrease of the pumping rate led to a slightly later arrival time and a slight drop in maximum concentration (scenarios 8 and 9).

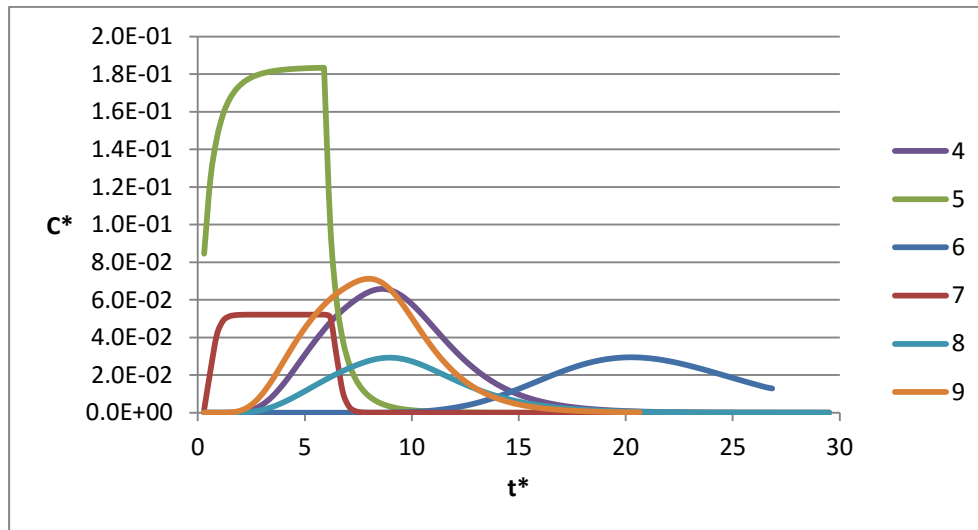


Figure 9 Breakthrough curves of test scenarios 4 to 9

In scenarios 4, 5 and 6 the horizontal distance between the well and the plume was varied. In scenario 7 the hydraulic gradient was increased and in scenarios 8 and 9 the pumping rate was varied.

Overall, in the BB scenarios, an increase in distance always led to an increase in arrival time and a decrease in concentration. A change in well screen length did not show a clear effect on arrival time and concentration. An increase in hydraulic gradient always led to a decrease in arrival time and sometimes to an increase in concentration depending on the other variables. An increase in pumping rate always led to a decrease in arrival time and almost always to an increase in concentration. A change in anisotropy did not led to a clear change in arrival time or concentration. An increase in retardation always led to an increase in arrival time. Finally an increase in decay rate led to a decrease in arrival time and concentration

The effect of the parameters on arrival time and concentration in the CCD was less clear since all parameters were varied at the same time and the variation was not as structured as in the BB (table 10). Variations in well screen length, hydraulic gradient and anisotropy did not led to a clear change in arrival time or concentration. Variations in distance led to an increase in arrival time and a decrease in concentration while an increase in pumping rate led to an opposite pattern. When the pumping rate was increased there was a decrease in arrival time and an increase in concentration. An increase in retardation led to an increase in arrival time and an increase in decay rate led to a decrease in arrival time and concentration.

Table 9 Concentrations and arrival times from the Box-Behnken scenarios

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor, Da is the Damkohler number (the dimensionless decay rate), C^*_{half} is half the maximum concentration, t^*_{half} is the arrival time, C^*_{max} is the maximum concentration and t^*_{max} is the time the maximum concentration is reached. The bold values are varied between their minimum and maximum values in that particular box.

Block	BB Scenario	X^*	Z_w^*	i	Q^*	m	R	Da	C^*_{half}	t^*_{half}	C^*_{max}	t^*_{max}
1	1	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	1	1	15.5	$6.77 \cdot 10^{-6}$	0.29	$1.35 \cdot 10^{-5}$	0.74
	2	0.24	0.0077	0.0078	0.0026	1	1	15.5	$8.31 \cdot 10^{-5}$	0.25	$1.66 \cdot 10^{-4}$	0.69

	3	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	50	1	15.5	$5.27 \cdot 10^{-6}$	0.29	$1.05 \cdot 10^{-5}$	0.74	
	4	0.24	0.0077	0.0078	0.0026	50	1	15.5	$4.40 \cdot 10^{-5}$	0.24	$8.80 \cdot 10^{-5}$	0.69	
	5	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	1	10	15.5	$4.09 \cdot 10^{-6}$	1.02	$8.18 \cdot 10^{-6}$	1.47	
	6	0.24	0.0077	0.0078	0.0026	1	10	15.5	$5.45 \cdot 10^{-5}$	0.94	$1.09 \cdot 10^{-4}$	1.33	
	7	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	50	10	15.5	$3.18 \cdot 10^{-6}$	1.02	$6.36 \cdot 10^{-6}$	1.47	
	8	0.24	0.0077	0.0078	0.0026	50	10	15.5	$2.86 \cdot 10^{-5}$	0.88	$5.72 \cdot 10^{-5}$	1.27	
	2	9	0.015	0.0077	0.0078	0.0013	25.5	1	0	$1.01 \cdot 10^{-2}$	0.11	$2.03 \cdot 10^{-2}$	0.58
		10	0.46	0.0077	0.0078	0.0013	25.5	1	0	$1.10 \cdot 10^{-2}$	0.97	$2.19 \cdot 10^{-2}$	1.38
11		0.015	0.0077	0.0078	0.0013	25.5	10	0	$6.19 \cdot 10^{-3}$	0.25	$1.24 \cdot 10^{-2}$	0.59	
12		0.46	0.0077	0.0078	0.0013	25.5	10	0	$5.70 \cdot 10^{-4}$	3.13	$1.14 \cdot 10^{-3}$	4.41	
13		0.015	0.0077	0.0078	0.0013	25.5	1	31	$1.53 \cdot 10^{-3}$	0.03	$3.06 \cdot 10^{-3}$	0.21	
14		0.46	0.0077	0.0078	0.0013	25.5	1	31	$2.58 \cdot 10^{-10}$	0.73	$5.16 \cdot 10^{-10}$	0.85	
15		0.015	0.0077	0.0078	0.0013	25.5	10	31	$1.52 \cdot 10^{-3}$	0.11	$3.04 \cdot 10^{-3}$	0.58	
16		0.46	0.0077	0.0078	0.0013	25.5	10	31	$2.25 \cdot 10^{-10}$	1.31	$4.51 \cdot 10^{-10}$	1.75	
3	17	0.24	0.0015	0.0078	0.0013	1	5.5	0	$1.46 \cdot 10^{-2}$	1.05	$2.92 \cdot 10^{-2}$	1.59	
	18	0.24	0.014	0.0078	0.0013	1	5.5	0	$1.37 \cdot 10^{-2}$	1.06	$2.73 \cdot 10^{-2}$	1.59	
	19	0.24	0.0015	0.0078	0.0013	50	5.5	0	$6.63 \cdot 10^{-3}$	1.07	$1.33 \cdot 10^{-2}$	1.68	
	20	0.24	0.014	0.0078	0.0013	50	5.5	0	$6.06 \cdot 10^{-3}$	1.09	$1.21 \cdot 10^{-2}$	1.65	
	21	0.24	0.0015	0.0078	0.0013	1	5.5	31	$1.83 \cdot 10^{-6}$	0.61	$3.65 \cdot 10^{-6}$	0.85	
	22	0.24	0.014	0.0078	0.0013	1	5.5	31	$1.73 \cdot 10^{-6}$	0.60	$3.47 \cdot 10^{-6}$	0.85	
	23	0.24	0.0015	0.0078	0.0013	50	5.5	31	$7.53 \cdot 10^{-7}$	0.63	$1.51 \cdot 10^{-6}$	0.85	
	24	0.24	0.014	0.0078	0.0013	50	5.5	31	$7.53 \cdot 10^{-7}$	0.60	$1.51 \cdot 10^{-6}$	0.85	
4	25	0.015	0.0015	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5	$5.30 \cdot 10^{-4}$	0.13	$1.06 \cdot 10^{-3}$	0.58	
	26	0.46	0.0015	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5	$4.52 \cdot 10^{-8}$	1.17	$9.04 \cdot 10^{-8}$	1.59	
	27	0.015	0.014	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5	$1.34 \cdot 10^{-4}$	0.14	$2.68 \cdot 10^{-4}$	0.58	
	28	0.46	0.014	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5	$3.14 \cdot 10^{-8}$	1.19	$6.28 \cdot 10^{-8}$	1.59	
	29	0.015	0.0015	0.0078	0.0026	25.5	5.5	15.5	$6.14 \cdot 10^{-3}$	0.08	$1.23 \cdot 10^{-2}$	0.48	
	30	0.46	0.0015	0.0078	0.0026	25.5	5.5	15.5	$2.48 \cdot 10^{-7}$	1.13	$4.96 \cdot 10^{-7}$	1.49	
	31	0.015	0.014	0.0078	0.0026	25.5	5.5	15.5	$4.86 \cdot 10^{-5}$	0.09	$9.72 \cdot 10^{-5}$	0.58	
	32	0.46	0.014	0.0078	0.0026	25.5	5.5	15.5	$5.34 \cdot 10^{-10}$	1.19	$1.07 \cdot 10^{-9}$	1.59	
5	33	0.24	0.0077	0.0055	$1.3 \cdot 10^{-5}$	25.5	5.5	0	$1.36 \cdot 10^{-3}$	1.53	$2.71 \cdot 10^{-3}$	2.36	
	34	0.24	0.0077	0.01	$1.3 \cdot 10^{-5}$	25.5	5.5	0	$2.48 \cdot 10^{-3}$	1.01	$4.96 \cdot 10^{-3}$	1.59	
	35	0.24	0.0077	0.0055	0.0026	25.5	5.5	0	$5.47 \cdot 10^{-3}$	1.37	$1.09 \cdot 10^{-2}$	2.12	
	36	0.24	0.0077	0.01	0.0026	25.5	5.5	0	$9.53 \cdot 10^{-3}$	0.97	$1.91 \cdot 10^{-2}$	1.42	
	37	0.24	0.0077	0.0055	$1.3 \cdot 10^{-5}$	25.5	5.5	31	$9.91 \cdot 10^{-9}$	0.78	$1.98 \cdot 10^{-8}$	1.00	
	38	0.24	0.0077	0.01	$1.3 \cdot 10^{-5}$	25.5	5.5	31	$4.58 \cdot 10^{-7}$	0.48	$9.17 \cdot 10^{-7}$	0.83	
	39	0.24	0.0077	0.0055	0.0026	25.5	5.5	31	$2.25 \cdot 10^{-7}$	0.66	$4.50 \cdot 10^{-7}$	0.88	
	40	0.24	0.0077	0.01	0.0026	25.5	5.5	31	$4.04 \cdot 10^{-6}$	0.41	$8.08 \cdot 10^{-6}$	0.83	
6	41	0.015	0.0077	0.0055	0.0013	1	5.5	15.5	$3.08 \cdot 10^{-2}$	0.08	$6.16 \cdot 10^{-2}$	0.58	
	42	0.46	0.0077	0.0055	0.0013	1	5.5	15.5	$2.10 \cdot 10^{-8}$	1.32	$4.19 \cdot 10^{-8}$	1.75	
	43	0.015	0.0077	0.01	0.0013	1	5.5	15.5	$3.66 \cdot 10^{-2}$	0.06	$7.32 \cdot 10^{-2}$	0.59	
	44	0.46	0.0077	0.01	0.0013	1	5.5	15.5	$1.65 \cdot 10^{-6}$	0.99	$3.30 \cdot 10^{-6}$	1.38	
	45	0.015	0.0077	0.0055	0.0013	50	5.5	15.5	$1.11 \cdot 10^{-3}$	0.12	$2.22 \cdot 10^{-3}$	0.58	
	46	0.46	0.0077	0.0055	0.0013	50	5.5	15.5	$1.30 \cdot 10^{-8}$	1.29	$2.60 \cdot 10^{-8}$	1.70	
	47	0.015	0.0077	0.01	0.0013	50	5.5	15.5	$1.38 \cdot 10^{-3}$	0.11	$2.77 \cdot 10^{-3}$	0.58	
	48	0.46	0.0077	0.01	0.0013	50	5.5	15.5	$8.43 \cdot 10^{-7}$	0.99	$1.69 \cdot 10^{-6}$	1.38	
7	49	0.24	0.0015	0.0055	0.0013	25.5	1	15.5	$9.01 \cdot 10^{-6}$	0.48	$1.80 \cdot 10^{-5}$	0.74	
	50	0.24	0.014	0.0055	0.0013	25.5	1	15.5	$8.72 \cdot 10^{-6}$	0.29	$1.74 \cdot 10^{-5}$	0.74	
	51	0.24	0.0015	0.01	0.0013	25.5	1	15.5	$8.27 \cdot 10^{-5}$	0.23	$1.65 \cdot 10^{-4}$	0.69	
	52	0.24	0.014	0.01	0.0013	25.5	1	15.5	$7.72 \cdot 10^{-5}$	0.23	$1.54 \cdot 10^{-4}$	0.69	
	53	0.24	0.0015	0.0055	0.0013	25.5	10	15.5	$1.12 \cdot 10^{-5}$	1.10	$2.24 \cdot 10^{-5}$	1.54	
	54	0.24	0.014	0.0055	0.0013	25.5	10	15.5	$5.20 \cdot 10^{-6}$	1.09	$1.04 \cdot 10^{-5}$	1.54	
	55	0.24	0.0015	0.01	0.0013	25.5	10	15.5	$5.65 \cdot 10^{-5}$	0.85	$1.13 \cdot 10^{-4}$	1.22	
	56	0.24	0.014	0.01	0.0013	25.5	10	15.5	$5.13 \cdot 10^{-5}$	0.85	$1.03 \cdot 10^{-4}$	1.22	
8	57	0.24	0.0077	0.0078	0.0013	25.5	5.5	15.5	$2.73 \cdot 10^{-5}$	0.62	$5.46 \cdot 10^{-5}$	0.96	

58	0.015	0.0077	0.0078	0.0013	25.5	5.5	15.5	$2.85 \cdot 10^{-3}$	0.10	$5.70 \cdot 10^{-3}$	0.59
59	0.46	0.0077	0.0078	0.0013	25.5	5.5	15.5	$1.64 \cdot 10^{-7}$	1.12	$3.28 \cdot 10^{-7}$	1.54
60	0.24	0.0015	0.0078	0.0013	25.5	5.5	15.5	$3.13 \cdot 10^{-5}$	0.64	$6.27 \cdot 10^{-5}$	1.01
61	0.24	0.014	0.0078	0.0013	25.5	5.5	15.5	$2.91 \cdot 10^{-5}$	0.62	$5.83 \cdot 10^{-5}$	0.96
62	0.24	0.0077	0.0055	0.0013	25.5	5.5	15.5	$7.21 \cdot 10^{-6}$	0.71	$1.44 \cdot 10^{-5}$	1.06
63	0.24	0.0077	0.01	0.0013	25.5	5.5	15.5	$6.46 \cdot 10^{-5}$	0.57	$1.29 \cdot 10^{-4}$	0.90
64	0.24	0.0077	0.0078	$1.3 \cdot 10^{-5}$	25.5	5.5	15.5	$4.69 \cdot 10^{-6}$	0.68	$9.38 \cdot 10^{-6}$	1.01
65	0.24	0.0077	0.0078	0.0026	25.5	5.5	15.5	$4.02 \cdot 10^{-5}$	0.61	$8.05 \cdot 10^{-5}$	0.96
66	0.24	0.0077	0.0078	0.0013	1	5.5	15.5	$6.31 \cdot 10^{-5}$	0.62	$1.26 \cdot 10^{-4}$	1.01
67	0.24	0.0077	0.0078	0.0013	50	5.5	15.5	$2.49 \cdot 10^{-5}$	0.62	$4.98 \cdot 10^{-5}$	0.96
68	0.24	0.0077	0.0078	0.0013	25.5	1	15.5	$3.10 \cdot 10^{-5}$	0.25	$6.20 \cdot 10^{-5}$	0.69
69	0.24	0.0077	0.0078	0.0013	25.5	10	15.5	$1.98 \cdot 10^{-5}$	0.94	$3.97 \cdot 10^{-5}$	1.38
70	0.24	0.0077	0.0078	0.0013	25.5	5.5	0	$6.69 \cdot 10^{-3}$	1.10	$1.34 \cdot 10^{-2}$	1.70
71	0.24	0.0077	0.0078	0.0013	25.5	5.5	31	$7.52 \cdot 10^{-7}$	0.60	$1.50 \cdot 10^{-6}$	0.85
72	0.24	0.0077	0.0078	0.0013	25.5	1	0	$1.47 \cdot 10^{-2}$	0.56	$2.93 \cdot 10^{-2}$	0.90

Table 10 Concentrations and arrival times from the Central Composite Design scenarios

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor, Da is the Damkohler number (the dimensionless decay rate), C^*_{half} is half the maximum concentration, t^*_{half} is the arrival time, C^*_{max} is the maximum concentration and t^*_{max} is the time the maximum concentration is reached.

CCD Scenario	X^*	Z_w^*	Q^*	i	m	R	Da	C^*_{half}	t^*_{half}	C^*_{max}	t^*_{max}
73	0.11	0.0046	0.00066	0.0067	13.25	3.25	23.25	$1.29 \cdot 10^{-4}$	0.22	$2.58 \cdot 10^{-4}$	0.64
74	0.11	0.0046	0.00066	0.0089	13.25	3.25	7.75	$2.57 \cdot 10^{-3}$	0.31	$5.13 \cdot 10^{-3}$	0.69
75	0.11	0.0046	0.00066	0.0067	37.75	3.25	7.75	$8.18 \cdot 10^{-5}$	0.22	$1.64 \cdot 10^{-4}$	0.64
76	0.11	0.0046	0.00066	0.0089	37.75	3.25	23.25	$1.56 \cdot 10^{-4}$	0.20	$3.11 \cdot 10^{-4}$	0.64
77	0.11	0.0046	0.00066	0.0067	13.25	7.75	7.75	$1.31 \cdot 10^{-3}$	0.58	$2.62 \cdot 10^{-3}$	0.96
78	0.11	0.0046	0.00066	0.0089	13.25	7.75	23.25	$2.36 \cdot 10^{-4}$	0.37	$4.72 \cdot 10^{-4}$	0.74
79	0.11	0.0046	0.00066	0.0067	37.75	7.75	23.25	$7.81 \cdot 10^{-5}$	0.41	$1.56 \cdot 10^{-4}$	0.74
80	0.11	0.0046	0.00066	0.0089	37.75	7.75	7.75	$1.23 \cdot 10^{-3}$	0.52	$2.47 \cdot 10^{-3}$	0.85
81	0.11	0.0046	0.00195	0.0067	13.25	3.25	7.75	$2.81 \cdot 10^{-3}$	0.33	$5.63 \cdot 10^{-3}$	0.69
82	0.11	0.0046	0.00195	0.0089	13.25	3.25	23.25	$4.38 \cdot 10^{-4}$	0.19	$8.75 \cdot 10^{-4}$	0.64
83	0.11	0.0046	0.00195	0.0067	37.75	3.25	23.25	$1.70 \cdot 10^{-4}$	0.20	$3.40 \cdot 10^{-4}$	0.64
84	0.11	0.0046	0.00195	0.0089	37.75	3.25	7.75	$2.69 \cdot 10^{-3}$	0.29	$5.38 \cdot 10^{-3}$	0.64
85	0.11	0.0046	0.00195	0.0067	13.25	7.75	23.25	$2.22 \cdot 10^{-4}$	0.39	$4.44 \cdot 10^{-4}$	0.74
86	0.11	0.0046	0.00195	0.0089	13.25	7.75	7.75	$3.10 \cdot 10^{-3}$	0.48	$6.20 \cdot 10^{-3}$	0.85
87	0.11	0.0046	0.00195	0.0067	37.75	7.75	7.75	$1.39 \cdot 10^{-3}$	0.52	$2.78 \cdot 10^{-3}$	0.90
88	0.11	0.0046	0.00195	0.0089	37.75	7.75	23.25	$2.89 \cdot 10^{-4}$	0.34	$5.78 \cdot 10^{-4}$	0.69
89	0.11	0.011	0.00066	0.0067	13.25	3.25	7.75	$1.48 \cdot 10^{-3}$	0.36	$2.96 \cdot 10^{-3}$	0.69
90	0.11	0.011	0.00066	0.0089	13.25	3.25	23.25	$1.90 \cdot 10^{-4}$	0.21	$3.80 \cdot 10^{-4}$	0.64
91	0.11	0.011	0.00066	0.0067	37.75	3.25	23.25	$6.52 \cdot 10^{-5}$	0.23	$1.30 \cdot 10^{-4}$	0.64
92	0.11	0.011	0.00066	0.0089	37.75	3.25	7.75	$1.37 \cdot 10^{-3}$	0.32	$2.75 \cdot 10^{-3}$	0.69
93	0.11	0.011	0.00066	0.0067	13.25	7.75	23.25	$9.64 \cdot 10^{-5}$	0.43	$1.93 \cdot 10^{-4}$	0.74
94	0.11	0.011	0.00066	0.0089	13.25	7.75	7.75	$1.55 \cdot 10^{-3}$	0.52	$3.10 \cdot 10^{-3}$	0.85
95	0.11	0.011	0.00066	0.0067	37.75	7.75	7.75	$6.80 \cdot 10^{-4}$	0.59	$1.36 \cdot 10^{-3}$	0.96
96	0.11	0.011	0.00066	0.0089	37.75	7.75	23.25	$4.02 \cdot 10^{-5}$	0.38	$8.04 \cdot 10^{-5}$	0.74
97	0.11	0.011	0.00195	0.0067	13.25	3.25	23.25	$2.08 \cdot 10^{-4}$	0.21	$4.17 \cdot 10^{-4}$	0.64
98	0.11	0.011	0.00195	0.0089	13.25	3.25	7.75	$3.52 \cdot 10^{-3}$	0.30	$7.04 \cdot 10^{-3}$	0.69
99	0.11	0.011	0.00195	0.0067	37.75	3.25	7.75	$1.68 \cdot 10^{-3}$	0.33	$3.36 \cdot 10^{-3}$	0.69
100	0.11	0.011	0.00195	0.0089	37.75	3.25	23.25	$2.58 \cdot 10^{-4}$	0.19	$5.15 \cdot 10^{-4}$	0.64
101	0.11	0.011	0.00195	0.0067	13.25	7.75	7.75	$1.82 \cdot 10^{-3}$	0.55	$3.63 \cdot 10^{-3}$	0.90
102	0.11	0.011	0.00195	0.0089	13.25	7.75	23.25	$3.66 \cdot 10^{-4}$	0.35	$7.31 \cdot 10^{-4}$	0.69
103	0.11	0.011	0.00195	0.0067	37.75	7.75	23.25	$1.42 \cdot 10^{-4}$	0.38	$2.84 \cdot 10^{-4}$	0.74
104	0.11	0.011	0.00195	0.0089	37.75	7.75	7.75	$1.75 \cdot 10^{-3}$	0.49	$3.51 \cdot 10^{-3}$	0.85

105	0.35	0.0046	0.00066	0.0067	13.25	3.25	7.75	$3.31 \cdot 10^{-5}$	0.87	$6.62 \cdot 10^{-5}$	1.27
106	0.35	0.0046	0.00066	0.0089	13.25	3.25	23.25	$2.81 \cdot 10^{-7}$	0.66	$5.62 \cdot 10^{-7}$	0.90
107	0.35	0.0046	0.00066	0.0067	37.75	3.25	23.25	$3.29 \cdot 10^{-8}$	0.74	$6.59 \cdot 10^{-8}$	1.01
108	0.35	0.0046	0.00066	0.0089	37.75	3.25	7.75	$8.00 \cdot 10^{-5}$	0.76	$1.60 \cdot 10^{-4}$	1.11
109	0.35	0.0046	0.00066	0.0067	13.25	7.75	23.25	$3.43 \cdot 10^{-8}$	1.03	$6.87 \cdot 10^{-8}$	1.43
110	0.35	0.0046	0.00066	0.0089	13.25	7.75	7.75	$5.87 \cdot 10^{-5}$	1.30	$1.17 \cdot 10^{-4}$	1.75
111	0.35	0.0046	0.00066	0.0067	37.75	7.75	7.75	$1.59 \cdot 10^{-5}$	1.51	$3.17 \cdot 10^{-5}$	2.07
112	0.35	0.0046	0.00066	0.0089	37.75	7.75	23.25	$2.02 \cdot 10^{-7}$	0.90	$4.04 \cdot 10^{-7}$	1.27
113	0.35	0.0046	0.00195	0.0067	13.25	3.25	23.25	$7.31 \cdot 10^{-8}$	0.75	$1.46 \cdot 10^{-7}$	0.96
114	0.35	0.0046	0.00195	0.0089	13.25	3.25	7.75	$1.52 \cdot 10^{-4}$	0.74	$3.03 \cdot 10^{-4}$	1.11
115	0.35	0.0046	0.00195	0.0067	37.75	3.25	7.75	$4.63 \cdot 10^{-5}$	0.85	$9.26 \cdot 10^{-5}$	1.22
116	0.35	0.0046	0.00195	0.0089	37.75	3.25	23.25	$4.48 \cdot 10^{-7}$	0.67	$8.96 \cdot 10^{-7}$	0.90
117	0.35	0.0046	0.00195	0.0067	13.25	7.75	7.75	$3.14 \cdot 10^{-5}$	1.46	$6.27 \cdot 10^{-5}$	2.02
118	0.35	0.0046	0.00195	0.0089	13.25	7.75	23.25	$4.32 \cdot 10^{-7}$	0.90	$8.64 \cdot 10^{-7}$	1.27
119	0.35	0.0046	0.00195	0.0067	37.75	7.75	23.25	$6.31 \cdot 10^{-8}$	1.00	$1.26 \cdot 10^{-7}$	1.38
120	0.35	0.0046	0.00195	0.0089	37.75	7.75	7.75	$8.05 \cdot 10^{-5}$	1.25	$1.61 \cdot 10^{-4}$	1.75
121	0.35	0.011	0.00066	0.0067	13.25	3.25	23.25	$3.32 \cdot 10^{-8}$	0.74	$6.64 \cdot 10^{-8}$	1.01
122	0.35	0.011	0.00066	0.0089	13.25	3.25	7.75	$8.43 \cdot 10^{-5}$	0.76	$1.69 \cdot 10^{-4}$	1.11
123	0.35	0.011	0.00066	0.0067	37.75	3.25	7.75	$2.51 \cdot 10^{-5}$	0.87	$5.02 \cdot 10^{-5}$	1.27
124	0.35	0.011	0.00066	0.0089	37.75	3.25	23.25	$8.38 \cdot 10^{-10}$	1.03	$1.68 \cdot 10^{-9}$	1.27
125	0.35	0.011	0.00066	0.0067	13.25	7.75	7.75	$1.65 \cdot 10^{-5}$	1.52	$3.30 \cdot 10^{-5}$	2.07
126	0.35	0.011	0.00066	0.0089	13.25	7.75	23.25	$2.10 \cdot 10^{-7}$	0.89	$4.19 \cdot 10^{-7}$	1.27
127	0.35	0.011	0.00066	0.0067	37.75	7.75	23.25	$2.68 \cdot 10^{-8}$	1.01	$5.36 \cdot 10^{-8}$	1.38
128	0.35	0.011	0.00066	0.0089	37.75	7.75	7.75	$4.46 \cdot 10^{-5}$	1.30	$8.92 \cdot 10^{-5}$	1.81
129	0.35	0.011	0.00195	0.0067	13.25	3.25	7.75	$4.88 \cdot 10^{-5}$	0.85	$9.76 \cdot 10^{-5}$	1.22
130	0.35	0.011	0.00195	0.0089	13.25	3.25	23.25	$4.57 \cdot 10^{-7}$	0.65	$9.13 \cdot 10^{-7}$	0.90
131	0.35	0.011	0.00195	0.0067	37.75	3.25	23.25	$7.30 \cdot 10^{-8}$	0.72	$1.46 \cdot 10^{-7}$	0.96
132	0.35	0.011	0.00195	0.0089	37.75	3.25	7.75	$1.22 \cdot 10^{-4}$	0.73	$2.44 \cdot 10^{-4}$	1.11
133	0.35	0.011	0.00195	0.0067	13.25	7.75	23.25	$6.02 \cdot 10^{-8}$	1.01	$1.20 \cdot 10^{-7}$	1.38
134	0.35	0.011	0.00195	0.0089	13.25	7.75	7.75	$8.53 \cdot 10^{-5}$	1.27	$1.71 \cdot 10^{-4}$	1.77
135	0.35	0.011	0.00195	0.0067	37.75	7.75	7.75	$2.62 \cdot 10^{-5}$	1.43	$5.24 \cdot 10^{-5}$	1.96
136	0.35	0.011	0.00195	0.0089	37.75	7.75	23.25	$3.79 \cdot 10^{-7}$	0.87	$7.57 \cdot 10^{-7}$	1.27

3.2 Empirical formula

The $\text{Ln}(C_{\text{max}}^*)$ concentration is calculated with equation 11 and depends on X^* , i , $\text{Ln}(Q)$, $\text{Ln}(m)$, Da , X^{*2} , Z_W^{*2} , Da^2 , $X^* \cdot i$, $X^* \cdot \text{Ln}(m)$ and $X^* \cdot Da$.

$$\text{Ln}(C_{\text{max}}^*) = 1.252 - 18.55 X^* + 43.29 i + 0.3901 \text{Ln}(Q^*) - 0.7015 \text{Ln}(m) - 0.1943 Da - 9.33 X^{*2} - 4643 Z_W^{*2} + 0.005047 Da^2 + 1673 X^* i + 1.525 X^* \text{Ln}(m) - 1.061 X^* Da \quad (\text{eq. 11})$$

Comparing the $\text{Ln}(C_{\text{max}}^*)$ empirical formula predictions with the $\text{Ln}(C_{\text{max}}^*)$ PMWIN results leads to a 1:1 trend with some off points (figure 10A). The Bland-Altman plot reveals that the PMWIN results and the empirical equation results were in the same range since the mean difference between the methods was very close zero (there was no bias of one of the methods). The lines of agreement (dotted lines that are the mean difference $\pm 1.96 \times$ standard deviation) were quite narrow (mean difference ± 1.68) and only few points lie outside these lines which indicates little variation between the methods. Most points that fall outside the lines of agreement show an overestimation of $\text{Ln}(C_{\text{max}}^*)$ for the empirical formula compared to the PMWIN results. These outliers are not specifically linked to one or two parameter variations but appear to be random. However, the outliers in figure 10A are the same as the outliers in figure 10B. The data points were scattered evenly and the range of the difference between the two methods was consistent across the mean of

the two methods, which means that there was no difference between the methods dependant on the magnitude of $\text{Ln}(C^*_{\text{max}})$ (figure 10B).

Figure 11 and 12 show that an increase in X^* , $\text{Ln}(m)$ and Da leads to a decrease in $\text{Ln}(C^*_{\text{max}})$ while an increase in Z_W^* , Q^* and i leads to an increase in $\text{Ln}(C^*_{\text{max}})$. R has no effect on $\text{Ln}(C^*_{\text{max}})$. A change in most parameters mostly leads to a small change in $\text{Ln}(C^*_{\text{max}})$ ($\text{Ln}(C^*_{\text{max}}) \pm 2.3$), but a change in X^* and Da lead to a large change in $\text{Ln}(C^*_{\text{max}})$ ($\text{Ln}(C^*_{\text{max}}) \pm 6.9$). Therefore $\text{Ln}(C^*_{\text{max}})$ was most sensitive for X^* and Da .

The confidence interval (red dotted line in figure 12) gives the confidence of the position of the empirical formula while the prediction interval (green dotted line in figure 12) gives the range in which future predictions may fall. The prediction interval lies 1.8 (on a natural logarithmic scale) above and below the empirical formula, while the confidence interval has a much narrower range. This means that future observations will fall within the range of the empirical formula result ± 1.8 (on a natural logarithmic scale). While the PMWIN data points (dots in figure 11) do not always lie within the confidence interval they do lie within the prediction interval. This means that the range of deviation of the empirical formula does encompass the PMWIN points.

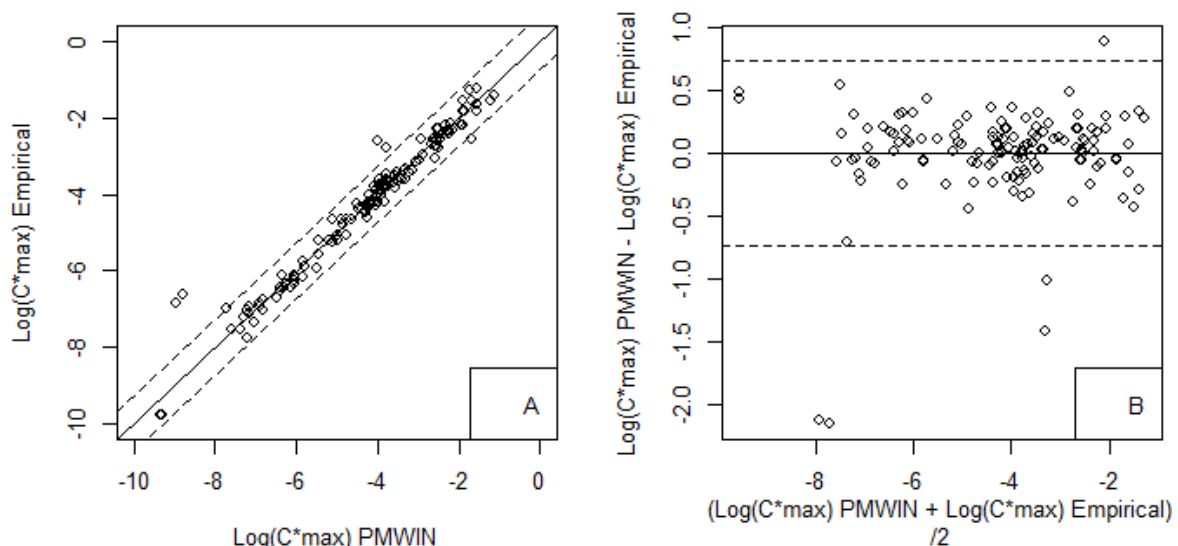


Figure 10 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{max}})$

A. The comparison between the PMWIN and empirical formula results for $\text{Log}(C^*_{\text{max}})$. The solid line is the $y=x$ line and the dotted line gives ± 1.96 standard deviation.

B. The Bland-Altman plot which shows the mean of the PMWIN and empirical formula results ($(\text{PMWIN result} + \text{empirical formula result})/2$) on the x-axis and the difference ($\text{PMWIN result} - \text{empirical formula result}$) on the y-axis. The solid line gives the mean difference and the dotted line gives ± 1.96 standard deviation.

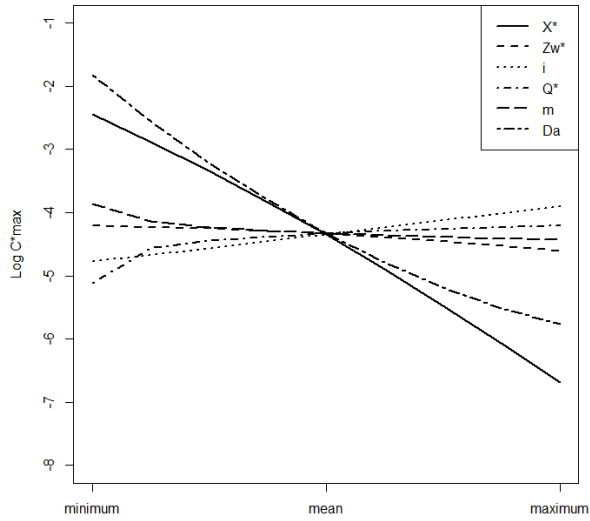


Figure 11 Sensitivity analysis for $\text{Log}(C^*_{\text{max}})$

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only one parameter was varied at a time (from its minimum to its maximum value), the rest was kept at their mean value (table 4). R did not have an influence on $\text{Log}(C^*_{\text{max}})$ and was therefore not included.

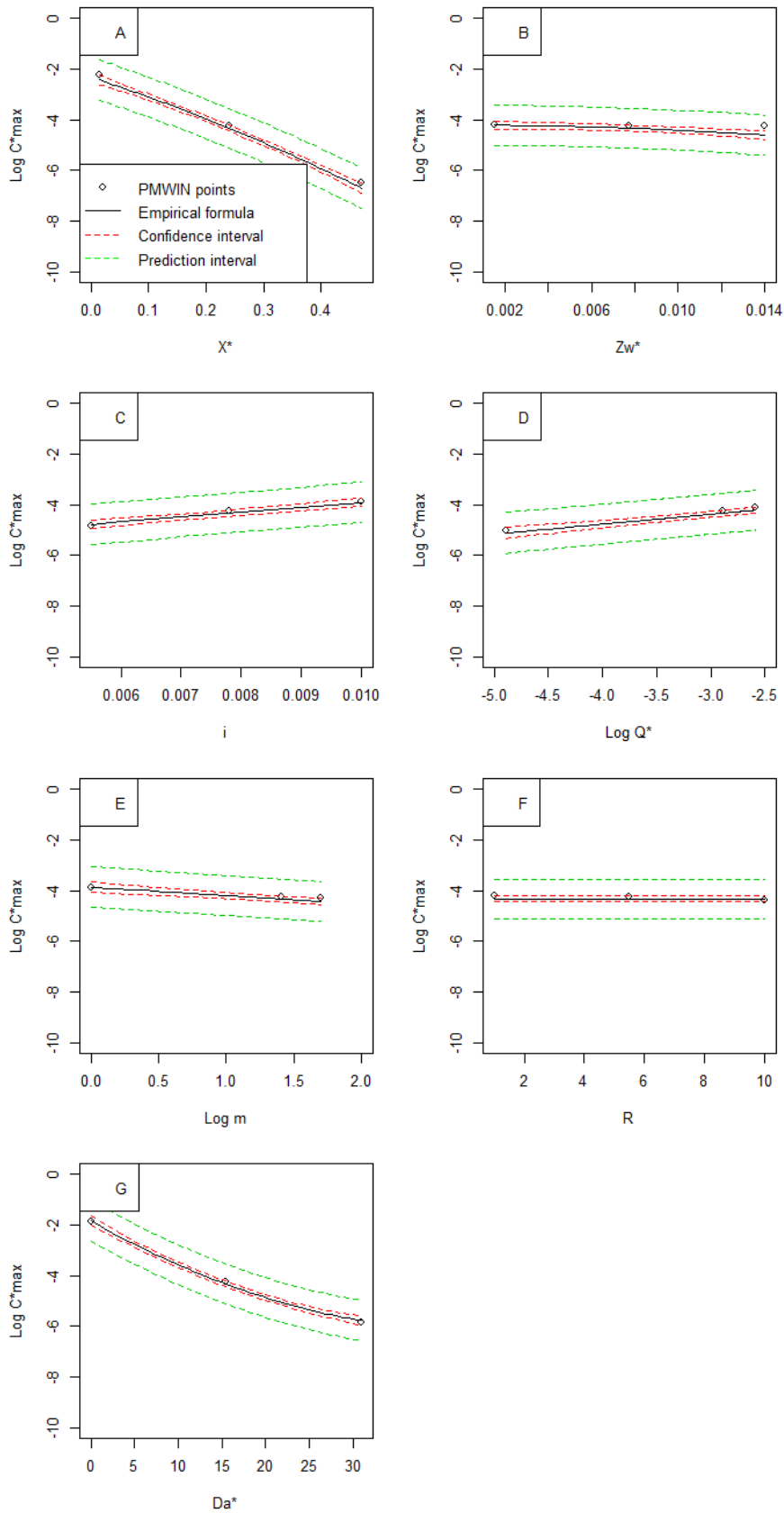


Figure 12 Influence of X^* , Z_w^* , i , $\text{Log}(Q^*)$, $\text{Log}(m)$, R and Da on $\text{Log}(C_{\max}^*)$

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only the parameter on the x-axis was varied, all other parameters were kept on their mean value (table 4). The solid line is the empirical formula, the red line the confidence interval and the green line the prediction interval. $C_{\max}^* = C_{\max}/C_0$

The vt_{\max}^* depends on X^* , i , $\ln(Q^*)$, R , Da , X^{*2} , i^2 , Da^2 , $X^* \cdot i$, $X^* \cdot R$, $X^* \cdot Da$, $i \cdot R$, $i \cdot Da$, $\ln(Q^*) \cdot Da$ and $R \cdot Da$.

$$\sqrt{t_{\max}^*} = 0.9014 + 2.019 X^* - 87.95 i - 0.02001 \ln(Q^*) + 0.06196 R - 0.05957 \sqrt{Da} - 0.6373 X^{*2} + 4325 i^2 - 0.001197 R^2 + 0.002629 Da - 70.01 X^* i + 0.1266 X^* R - 0.1753 X^* \sqrt{Da} - 2.949 i R + 9.159 i \sqrt{Da} + 0.003023 \ln(Q^*) \sqrt{Da} - 0.004874 R \sqrt{Da}$$

(eq. 12)

Comparing the vt_{\max}^* empirical formula predictions with the vt_{\max}^* PMWIN results leads to a 1:1 trend with some off points (figure 13A). The Bland-Altman plot reveals that the vt_{\max}^* PMWIN results and the vt_{\max}^* empirical equation results were in the same range since the mean difference between the methods was very close to zero (there was no bias of one of the methods). The lines of agreement were quite narrow (mean difference ± 0.075) and only few points lie outside these lines which indicates little variation between the methods. These outliers are not specifically linked to one or two parameter variations but appear to be random. The data points were scattered evenly and the range of difference between the methods was consistent across the mean of the two methods, which means that there was no difference between the methods dependant on the magnitude of vt_{\max}^* (figure 13B).

Figure 14 and 15 show that an increase i , $\ln(Q^*)$ and Da leads to a decrease in vt_{\max}^* while an increase in X^* and R leads to an increase in vt_{\max}^* . Z_W^* and m have no effect on vt_{\max}^* . A variation in most parameters only leads to a small change in vt_{\max}^* (less than $vt_{\max}^* \pm 0.01$), but a change in X^* , R and Da lead to a large change in vt_{\max}^* (more than $vt_{\max}^* \pm 0.025$). Therefore vt_{\max}^* was most sensitive for X^* , R and Da .

For vt_{\max}^* the prediction interval lies 0.082 (on a square root scale) above and below the empirical formula, while the confidence has a much narrower range. While the PMWIN data points (dots in figure 15) do not always lie within the confidence interval they do lie within the prediction interval. This means that the range of deviation of the empirical formula does encompass the PMWIN points.

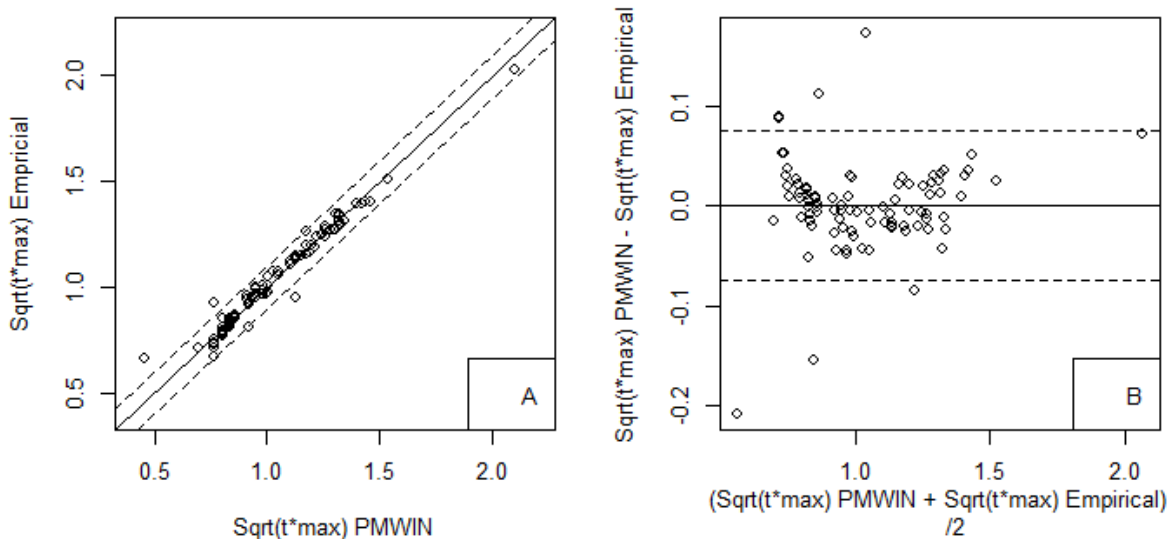


Figure 13 Comparison of results and Bland-Altman plot for vt_{\max}^*

A. The comparison between the PMWIN and empirical formula results for $\sqrt{t_{\max}^*}$. The solid line is the $y=x$ line and the dotted line gives ± 1.96 standard deviation.

B. The Bland-Altman plot which shows the mean of the PMWIN and empirical formula results $((\text{PMWIN result} + \text{empirical formula result})/2)$ on the x-axis and the difference $(\text{PMWIN result} - \text{empirical formula result})$ on the y-axis. The solid line gives the mean difference and the dotted line gives ± 1.96 standard deviation.

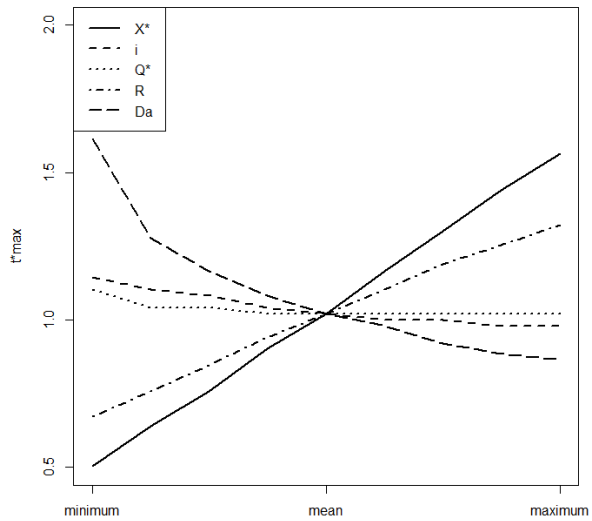


Figure 14 Sensitivity analysis for vt^*_{max}

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only one parameter was varied at a time (from its minimum to its maximum value), the rest was kept at their mean value (table 4). Z_w^* and m did not have an influence on vt^*_{max} and were therefore not included.

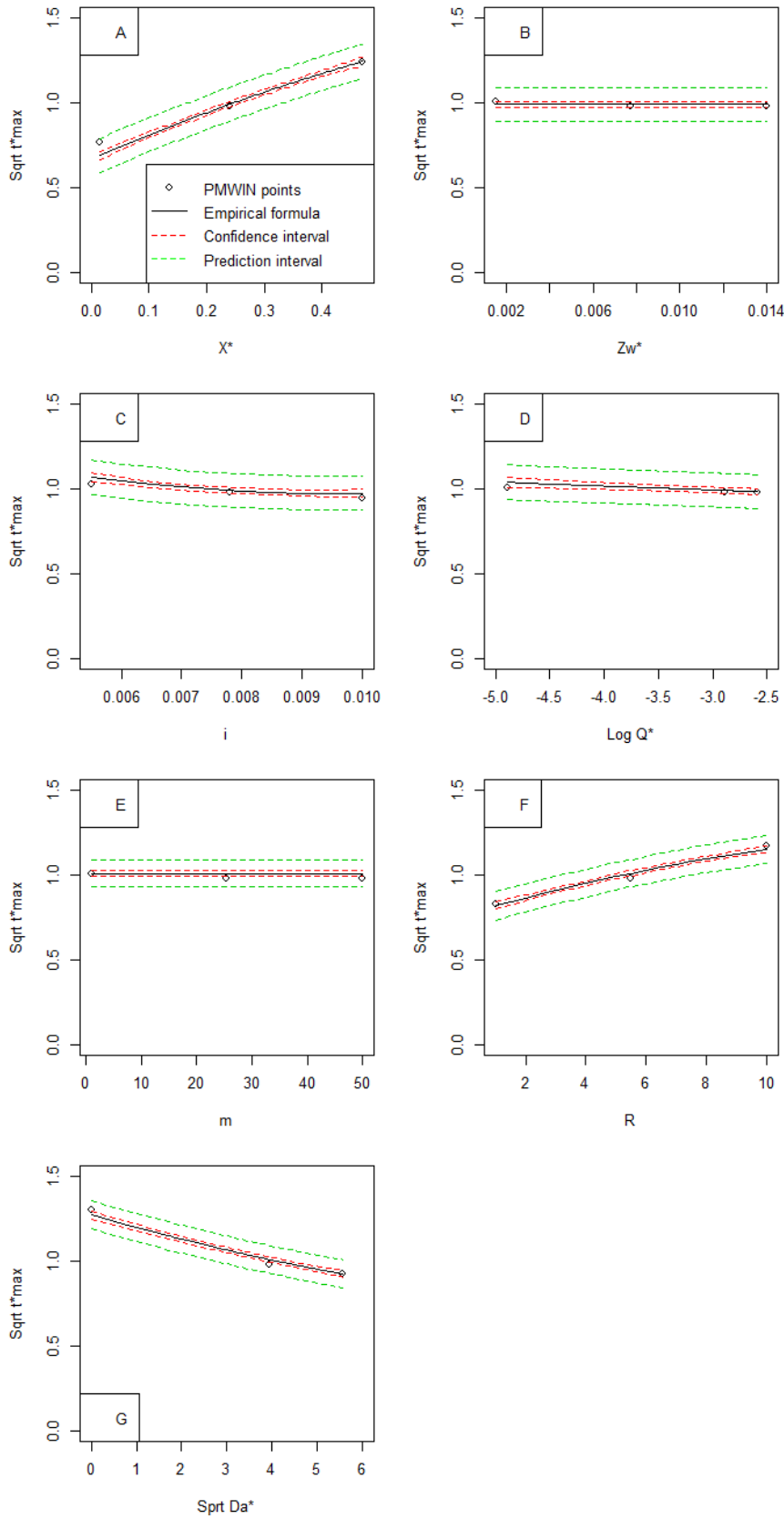


Figure 15 Influence of X^* , Z_w^* , i , $\text{Log}(Q^*)$, m , R and Da on $\sqrt{t^*_{max}}$

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only the parameter on the x-axis was varied, all other parameters were kept on their mean value (table 4). The solid line is the empirical formula, the red line the confidence interval and the green line the prediction interval. $t^*_{max} = \frac{k_x h_d t_{max}}{n L^2}$

The $\text{Ln}(C^*_{\text{half}})$ concentration is given by equation 13 and it depends on X^* , i , $\text{Ln}(Q)$, $\text{Ln}(m)$, Da , X^{*2} , Z_W^{*2} , Da^2 , $X^* \cdot i$, $X^* \cdot \text{Ln}(m)$ and $X^* \cdot Da$.

$$\text{Ln}(C^*_{\text{half}}) = 0.5572 - 18.53 X^* + 43.12 i + 0.3901 \text{Ln}(Q^*) - 0.7018 \text{Ln}(m) - 0.1942 Da - 9.345 X^{*2} - 4645 Z_W^{*2} + 0.005047 Da^2 + 1673 X^* i + 1.525 X^* \text{Ln}(m) - 1.061 X^* Da \quad (\text{eq. 13})$$

Comparing the $\text{Ln}(C^*_{\text{half}})$ empirical formula predictions with the $\text{Ln}(C^*_{\text{half}})$ PMWIN results leads to a 1:1 trend with some off points (figure 16A). The Bland-Altman plot reveals that the $\text{Ln}(C^*_{\text{half}})$ PMWIN results and the $\text{Ln}(C^*_{\text{half}})$ empirical equation results were in the same range since the mean difference between the methods was very close to zero (there was no bias of one of the methods). The lines of agreement were quite narrow (mean difference ± 1.7) and only few points lie outside these lines which indicates little variation between the methods. Most points that fall outside the lines of agreement show an overestimation of $\text{Ln}(C^*_{\text{half}})$ for the empirical formula compared to the PMWIN results. These outliers are not specifically linked to one or two parameter variations but appear to be random. The data points were scattered evenly and the range of difference between the methods was consistent across the mean of the methods, which means that there was no difference between the methods dependant on the magnitude of $\text{Ln}(C^*_{\text{half}})$ (figure 16B).

Figure 17 and 18 show that an increase X^* , $\text{Ln}(m)$ and Da leads to a decrease in $\text{Ln}(C^*_{\text{half}})$ while an increase in Z_W^* , Q^* and i leads to an increase in $\text{Ln}(C^*_{\text{half}})$. R has no effect on $\text{Ln}(C^*_{\text{half}})$. A change in most parameters mostly leads to a small change in $\text{Ln}(C^*_{\text{half}})$ ($\text{Ln}(C^*_{\text{half}}) \pm 2.3$), but a change in X^* and Da lead to a large change in $\text{Ln}(C^*_{\text{half}})$ ($\text{Ln}(C^*_{\text{half}}) \pm 6.9$). Therefore $\text{Ln}(C^*_{\text{half}})$ was most sensitive for X^* and Da .

For $\text{Ln}(C^*_{\text{half}})$ the prediction interval lies 1.8 (on a natural logarithmic scale) above and below the empirical formula, while the confidence has a much narrower range. While the PMWIN data points (dots in figure 18) do not always lie within the confidence interval they do lie within the prediction interval. This means that the range of deviation of the empirical formula does encompass the PMWIN points.

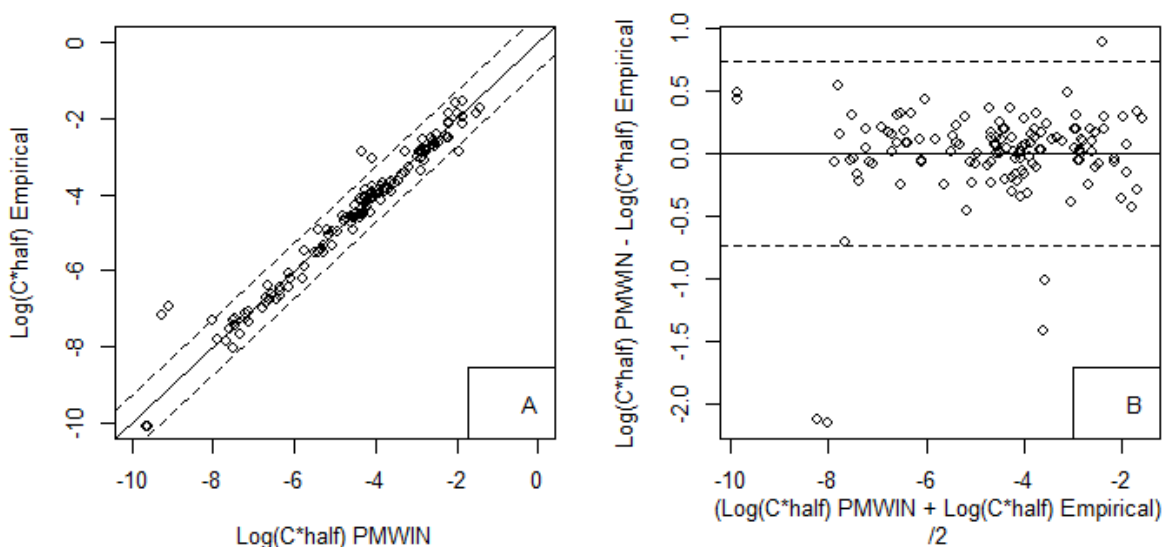


Figure 16 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{half}})$

A. The comparison between the PMWIN and empirical formula results for $\text{Log}(C^*_{\text{half}})$. The solid line is the $y=x$ line and the dotted line gives ± 1.96 standard deviation.

B. The Bland-Altman plot which shows the mean of the PMWIN and empirical formula results $((\text{PMWIN result} + \text{empirical formula result})/2)$ on the x-axis and the difference $(\text{PMWIN result} - \text{empirical formula result})$ on the y-axis. The solid line gives the mean difference and the dotted line gives ± 1.96 standard deviation.

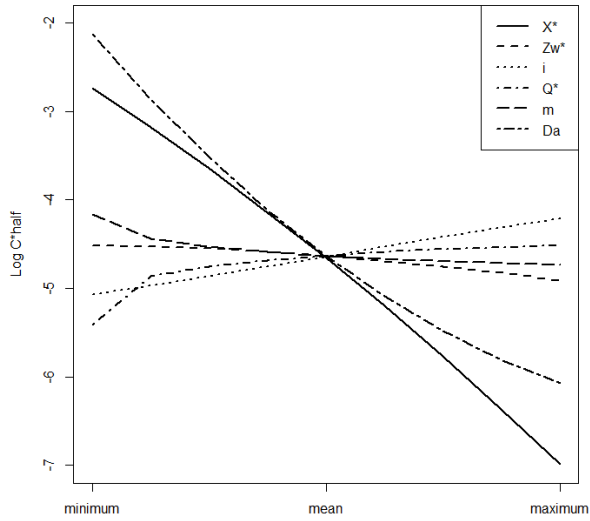


Figure 17 Sensitivity analysis for $\text{Log}(C^*_{\text{half}})$

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only one parameter was varied at a time (from its minimum to its maximum value), the rest was kept at their mean value (table 4). R did not have an influence on $\text{Log}(C^*_{\text{half}})$ and was therefore not included.

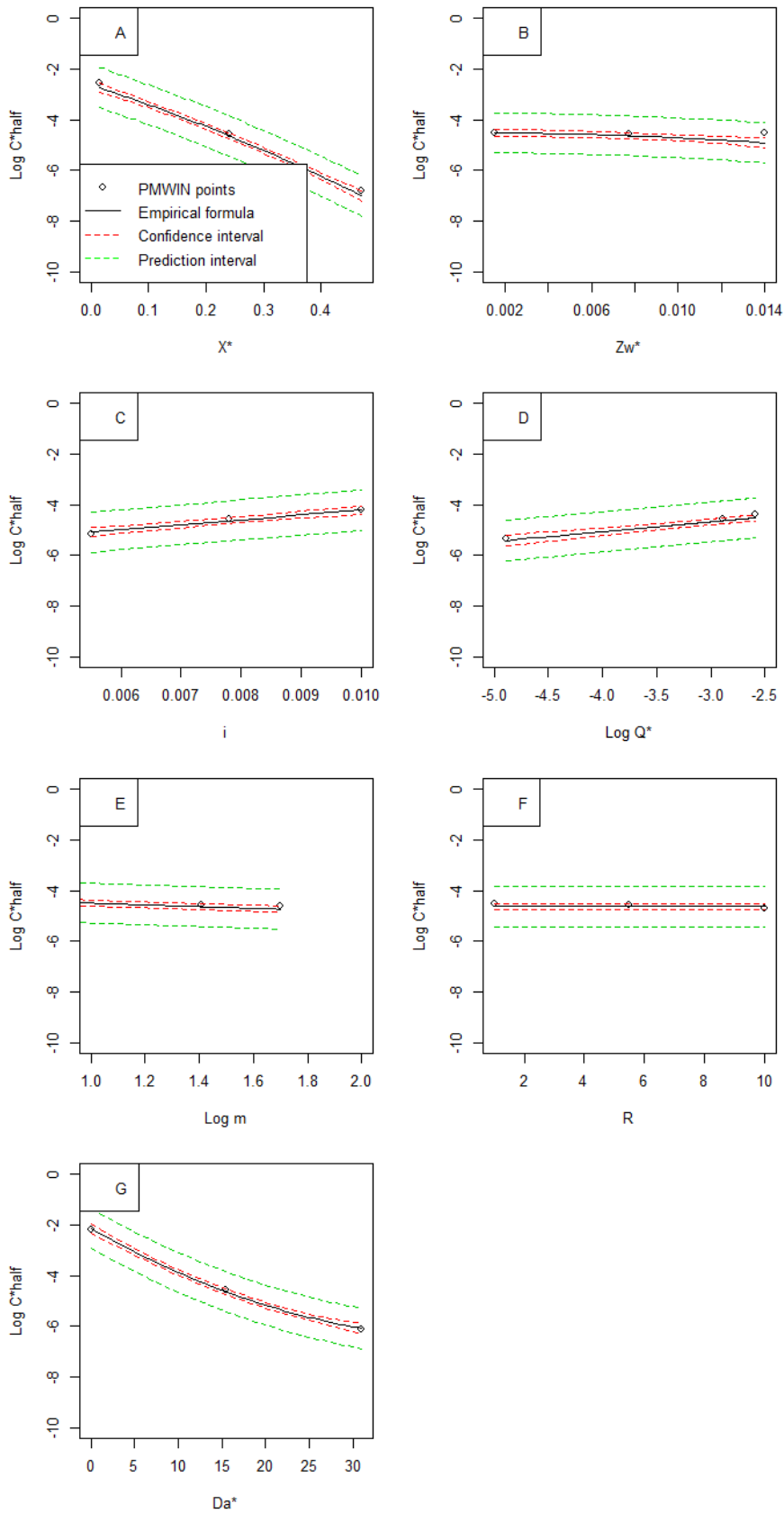


Figure 18 Influence of X^* , Z_w^* , i , $\text{Log}(Q^*)$, $\text{Log}(m)$, R and Da on $\text{Log}(C^*_{\text{half}})$

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only the parameter on the x-axis was varied, all other parameters were kept on their mean value (table 4). The solid line is the empirical formula, the red line the confidence interval and the green line the prediction interval. $C^*_{\text{half}} = C_{\text{half}}/C_0$

The vt_{half}^* depends on $\ln(X^*)$, i , $\ln(Q^*)$, R , Da , $\ln(X^*)^2$, $\ln(Q^*)^2$, R^2 , Da^2 , $\ln(X^*) \cdot i$, $\ln(X^*) \cdot R$, $\ln(X^*) \cdot Da$ and $R \cdot Da$.

$$\sqrt{t_{half}^*} = 1.246 + 0.585 \ln(X^*) - 40.24 i - 0.07178 \ln(Q^*) + 0.1037 R - 0.02029 Da + 0.06524 \ln(X^*)^2 - 0.003687 \ln(Q^*)^2 - 0.002363 R^2 + 0.0004749 Da^2 - 8.924 \ln(X^*) i + 0.01051 \ln(X^*) R - 0.001512 \ln(X^*) Da - 0.001 R Da \quad (\text{eq. 14})$$

Comparing the vt_{half}^* empirical formula predictions with the vt_{half}^* PMWIN results leads to a 1:1 trend with some off points (figure 19A). The Bland-Altman plot reveals that the vt_{half}^* PMWIN results and the vt_{half}^* empirical equation results were in the same range since the mean difference between the methods was very close to zero (there was no bias of one of the methods). The lines of agreement were quite narrow (mean difference ± 0.078) and only few points lie outside these lines which indicates little variation between the methods. Most points that fall outside the lines of agreement show an overestimation of vt_{half}^* for the PMWIN results compared to the empirical formula. These outliers are not specifically linked to one or two parameter variations but appear to be random. The data points were scattered evenly and the range of difference between the methods was consistent across the mean of the methods, which means that there was no difference between the methods dependant on the magnitude of vt_{half}^* (figure 19B).

Figure 20 and 21 show that an increase, $\ln(Q^*)$ and Da leads to a decrease in vt_{half}^* while an increase in $\ln(X^*)$ and R leads to an increase in vt_{half}^* . Z_W^* and m have no effect on vt_{half}^* . A variation in most parameters only leads to a small change in vt_{half}^* (less than $vt_{half}^* \pm 0.1$), but a change in $\ln(X^*)$, R and Da lead to a large change in vt_{half}^* (more than $vt_{half}^* \pm 0.25$). Therefore vt_{half}^* was most sensitive for $\ln(X^*)$, R and Da .

For vt_{half}^* the prediction interval lies 0.084 (on a square root scale) above and below the empirical formula, while the confidence has a much narrower range. While the PMWIN data points (dots in figure 21) do not always lie within the confidence interval they do lie within the prediction interval. This means that the range of deviation of the empirical formula does encompass the PMWIN points.

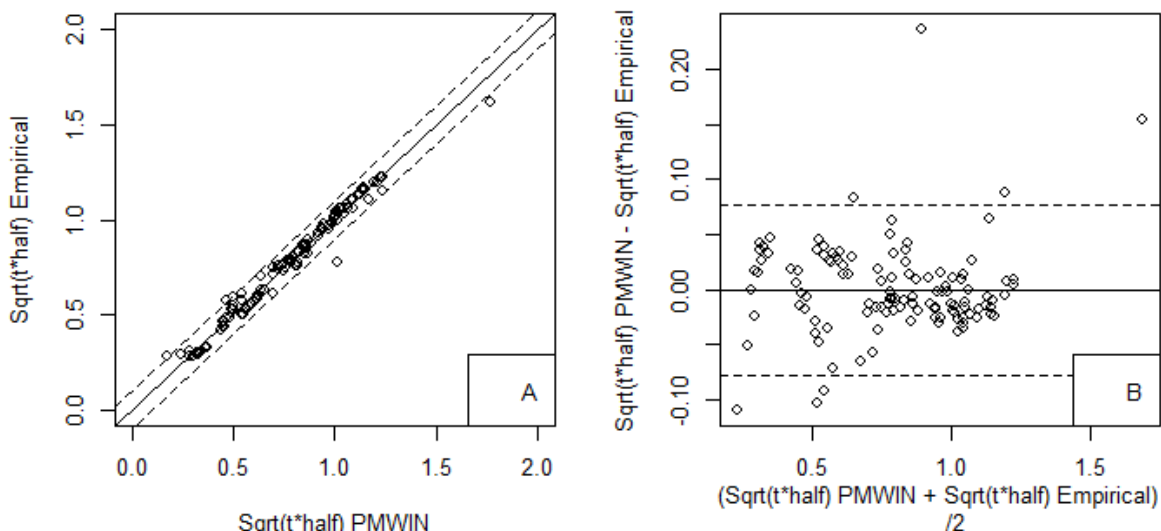


Figure 19 Comparison of results and Bland-Altman plot for vt_{half}^*

A. The comparison between the PMWIN and empirical formula results for $\sqrt{t_{half}^*}$. The solid line is the $y=x$ line and the dotted line gives ± 1.96 standard deviation.

B. The Bland-Altman plot which shows the mean of the PMWIN and empirical formula results $((\text{PMWIN result} + \text{empirical formula result})/2)$ on the x-axis and the difference $(\text{PMWIN result} - \text{empirical formula result})$ on the y-axis. The solid line gives the mean difference and the dotted line gives ± 1.96 standard deviation.

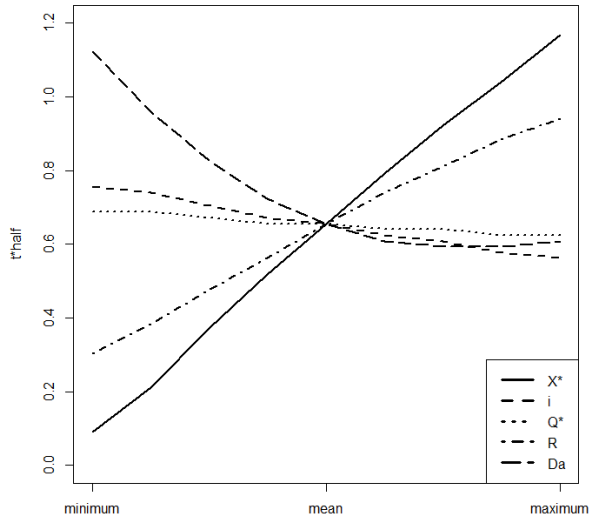


Figure 20 Sensitivity analysis for vt^*_{half}

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only one parameter was varied at a time (from its minimum to its maximum value), the rest was kept at their mean value (table 4). Z_w^* and m did not have an influence on vt^*_{half} and were therefore not included.

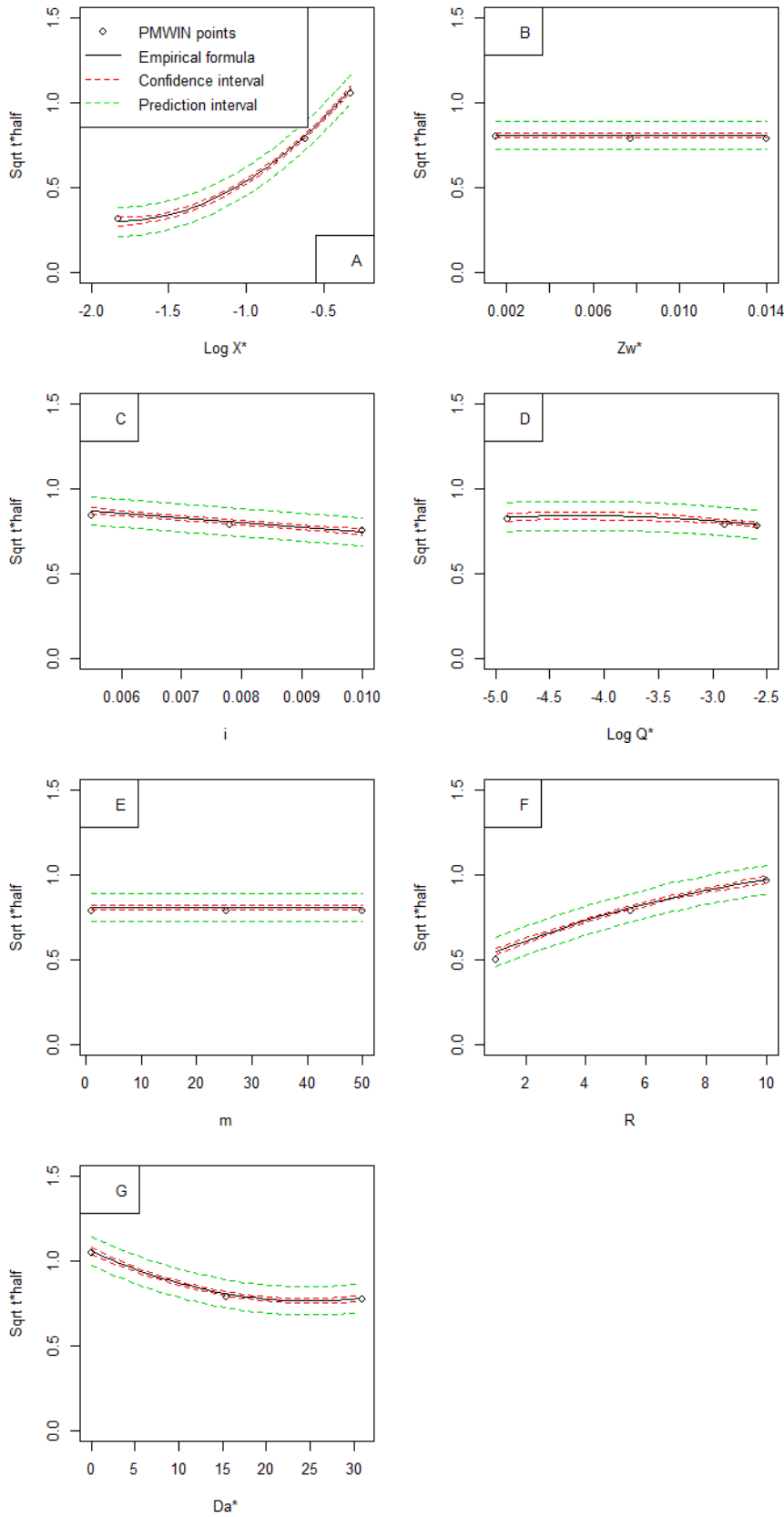


Figure 21 Influence of X^* , Z_w^* , i , $\text{Log}(Q^*)$, m , R and Da on $\text{vt}_{\text{half}}^*$

In which X^* is the horizontal distance from the plume to the well, Z_w^* is the well screen length, i is the hydraulic gradient, Q^* is the pumping rate, m is the anisotropy, R is the retardation factor and Da is the Damkohler number. Only the parameter on the x-axis was varied, all other parameters were kept on their mean value (table 4). The solid line is the empirical formula, the red line the confidence interval and the green line the prediction interval. $t_{\text{half}}^* = \frac{k_x h_d t_{\text{half}}}{n L^2}$

Figure 22 shows the comparison between the empirical formula and the analytical 3D solution. The solid line is the 1:1 line. The analytical solution gives higher concentrations than the empirical equation which becomes apparent when looking at the displacement of the mean of the two methods (-1.1 on a logarithmic scale) in the Bland-Altman plot (figure 18B, the solid line). The deviation between the two methods increased the lower the concentrations were. When omitting all scenarios where $R=1$ there was less deviation between the empirical formula and the analytical solution (figure 23) but the bias increased (-1.6 on a logarithmic scale). However the deviation between the two concentrations still increases the lower the concentrations were.

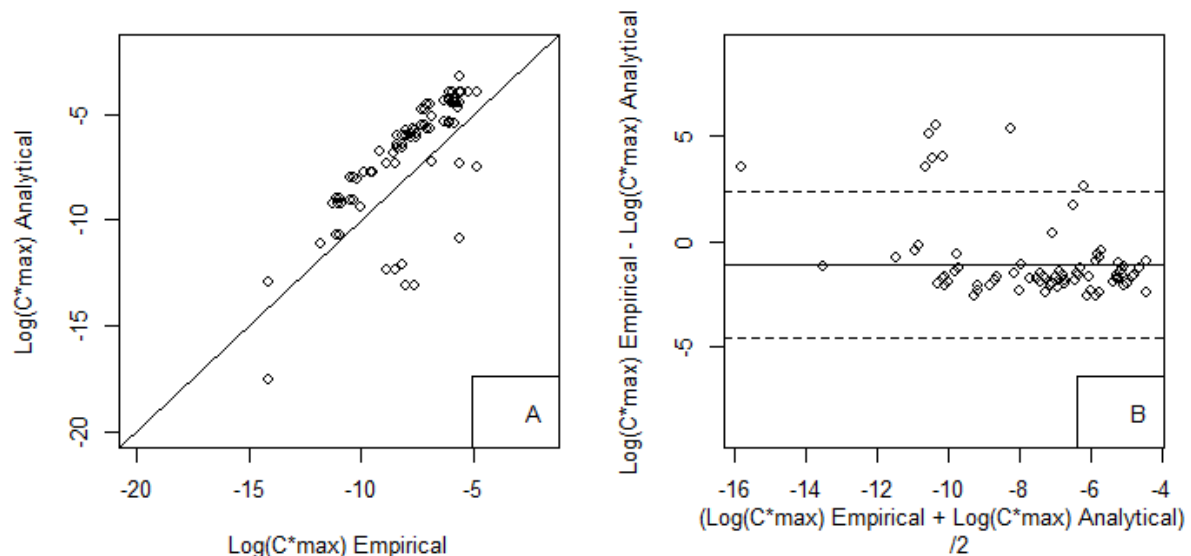


Figure 22 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{max}})$
A. The comparison between the PMWIN and empirical formula results for $\text{Log}(C^*_{\text{max}})$. The solid line is the $y=x$ line.
B. The Bland-Altman plot which shows the mean of the PMWIN and empirical formula results $((\text{PMWIN result} + \text{empirical formula result})/2)$ on the x-axis and the difference $(\text{PMWIN result} - \text{empirical formula result})$ on the y-axis. The solid line gives the mean difference and the dotted line gives $\pm 1.96 \times \text{standard deviation}$.

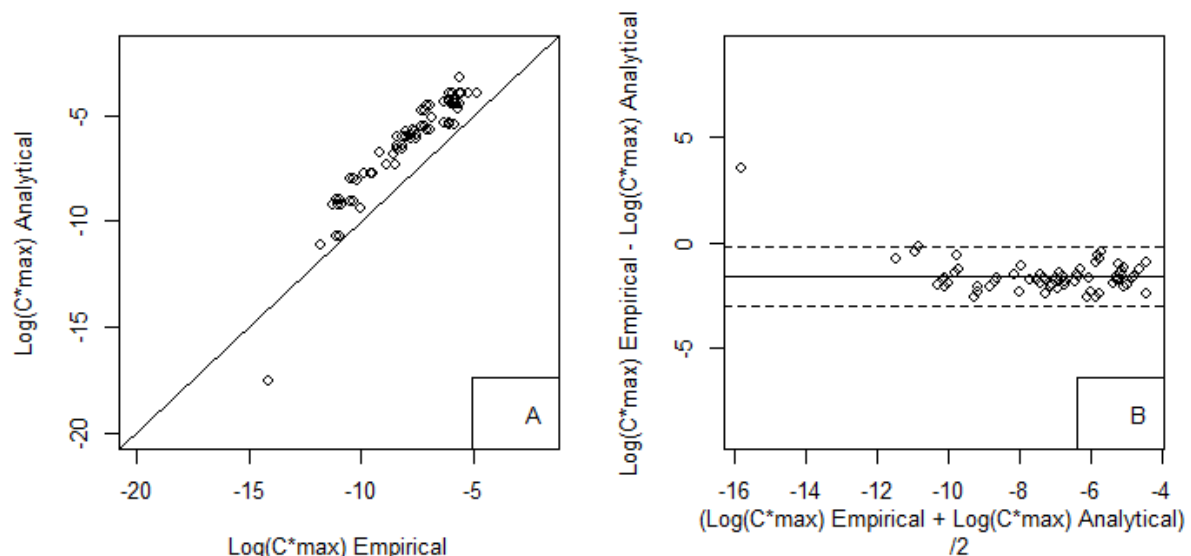


Figure 23 Comparison of results and Bland-Altman plot for $\text{Log}(C^*_{\text{max}})$ with $R > 1$
A. The comparison between the PMWIN and empirical formula results for $\text{Log}(C^*_{\text{max}})$. The solid line is the $y=x$ line.
B. The Bland-Altman plot which shows the mean of the PMWIN and empirical formula results $((\text{PMWIN result} + \text{empirical formula result})/2)$ on the x-axis and the difference $(\text{PMWIN result} - \text{empirical formula result})$ on the y-axis. The solid line gives the mean of the difference and the dotted line gives $\pm 1.96 \times \text{standard deviation}$.

To check the empirical formula it was also compared with field data from a study near Pratteln in Switzerland (Goldscheider et al., 2008). In this study uranine was used as a tracer to check if heavy metal contaminants in a drinking water production well originated from a waste incineration site. The distance between the incineration site and the well was 760m and the hydraulic gradient was 0.006m/m. Groundwater flow was from the incineration site past the pumping well and not directly to the well. The unconfined aquifer was about 20 to 30m thick, with a saturated zone of 5 to 10m thick. Since it was a gravel aquifer, the high hydraulic conductivity (0.03m/s=2592m/d) led to a high flow velocity of $v_{x,max}=127\text{m/d}$. The prevailing groundwater flow velocity was between 63 and 76m/d. The pumping rate of the extraction well varied over time, but was averaged at $Q=21.2\text{l/s}=1832\text{m}^3/\text{d}$. The study did not supply the well length and the anisotropy, therefore $Z_w=10\text{m}$ and $m=1$ was assumed. 1kg of uranine was diluted in 10L of water and injected into the ground, after that it was flushed with 120L of water. This led to $C_0=833333\mu\text{g/L}$. Uranine is a tracer and therefore $R=1$ and $\mu=0\text{d}^{-1}$. The maximum concentration of $0.33\mu\text{g/L}$ reached the well after 10-12d (Goldscheider et al., 2008).

Inserting the dimensionless values ($X^*=0.117$, $Z_w=0.0015$, $Q^*=0.00172$ and $Da=0$) into the formulas for C^*_{max} and t^*_{max} led to $\text{Ln}(C^*_{max})=-2.11$ and $\text{vt}^*_{max}=0.914$. A $\text{Ln}(C^*_{max})$ of -2.11 is the same as $C^*_{max}=0.1212$. Since $C^*_{max}=C_{max}/C_0$ thus $C_{max}=C^*_{max} \times C_0=0.1212 \times 833,333=101,4\mu\text{g/L}$. The maximum concentration at the well is $0.33\mu\text{g/L}$ while the formula predicts a maximum concentration of $101,446\mu\text{g/L}$. The concentration reduction is much larger than predicted and falls outside the prediction interval of the formula. However this may be due to the flow direction. Only 0.93% of the uranine mass was recovered in the pumping well because the plume only grazed the well (Goldscheider et al., 2008). The concentration at the edge of the plume is much lower than the concentration at the centre and thus the apparent reduction is larger than it really is. To calculate the vt^*_{max} , first the v_{res} needs to be calculated where $v_{res} = k_x \frac{h_d}{nL}$. The porosity is set at 0.2 for model purposes in the paper therefore $v_{res}=19.92\text{m/d}$. A $\text{vt}^*_{max}=0.914$ is the same as $t^*_{max}=0.8354$ and $t^*_{max}=t_{max}/v_{res}$. Therefore $t_{max}=t^*_{max} \times v_{res}=0.8354 \times 19.92=16.7$ days. The predicted and actual t_{max} differ a little bit, but this may be caused by the difference in soil type and flow velocities.

4 Discussion

4.1 Discussion results

The larger the distance between the well and the plume the longer the time for dilution, dispersion, adsorption and decay is. Therefore the larger the distance the more the concentration may decrease. When the anisotropy is increased, the horizontal hydraulic conductivity is larger than the vertical hydraulic conductivity, which leads to a decrease in vertical flow and thus a decrease in vertical transport. Therefore, the concentration will decrease faster with a larger anisotropy. Finally the larger the decay rate is the larger the reduction in concentration will be.

An increase in well screen length, pumping rate and hydraulic gradient leads to an increase in concentration. When the pumping rate or the hydraulic gradient is increased the flow velocity increases leading to faster transport. The faster the transport the smaller the effect of adsorption and decay and thus the concentration at the well becomes higher. The effect of dilution due to radial flow around the well should lower the concentration, but this is not the visible trend in this study, therefore the effect of dilution is smaller than that of the increase in flow velocity due to an increase in pumping rate. An increase in the well screen length leads to a larger area through which water is pumped and will thus capture more of the contaminant.

Concentration is not dependant on retardation according to the PMWIN results and the empirical formulas. While a larger retardation factor leads to a later arrival of the plume the drop in concentration it causes is negligible. This is not as expected since an increase in retardation allows for more time for dispersion and decay.

The effect of hydraulic gradient and decay separately is increased when these parameters are combined with distance. The larger the distance the more time there is for decay and the bigger the effect of flow velocity on the concentration will be. When the distance is increased the anisotropy no longer causes a decrease in concentration but rather an increase. When the distance between the well and the plume is larger the bigger the role of horizontal transport will be compared to vertical transport. A high anisotropy causes a preference for horizontal groundwater flow and contaminant transport, compared to vertical groundwater flow and contaminant transport. At longer distances, when horizontal transport is more significant, the anisotropy will increase the concentration due to the preferential flow field, while at smaller distance the vertical transport will be constrained by the anisotropy leading to a decrease in concentration.

The concentration is most sensitive to a change in travel distance and decay rate.

An increase in hydraulic gradient and the pumping rate both led to a decrease of travel time. This is due to the increase in groundwater flow velocity both parameters cause. Another parameter to decrease the travel time is the decay rate. An increase in the decay rate makes the breakthrough curve more asymmetrical, with an earlier maximum concentration and longer tailing. The larger the concentration, the more will be decayed. Therefore the diluted front of the plume will be reduced less than the peak concentration at the centre or at the tail. This means that the centre and tail of the plume now have a lower concentration than the front and therefore there is an earlier arrival of the maximum concentration.

An increase in distance and retardation has the opposite effect on travel time. The larger the distance the longer a contaminant will take to travel. An increase in retardation means that more of the contaminant is adsorbed, to be released at a later time. This causes a delay in travel time.

The size of the well screen and the anisotropy do not have an effect on the travel time. The well screen size and anisotropy both affect the vertical flow rate and since they have no effect on travel time this means that the vertical flow rate is negligible compared to the horizontal flow rate.

This effect of the individual parameters on travel time is both visible in the PMWIN results and in the empirical equation.

The individual effect of hydraulic gradient, retardation and decay is increased when they are coupled with distance. A larger distance will lead to a bigger effect of flow velocity and allow for more time

for retardation and decay. The effect of decay on travel time is increased when retardation is increased. An increase in retardation increases the travel time and thus leaves more time for decay. The effect of retardation and decay on t_{max}^* are reversed the larger the hydraulic gradient is. An increase in flow velocity causes faster transport and thus less time for retardation and decay. An increase in pumping rate also leads to the reversal of the effect of decay on t_{max}^* as it increases the flow velocity as well.

The travel time is most sensitive to a change in travel distance, retardation and decay rate.

When the distance is larger than 3km the pollutant will not arrive at the well within 50 years at concentrations above detection limit, unless it is a tracer ($R=1$, $\mu=0d^{-1}$). The same goes for a hydraulic gradient under 0.0055m/m, a retardation above 10 and a decay rate above $0.005d^{-1}$.

The prediction intervals show that the actual concentration may deviate a factor 6 from the one calculated by the empirical formula. This difference is acceptable. The time only deviates about 41 days.

4.2 Comparison

When comparing the empirical formula with the analytical 3D solution the empirical formula shows an underestimation of the concentration (+1.1 on a logarithmic scale) compared to analytical 3D solution. This overestimation can be caused by the difference in complexity of both methods. Next to that the empirical equation was extrapolated to a situation with a very small pumping rate while the analytical solution does not account for pumping wells. As discussed above, the higher the pumping rate, the higher the concentration. Thus even the small pumping rate may lead to an increase in concentration compared to having no pumping well at all.

The comparison of the empirical formula with a field study showed that the empirical formula overestimated the concentration and arrival time. The predicted concentration was a factor 1000 higher and the arrival time was 7 days longer. However this overestimation may be caused by the flow direction which is not directly from the plume to the well but at an angle past the well.

4.3 Restrictions

These empirical formulas are only applicable within the parameter space and the list of assumptions. When the natural groundwater flow is at an angle with the well, as it was in the abovementioned study, the predicted concentration in the well will be higher than the actual concentration since the travel pathway and travel time will be longer and the center of the plume will not reach the well. The formulas are not applicable when groundwater flow is away from the well.

Another assumption is that of the unconfined homogeneous aquifer. Though the presence of clay lenses and confining layers may lower the concentration and increase the arrival time it would be interesting to see to what extent they affect concentration and arrival time.

Not only the assumptions, but the constants influence the formula outcomes as well. The dispersivity was chosen at a value of 130m. This means that there is relatively more advection than dispersion.

To see what happens when dispersion plays a larger role, future studies could lower the α_L value.

In this study the top of the well screen is mostly kept at 50m below the contaminant plume, however some contaminants are at the same height as the well screen and will therefore travel faster and reach the well at higher concentrations than predicted by the formulas.

Other constants like the volume of the plume could be changed but this is not necessary. As mentioned in the tiered approach; if a contaminant has a volume of more than $6000m^3$ soil remediation measures must be taken. If the plume size is lower than $1000m^3$ the formulas predict a higher concentration and later arrival time at the well and therefore will be on the safe side.

5 Conclusion and recommendations

Four empirical formulas were developed to predict the maximum concentration, half the maximum concentration, the arrival time of the maximum concentration and the arrival time of half the maximum concentration. These formulas are dependent on the distance between the plume and the well, the well screen length, the hydraulic gradient, the pumping rate, the anisotropy, the retardation factor and the decay rate. The concentration is influenced by all parameters except the retardation factor and is most sensitive to a change in distance and decay rate. The arrival time is influenced by all parameters except the well screen length and anisotropy. It is most sensitive to a change in distance, retardation and decay rate.

The empirical formulas for maximum concentration and half the maximum concentration allow the prediction of the contaminant concentrations in the drinking water abstraction well with a factor 6 error. The predicted concentration can then be made dimensional and compared with the Drinking Water Quality Standard. The empirical formulas for arrival time of the maximum concentration and the arrival time of half the maximum concentration predict the arrival time of said concentrations within an error of 41 days.

The empirical formula predictions and the PMWIN model results show a good correlation but a comparison with a 3D analytical solution shows an overestimation (+1.5 on a log scale) by the empirical formulas. When comparing the empirical formulas with a field study there was a discrepancy between the formula predictions and the field measurements. This discrepancy was roughly a factor 1000 for the maximum concentration and 7 days for the arrival time of the maximum concentration. The discrepancy of the arrival time falls within the error but the discrepancy of the maximum concentration is very large. This may be caused by the deviating circumstances, like the flow direction.

The empirical formulas are only applicable when the parameters are within their prescribed range and the conditions do not differ from the assumed conditions. This means that the empirical equation is applicable in fully saturated unconfined sand aquifers with groundwater flow from the plume in a straight line to the well. The contaminant only undergoes anaerobic decay and equilibrium sorption and the adsorbed phase does not decay. Any deviation of the assumed circumstances leads to a faulty prediction of the empirical formulas.

This is a restriction to the use of the empirical formulas and therefore future studies may consider a larger range of parameters, or different assumed circumstances. Some examples of what future studies may consider are the effect of a lower dispersivity, the presence of clay lenses or confining layers, the position of the top of the well screen and the direction of the groundwater flow to increase the applicability of the formulas. This study only considers advection dominated transport and by decreasing the dispersivity, dispersion becomes more significant. Aquifers in the Netherlands are never homogeneous and often have clay lenses and confining layers therefore it they may be added to the empirical formula in the future. As shown with the comparison with the field study, a groundwater flow direction that is not in a straight line from the plume to the well will lead to deviating concentrations and arrival times than the formulas predict. Therefore it is advisable to include this in future studies.

When a contaminant is at a greater distance than 3km from the well, has a retardation factor of more than 10 and a decay rate above 0.005d^{-1} it will not reach the well within 50 years above the detection limit unless it is a tracer. Only conservative tracers injected more than 3 km from the well would reach it.

The empirical formulas are user friendly and require little in-depth knowledge of groundwater hydrology and contaminant transport. To increase the user friendliness of the formulas they might be combined in a tool. In this tool the parameters can be supplied. Behind the screens the parameters are then made dimensionless and the formulas are solved. The dimensional concentration at the well and arrival time will then be given back to the user as a result.

References

Papers, books and reports

- Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control. AC-19, 716-723
- Bear, J., 1988. Dynamics of fluids in porous media. Dover publications, New York.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurements. The Lancet 8476, 307-310
- Box, G.E.P., and Behnken, D.W., 1960. Some new three level designs for the study of quantitative variables. Technometrics 2, 455-475
- Cavazzuti, M., 2013. Optimization methods. From theory to design scientific and technological aspects in mechanics. Chapter 2 Design of experiments. Springer Berlin, Heidelberg
- Council of the European Union, 1998. Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption. Official journal of the European communities L330, 32-54
- David, P., Wilson, S., 2011. Groundwaters of Marlborough. The Caxton Press, Christchurch, New Zealand, Chapter 9
- Fetter, C.W., 1999. Contaminant hydrology, Second edition. Macmillan New York, chapters 2 and 3
- Goldscheider, N., Milnes, E., Fries, W., Joppen, M., 2008. Markierungsversuche und Modellierung zur Bewertung der Gefährdung eines Trinkwasserbrunnens. Grundwasser -Zeitschrift der Fachsektion Hydrogeologie 13, 220-230
- Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW 2000, the U.S. geological survey modular ground-water model—user guide to modularization concepts and the ground-water flow process. U.S.G.S. open-file report 00-92
- Hendriks, M.R., 2010. Introduction to physical hydrology. Oxford University press. New York
- Hoogeveen, R.J.A., Witjes, Th.G.J., 2012. Gebiedsdossier. Provincie Limburg. Report MT900-14
- Koorevaar, P., Menelik, G., Dirksen, C., 1983. Element of Soil Physics. Developments in Soil Science 13, Elsevier, Amsterdam
- Lodder, A., Steinweg, C.M., 2013. Gebiedsdossier grondwaterbeschermingsgebieden in Fryslân. Provincie Friesland. Report 9W7289
- Medina, M.A., Jr, M.M.A., Huber, W.C., Heaney, J.P., 1981. River quality model for urban stormwater impacts. Journal of the water resources planning and management division. American society of civil engineers 107, 263-280
- Ministry of IenM, 2011. Besluit van 23 mei 2011, houdende bepalingen inzake de productie en distributie van drinkwater en de organisatie van de openbare drinkwatervoorziening (Drinkwaterbesluit). Staatsblad van het Koninkrijk der Nederlanden 2011, 293
- Ministry of Volksgezondheid en Milieuhygiëne, 1979. Wet van 13 juni 1979, houdende regelen met betrekking tot een aantal algemene onderwerpen op het gebied van de milieuhygiëne. Staatsblad van het Koninkrijk der Nederlanden 1979, 422
- Ministry of VROM, 2009. Wet van 8 juli 2009, houdende nieuwe bepalingen met betrekking tot de productie en de distributie van drinkwater en de organisatie van openbare drinkwatervoorzieningen (Drinkwaterwet). Staatsblad van het Koninkrijk der Nederlanden 2009, nr. 370.
- Ministry of VROM, 2013. Circulaire bodemsanering per 1 juli 2013. Staatscourant 2013 nr 16675
- Ministry of VROM and Ministry of Landbouw en Visserij, 1986. Wet van 3 juli 1986, houdende regelen inzake bescherming van de bodem (Wet bodembescherming). Staatsblad van het Koninkrijk der Nederlanden 1986 nr. 374
- Otte, P.F., Lijzen, J.P.A., Otte, J.G., Swartjes, F.A., Versluijs, C.W., 2001. Evaluation and revision of the CSOIL parameter set; Proposed parameter set for human exposure modeling and deriving Intervention Values for the first series of compounds. RIVM report 711701021

Otte, P.F., Zijp, M.C., Kovar, K., Lijzen, J.P.A., Swartjes, F.A., Verschoor, A.J., 2007. A tiered procedure to assess risk due to contaminant migration in groundwater. RIVM report 711701056

van Rooijen, P., 1989. Grondwater in Limburg. Grondboor en Hamer 43, 377-386

Schijven, J.F., Hassanizadeh, S.M., De Roda Husman, A.M., 2010. Vulnerability of unconfined aquifers to virus contamination. Water research 44, 1170-1181

Swartjes, F.A., Wuijts, S., Otte, P.F., 2014. Beoordeling bodem- en grondwaterkwaliteit in grondwaterbeschermingsgebieden. RIVM rapport 2014-0066

WMD-water, 2011. Gebiedsdossier grondwaterbescherming in Drenthe. Provincie Drenthe, report D011012702.

Wuijts, S., Schijven, J.F., Van der Aa, N.G.F.M., Dik, H.H.J., Versluijs, C.W., Van Wijnen, H.J., 2007. Bouwstenen Leidraad Grondwaterbescherming. RIVM rapport 734301029/2007

Wuijts, S., Bogte, J.J., Dik, H.H.J., Verweij, W.H.J., Van der Aa, N.G.F.M., 2014. Eindevaluatie gebiedsdossiers drinkwaterwinningen. RIVM Rapport 270005001/2014

Zheng, C., Wang, P.P., 1999. MT3DMS, A modular three-dimensional multispecies transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. Strategic environmental research and development program (SERDP), report SERDP-99

Zwolsman, J.J.G., 2011. Basisinformatie ten behoeve van de knelpuntenanalyse drinkwater en industriewater (fase 2). KWR report 2012.055

Websites

TNO, 2015

Developed by TNO, Geologiesche Dienst Nederland

Dinoloket – ondergrond modellen – REGIS II

<https://www.dinoloket.nl/ondergrondmodellen>

Last checked 19-11-2015

Map groundwater protection zones Utrecht

[https://webkaart.provincie-](https://webkaart.provincie-utrecht.nl/viewer/app/Webkaart?bookmark=680bb7a8c19548608a5bbccb8ff97be4)

[utrecht.nl/viewer/app/Webkaart?bookmark=680bb7a8c19548608a5bbccb8ff97be4](https://webkaart.provincie-utrecht.nl/viewer/app/Webkaart?bookmark=680bb7a8c19548608a5bbccb8ff97be4)

Last checked: 20-11-2015

Map groundwater protection zones Gelderland

<http://flamingo.prvgld.nl/viewer/app/Drinkwater>

Last checked: 20-11-2015

Map groundwater protection zones Overijssel

<http://gisopenbaar.overijssel.nl/website/wateratlas/wateratlas.html>

Last checked: 20-11-2015

Environmental decree Noord-Brabant

Provinciale milieu verordening Noord-Brabant – appendix 5

http://www.brabant.nl/applicaties/regelingen/838_provinciale_milieuverordening_noord_brabant_2010.aspx

Last checked: 20-11-2015

Valocchi et al., 2016

Developed by A.J. Valocchi, C.J. Werth, J.J. Decker, G. Hammond, P. Zhou and M. Hafiz

Continuous input in a semi-infinite domain – “third type” boundary condition

http://hydrolab.illinois.edu/gw_applets/gui/GUI.html

Last checked: 25-02-2016

Programs

Processing Modflow for Windows version 8.0.40
©1991-2013, Simcore Software

R Studio version 0.99.491
© 2009-2015, RStudio Inc.

Appendix

List of Appendixes:

- A – List of phreatic aquifers in the Netherlands
- B – Reproducing the MODFLOW and MT3DMS model
- C – R script

A – List of phreatic aquifers in the Netherlands

Ameland Buren	Ellecom	Hertenbos	Nieuwe Marktstraat	T Klooster
Ameland Hollum	Epe	Hoenderloo Mix	Noordbargeres	Terschelling
Amersfoort Berg	Gasselte	Hoge Hexel	Oirschot	Terwisscha
Amersfoortseweg	Goor	Hoogeveen	Olden Eibergen	Tilburg middeldiep
Beilen	Groenekan	Kruidhaars	Oosterbeek	Tilburg ondiep
Cothen	Groote Heide	Laren Mix	Pinkenberg	Tull En T Waal
Dalen	Grubbenvorst	Leidsche Rijn	Prinsenbosch	Van Heek
De Haere	Harderwijk I	Lieshout	Putten	Vlieland Bokkendal
De Pol	Hasselo	Linschoten	Rodenmors	Vorden
Diepenveen	Havelterberg	Lochem	Ruinerwold	Wierden
Dinxperlo	Hanik	Manderveen	Someren	Woudenberg
Eemdijk	Herikerberg	Muntberg	Speuld	Zutphen Vierakker
Eerbeek				

B – Reproducing the MODFLOW and MT3DMS model

This appendix contains an explanation on how to recreate the MODFLOW and MT3DMS model used. It is split up in a MODFLOW part and a MT3DMS part.

MODFLOW

1. Create a model
2. Assign data
 - 2.1 Grid
 - 2.2 Parameters
 - 2.3 MODFLOW
3. Run simulation
4. Check results

MTR3MS

1. Assign model data
 - 2.1 MT3DMS
2. Run simulation
3. Check results

Reproducing MODFLOW

1. Create a model

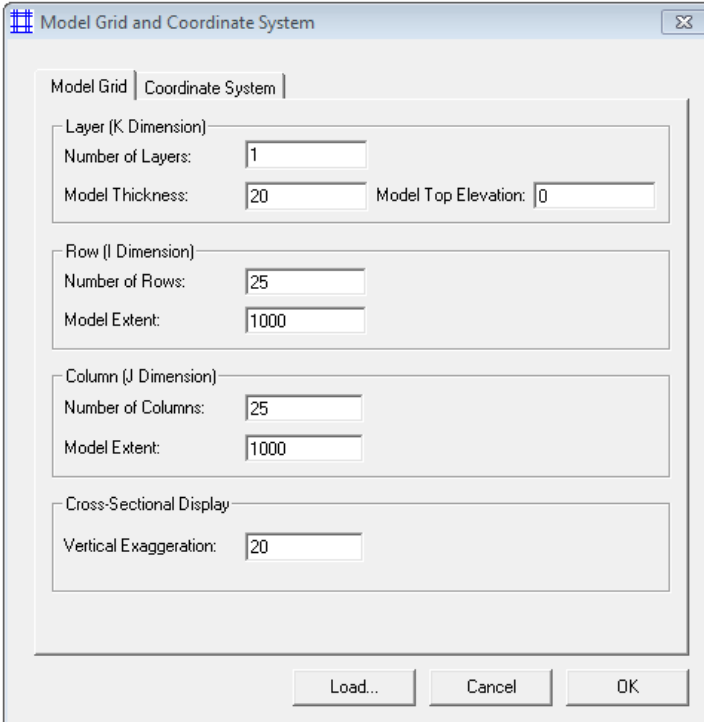
To make a new model select the “*New Model*” option under File. Choose the location where you want to save the model and name it. The model name cannot contain any spaces.

2. Assign data

2.2 Grid

Under the tab grid the outline of the model is created.

→ Select “*Mesh size*” and fill in the “*Model Grid and Coordinate System*” dialog box




The model has 12 layers and a total thickness of 0.0018. The model top elevation is 0.0018.

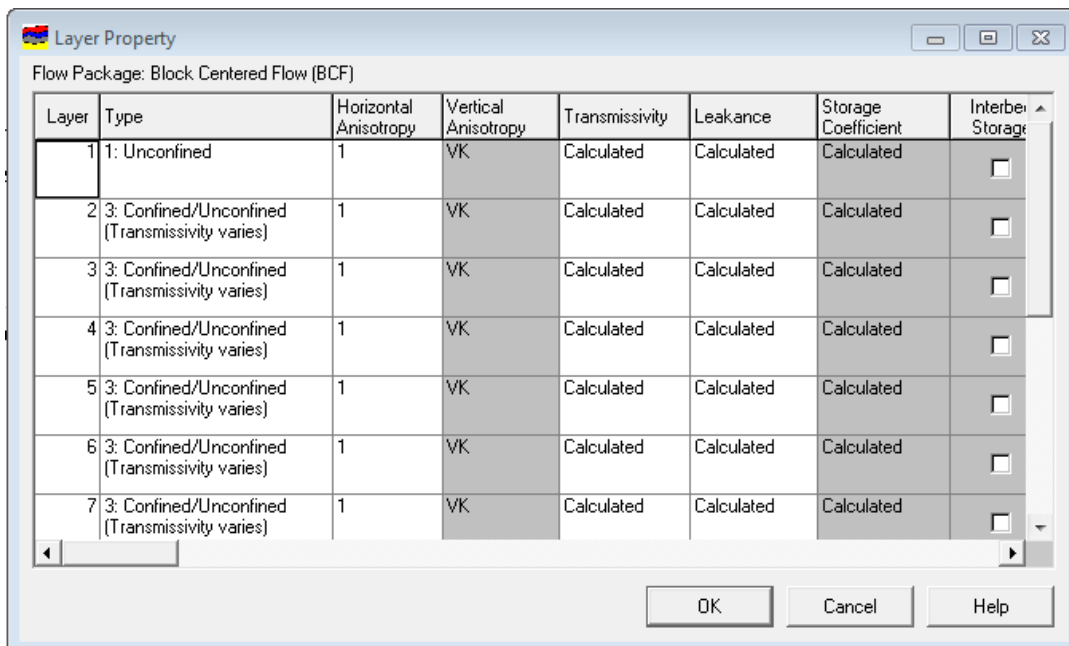
The model has 10 rows and an extent of 0.15

The model has 65 columns and an extent of 1

Leave the vertical exaggeration on 20 and press OK


You now see the grid in the editor. Leave the editor by pressing the “*Leave Editor*” button  and save changes


→ Then select “*Layer properties*” to assign a layer type in the layer property dialog box.



Set all layers except the first to type 3 (Confined/Unconfined). Set the first layer to type 1 (Unconfined) and press OK.

→ Select “Cell Status” and then “IBOUND” to set boundary conditions. Again you see the grid in the editor.

Select the top left grid cell and open the dialog box by right clicking the cell. Assign a value of -1 and press OK. Select the “Cell-by-cell duplication” button  and select the left most row with the arrow keys. Then select the top right cell with the mouse and select the right most row with the arrow keys. Turn off the duplication button.

To apply this to all the layers select the “Layer copy” button  and duplicate to all layers by viewing all layers with the page down key.

This way the left most row and the right most row have a constant head boundary condition. The edges of the domain automatically have a no-flow boundary condition unless supplied otherwise.

Leave the editor and save the changes.

→ Select “Cell Status” and then “ICBUND” to set boundary conditions.

In the editor select “Reset Matrix” under Value and fill in 1. Then check “Apply to entire model” and press Ok.

Leave the editor and save the changes.

→ Select “Top of layers” to assign an upper height to all layers.

In the editor select “Reset Matrix” under Value and fill in 0.0018 and then check “Apply to current layer” and press Ok. The second layer must have a value of 0.00015 lower and so on until you reach the bottom layer

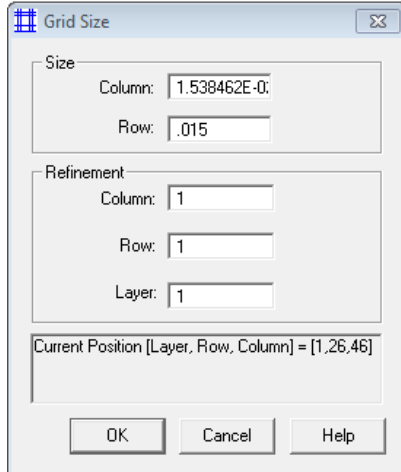
Leave the editor and save the changes

→ Select “Bottom of layers” to assign a bottom height to all layers.

Assign the top layer a value of 0.0165 by using reset matrix. The second layer must have a value of 0.00015 lower and so on until you reach the bottom layer.

Leave the editor and save the changes.

→ Select “Mesh size” once again to make the grid finer.



Select the 5th row and the 3rd column and right click. Under refinement fill in 10 for both rows and columns. The individual cells will now be split up in 10×10 cells. Press OK.

Do the same for the 6th row and the 4th column.

Now select the 5th column and only fill in 10 columns under refinement.

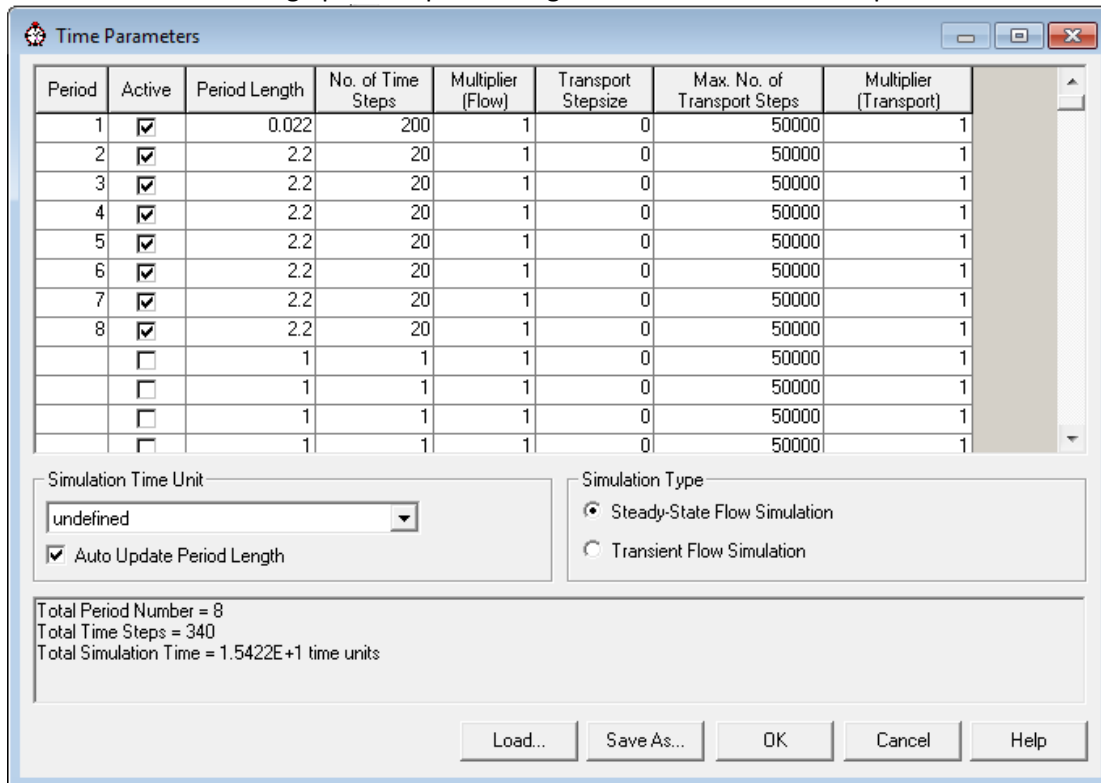
Now the cells have a 10×10×10 size around the contamination plume and a 10×10×100 size around the wells.

Leave the editor and save changes

2.2 Parameters


Under the parameter tab values for time, flow parameters and soil parameters


→ Select “Time” to assign periods, period length and number of time steps



Set simulation time unit to undefined and uncheck auto update period length. Activate 10 periods by checking the boxes. The first period is the spin-up time of the model to reach steady-state flow. This period can have a short number of time steps but have many time steps.

Using $t^* = \frac{v_{res} t}{L}$ the conversion from dimensions to dimensionless is made. Period one has a duration of 1 day and the other time steps have a length of 10 years.
For period 1 assign a period length of 0.00016 and 800 time steps. For the other periods assign a period length of 0.59 and 20 time steps.
Also select a steady-state flow simulation as simulation type and press OK.

→ Select “*Initial hydraulic head*” to prescribe a hydraulic gradient.
In the editor use reset matrix assign a value of 1 to the entire model. Then select the top left cell and right click it to assign a value of 1.05. Use the “*Cell-by-cell duplication*” button  and select the left most row with the arrow keys. Turn off the duplication button.

To apply this to all the layers select the “*Layer copy*” button  and duplicate to all layers by viewing all layers with the page down key. Now $i=0.001$.
Leave the editor and save the changes.

→ Select “*Horizontal hydraulic conductivity*” to give a k_x value.
In the editor use reset matrix assign a value of 1 to the entire model.
Leave the editor and save the changes.

→ Select “*Vertical hydraulic conductivity*” to give a k_z value.
In the editor use reset matrix assign a value of 1 to the entire model. Now $m=1$.
Leave the editor and save the changes.


→ Select “*Porosity*” to give a n value.
In the editor use reset matrix assign a value of 0.35 to the entire model.
Leave the editor and save the changes.

→ Select “*Bulk density*” and then “*Cell-by-cell*” to give a ρ^b value.
In the editor use reset matrix assign a value of 2000 to the entire model.
Leave the editor and save the changes.

2.3 MODFLOW

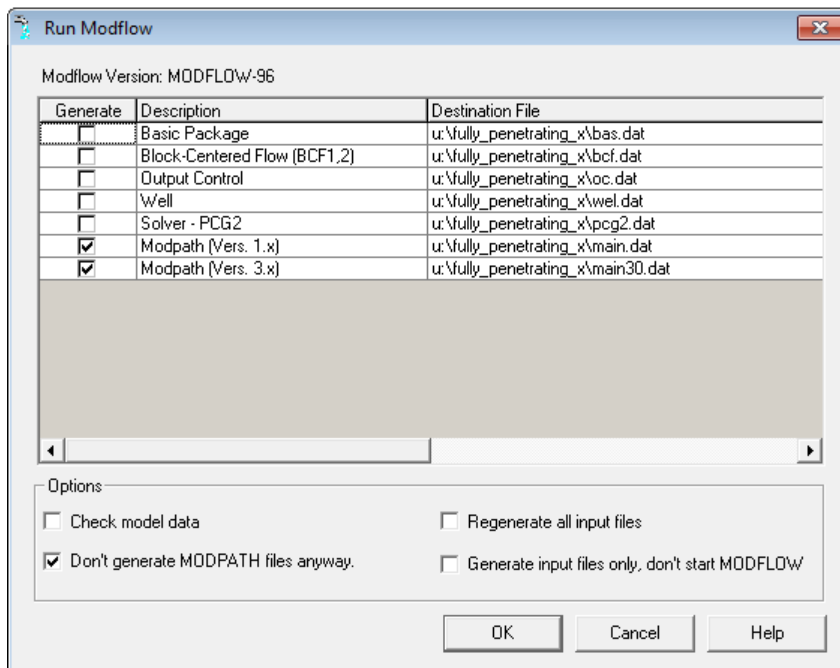
Under the Models tab MODFLOW can be selected. Here extra packages can be selected

→ Select “*Well*” in “*Flow packages*” to add a well to the model.
Select the cell at the 28th row and the 14th row and the 60th column in the 6th layer and give it a value of $-0.00032/5$ ($Q=200 \text{ m}^3/\text{d}$ over 5 layers). Do the same for the next 4 layers to create a well of 50 m in length.

Then press the Change Stress Period button  and copy all data from stress period 1 to the other stress periods and press close.
Leave the editor to save the changes.

3. Run model

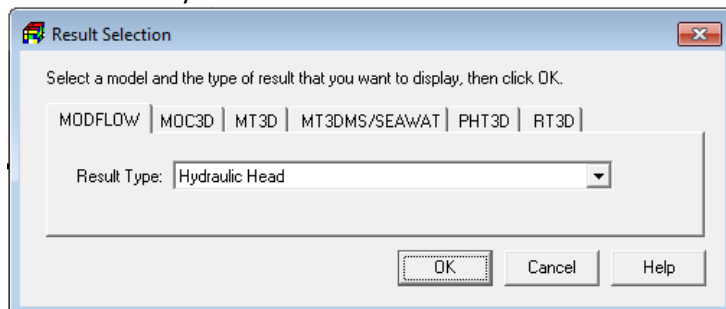
To run the model select “*run...*” in “*MODFLOW*”. This opens the run Modflow dialog box.



The first time the model is run all the boxes under generate should be checked. Under options the don't generate MODPATH files anyway option should be checked as well. Press OK. This will open the command prompt. Once the model is run press any key to leave the command prompt.

4. Check results

To check the hydraulic head select "2D-visualization" under the Tools tab.



Choose the MODFLOW tab and select the Hydraulic Head and press OK. Now the editor will open showing the domain and the hydraulic heads. By looking through the different time steps you can check whether the hydraulic head is in steady state. If not the number of time steps of the first time period should be increased.

Reproducing MT3DMS

1. Assign data

1.1 MT3DMS

→ Select "Simulation Settings" under the tab "MT3DMS" in the tab "Models".

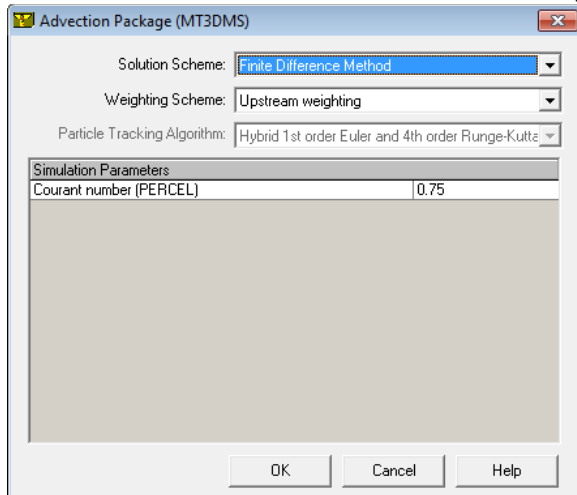
Check a box to activate one species and name it. For the simulation type select constant density transport from the dropdown menu. For the reaction type select first-order irreversible reaction. This way the species can undergo decay. Press OK

→ Select "Initial concentration" to set initial conditions.

Using reset matrix assign the whole model a value of 0. Since there is a spin-up time for the hydraulic head there is no concentration in the first time step.

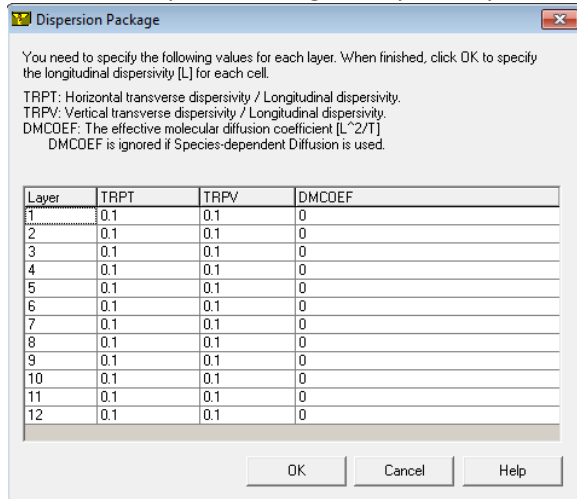
Leave the editor and save the changes

→ Select “Advection” to activate the advection package.



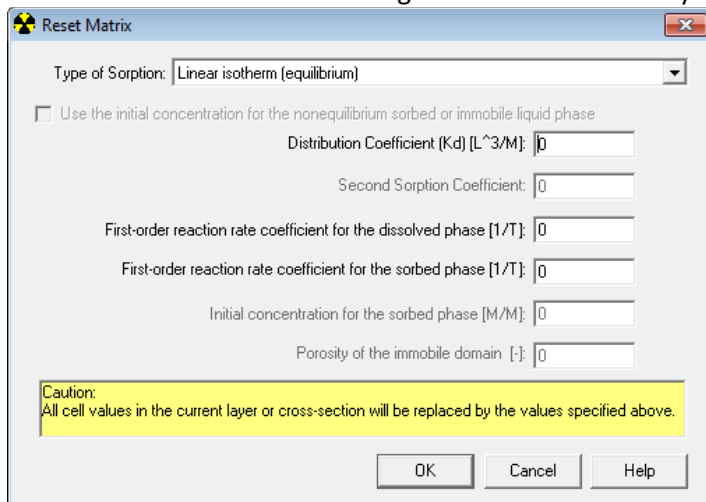
Choose the finite difference method as the solution scheme and select upstream weighting as the weighting scheme. Take a courant number of 0.1. Press OK

→ Select “Dispersion” to give dispersivity values.



TRPT= α_{Ty}/α_L and TRPV= α_{Tz}/α_L . Give both a value of 0.1 in all layers and then press OK. In the editor use reset matrix to assign a value of 0.02 to the entire model. Leave the editor and save the changes.

→ Select “Chemical reaction” to give values for the decay rate and the distribution coefficient.




Select the correct species and press edit. In the editor select reset matrix and select linear isotherm (equilibrium) as the type of sorption. Fill in the K_D value next to the distribution coefficient and the decay rate next to the first-order reaction rate coefficient for the dissolved phase. Always leave the first-order reaction rate coefficient for the sorbed phase as 0. Press OK and then duplicate all values for the other layers with the layer copy button.

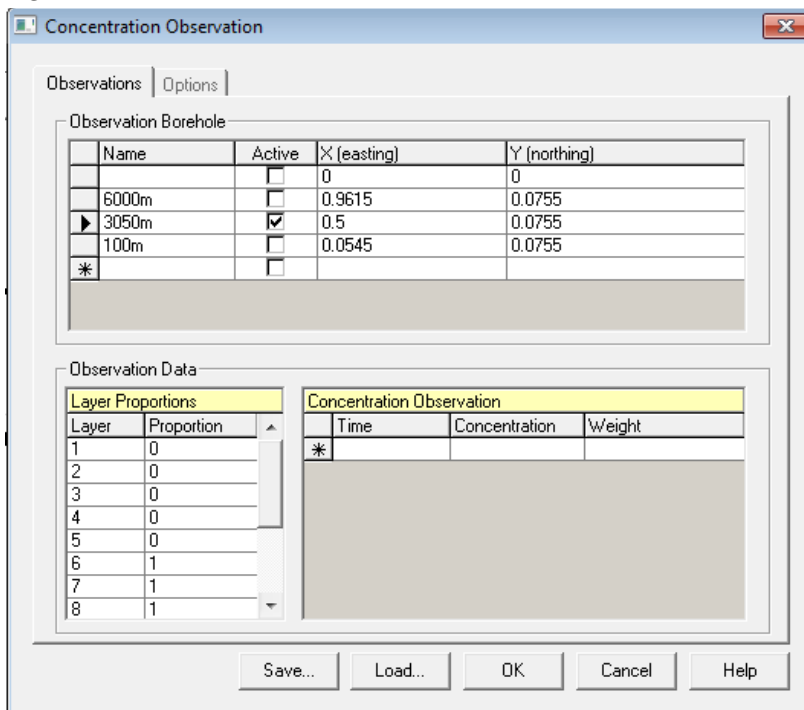
Leave the editor and save the changes

→ Select “Sink/Source Concentrations” and the “Time-Variant Specified-Concentration” to add a the contaminant.

Select the species and press edit to go to the editor. Right click the cell at row 14 and column 8 in the first layer and give it a value 1 at Flag. The specified concentration is 0 at time step 1. Press Ok and

then press the Change Stress Period button  to select period 2. Again right click the same cell and give it a flag of 1 and a specified concentration of 1. Do the same for time periods 3 to 8 but give a specified concentration of 0 as to create a plume instead of a continuous input of contaminant.

→ Select “Concentration Observation” to add observation points to view the concentration in the well.



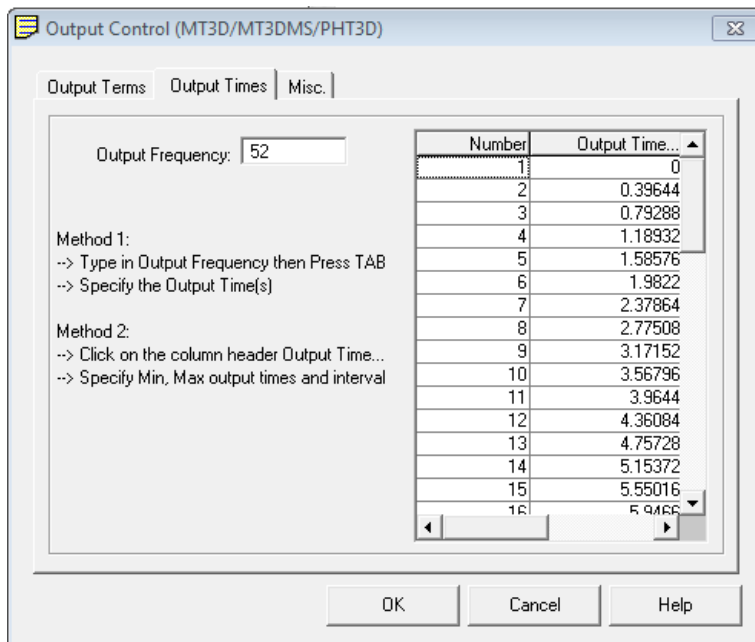
Add an observation point. Name in 100m and give it the coordinates (0.0545. 0.0755). In layer proportions give the layer with the well proportion 1 and the others proportion 0.

Add a second observation point. Name it 3050m and give it the coordinates (0.5. 0.0755). Again give the layers in which the well is located proportion 1 and the rest proportion 0.

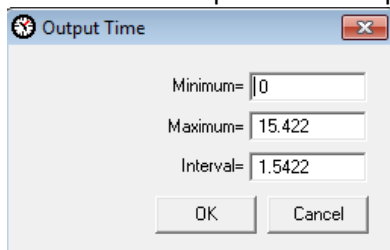
Add the last well. Name it 6000m and give it the coordinates (0.9615. 0.0755). Again give the layers in which the well is located proportion 1 and the rest proportion 0.

Activate the observation point in which the well is located and press OK.

→ Select “Output control” to change the output times.



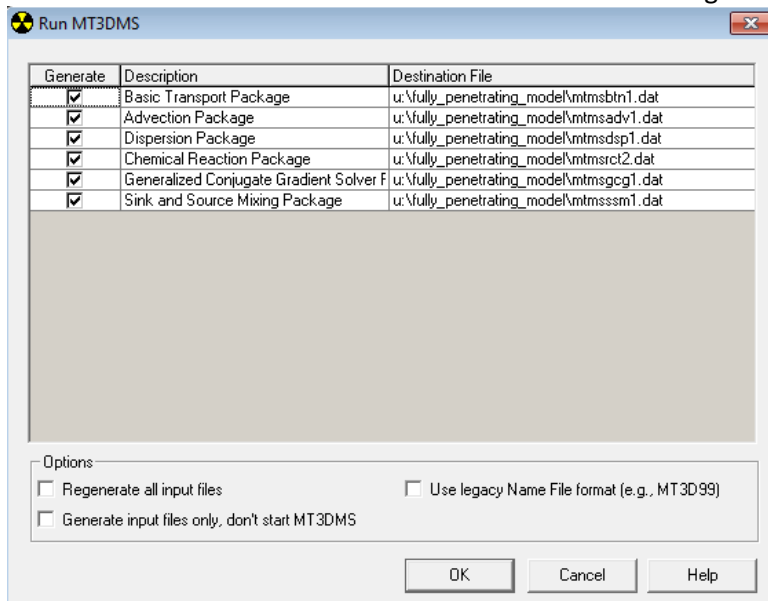
Go to the tab output times and press output time in the table. A new dialog box will open.



Give the interval a value which is 1% of the maximum value to get an output frequency of 102. Press Ok and press Ok again.

2. Run model

To run MT3DMS select "Run..." and the run MT3DMS dialog box opens.



Make sure all the boxes under generate are checked. Press OK. This will open the command prompt. Once the model is run press any key to leave the command prompt.

3. Check results

To see the breakthrough curves select “*MT3DMS*” under “*Models*” . then select “*View*” and finally “*Concentration-time curves...*” and select the species. Press OK. The Data tab shows a table with all calculated concentrations at the activated observation point. To save the table press Save Table. The tab Chart shows the breakthrough curve at the observation point.

To see the concentration plume in the different layers at different times select “*2-D Visualization*” under “*Tools*” . Select the *MT3DMS/SEAWAT* tab and choose the right species. Press OK. This will open the editor.

C – R-script

```
# read data
par(mfrow=c(1,1))
par(mar=c(4,4,2,2))
tracer=0
setwd("C:/Users/tan_i/Documents/Scriptie M/R")
obs <- read.table("RdataNew.txt", header = T, sep = "\t")
pairs(obs,gap=0.5)

#subset centre
obsX <- subset(obs, Scenario==181:184 & Zwstar>0.0015, select=c(Scenario:tmax))
obsZw <- subset(obs, Scenario==181:188 & Xstar==0.24 & i==0.0078 & Qstar==0.0013,
select=c(Scenario:tmax))
obsi <- subset(obs, Scenario==181:188 & Xstar==0.24 & Zwstar==0.0077 &
Qstar==0.0013 & m==25.5, select=c(Scenario:tmax))
obsQ <- subset(obs, Scenario>180 & Xstar==0.24 & Zwstar==0.0077 & i==0.0078 &
m==25.5 & R==5.5 & Da==15.5, select=c(Scenario:tmax))
obsm <- subset(obs, Scenario>180 & Xstar==0.24 & Zwstar==0.0077 & i==0.0078 &
Qstar==0.0013 & R==5.5 & Da==15.5, select=c(Scenario:tmax))
obsR <- subset(obs, Scenario>180 & Xstar==0.24 & Zwstar==0.0077 & i==0.0078 &
Qstar==0.0013 & m==25.5 & Da==15.5, select=c(Scenario:tmax))
obsDa <- subset(obs, Scenario>180 & Xstar==0.24 & Zwstar==0.0077 & i==0.0078 &
Qstar==0.0013 & m==25.5 & R==5.5, select=c(Scenario:tmax))

#####
# stepwise regression Cmax
modCmax <- lm(log(Cmax) ~ (Xstar+Zwstar+i+log(Qstar)+log(m)+R+Da)^2+
I(Xstar^2)+I(Zwstar^2)+I(i^2)+I((log(Qstar))^2)+I((log(m))^2)+
I(R^2)+I(Da^2), data =obs)
modCmax1<-step(modCmax, direction = "both", k = 3.84,trace=tracer)
plot(fitted(modCmax1),resid(modCmax1))
abline(h=0)
qqnorm(resid(modCmax1))
qqline(resid(modCmax1))
anova(modCmax1)
summary(modCmax1)

# Residuals
Cmax <- na.omit(obs$Cmax)
Cmax.fit <- fitted(modCmax1)
plot(log(Cmax), Cmax.fit, xlab="Ln C*max PMWIN", ylab="Ln C*max R",
xlim=c(-25,0), ylim=c(-25,0))
abline(0,1)
abline(2,1, lty=2)
abline(-2,1, lty=2)
legend("topleft", legend="A")

# Bland-Altman plot
mean <- (log(Cmax)+Cmax.fit)/2
diffs <- log(Cmax)-Cmax.fit
mdiff <- mean(diffs)
sddiff <- sd(diffs)
plot(mean, diffs, xlab="Mean", ylab="Difference")
abline(h=mdiff)
abline(h = mdiff + 1.96 * sddiff, lty = 2)
abline(h = mdiff - 1.96 * sddiff, lty = 2)
legend("topleft", legend="B")

# data frame Cmax, Xstar
newobs <- data.frame(Xstar=seq(0.015, 0.47, by=0.005),
Zwstar=0.0077,
i=0.0078,
Qstar=0.0013,
m=25.5,
R=5.5,
Da=15.5)
newobs <- within (newobs, {
conf <- predict (modCmax1, newdata=newobs, interval="confidence")
pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#Xstar
with (obsX, plot(Xstar, log(Cmax), xlim=c(0,0.47), ylim=c(-30,0), xlab="X*",
ylab="Ln C*max"))
```

```

with (newobs, {
  matlines(Xstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Xstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("bottomleft", legend=c("PMWIN points","Empirical formula",
                              "Confidence interval","Prediction interval"),
      col=c("black","black","red","green"), pch=c(1,NA,NA,NA),
      lty=c(NA, 1, 2, 2, 2), text.font=)
legend("topleft", legend="A")

# data frame Cmax, Zwstar
newobs <- data.frame(Xstar=0.24,
                    Zwstar=seq(0.0015, 0.014, by=0.0005),
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modCmax1, newdata=newobs, interval="confidence")
  pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#Zwstar
with (obsZw, plot(Zwstar, log(Cmax), xlim=c(0.0015, 0.014), ylim=c(-30,0),
                 xlab="Zw*", ylab="Ln C*max"))
with (newobs, {
  matlines(Zwstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Zwstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="B")

# data frame Cmax, i
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=seq(0.0055, 0.01, by=0.0005),
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modCmax1, newdata=newobs, interval="confidence")
  pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#i
with (obsi, plot(i, log(Cmax), xlim=c(0.0055, 0.01), ylim=c(-30,0), xlab="i",
                ylab="Ln C*max"))
with (newobs, {
  matlines(i, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(i, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="C")

# data frame Cmax, Qstar
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=seq(0.000013, 0.0026, by=0.000005),
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modCmax1, newdata=newobs, interval="confidence")
  pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#Qstar
with (obsQ, plot(log(Qstar), log(Cmax), xlim=c(-5,-2.5), ylim=c(-30,0),
                 xlab="Ln Q*", ylab="Ln C*max"))
with (newobs, {
  matlines(log(Qstar), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(Qstar), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="D")

```

```

# data frame Cmax, m
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=seq(1, 50, by=0.5),
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modCmax1, newdata=newobs, interval="confidence")
  pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#m
with (obsm, plot(log(m), log(Cmax), xlim=c(0,2), ylim=c(-30,0), xlab="Ln m",
               ylab="Ln C*max"))
with (newobs, {
  matlines(log(m), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(m), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="E")

# data frame Cmax, R
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=seq(1, 10, by=0.5),
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modCmax1, newdata=newobs, interval="confidence")
  pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#R
with (obsR, plot(R, log(Cmax), xlim=c(1,10), ylim=c(-30,0),
               xlab="R", ylab="Ln C*max"))
with (newobs, {
  matlines(R, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(R, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="F")

# data frame Cmax, Da
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=seq(0, 31, by=0.5))
newobs <- within (newobs, {
  conf <- predict (modCmax1, newdata=newobs, interval="confidence")
  pred <- predict (modCmax1, newdata=newobs, interval="prediction")
})
#Da
with (obsDa, plot(Da, log(Cmax), xlim=c(0, 31), ylim=c(-30,0), xlab="Da*",
                 ylab="Ln C*max"))
with (newobs, {
  matlines(Da, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Da, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="G")

#####
# stepwise regression tmax
modtmax <- lm(sqrt(tmax) ~ (Xstar+Zwstar+i+log(Qstar)+m+R+(sqrt(Da)))^2+
              I(Xstar^2)+I(Zwstar^2)+I(i^2)+I((log(Qstar))^2)+I(m^2)+
              I(R^2)+I((sqrt(Da))^2), data = obs)
modtmax1<-step(modtmax, direction = "both", k = 3.84,trace=tracer)
plot(fitted(modtmax1),resid(modtmax1))
abline(h=0)
qqnorm(resid(modtmax1))
qqline(resid(modtmax1))
anova(modtmax1)

```



```

summary(modtmax1)

# Residuals
tmax <- na.omit(obs$tmax)
tmax.fit <- fitted(modtmax1)
plot(sqrt(tmax), tmax.fit, xlab="Sqrt t*max PMWIN", ylab="Sqrt t*max R",
      xlim=c(0.4,2.2), ylim=c(0.4,2.2))
abline(0,1)
abline(0.1,1, lty=2)
abline(-0.1,1, lty=2)
legend("topleft", legend="A")

# Bland-Altman plot
mean <- (sqrt(tmax)+tmax.fit)/2
diffs <- sqrt(tmax)-tmax.fit
mdiff <- mean(diffs)
sddiff <- sd(diffs)
plot(mean, diffs, xlab="Mean", ylab="Difference")
abline(h=mdiff)
abline(h = mdiff + 1.96 * sddiff, lty = 2)
abline(h = mdiff - 1.96 * sddiff, lty = 2)
legend("topleft", legend="B")

# data frame tmax, Xstar
newobs <- data.frame(Xstar=seq(0.015, 0.47, by=0.005),
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#Xstar
with (obsX, plot(Xstar, sqrt(tmax), xlim=c(0,0.47), ylim=c(0,1.5), xlab="X*",
                ylab="Sqrt t*max"))
with (newobs, {
  matlines(Xstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Xstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("bottomright", legend=c("PMWIN points","Empirical formula",
                               "Confidence interval","Prediction interval"),
      col=c("black","black","red","green"), pch=c(1,NA,NA,NA),
      lty=c(NA, 1, 2, 2, 2), text.font=)
legend("topleft", legend="A")

# data frame tmax, Zwstar
newobs <- data.frame(Xstar=0.24,
                    Zwstar=seq(0.0015, 0.014, by=0.0005),
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#Zwstar
with (obsZw, plot(Zwstar, sqrt(tmax), xlim=c(0.0015, 0.014), ylim=c(0,1.5),
                xlab="Zw*", ylab="Sqrt t*max"))
with (newobs, {
  matlines(Zwstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Zwstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="B")

# data frame tmax, i
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=seq(0.0055, 0.01, by=0.0005),
                    Qstar=0.0013,

```

```

        m=25.5,
        R=5.5,
        Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#i
with (obsi, plot(i, sqrt(tmax), xlim=c(0.0055, 0.01), ylim=c(0,1.5), xlab="i",
  ylab="Sqrt t*max"))
with (newobs, {
  matlines(i, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(i, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="C")

# data frame tmax, Qstar
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=seq(0.000013, 0.0026, by=0.000005),
  m=25.5,
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#Qstar
with (obsQ, plot(log(Qstar), sqrt(tmax), xlim=c(-5, -2.5), ylim=c(0,1.5),
  xlab="Ln Q*", ylab="Sqrt t*max"))
with (newobs, {
  matlines(log(Qstar), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(Qstar), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="D")

# data frame tmax, m
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=0.0013,
  m=seq(1, 50, by=0.5),
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#m
with (obsm, plot(m, sqrt(tmax), xlim=c(1, 50), ylim=c(0,1.5), xlab="m",
  ylab="Sqrt t*max"))
with (newobs, {
  matlines(m, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(m, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="E")

# data frame tmax, R
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=0.0013,
  m=25.5,
  R=seq(1, 10, by=0.5),
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#R
with (obsR, plot(R, sqrt(tmax), xlim=c(1,10), ylim=c(0,1.5),
  xlab="R", ylab="Sqrt t*max"))
with (newobs, {

```

```

    matlines(R, conf, col=c(1,2,2), lty=c(1,2,2))
    matlines(R, pred, col=c(1,3,3), lty=c(1,2,2))
  })
  legend("topleft", legend="F")

# data frame tmax, Da
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=seq(0, 31, by=0.5))
newobs <- within (newobs, {
  conf <- predict (modtmax1, newdata=newobs, interval="confidence")
  pred <- predict (modtmax1, newdata=newobs, interval="prediction")
})
#Da
with (obsDa, plot(sqrt(Da), sqrt(tmax), xlim=c(0,6), ylim=c(0,1.5), xlab="Da*",
                 ylab="Sqrt t*max"))
with (newobs, {
  matlines(sqrt(Da), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(sqrt(Da), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="G")

#####
# stepwise regression Chalf
modChalf <- lm(log(Chalf) ~ (Xstar+Zwstar+i+log(Qstar)+log(m)+R+Da)^2+
              I(Xstar^2)+I(Zwstar^2)+I(i^2)+I((log(Qstar))^2)+I((log(m))^2)+
              I(R^2)+I(Da^2), data =obs)
modChalf1<-step(modChalf, direction = "both", k = 3.84,trace=tracer)
plot(fitted(modChalf1),resid(modChalf1))
abline(h=0)
qqnorm(resid(modChalf1))
qqline(resid(modChalf1))
anova(modChalf1)
summary(modChalf1)

# Residuals
Chalf <- na.omit(obs$Chalf)
Chalf.fit <- fitted(modChalf1)
plot(log(Chalf), Chalf.fit, xlab="Ln C*half PMWIN", ylab="Ln C*half R",
     xlim=c(-25,0), ylim=c(-25,0))
abline(0,1)
abline(2,1, lty=2)
abline(-2,1, lty=2)
legend("topleft", legend="A")

# Bland-Altman plot
mean <- (log(Chalf)+Chalf.fit)/2
diffs <- log(Chalf)-Chalf.fit
mdiff <- mean(diffs)
sddiff <- sd(diffs)
plot(mean, diffs, xlab="Mean", ylab="Difference")
abline(h=mdiff)
abline(h = mdiff + 1.96 * sddiff, lty = 2)
abline(h = mdiff - 1.96 * sddiff, lty = 2)
legend("topleft", legend="B")

# data frame Chalf, Xstar
newobs <- data.frame(Xstar=seq(0.015, 0.47, by=0.005),
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#Xstar
with (obsX, plot(Xstar, log(Chalf), xlim=c(0,0.47), ylim=c(-10,0), xlab="X*",

```

```

        ylab="Ln C*half"))
with (newobs, {
  matlines(Xstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Xstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("bottomleft", legend=c("PMWIN points", "Empirical formula",
  "Confidence interval", "Prediction interval"),
  col=c("black", "black", "red", "green"), pch=c(1,NA,NA,NA),
  lty=c(NA, 1, 2, 2, 2), text.font=)
legend("topleft", legend="A")

# data frame Chalf, Zwstar
newobs <- data.frame(Xstar=0.24,
  Zwstar=seq(0.0015, 0.014, by=0.0005),
  i=0.0078,
  Qstar=0.0013,
  m=25.5,
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#Zwstar
with (obsZw, plot(Zwstar, log(Chalf), xlim=c(0.0015, 0.014), ylim=c(-10,0),
  xlab="Zw*", ylab="Ln C*half"))
with (newobs, {
  matlines(Zwstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Zwstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="B")

# data frame Chalf, i
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=seq(0.0055, 0.01, by=0.0005),
  Qstar=0.0013,
  m=25.5,
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#i
with (obsi, plot(i, log(Chalf), xlim=c(0.0055, 0.01), ylim=c(-10,0), xlab="i",
  ylab="Ln C*half"))
with (newobs, {
  matlines(i, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(i, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="C")

# data frame Chalf, Qstar
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=seq(0.000013, 0.0026, by=0.000005),
  m=25.5,
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#Qstar
with (obsQ, plot(log(Qstar), log(Chalf), xlim=c(-5,-2.5), ylim=c(-10,0),
  xlab="Ln Q*", ylab="Ln C*half"))
with (newobs, {
  matlines(log(Qstar), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(Qstar), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="D")

```

```

# data frame Chalf, m
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=seq(1, 50, by=0.5),
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#m
with (obsm, plot(log(m), log(Chalf), xlim=c(1,2), ylim=c(-10,0), xlab="Ln m",
                ylab="Ln C*half"))
with (newobs, {
  matlines(log(m), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(m), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="E")

# data frame Chalf, R
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=seq(1, 10, by=0.5),
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#R
with (obsR, plot(R, log(Chalf), xlim=c(1,10), ylim=c(-10,0),
                xlab="R", ylab="Ln C*half"))
with (newobs, {
  matlines(R, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(R, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="F")

# data frame Chalf, Da
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=seq(0, 31, by=0.5))
newobs <- within (newobs, {
  conf <- predict (modChalf1, newdata=newobs, interval="confidence")
  pred <- predict (modChalf1, newdata=newobs, interval="prediction")
})
#Da
with (obsDa, plot(Da, log(Chalf), xlim=c(0,31), ylim=c(-10,0), xlab="Da*",
                ylab="Ln C*half"))
with (newobs, {
  matlines(Da, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Da, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="G")

#####
# stepwise regression thalf
modthalf <- lm(sqrt(thalf) ~ (log(Xstar)+Zwstar+i+log(Qstar)+m+R+Da)^2+
              I((log(Xstar))^2)+I(Zwstar^2)+I(i^2)+I((log(Qstar))^2)+I(m^2)+
              I(R^2)+I(Da^2), data =obs)
modthalf1<-step(modthalf, direction = "both", k = 3.84,trace=tracer)
plot(fitted(modthalf1),resid(modthalf1))
abline(h=0)
qqnorm(resid(modthalf1))
qqline(resid(modthalf1))
anova(modthalf1)

```

```

summary(modthalf1)

# Residuals
thalf <- na.omit(obs$thalf)
thalf.fit <- fitted(modthalf1)
plot(sqrt(thalf), thalf.fit, xlab="Sqrt t*half PMWIN", ylab="Sqrt t*half R",
      xlim=c(0,2), ylim=c(0,2))
abline(0,1)
abline(0.1,1, lty=2)
abline(-0.1,1, lty=2)
legend("topleft", legend="A")

# Bland-Altman plot
mean <- (sqrt(thalf)+thalf.fit)/2
diffs <- sqrt(thalf)-thalf.fit
mdiff <- mean(diffs)
sddiff <- sd(diffs)
plot(mean, diffs, xlab="Mean", ylab="Difference")
abline(h=mdiff)
abline(h = mdiff + 1.96 * sddiff, lty = 2)
abline(h = mdiff - 1.96 * sddiff, lty = 2)
legend("topleft", legend="B")

# data frame thalf, Xstar
newobs <- data.frame(Xstar=seq(0.015, 0.47, by=0.005),
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#Xstar
with (obsX, plot(log(Xstar), sqrt(thalf), xlim=c(-2,-0.25), ylim=c(0,1.5),
                xlab="Ln X*", ylab="t*half"))
with (newobs, {
  matlines(log(Xstar), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(Xstar), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend=c("PMWIN points", "Empirical formula",
                           "Confidence interval", "Prediction interval"),
      col=c("black", "black", "red", "green"), pch=c(1,NA,NA,NA),
      lty=c(NA, 1, 2, 2, 2), text.font=)
legend("bottomright", legend="A")

# data frame thalf, Zwstar
newobs <- data.frame(Xstar=0.24,
                    Zwstar=seq(0.0015, 0.014, by=0.00005),
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#Zwstar
with (obsZw, plot(Zwstar, sqrt(thalf), xlim=c(0.0015, 0.014), ylim=c(0,1.5),
                xlab="Zw*", ylab="t*half"))
with (newobs, {
  matlines(Zwstar, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Zwstar, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="B")

# data frame thalf, i
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=seq(0.0055, 0.01, by=0.0005),
                    Qstar=0.0013,

```

```

        m=25.5,
        R=5.5,
        Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#i
with (obsi, plot(i, sqrt(thalf), xlim=c(0.0055, 0.01), ylim=c(0,1.5), xlab="i",
  ylab="t*half"))
with (newobs, {
  matlines(i, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(i, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="C")

# data frame thalf, Qstar
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=seq(0.000013, 0.0026, by=0.000005),
  m=25.5,
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#Qstar
with (obsQ, plot(log(Qstar), sqrt(thalf), xlim=c(-5,-2.5), ylim=c(0,1.5),
  xlab="Ln Q*", ylab="t*half"))
with (newobs, {
  matlines(log(Qstar), conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(log(Qstar), pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="D")

# data frame thalf, m
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=0.0013,
  m=seq(1, 50, by=0.5),
  R=5.5,
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#m
with (obsm, plot(m, sqrt(thalf), xlim=c(1, 50), ylim=c(0,1.5), xlab="m",
  ylab="t*half"))
with (newobs, {
  matlines(m, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(m, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="E")

# data frame thalf, R
newobs <- data.frame(Xstar=0.24,
  Zwstar=0.0077,
  i=0.0078,
  Qstar=0.0013,
  m=25.5,
  R=seq(1, 10, by=0.5),
  Da=15.5)
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#R
with (obsR, plot(R, sqrt(thalf), xlim=c(1,10), ylim=c(0,1.5),
  xlab="R", ylab="t*half"))
with (newobs, {

```

```

    matlines(R, conf, col=c(1,2,2), lty=c(1,2,2))
    matlines(R, pred, col=c(1,3,3), lty=c(1,2,2))
  })
  legend("topleft", legend="F")

# data frame thalf, Da
newobs <- data.frame(Xstar=0.24,
                    Zwstar=0.0077,
                    i=0.0078,
                    Qstar=0.0013,
                    m=25.5,
                    R=5.5,
                    Da=seq(0, 31, by=0.805))
newobs <- within (newobs, {
  conf <- predict (modthalf1, newdata=newobs, interval="confidence")
  pred <- predict (modthalf1, newdata=newobs, interval="prediction")
})
#Da
with (obsDa, plot(Da, sqrt(thalf), xlim=c(0,31), ylim=c(0,1.5), xlab="Da*",
                 ylab="t*half"))
with (newobs, {
  matlines(Da, conf, col=c(1,2,2), lty=c(1,2,2))
  matlines(Da, pred, col=c(1,3,3), lty=c(1,2,2))
})
legend("topleft", legend="G")

```