Identification of Parameter Importance for Benzene Transport in the Unsaturated Zone Using Global Sensitivity Analysis

Meirav Cohen[1], Nimrod Schwartz[2], and Ravid Rosenzweig[3]

[1] The Dead Sea & Arava Science Center (DSASC), Mitzpe Ramon, Israel

[2] The Institute of Environmental Sciences, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot 76100 , Israel

[3] Geological Survey of Israel, Jerusalem, Israel

This document provides additional information regarding the paper's methodology and results.

A1. Literature values for first order biodegradation rate constant (day^{-1})

Table S1 – Literature values for benzene first order biodegradation rate constant (day^{-1})

First order Biodegradation Rate Constant (day^{-1})	Benzene conc. (mg/l)	Saturation level WC (%)	Calculation form	Comments	Reference
174*	0.000104	Unsaturated zone 5-10% WC	Calculated and model-calibrated	Field data - upward flux from NAPL	Lahvis et al., (1999)
4.01	0.00443				
1.48	0.0197				
0.066	0.231	Capillary zone 12-36% WC			
0.31	0.231				
0.48	360-850	Unsaturated zone 3% WC	Calculated and model-calibrated	Column experiments upward flux from NAPL	DeVaull et al., (1997)
0.144	1550	18% WC			
6.48			Calculated and model-calibrated	Summary of measured data	
28.8	~0.005-2	Unsaturated	Calculated and model-calibrated	Field data - upward flux from NAPL	Hers et al., (2000)
9.6		Unsaturated		Field data	Ririe & Sweeney, (1995)
0.035	1770	Unsaturated	Modeled	Lower modeled value	Berlin et al., (2016)
0.05				Upper modeled value	
0.98	**Overall median value**				
0.48	**Overall median value excluding the highest value (174)**				

A2. Literature values for benzene sorption coefficient

Table S2 - Literature values of benzene sorption constant

K_{OW} Octanol-water partition coefficient (log (m^3/kg)	K_{OC} Organic-carbon partition coefficient (log m^3/kg)	K_{OW} Octanol-water partition coefficient (m^3/kg)	K_{OC} Organic-carbon partition coefficient (m^3/kg)	Freundlich sorption coefficients		Fraction organic carbon (f_{oc})	**Kd Soil Sorption coefficient (m^3/kg) ($Kd=f_{oc}K_{oc}$)**	Water content	Soil type	Comments	Reference
				$1/n_f$	K_f (kg/kg) $(kg/m^3)^{-}$ 1/nf						
-	1.72E-03	-	5.25E-02			0.05*	**2.62E-03**	-	-	Calculated from K_{ow}	Karickhoff, (1981)
-	1.82E-03	-	6.61E-02			0.05*	**3.30E-03**	-	-	Calculated from solubility	
2.11E-03	1.78E-03	1.30E-01	6.00E-02			0.05*	**3.00E-03**	Saturated		Literature	
-	1.92E-03		8.30E-02			0.05*	**4.15E-03**	-	-	Measured	Kenaga, (1980)
-	1.85E-03		7.10E-02			0.05*	**3.55E-03**	-	-	Calculated	
	2.02E-03		1.00			0.001-0.004	**7E-05 - 4.2E-4**	Saturated	Sandy-loam contaminated with (BTEX)	Measured	Mackay et al., (1996)
	-1.10		0.079			0.05*	**3.95E-03**	-		Literature	Wiedemeier et al., (1996)
						0.05	**2.86E-03**	-	Sandy aquifer materials	Measured	Baek et al., (2003)
				1.11E+00	2.00E-07			-	Sand	Measured	Du et al., (2010)
				1.08E+00	2.30E-06	0.03		Unsaturated	Fine	Batch	English & Loehr,

									sandy loam	sorption test	(1991)
5.10E-04							**5.10E-04**		Clay	Measured	Osagie & Owabor, (2015)
				1.11E+00	1.11E-03				Sand	Literature	Donahue et al., (1999)
				9.17E-01	4.88E-04						
							3.33E-04 - 3.09E-02		Various clays	Measured	
							2.00E-04 - 2.23E-02		Silty clay		
						0.043	**2.22E-03**		Clay	Measured	Redding et al., (2002)
						0.075	**8.39E-03**		Clay		
							3.70E-05 - 4.6E-05		Clay	Measured	Karickhoff, (1981)

A3. Sobol sampling and confidence interval

Parameter spaces were sampled using the Sobol quasi-random, cross-sampling strategy (Sobol, 2001). Rather than generating random numbers, this technique generates a uniform distribution in the probability space. The distribution appears qualitatively random, but sampling only takes place in regions of the probability function that were not previously sampled.

In order to assess the accuracy of the Sobol indices, confidence intervals of the indices should be constructed. The analytical procedure for confidence interval calculation involves repeating the $r(2k+2)$ model runs several times, which is too time consuming and computationally demanding in this case. Therefore the bootstrapping approach was used instead (Efron & Tibshirani, 1986). Archer et al. (1997) suggested using bootstrap confidence intervals to produce confidence intervals of complicated data structures. The bootstrapping approach is based on resampling the parameter space of the already available data many times with replacement (randomly selecting values and allowing for duplicates), and constructing a distribution of the output (Efron & Tibshirani, 1986). Here, resamples were taken from the existing dataset Y with replacement, and the indices' values were recalculated. That gives an estimate of the mean and variance of each of the indices and allows calculation of the confidence interval. The method thus relies on computational cost rather than on an analytical cost (running the model again). Here, the samples used for the model evaluation were resampled 1000 times with replacement, and 95% confidence intervals were constructed (Archer et al., 1997).

Still, confidence intervals for the first-order indices (S_1), with the Sobol sampling method gave values of more than 100%. This was also observed by Brunetti et al. (2016) and Hartmann et al. (2018) who also studied transport in unsaturated media. This may be a result of insufficient sample size, since Sobol's convergence requires a very large sample size (Saltelli et al., 2004). Therefore, here the S_1 values were extracted using the delta method of Plischke et al. (2013), calculating S_1 values from a given data through emulators and bootstrapping rather than running the model itself multiple times.

A4. The RBF approximation equations:

The RBF approximation is a weighted summation of n_a basis functions (and a polynomial or constant value) that can approximate the predictive response $Y(x)$ at a point $\boldsymbol{x_i}$ as follows:

1. $$Y(\boldsymbol{x}) = \sum_{i=1}^{n_a} \omega_i \, f(\|\boldsymbol{x} - \boldsymbol{x}_i\|) = f(\boldsymbol{x})\boldsymbol{\omega}$$

where $f = \{f_1, f_2, \dots, f_{n_a}\}$ is the vector of the basis functions, ω_i is the i^{th} component of the radial basis coefficient vector $\boldsymbol{\omega} = \{\omega_1, \omega_2, \dots, \omega_{n_a}\}^T$ and $\|\boldsymbol{x} - \boldsymbol{x}_i\|$ is the Euclidian distance between two sample points.

There are various forms of basis functions; the most popular is the Gaussian kernel function:

2. $$f(\|\boldsymbol{x} - \boldsymbol{x}_i\|) = exp\left(\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|^2}{\varepsilon^2}\right)$$

where ε is the shape parameter which determines the spread of the kernel function f_i. The coefficient vector $\boldsymbol{\omega}$ can be obtained by enforcing the accurate interpolation condition, i.e.,

3. $$\begin{bmatrix} y(\boldsymbol{x}_1) \\ y(\boldsymbol{x}_2) \\ \vdots \\ y(\boldsymbol{x}_{n_a}) \end{bmatrix} = \begin{bmatrix} f(x_{11}) & f(x_{12}) & \dots & f(x_{1n_a}) \\ f(x_{21}) & f(x_{22}) & \cdots & f(x_{2n_a}) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_{n_a1}) & f(x_{n_a2}) & \cdots & f(x_{n_an_a}) \end{bmatrix}$$

where $f_{uv} = f(\|\boldsymbol{x_u} - \boldsymbol{x}_v\|)$. In a matrix form, Equation 3 can be rewritten as $\boldsymbol{Y_a} = \boldsymbol{F\omega}$. This equation has a unique solution $\boldsymbol{\omega} = \boldsymbol{F^{-1}Y_a}$ if and only if all the sample points are different from each other. Therefore, the filled values for the remaining n_m locations, for which the model responses are missing due to simulation crashes, can be approximated by Equation 4:

4. $$Y(\boldsymbol{x}_i) = f(\boldsymbol{x}_i)\boldsymbol{F^{-1}Y_a} \quad (i = 1,2, \dots, n_a)$$

In this study n_a was set to the number of non-missing sample points. The shape parameter ε was changed from the default of the mean distance between neighboring nodes, to the average KNN of the response function, where k was set to five.

A5. Morris homogenous sandy soil results:

Tables S3 and S4 present μ*, μ* confidence interval and σ results of the Morris analysis in homogeneous media for the flux and concentration respectively, with different methods for missing data handling.

Table S3: Morris indices values of total benzene flux to the aquifer, indicating the different methods for missing parameter handling

Method	Output value	Input parameter						
		α	**n**	**K_s**	**α_l**	**k_d**	**λ_k**	**z**
Zero replaced	μ*	1.001	3.080	0.444	0.043	1.941	6.791	4.461
	μ* confidence interval	0.756	1.160	0.181	0.017	0.812	2.000	1.679
	σ	5.675	10.343	1.595	0.143	6.253	16.472	13.494
5NN	μ*	1.271	3.379	0.614	0.077	1.941	6.791	4.638
	μ* confidence interval	0.734	1.498	0.317	0.054	0.811	1.908	1.663
	σ	6.480	10.706	2.629	0.459	6.253	16.472	13.512
45NN	μ*	1.161	3.399	0.587	0.057	1.941	6.793	4.737
	μ* confidence interval	0.697	1.406	0.271	0.026	0.658	1.882	1.946
	σ	5.843	10.445	2.056	0.193	6.253	16.472	13.531
RBF	μ*	1.288	4.269	0.787	0.153	1.973	6.850	5.305
	μ* confidence interval	0.727	1.502	0.416	0.152	0.723	1.998	1.944
	σ	6.052	13.149	3.564	1.237	6.249	16.458	15.322
Full trajectories removed	μ*	0.713	3.069	0.451	0.045	1.913	6.544	4.331
	μ* confidence interval	0.546	1.440	0.170	0.021	0.787	2.258	1.726
	σ	4.292	10.492	1.614	0.148	6.167	16.027	12.787
Average	μ*	1.087	3.439	0.576	0.075	1.941	6.754	4.694
	μ* STDEV	0.238	0.490	0.141	0.046	0.021	0.120	0.376
	μ* confidence interval	0.692	1.401	0.271	0.054	0.758	2.009	1.792
	μ* confidence interval STDEV	0.084	0.141	0.102	0.057	0.067	0.149	0.142
	σ	5.668	11.027	2.292	0.436	6.235	16.380	13.729
	σ STDEV	0.826	1.194	0.827	0.466	0.038	0.198	0.944

Table S4: Morris indices values of final concentration of benzene in aquifer, indicating the different methods for missing parameter handling

Method	Output value	Input parameter						
		α	**n**	**K_s**	**α_l**	**k_d**	**Λ_k**	**Z**
Zero replaced	μ*	0.0169	0.0894	0.0286	0.0040	0.1516	0.5540	0.2188
	μ* confidence interval	0.0230	0.0413	0.0142	0.0025	0.0643	0.1273	0.0755
	σ	0.1886	0.3897	0.1172	0.0217	0.4891	1.0351	0.6164
5NN	μ*	0.0294	0.1004	0.0356	0.0060	0.1516	0.5540	0.2290
	μ* confidence interval	0.0258	0.0454	0.0168	0.0046	0.0635	0.1339	0.0740
	σ	0.2295	0.3880	0.1446	0.0314	0.4891	1.0351	0.6188
45NN	μ*	0.0258	0.1049	0.0356	0.0048	0.1516	0.5542	0.2337
	μ* confidence interval	0.0242	0.0511	0.0148	0.0028	0.0631	0.1274	0.0728
	σ	0.1972	0.3846	0.1311	0.0230	0.4891	1.0350	0.6185
RBF	μ*	0.0189	0.1046	0.0306	0.0076	0.1545	0.5575	0.2536
	μ* confidence interval	0.0207	0.0606	0.0135	0.0054	0.0672	0.1270	0.0835
	σ	0.1972	0.3846	0.1311	0.0230	0.4891	1.0350	0.6185
Full trajectories removed	μ*	0.0046	0.0865	0.0300	0.0041	0.1511	0.5416	0.2195
	μ* confidence interval	0.0030	0.0465	0.0184	0.0031	0.0636	0.1198	0.0789
	σ	0.0241	0.3702	0.1216	0.0225	0.4892	1.0207	0.6095
Average	μ*	0.019	0.097	0.032	0.005	0.152	0.552	0.231
	μ* STDEV	0.010	0.009	0.003	0.002	0.001	0.006	0.014
	μ* confidence interval	0.019	0.049	0.016	0.004	0.064	0.127	0.077
	μ* confidence interval STDEV	0.009	0.007	0.002	0.001	0.002	0.005	0.004
	σ	0.167	0.383	0.129	0.024	0.489	1.032	0.616
	σ STDEV	0.082	0.008	0.011	0.004	0.000	0.006	0.004

A6. Sobol homogenous sandy soil results:

Tables S5 and S6 present first order results of Sobol indices values determined by the Sobol analysis in homogeneous media for benzene total flux and final concentration in the aquifer, respectively, with the different methods for missing data imputation.

Table S5: First order Sobol indices values for benzene total flux to the aquifer with the different methods for missing parameter handling:

Method	**Output value**	**Input parameter**			
		n	k_d	λ_k	z
Zero replaced	S_1	0.003	0.001	0.107	0.019
	S_1 confidence interval	0.001	0.001	0.014	0.005
5 KNN	S_1	0.003	0.001	0.111	0.021
	S_1 confidence interval	0.001	0.001	0.015	0.005
233 KNN	S_1	0.003	0.001	0.109	0.022
	S_1 confidence interval	0.001	0.001	0.013	0.005
RBF	S_1	0.002	0.001	0.135	0.012
	S_1 confidence interval	0.001	0.001	0.017	0.003
Average	S_1	0.003	0.001	0.115	0.018
	STDEV S_1	0.00049	0.00012	0.01307	0.00453
	S_1 confidence interval	0.001	0.001	0.015	0.005
	STDEV S_1 confidence interval	0.00022	0.00006	0.00173	0.00130

Table S6: First order Sobol indices values for final concentration of benzene in the aquifer with the different methods for missing parameter handling:

Method	**Output value**	**Input parameter**			
		n	k_d	λ_k	z
Zero replaced	S_1	0.002	0.001	0.130	0.011
	S_1 confidence interval	0.001	0.001	0.014	0.003
5 KNN	S_1	0.002	0.001	0.133	0.012
	S_1 confidence interval	0.001	0.001	0.014	0.003
233 KNN	S_1	0.002	0.001	0.131	0.012
	S_1 confidence interval	0.001	0.001	0.013	0.003
RBF	S_1	0.002	0.001	0.135	0.012
	S_1 confidence interval	0.001	0.001	0.017	0.003
Average	S_1	0.002	0.001	0.132	0.012
	STDEV S_1	0.00009	0.00006	0.00233	0.00057

	S_1 confidence interval	0.001	0.001	0.015	0.003
	STDEV S_1 confidence interval	0.00009	0.00005	0.00147	0.00016

Tables S7 and S8 present total Sobol indices values of the Sobol analysis in homogeneous media for benzene total flux and final concentration in the aquifer respectively, with the different methods for missing data imputation.

Table S7: Total Sobol indices values for benzene total flux to the aquifer with the different methods for missing parameter handling:

Method	**Output value**	**Input parameter**			
		n	k_d	λ_k	z
Zero replaced	S_T	0.226	0.107	1.046	0.781
	S_T confidence interval	0.160	0.095	0.463	0.421
5 KNN	S_T	0.215	0.105	1.047	0.785
	S_T confidence interval	0.208	0.104	0.588	0.409
233 KNN	S_T	0.219	0.106	1.045	0.782
	S_T confidence interval	0.178	0.107	0.599	0.450
RBF	S_T	0.225	0.105	1.046	0.776
	S_T confidence interval	0.189	0.114	0.573	0.459
Average	S_T	0.221	0.106	1.046	0.781
	STDEV S_T	0.00534	0.00065	0.00095	0.00360
	ST confidence interval	0.184	0.105	0.556	0.435
	STDEV S_T confidence interval	0.02029	0.00788	0.06268	0.02372

Table S8: Total Sobol indices values for benzene final concentration in the aquifer with the different methods for missing parameter handling:

Method	**Output value**	**Input parameter**			
		n	k_d	λ_k	z
Zero replaced	S_T	0.173	0.087	1.204	0.583
	S_T confidence interval	0.121	0.084	0.585	0.346
5 KNN	S_T	0.164	0.086	1.202	0.588
	S_T confidence interval	0.136	0.087	0.635	0.355
233 KNN	S_T	0.168	0.087	1.203	0.583
	S_T confidence interval	0.101	0.075	0.499	0.304
RBF	S_T	0.164	0.086	1.200	0.581
	S_T confidence interval	0.110	0.091	0.489	0.312
Average	S_T	0.167	0.087	1.202	0.584

	STDEV S_T	0.00415	0.00055	0.00162	0.00301
	ST confidence interval	0.117	0.084	0.552	0.329
	STDEV S_T confidence interval	0.01500	0.00672	0.07055	0.02498

A7. Morris analysis for heterogeneous media

a. Clay layer characterization

Table S9 presents overall hydraulic conductivities and soil types of the different soil categories used for characterization of field site clay layers and Table S10 presents an example from one such site.

Table S9: Soil type characterization according to hydraulic conductivities values:

Category	Name	Soil type	K (m/day)	Source
Gravel	gravel	gravel	8640	Bear (1972)
	gravel sand	sandy gravel		
Sand	sand	sand	7.128	Carsel & Parrish (1988)
	sand gravel	gravely sand	7.128	
	kurkar sand		7.128	
	filling sub		7.128	
	sand silt	silty sand	3.502	Carsel & Parrish (1988)
	silt sand	sandy silt	1.061	Carsel & Parrish (1988)
Clayey Sand	sand clay	clayey sand	0.1888	Rosetta assuming 55% sand and 45% clay
	sand with clay		0.1888	
	sand clay gravel	clayey gravely sand	0.1888	
Clay	silt	silt	0.06	Carsel & Parrish (1988)
	clay	clay	0.048	Carsel & Parrish (1988)
	limestone	limestone	8.64E-06	Bear (1972)
	clay sand	sandy clay	0.0288	Carsel & Parrish (1988)
	clay silt	silty clay	0.0048	Carsel & Parrish (1988)
	gravel sand clay	gravely sandy clay	0.0288	

Table S10: Example of clay layers data extraction from one site

Data received						Data extrapolated			
From depth (m)	To depth (m)	Sand /sand-gravel (%)	Sand clay/ clayey sand (%)**	Clay (%)	Number of boreholes	Sand (%)	Clay (%)	Thickness of clay (m)	Overall clay layer thickness (m)
1	3	33%	0%	67%	3	33.3%	67%	2	3.5
	4	0%	50%	50%	3	27.5%	73%	1	
	5	25%	50%	25%	2	52.5%	48%	0.5	
	5.5	100%	0%	0%	4	100.0%	0%	1	
	6.5	0%	0%	100%	2	0.0%	100%	0.5	4
	7	0%	0%	100%	2	0.0%	100%	2	
	9	0%	67%	33%	1	0.0%	100%	1	
	10	0%	0%	100%	2	36.7%	63%	0.5	
10.5	11	100%	0%	0%	1	100.0%	0%	0.5	
	11.5	0%	100%	0%	1	55.0%	45%	0.5	1
	12	50%	0%	50%	3	50.0%	50%	0.5	
	12.5	100%	0%	0%	1	100.0%	0%		

*Other sites also contained other soil types **Sand clay and clayey sand soil were classified as 55% sand and 45% clay (Table S10) according to Rosetta (Schaap et al., 2001).

b. Morris analysis results

Tables S11 and S12 present μ*, μ* confidence interval and σ results of the Morris analysis in heterogeneous media for benzene total flux to the aquifer and final concentration in the aquifer respectively, with the different methods for missing data handling.

Table S11: Morris indices values of benzene flux to the aquifer for heterogeneous media received with the different methods for missing parameter handling

Method	Output value	Input parameter														
		α_1	n_1	K_{s1}	α_2	n_2	K_{s2}	α_{l1}	α_{l2}	k_{d1}	λ_{k1}	k_{d2}	λ_{k2}	z	N	b
Zero replaced	μ*	0.105	0.000	0.560	0.009	0.259	0.022	0.022	0.094	0.210	0.437	0.820	0.520	0.170	0.326	0.266
	μ* confidence interval	0.091	0.000	0.418	0.011	0.219	0.037	0.026	0.110	0.182	0.401	0.745	0.449	0.162	0.269	0.341
	σ	0.741	3.942	1.728	0.001	0.101	0.321	0.208	0.842	1.551	6.008	3.779	3.944	1.305	2.590	2.516
5NN	μ*	0.149	0.001	0.550	0.009	0.355	0.009	0.022	0.216	0.210	0.437	0.821	0.520	0.172	0.338	0.226
	μ* confidence interval	0.104	0.002	0.464	0.011	0.254	0.014	0.023	0.219	0.194	0.433	0.734	0.503	0.158	0.343	0.245
	σ	0.855	0.017	3.923	0.102	2.021	0.108	0.208	1.637	1.551	3.779	6.008	3.944	1.305	2.599	2.078
63NN	μ*	0.131	0.009	0.576	0.022	0.305	0.030	0.028	0.121	0.210	0.437	0.828	0.521	0.211	0.347	0.272
	μ* confidence interval	0.085	0.010	0.575	0.020	0.232	0.033	0.023	0.102	0.207	0.516	0.746	0.500	0.167	0.328	0.325
	σ	0.750	3.928	1.755	0.067	0.151	0.264	0.214	0.856	1.551	6.008	3.779	3.944	1.322	2.595	2.430
RBF	μ*	0.247	0.059	0.688	0.030	0.486	0.007	0.060	0.280	0.221	0.481	0.863	0.559	0.331	0.408	0.225
	μ* confidence interval	0.117	0.036	0.554	0.018	0.323	0.004	0.039	0.121	0.181	0.439	0.705	0.472	0.193	0.336	0.221
	σ	1.101	0.295	3.987	0.151	2.397	0.031	0.317	1.370	1.555	3.791	6.011	3.947	1.586	2.647	1.702
Full	μ*	0.152	0.000	0.822	0.013	0.213	0.003	0.031	0.118	0.308	0.649	0.588	0.709	0.261	0.282	0.165

trajectories removed	μ* confidence interval	0.147	0.000	0.676	0.018	0.188	0.005	0.043	0.170	0.301	0.645	0.699	0.857	0.231	0.341	0.308
	σ	0.911	4.864	1.463	0.000	0.126	0.026	0.258	1.005	1.915	4.915	4.693	4.854	1.621	2.295	1.797
Average	μ*	0.157	0.014	0.639	0.017	0.323	0.014	0.032	0.166	0.232	0.488	0.784	0.566	0.229	0.340	0.231
	μ* STDE	0.054	0.025	0.116	0.009	0.105	0.011	0.016	0.079	0.043	0.092	0.111	0.082	0.068	0.046	0.043
	μ* confidence interval	0.109	0.010	0.537	0.015	0.243	0.018	0.031	0.144	0.213	0.487	0.726	0.556	0.182	0.323	0.288
	μ* confidence interval STDE	0.025	0.015	0.101	0.004	0.051	0.016	0.009	0.049	0.050	0.098	0.022	0.170	0.031	0.031	0.052
	σ	0.872	2.609	2.571	0.064	0.959	0.150	0.241	1.142	1.624	4.900	4.854	4.127	1.428	2.545	2.105
	σ STDE	0.147	2.273	1.269	0.065	1.149	0.136	0.047	0.349	0.163	1.111	1.119	0.407	0.161	0.142	0.365

Table S12: Morris indices values for benzene final concentration in the aquifer in heterogeneous media received with the different methods for missing parameter handling

Method	Output value	Input parameter														
		α_1	n_1	K_{s1}	α_2	n_2	K_{s2}	α_{l1}	α_{l2}	k_{d1}	λ_{k1}	k_{d2}	λ_{k2}	z	N	b
Zero replaced	μ*	0.0014	0.0000	0.0331	0.0006	0.0131	0.0033	0.0042	0.0095	0.0243	0.0295	0.0580	0.0428	0.0186	0.0185	0.0221
	μ* confidence interval	0.0013	0.0000	0.0244	0.0008	0.0137	0.0067	0.0069	0.0093	0.0192	0.0298	0.0485	0.0283	0.0171	0.0171	0.0237
	σ	0.0112	0.2071	0.0921	0.0002	0.0058	0.0518	0.0536	0.0749	0.1630	0.3312	0.2145	0.2603	0.1442	0.1296	0.2004
5NN	μ*	0.0067	0.0002	0.0331	0.0006	0.0200	0.0013	0.0042	0.0157	0.0243	0.0295	0.0580	0.0428	0.0198	0.0195	0.0188
	μ* confidence interval	0.0054	0.0004	0.0248	0.0008	0.0151	0.0020	0.0062	0.0127	0.0193	0.0257	0.0429	0.0329	0.0187	0.0167	0.0197
	σ	0.0416	0.0032	0.2058	0.0058	0.1169	0.0168	0.0536	0.0978	0.1630	0.2145	0.3312	0.2603	0.1449	0.1307	0.1681
63NN	μ*	0.0042	0.0007	0.0352	0.0019	0.0169	0.0037	0.0046	0.0113	0.0243	0.0295	0.0585	0.0429	0.0221	0.0203	0.0226
	μ* confidence interval	0.0021	0.0005	0.0241	0.0013	0.0102	0.0049	0.0073	0.0086	0.0213	0.0261	0.0395	0.0361	0.0188	0.0165	0.0243
	σ	0.0157	0.2063	0.0959	0.0043	0.0112	0.0406	0.0537	0.0756	0.1630	0.3312	0.2145	0.2603	0.1449	0.1303	0.1940
RBF	μ*	0.0146	0.0051	0.0400	0.0024	0.0285	0.0004	0.0071	0.0198	0.0254	0.0324	0.0610	0.0465	0.0331	0.0249	0.0190
	μ* confidence interval	0.0083	0.0026	0.0241	0.0011	0.0154	0.0002	0.0055	0.0110	0.0185	0.0254	0.0403	0.0295	0.0191	0.0143	0.0162
	σ	0.0706	0.0249	0.2080	0.0105	0.1312	0.0017	0.0559	0.0912	0.1632	0.2152	0.3313	0.2608	0.1611	0.1361	0.1410

Full trajectories removed	μ*	0.0014	0.0000	0.0471	0.0008	0.0132	0.0001	0.0062	0.0107	0.0365	0.0399	0.0479	0.0524	0.0281	0.0173	0.0156
	μ* confidence interval	0.0012	0.0000	0.0335	0.0011	0.0141	0.0001	0.0113	0.0147	0.0313	0.0402	0.0477	0.0465	0.0284	0.0208	0.0223
	σ	0.0093	0.2536	0.0872	0.0000	0.0072	0.0008	0.0668	0.0852	0.2016	0.2927	0.2616	0.3085	0.1790	0.1359	0.1559
Average	μ*	0.006	0.001	0.038	0.001	0.018	0.002	0.005	0.013	0.027	0.032	0.057	0.046	0.024	0.020	0.020
	μ* STDE	0.005	0.002	0.006	0.001	0.006	0.002	0.001	0.004	0.005	0.004	0.005	0.004	0.006	0.003	0.003
	μ* confidence interval	0.004	0.001	0.026	0.001	0.014	0.003	0.007	0.011	0.022	0.029	0.044	0.035	0.020	0.017	0.021
	μ* confidence interval STDE	0.003	0.001	0.004	0.000	0.002	0.003	0.002	0.002	0.005	0.006	0.004	0.007	0.005	0.002	0.003
	σ	0.030	0.139	0.138	0.004	0.054	0.022	0.057	0.085	0.171	0.277	0.271	0.270	0.155	0.133	0.172
	σ STDE	0.026	0.116	0.063	0.004	0.064	0.023	0.006	0.010	0.017	0.059	0.059	0.021	0.015	0.003	0.025

Figure S1 presents the ratio of μ^* confidence interavl to μ^* values. Note that the least influential parameters to the right of the graph show higher ratio of μ^* confidence interavl to μ^*. The RBF method for missing data imputation preformed better than the other method tested.

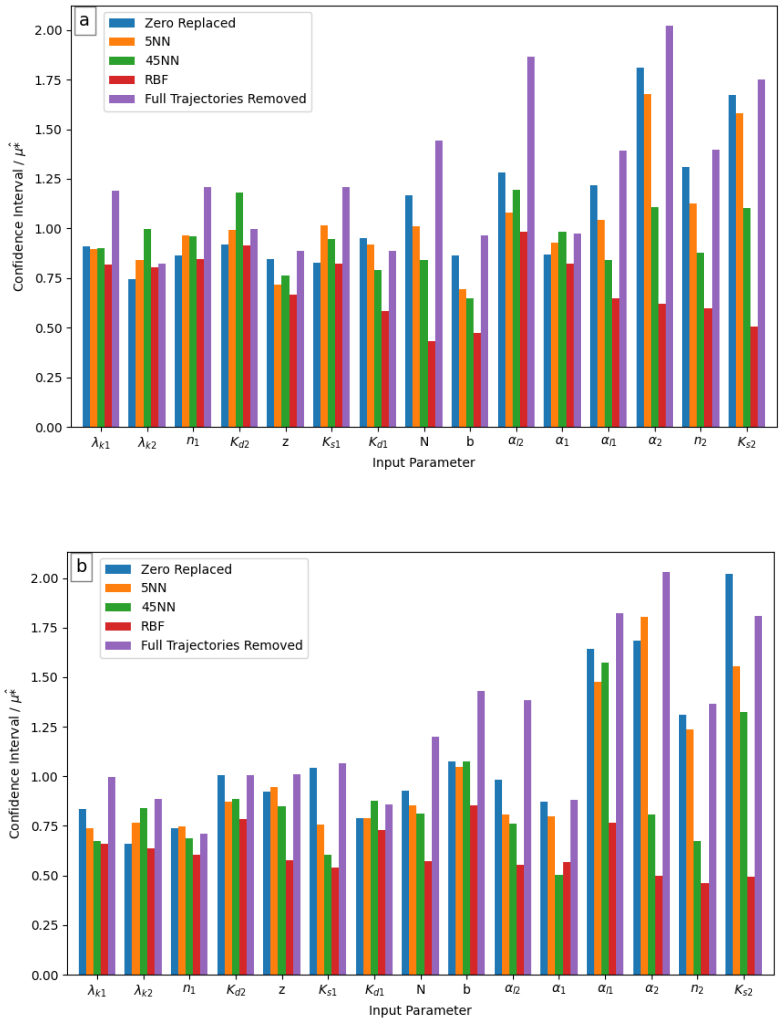

Figure S1 – Morris analysis results for heterogeneous media of the confidence interval to μ^* ratio for: a. benzene flux to the aquifer; and b. final benzene concentration in the aquifer.

References

Archer, G. E. B., Saltelli, A., & Sobol, I. M. (1997). Sensitivity measures ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, *58*(2), 99–120. https://doi.org/10.1080/00949659708811825

Baek, D. S., Kim, S. B., & Kim, D. J. (2003). Irreversible sorption of benzene in sandy aquifer materials. *Hydrological Processes*, *17*(6), 1239–1251. https://doi.org/10.1002/hyp.1181

Bear, J. (1972). *Dynamics of Fluids in Porous Media*. Dover Publications.

Berlin, M., Vasudevan, M., Kumar, G. S., & Nambi, I. M. (2016). Numerical modelling on fate and transport of coupled adsorption and biodegradation of pesticides in an unsaturated porous medium. *ISH Journal of Hydraulic Engineering*, *22*(3), 236–246. https://doi.org/10.1080/09715010.2016.1166073

Brunetti, G., Šimůnek, J., & Piro, P. (2016). A comprehensive numerical analysis of the hydraulic behavior of a permeable pavement. *Journal of Hydrology*, *540*, 1146–1161. https://doi.org/10.1016/j.jhydrol.2016.07.030

Carsel, R. F., & Parrish, R. S. (1988). Developing joint probability distributions of soil waterretention characteristics. *Water Resources Research*, *24*, 755–769. https://doi.org/10.1029/88WR01772.

DeVaull, G. E., Ettinger R. A., Salanitro J. P., & B., G. J. (1997). Degradation in vadose zone soils during vapor transport: first-order rate constants. *Petroleum Hydrocarbons and Organic Chemicals in Ground Water -- Prevention, Detection, and Remediation Conference*.

Donahue, R. B., Barbour, S. L., & Headley, J. V. (1999). Diffusion and adsorption of benzene in Regina clay. *Canadian Geotechnical Journal*, *36*(3), 430–442. https://doi.org/10.1139/cgj-36-3-430

Du, P., Sagehashi, M., Terada, A., & Hosomi, M. (2010). Numerical modeling of benzene transport in andosol and sand: Adequacy of diffusion and equilibrium adsorption equations. *World Academy of Science, Engineering and Technology*, *66*(6), 38–42. https://doi.org/10.5281/zenodo.1330975

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54–75.

English, C. W., & Loehr, R. C. (1991). Degradation of organic vapors in unsaturated soils. *Journal of Hazardous Materials*, *28*(1–2), 55–64. https://doi.org/10.1016/0304-3894(91)87005-M

Hartmann, A., Šimůnek, J., Aidoo, M. K., Seidel, S. J., & Lazarovitch, N. (2018). Implementation and Application of a Root Growth Module in HYDRUS. *Vadose Zone Journal*, *17*(1), 170040. https://doi.org/10.2136/vzj2017.02.0040

Hers, I., Atwater, J., Li, L., & Zapf-Gilje, R. (2000). Evaluation of vadose zone biodegradation of BTX vapours. *Journal of Contaminant Hydrology*, *46*(3–4), 233–264. https://doi.org/10.1016/S0169-7722(00)00135-2

Karickhoff, S. W. (1981). Semi-empirical estimation of sorption of hydrophobic pollutants on

natural sediments and soils. *Chemosphere*, *10*(1), 833–846.

Kenaga, E. E. (1980). Predicted bioconcentration factors and soil sorption coefficients of pesticides and other chemicals. *Ecotoxicology and Environmental Safety*, *4*(1), 26–38. https://doi.org/10.1016/0147-6513(80)90005-6

Lahvis, M. A., Baehr, A. L., & Baker, R. J. (1999). Quantification of aerobic biodegradation and volatilization rates of gasoline hydrocarbons near the water table under natural attenuation conditions. *Water Resources Research*, *35*(3), 753–765. https://doi.org/10.1029/1998WR900087

Mackay, A. A., Chin, Y. P., Macfarlane, J. K., & Gschwend, P. M. (1996). Laboratory assessment of BTEX soil flushing. *Environmental Science and Technology*, *30*(11), 3223–3231. https://doi.org/10.1021/es950963b

Osagie, E., & Owabor, C. N. (2015). Adsorption of benzene in batch system in natural clay and sandy soil. *Advances in Chemical Engineering and Science*, *05*(03), 352–361. https://doi.org/10.4236/aces.2015.53037

Plischke, E., Borgonovo, E., & Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, *226*(3), 536–550. https://doi.org/10.1016/j.ejor.2012.11.047

Redding, A. Z., Burns, S. E., Upson, R. T., & Anderson, E. F. (2002). Organoclay sorption of benzene as a function of total organic carbon content. *Journal of Colloid and Interface Science*, *250*(1), 261–264. https://doi.org/10.1006/jcis.2001.8205

Ririe, T., & Sweeney, R. (1995). Fate and transport of volatile hydrocarbons in the vadose zone. *1995 Petroleum Hydrocarbon and Organic Chemicals in Groundwater*, 529–542.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models (Google eBook)*. http://books.google.com/books?id=NsAVmohPNpQC&pgis=1

Schaap, M. G., Leij, F. J., & Van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*(3–4), 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Wiedemeier, T. H., Swanson, M. A., Wilson, J. T., Kampbell, D. H., Miller, R. N., & Hansen, J. E. (1996). Approximation of biodegradation rate constants for monoaromatic hydrocarbons (BTEX) in ground water. *Ground Water Monitoring and Remediation*, *16*(3), 186–194. https://doi.org/10.1111/j.1745-6592.1996.tb00149.x