

DISS. ETH NO. 23451

UNDERSTANDING ENGINEERED NANOMATERIAL
TRANSPORT IN SOIL: RECONCILING WEAKNESSES
IN ADVECTION DISPERSION MODELING WITH
MACHINE LEARNING

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

ELI GOLDBERG

MSc ETH in Environmental Engineering, ETH Zurich

born on 15.02.1984

citizen of USA

accepted on the recommendation of

Prof. Dr. Konrad Hungerbühler, examiner

Prof. Dr. Martin Scheringer, co-examiner

Prof. Dr. Alexander Wokaun, co-examiner

2016

ABSTRACT

It is not yet analytically feasible to determine the concentration of engineered nanomaterials (ENMs) in real soils. This is because many of the metals that comprise the most commonly produced ENMs (e.g., TiO_2 , FeO_x , ZnO , and CuO) occur naturally as nanoparticulates with similar size and shape. These natural analogues are also present in concentrations within the range of predicted environmental concentrations for ENMs in soil, which make it difficult to distinguish ENMs from the background concentration of natural nanomaterials. As a result, investigation of the transport behavior of ENMs is constrained to bench-top soil column transport experiments performed in simple, synthetic soil systems (i.e., monodisperse, fully saturated, single-media soils).

In Chapter 2, it is shown that most advection-dispersion based particle transport models (PTMs) commonly employed to track ENMs in soil are fundamentally flawed. The fundamental flaw is not that they lack descriptive ability entirely, as they can well describe a large portion of observed ENM transport behavior. It is that the mathematical construction of these models considers neither the physicochemical properties of the ENMs, nor those of the system as a whole, explicitly. Instead, and because the mechanistic understanding of how physicochemical conditions influence ENM transport is incomplete, researchers employ quasi-mechanistic PTMs that involve a great deal of empirical specification and ad hoc adjustment. As a result, the applicability of these models beyond the exact scenario under which they were conditioned is low and current PTMs are merely descriptive tools without predictive power.

In Chapter 3, machine learning is employed to enable the prediction of ENM transport behavior and overcome the limitations established in Chapter 2. To support the machine learning effort, a database of physicochemical conditions (explanatory variables) extracted from more than 200 ENM transport experiments conducted in saturated porous media is developed. This database is used to train machine learning models to predict the fraction of ENMs retained over a soil column length, a measure of transport distance, and to predict the retention profile shape, a measure of transport behavior. The retained

fraction is able to be predicted with a mean squared error between 0.025–0.033, and the retention profile shape is predicted with an F1-score (the weighted harmonic mean of precision and recall) between 60–70%. Further, by recursively removing physical and chemical features to optimize model predictive performance, the importance of the physicochemical features to predict transport and behavior is quantitatively determined and ranked - thereby establishing strategic targets for inclusion in new mechanistic PTMs.

In Chapter 4, the database developed in Chapter 3 is expanded and enhanced by combining many of the experimental physicochemical descriptors to derive common dimensionless fluid dynamics and particle interaction parameters. Evaluation of the investigated range of experimental conditions within the database shows that the environmental relevance of the studies to date is low and notable gaps in the examined range are identified. Then, a machine learning decision tree classification is employed to quantitatively, and visually, disentangle the relationship between physicochemical conditions and ENM transport behavior within the database. This approach allows the identification of the order of importance and range of influence of the physical, chemical, and electrical conditions that control the retention profile shape. This is important because it offers an alternative to the current approach and opens a path toward a deeper level of mechanistic understanding of ENM transport in saturated porous media. The current approach has not advanced the mechanistic understanding of ENM transport because it relies on PTMs that have no predictive power and, for the case of nonexponential profiles, even little descriptive power. Through the presented alternative approach, it is possible, for the first time, to identify domains of conditions where nonexponential retention is frequent/dominant, and thus where entirely new mechanistic approaches can be developed, and domains of conditions where exponential retention is frequent/dominant, thus existing approaches may be sufficient.

In Chapter 5, the machine learning methods in Chapter 3 are employed to create, evaluate, and validate a model to predict the ENM attachment efficiency, α , using a database of physicochemical conditions extracted from quartz crystal microbalance with dissipation (QCM-D) experiments. Unlike traditional soil column experiments, QCM-D experiments directly monitor the deposition rate of particles onto an oscillating sensor surface in real-time, isolating deposition kinetics, and allowing for a direct measure of α without contributions from convection or filtration. α is an important kinetic transport

parameter and is widely used in colloid filtration theory (CFT), which is described in Chapter 2 of this thesis, to quantify the likelihood of a particle attaching to a surface after a collision.

In Chapter 6, a brief outlook is provided.

PUBLICATIONS

Chapters 2 and 3 in the main part of this thesis are reprinted with permission from the following publications:

- Eli Goldberg, Martin Scheringer, Thomas D Bucheli, and Konrad Hungerbühler. Critical assessment of models for transport of engineered nanoparticles in saturated porous media. *Environmental Science & Technology*, 48(21):1273212741, 2014. DOI: 10.1021/es502044k; Publication Date (Web): September 26, 2014; ©2014 American Chemical Society.
- Eli Goldberg, Martin Scheringer, Thomas D Bucheli, and Konrad Hungerbühler. Prediction of nanoparticle transport behavior from physicochemical properties: machine learning provides insights to guide the next generation of transport models. Reprinted with permission from *Environmental Science: Nano*, 2(4):352360, 2015. DOI: 10.1039/C5EN00050E; ©2015 Royal Society of Chemistry.

Chapters 4 and 5 are part of collaborative efforts and presented herein with permission from the coauthors:

- Eli Goldberg, Coy P. McNew, Martin Scheringer, Thomas D. Bucheli, Peter Nelson, and Konrad Hungerbühler. What factors determine the retention behavior of engineered nanomaterials in saturated porous media?. *Manuscript submitted*.
- Eli Goldberg, Coy P. McNew, Martin Scheringer, Eugene J. LeBoeuf, and Konrad Hungerbühler. "Predicting particle attachment efficiency from physicochemical characteristics: a machine learning approach".

ZUSAMMENFASSUNG

Es ist noch nicht praktikabel, die Konzentration von engineered nanomaterials (ENMs) in Böden zu bewerten. Dies ist darauf zurückzuführen, dass viele der Metalle, die in den am häufigsten hergestellten ENMs vorkommen, auch in der Natur als Nanopartikel erscheinen. Diese natürlichen Analoge sind auch in Konzentrationen vorhanden, die noch innerhalb der vorhergesagten Bandbreite liegen. Dies erschwert die Arbeit, jene ENMs von der Hintergrundkonzentration der natürlich vorkommenden Nanomaterialien zu unterscheiden. Aufgrund dessen beschränkt sich die Untersuchung auf Bench-Top Bodensäulen Transport Experimente in einfachen, synthetischen Bodensystemen (sprich monodisperse, vollgesättigte, single-media Böden).

Im 2. Kapitel wird gezeigt, dass advektions/diffusionsbasierte Partikeltransportmodelle (PTMs), die häufig gebraucht werden um ENMs in Böden zu folgen, grundlegende Mängel aufzeigen. Obwohl PTMs ein Grossteil des ENM Transportverhaltens beschreiben können, werden die physikalisch-chemischen Eigenschaften der ENMs, sowie die des Systems als Ganzes, ignoriert. Stattdessen, verwenden Forscher quasi-mechanistische PTMs die durch Ad-hoc-Anpassungen sowie empirische Spezifikation optimiert werden müssen. Dem zur Folge ist die breitere Anwendbarkeit der Modelle fast ausgeschlossen. Heutige PTMs sind lediglich als deskriptive Werkzeuge anwendbar, ohne echte Vorhersagekraft zu besitzen.

In 3 Kapitel werden maschinelle Lernverfahren angewendet um das Vorhersagen von ENM Transportverhalten zu ermöglichen, und ferner, die Beschränkungen der Themen aus Kapitel 2 zu überwinden. Um dieses Verfahren zu unterstützen, wird eine Datenbank der physikalisch-chemischen Eigenschaften aus über 200 ENM Transportversuchen, die in teilgesättigten, poröse Medien durchgeführt wurden, aufgesetzt. Die Datenbank wird verwendet um Machine-Learning-Modelle zu trainieren, die erstens die Verteilung der behaltene ENMs über die Länge einer Bodensäule – ein Mass für Transportlänge, und zweitens die Profilhaltigkeit – ein Mass für Transportverhalten – vorhersagen. Der Anteil gespeicherter ENMs kann mit einem mittleren quadrierten Fehler von 0.025-0.033, und die

Profilhaltigkeit mit einem F₁-Mass von 60-70% vorhergesagt werden. Zudem wird die Wichtigkeit der einzelnen physikalisch-chemischen Eigenschaften eingestuft, anhand eines quantitativen Verfahrens – nämlich durch das rekursive löschen der Merkmale. Dadurch werden statische Ziele für neue mechanistische PTMs etabliert.

Im 4. Kapitel wird die Datenbank aus Kapitel 3 erweitert. Die Kombination der einzelnen physikalisch-chemischen Merkmale erlaubt die Ableitung von gemeinsamen Parametern, die für dimensionslose Strömungsdynamik sowie Teilcheninteraktion gelten. Die Bewertung der Versuchsbedingungen zeigt, dass bisherige Experimente wenig Relevanz aufzeigen. Durch Anwendung einer Machine-Learning Entscheidungsbaum-Klassifizierung entflechtet sich die Beziehung zwischen den physikalisch-chemischen Eigenschaften und das ENM Transportverhalten.

In 5 Kapitel werden die Machine-Learning-Modelle, die im 3. Kapitel beschrieben wurden, angewendet um ein Model der ENM Bindungseffizienz, α , zu validieren. Zu diesem Zweck wurde eine Datenbank mit den Resultaten mehrere Quarzkristall-Mikrowaagen mit Dissipationsmessungs-(QCM-D) Experimenten aufgestellt. Im Gegensatz zu traditionellen Bodensäule-Experimenten, bietet QCM-D die Möglichkeit, Depositionsraten von Partikeln auf einer Sensorfläche in Echt-Zeit zu überwachen, was eine direkte Messung von α ohne Einfluss von Filtrations- oder Konvektionsprozessen ermöglicht. α ist ein wichtiger Transportparameter, der in der Kolloid-Filtrations Theorie (in Kapitel 2 ausführlich beschrieben) benutzt wird um die Bindungswahrscheinlichkeit eines Partikels nach der Kollision mit einer Fläche zu quantifizieren.

Im 6. Kapitel, wird ein kurzer Ausblick vorgestellt.

For my wife, who helped me to achieve. For my brothers, who convinced me I could. For my parents, who said that I shouldn't, which was exactly what I needed to ensure that I did.

ACKNOWLEDGEMENTS

As is my style, I will keep the acknowledgments brief.

To Konrad: It is difficult to describe the gratitude I feel for extending me this opportunity. Your abilities as a leader and scientist are only outweighed by your kindness and compassion. Thanks to you, my time in Switzerland has been the best in my life.

To Martin: Your curiosity, patience, and insight has helped me to become a better writer, a better scientist, and a more insightful man. You taught me how to organize my thoughts, oftentimes by hand, and how to tell a story. Although I leave ETH, I hope to continue our friendship, our insightful discussions, and our never ending contempt for the status quo.

To Antonia, Greta, Eva, Sami, Nicole, Coy, Kurt, and Fabian: Doctoral studies are not *all* doctoral. Whether we were starting businesses, playing video games, or taking Reason for a walk, your friendship made the tedious exciting, the arduous bearable, and the mundane shockingly hilarious.

To my wife, Sarah: Without your support and love, none of this would be possible. You are my best friend and my confidant. Thank you for your patience and understanding.

I would also like to thank Alexander Wokaun for your interest in my work, your time to review my thesis, and your support as a co-examiner.

CONTENTS

1	INTRODUCTION	31
1.1	What are engineered nanomaterials (ENMs)?	31
1.2	ENM transport to soil	32
1.3	ENM transport through soil	34
1.3.1	Investigating ENM transport with soil columns . .	34
1.4	How does this dissertation add value to the field?	38
2	CRITICAL ASSESSMENT OF ENM TRANSPORT MODELS	41
2.1	Abstract	42
2.2	Introduction	42
2.3	Methods	45
2.3.1	Particle Transport Models (PTMs)	45
2.3.2	Nanospecific model considerations	48
2.3.3	Selection of experimental data	49
2.3.4	Parameter fitting	50
2.4	Results and discussion	51
2.4.1	Hyperexponential retention	54
2.4.2	Linearly decreasing retention	55
2.4.3	Nonmonotonic retention	57
2.4.4	Monotonically increasing retention	59
2.5	Environmental implications	60
3	PREDICTION OF ENM TRANSPORT FROM PHYSICOCHEMICAL PROPERTIES	63
3.1	Abstract	64
3.2	Introduction	64
3.3	Methods	67
3.3.1	Nanomaterial Transport Experiment Database . . .	67
3.3.2	Random Forest Regression and Classification . . .	69
3.3.3	Feature selection with recursive feature elimination with 5-fold cross validation (RFECV) .	70
3.3.4	Database partitioning and model structure	72
3.4	Results and discussion	75
3.4.1	Prediction of the retained fraction (RF)	75
3.4.2	Prediction of the retention profile (RP) shape . . .	80
3.5	Conclusions	84
4	WHAT FACTORS DETERMINE THE RETENTION BEHAVIOR OF ENMS?	87

4.1	Abstract	87
4.2	Introduction	87
4.3	Methods	89
4.3.1	Database description	89
4.3.2	Machine learning classification	93
4.4	Results	94
4.4.1	Effects of individual factors on retention behavior	94
4.4.2	Disentangling and visualizing the influence of different factors	97
4.5	Conclusions	102
5	A MACHINE LEARNING MODEL TO PREDICT ENM ATTACHMENT EFFICIENCY	105
5.1	Introduction	105
5.2	Materials and Methods	108
5.2.1	Database	108
5.2.2	Machine learning	110
5.2.3	Model structure	112
5.3	Results and Discussion	112
5.3.1	Recursive feature elimination with cross validation	112
5.3.2	Attachment efficiency prediction performance	114
5.3.3	Optimum Result	114
5.3.4	Implications of model findings	117
5.4	Conclusions	120
6	A BRIEF OUTLOOK	123
7	APPENDIX I: SUPPORTING INFORMATION FOR CHAPTER 2	147
7.1	Supporting model information	147
7.1.1	Model 1	148
7.1.2	Model 2	149
7.1.3	Model 3	150
7.1.4	Model 4	151
7.1.5	Model 5	151
7.1.6	Model 6	152
7.1.7	Model 7	152
7.2	Supporting methods information	152
7.2.1	Programming details	152
7.2.2	Bias-adjusted Akaike Information Criteria	154
7.2.3	Analysis of experimental data availability for ENP column experiments	156
7.3	Supporting simulation results	168
7.3.1	Supplementary archetypal simulations	168
7.3.2	Results from the parameter fitting for the archetypal retention profiles	169

7.3.3 Hyperexponentiality influence on model plausibility 174

8 APPENDIX II: SUPPORTING INFORMATION FOR CHAPTER 3 179

8.1 Database development and distributed computing 179

8.2 Controlling randomness 179

8.3 RFECV method and modification 180

8.3.1 Regression RFECV Modification 181

8.3.2 Classification RFECV Modification 181

8.4 Publications mined for the database 182

9 APPENDIX III: SUPPORTING INFORMATION FOR CHAPTER 4 185

9.1 Inclusion, exclusion, and quality criteria 185

9.2 Decision tree performance 186

10 CURRICULUM VITAE 189

LIST OF FIGURES

Figure 1	Estimated global mass flow of ENMs (in metric tons per year).	33
Figure 2	Soil column experiments in three steps.	36
Figure 3	A visualization of the physical, chemical, and electrical parameters for 214 ENM transport experiments.	37
Figure 4	Digitized retention profiles (dots) and fitted model results (lines) are presented in panels 1a, 2a, 3a, and 4a; digitized breakthrough profiles (dots) and fitted model results (lines) are presented in panels 1b, 2b, 3b, and 4b.	52
Figure 5	Nanomaterial transport database structure for machine learning regression and classification. . .	69
Figure 6	Graphical depiction of data partitioning scheme and model structure for the classification and regression problems.	74
Figure 7	Aggregate model performance to predict the RF of the cross validation test fold.	76
Figure 8	Aggregate model performance in predicting the RF of the holdout set as a function of the number of RFECV-selected features employed to train the model.	78
Figure 9	The frequency of features included through RFECV (in percent of number of model runs) for the prediction of the RF.	79
Figure 10	Aggregate model performance (F1-score) in predicting the RP shape class of the cross-validation test fold as a function of the number of features employed to train.	81
Figure 11	Aggregate model performance to predict the RP shape for the holdout set as a function of the number of RFECV-selected features employed to train the model.	82
Figure 12	The frequency of features included through RFECV (in percent of number of model runs) for the prediction of RP shape class.	84

Figure 13	Histograms for individual factors showing the number of experiments with either exponential (blue) or nonexponential (orange) retention behavior.	95
Figure 14	Blue and orange bars: histograms for categorical factors showing the number of experiments with either exponential (blue) or nonexponential (orange) retention behavior; colored stacked bars: histograms showing the number of experiments conducted with each NOM type and ENM type over the ranges of NOM concentrations and influent ENM concentrations, respectively.	96
Figure 15	The optimal decision tree to classify the ENM transport experiment database.	99
Figure 16	A radar chart of the coefficient of variation ($CV = \frac{\sigma}{\mu}$) for non-categorical (ENM type, NOM type, ENM coating), non-preceding (M_{inf} and N_{Pe}), non inherently low CV (N_{por} , $N_{asp,c}$, $N_{\zeta 1}$, and $N_{\zeta 2}$) physicochemical descriptors.	100
Figure 17	A visualization of the physical, chemical, and electrical parameters for the α database.	111
Figure 18	Graphical depiction of data partitioning scheme and model structure for the classification and regression problems.	113
Figure 19	Aggregate model performance in predicting α in the cross validation test fold during the 500 RFECV model runs.	115
Figure 20	Aggregate model performance in predicting α in the holdout set as a function of the number of RFECV selected features.	116
Figure 21	Relative feature importance, reported as percentage of total model runs in which each feature was included in the optimal model to predict α	118
Figure 22	Supplementary Archetypal Retention Profiles and Model results.	168

Figure 23 Digitized retention profiles (dots) and fitted model results (lines) are presented in panels 1a, 2a, and 3a, digitized breakthrough profiles (dots) and fitted model results are presented in panels 1b, 2b, and 3b. Note that the y-axis of panel 3a is presented at a 100x reduced scale in comparison to panels 1a and 2a. From a fit of the RP alone, the most plausible models that fit both the RP and BTC are as follows: In panels 1a and 1b, M6 provides the most plausible RP and BTC. In panels 2a and 2b, M2 provides the most plausible RP and BTC. In panels 3a and 3b, M7 provides the most plausible RP and BTC. 175

LIST OF TABLES

Table 1	Routinely employed particle transport models. . .	47
Table 2	Modeling parameters for the retention profiles in Figure 4 as reported in the original studies. . . .	53
Table 3	Investigated domain of physicochemical training and target features employed for the machine learning effort.	68
Table 4	Distribution of observations by RP shape within the database. RP shapes that can be modeled using PTMs (HE, EXP) are separated from those that cannot (SIR, IRwD, LD).	75
Table 5	Domain of properties of the nanomaterial, the soil column, and the fluid for the investigated transport experiments.	90
Table 6	Descriptors employed to train the decision tree classification.	92
Table 7	Investigated domain of physicochemical training and target features.	109
Table 8	Summary of values utilized in determining the highest performing model with the fewest number of features and resulting optimum values using Equation 4.	117
Table 9	Publications on ENP transport in porous media (2008–2014)	156
Table 10	Model parameter estimation results for the archetypal retention profiles Figure 4 in the main text	170
Table 11	Model parameter estimation results for the archetypal retention profiles Figure 4 in the main text	171
Table 12	Model parameter estimation results for the higher error tolerance estimation for the nonmonotonic retention profile in Figure 4, panel 3a	172
Table 13	Model plausibility comparison results for the archetypal retention profiles in Figure 4	173

Table 14	Model plausibility results for the higher error tolerance estimation for the NM retention profile in Figure 4	174
Table 15	Model parameter estimation results for the supplementary hyperexponential profiles in Figure 23	176
Table 16	Model plausibility comparison results for the archetypal retention profiles in Figure 23	177
Table 17	Publications employed for this study organized by types of nanoparticles.	183
Table 18	Sensitivity, selectivity, and precision of the presented decision tree.	187

NOMENCLATURE

$20h - UV$	20 hours under UV exposure
$7D - UV$	7 days under UV exposure
α	attachment efficiency [-]
α_f	fast attachment efficiency associated [-]
α_L	longitudinal dispersivity [m]
α_s	slow attachment efficiency associated [-]
β	empirical depth-dependant retention parameter [-]
η_0	single-collector contact efficiency [-]
γ	Happel porosity factor [-]
v_p	pore water velocity [$\frac{m}{s}$]
ψ_b	dynamic Langmuirian blocking function [-]
ψ_s	depth-dependant retention function [-]
ρ_b	bulk soil density [$\frac{kg}{m^3}$]
ρ_f	fluid density [$\frac{kg}{m^3}$]
ρ_p	particle density [$\frac{kg}{m^3}$]
θ	porosity [-]
ζ_c	collector ζ -potential [mV]
ζ_p	particle ζ -potential [mV]
A	Hamaker constant for particle-fluid-collector interaction [J]

<i>ADA</i>	adaptive boosting
<i>AIC_c</i>	adjusted Akaike information criteria [-]
<i>Alg</i>	algae
<i>BTC</i>	breakthrough curve
<i>C</i>	concentration [$\frac{kg}{m^3}$]
<i>C60</i>	Fullerene
<i>C_l</i>	column length [m]
<i>C_w</i>	column inner diameter [m]
<i>C_{inf}</i>	influent concentration [$\frac{kg}{m^3}$]
<i>C_{nom}</i>	concentration of NOM in solution [$\frac{kg}{m^3}$]
<i>CdTe</i>	cadmium telluride
<i>CeO₂</i>	nano cerium oxide
<i>CFT</i>	colloid filtration theory
<i>CV</i>	coefficient of variation [-]
<i>D_∞</i>	Dispersion coefficient for an infinite dilution [$\frac{m^2}{s}$]
<i>d_c</i>	collector diameter [m]
<i>D_p</i>	longitudinal dispersion coefficient [$\frac{m^2}{s}$]
<i>d_p</i>	particle diameter [m]
<i>DLVO</i>	Derjaguin, Landau, Verwey, and Overbeek
<i>DTC</i>	decision tree classification
<i>EHA</i>	Elliott humic acid
<i>ENM</i>	engineered nano material
<i>ENP</i>	engineered nano particle

<i>EU</i>	European Union
<i>EXP</i>	exponential retention profile shape
<i>f</i>	particle population associated with α_f [-]
<i>F_{observation%}</i>	fraction of model runs that result in a given RFECV-selected feature set size [-]
<i>FA</i>	fulvic acid
<i>Fe</i>	nano iron (unspecified)
<i>Fe(OH)₃</i>	iron(III) oxide-hydroxide
<i>FeO_x</i>	nano iron oxide
<i>FeOOH</i>	goethite
<i>g</i>	gravitational constant [$\frac{m}{s^2}$]
<i>GBC</i>	gradient boosting decision tree classification
<i>GSHPOCV</i>	grid search hyperparameter optimization with cross-validation
<i>HA</i>	humic acid
<i>HE</i>	hyperexponential
<i>HFA</i>	Harpeth fulvic acid
<i>HHA</i>	Harpeth humic acid
<i>HS</i>	humic substance
<i>I</i>	ionic strength [mM]
<i>IEP</i>	isoelectric point [-]
<i>IRwD</i>	increasing retention with depth retention profile shape
<i>k_b</i>	Boltzmann constant [$\frac{1}{K}$]
<i>k_d</i>	first-order kinetic deposition rate constant [$\frac{1}{s}$]

k_r	first-order kinetic remobilization rate constant [$\frac{1}{s}$]
$k_{d,2}$	second site first-order kinetic deposition rate constant [$\frac{1}{s}$]
$k_{d,f}$	first-order kinetic deposition rate constant associated with α_f [$\frac{1}{s}$]
$k_{d,s}$	first-order kinetic deposition rate constant associated with α_s [$\frac{1}{s}$]
$k_{r,2}$	second site first-order kinetic remobilization rate constant [$\frac{1}{s}$]
L_d	Debye length [m]
Lbl	layer-by-layer
LD	linearly decreasing retention profile shape
LO/HO	lowly oxidized/ highly oxidized
M_{inf}	total influent mass of ENM particles
MI	monotonically increasing retention profile shape
MMF	multimedia fate
$MS2/BL21RNA$	RNA extracted from bacteriophage MS2 and <i>E. coli</i> BL21
MSE	mean squared error
$N_{\zeta 1}$	dimensionless first electrokinetic parameter
$N_{\zeta 2}$	dimensionless second electrokinetic parameter
$N_{asp,c}$	dimensionless aspect ratio between length of the column and the inner diameter
N_{asp}	dimensionless aspect ratio between particle and collector
N_{att}	dimensionless attraction number

N_{dl}	dimensionless electrical double-layer force parameter
N_g	dimensionless gravity number
N_{lon}	dimensionless London force parameter
N_{Pe}	dimensionless Peclet number
N_{por}	dimensionless porosity dependant parameter
nAg/Ag	nano silver
$nCuO/CuO$	nano copper oxide
$nHAP/HAP$	nano hydroxyapatite
NM	nonmonotonic retention profile shape
NOL	natural organic ligand
NOM	natural organic matter
$nTiO_2/TiO_2$	nano titanium dioxide
$NZVI$	nano zero valent iron
PMF	probabilistic mass flow
PPP	plant-protection-product
PS	polystyrene
PTM	advection-dispersion particle transport model
$QCM - D$	quartz crystal microbalance with dissipation
QD	quantum dot
R_{IQ}	inner-quartile range
RF	retained fraction
$RFECV$	recursive feature elimination with cross-validation
RP	retention profile

RSS	residual sum of squares
S	attachment site solid-phase concentration [$\frac{kg_{nano}}{kg_{soil}}$]
S_2	second site solid-phase concentration [$\frac{kg_{nano}}{kg_{soil}}$]
S_m	maximum site solid-phase concentration [$\frac{kg_{nano}}{kg_{soil}}$]
SiO_2	nano silicon dioxide
SIR	suppressed inlet retention profile shape
$SRHA$	Sewanee River humic acid
$SWCNT/MWCNT$	single/double-wall carbon nanotube
T	temperature [K]
t	time [s]
$TP/TN/FP/FN$	true positives / true negatives / false positives / false negatives
$TRIZMA$	Tris(hydroxymethyl)aminomethane
w_i	Akaiki weight [-]
x	distance from soil column inlet [m]
$ZnO/nZnO$	nano zinc oxide

INTRODUCTION

Contents

1.1	What are engineered nanomaterials (ENMs)?	31
1.2	ENM transport to soil	32
1.3	ENM transport through soil	34
1.3.1	Investigating ENM transport with soil columns	34
1.4	How does this dissertation add value to the field? .	38

1.1 WHAT ARE ENGINEERED NANOMATERIALS (ENMS)?

The EU definition of a nanomaterial (2011/696/EU) is:

“A natural, incidental or manufactured material containing particles, in an unbound state or as an aggregate or as an agglomerate and where, for 50% or more of the particles in the number size distribution, one or more external dimensions is in the size range 1 nm–100 nm.”

The market for the development and use of engineered nanomaterials (ENMs) is increasing rapidly. In 2014, the global market for environmental applications of nanotechnology was estimated to be \$21.8 billion.¹ By 2020, the total sales of nanotechnology-based products are anticipated to reach \$48.9 billion, with a total market cap in excess of a \$1 trillion.¹

ENMs are used in numerous consumer and industrial products, such as cosmetics, food, hybrid organic-inorganic light-emitting diodes, solar cells, paints, and remediation technologies.^{2–6} While no ENM agricultural products have made it to market⁷, the use of nano-based plant protection products (PPP), such as nanopesticides and

nanofertilizers, is expected to significantly contribute to the use of nano products for agricultural applications in the future.⁸

1.2 ENM TRANSPORT TO SOIL

Multimedia fate models (MMF) are a common technique to estimate the predicted environmental concentration of organic chemicals in the environment.⁹⁻¹² However, MMF models cannot be used for nanomaterials, as the behavior of particles in the environment cannot be described in the same manner as organic chemicals.¹³ For organic chemicals, there exists a known thermodynamic relationship between environmental compartments. This relationship is expressed in terms of an equilibrium partitioning coefficient, which establishes a physically meaningful relationship between the equilibrium concentration in one or more phases (e.g., between air and water).¹⁰ Using these partition coefficients, the concentration of chemicals can be determined by analyzing a series of mass balances, where soil, water, and air compartments are represented by interconnected boxes.¹⁰ ENMs, as opposed to chemicals, cannot be assumed to be in thermodynamic equilibrium between environmental compartments, and partitioning coefficients do not exist for ENMs.¹⁴ Thus, there is no thermodynamic justification for modeling the transport of particles in terms of partition coefficients.

In place of MMF models, probabilistic mass flow (PMF) models are used to predict the environmental concentration of ENMs by estimating the flux of materials from production to one or more environmental compartments.^{15,16} Several PMF estimates are available.^{8,17,18} However, most estimate the flux of nano titanium dioxide (TiO₂), nano silver (Ag), and carbon nanotubes (CNTs) to the soil compartment are between 0.56 ng kg⁻¹ a⁻¹ and 1.28 μg kg⁻¹ a⁻¹ for the U.S., Europe, and Switzerland.^{16,18} A breakdown of global estimates for ENM release to the environment based on production volumes and uses for the top 10 ENMs (> 100 metric tons per year) are presented in Figure 1.¹⁹

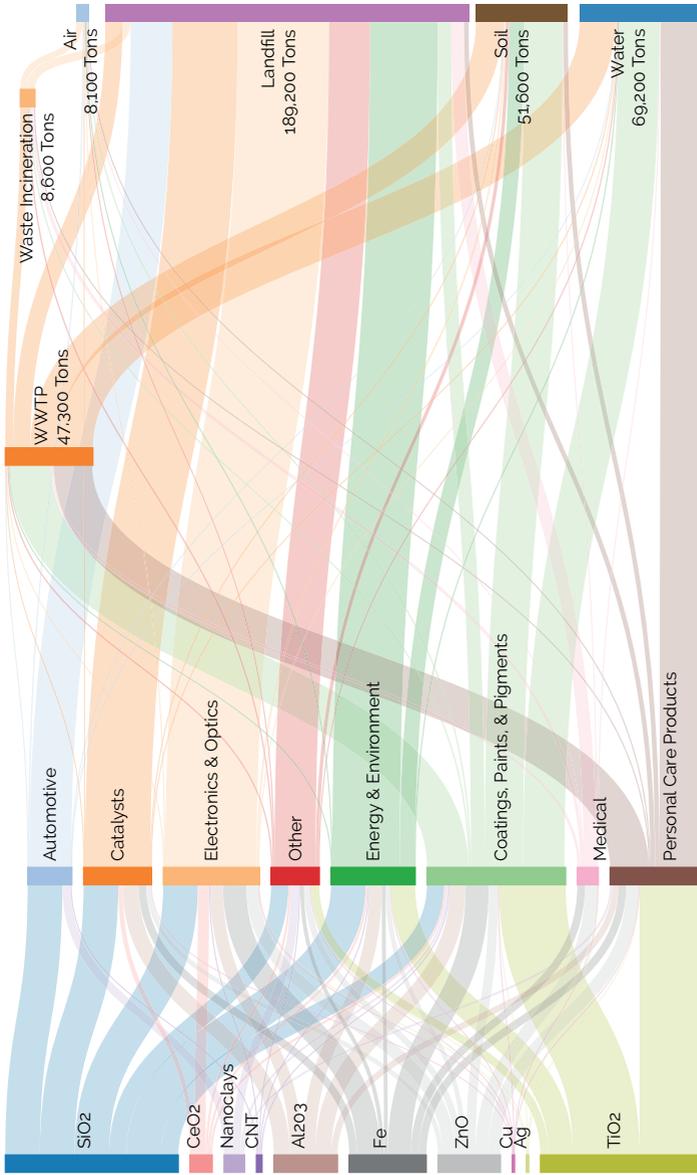


Figure 1: Estimated global mass flow of ENMs (in metric tons per year) from production to disposal or release for the 10 highest production volume ENMs in 2010. Figure adapted from Keller and Lazareva ¹⁹.

1.3 ENM TRANSPORT THROUGH SOIL

At present, the earth's soil is not critically threatened by ENMs although more than 240 metric tons of ENMs end up in soil or landfills.¹⁹ However, the rapidly growing market for, and use of, ENMs is likely to result in increasing environmental exposure.^{16,19} Specifically, future uses of ENMs as PPPs or fertilizers could significantly increase the mass of ENMs to soil.⁸ In addition, indirect releases of ENMs from consumer products, which occur from leaching of ENM-containing products disposed in landfills or from application of sludge to agricultural land, could present risks to groundwater. Therefore, it is also important to understand how these ENMs are transported through soil.

1.3.1 *Investigating ENM transport with soil columns*

Soil column experiments yield a breakthrough curve and a retention profile (Figure 2). A breakthrough curve provides temporal information concerning the concentration of ENMs leaving the column in relation to the concentration of ENMs entering the column. A retention profile provides information concerning the spatial distribution of ENMs in relation to the depth of soil at the end of the experiment. Qualitatively, ENM transport behavior is influenced by a number of physical, electric, and chemical factors.²⁰⁻⁴⁵ Physical factors include: particle and collector shape, size, and surface potential, hydrodynamic characteristics, and Hamaker constants (which do not exist for many particle-particle/particle-surface interactions and limit theoretically validation with DLVO theory). Electric factors include: the material isoelectric point, and the surface charge and surface charge heterogeneity of the particle and collector (which affect the binding affinity and particle stability). Chemical factors include the pH and ionic strength conditions, which influence the van der Waals and electrical double layer forces, the surface chemistry of the particle and collector, which can influence stability and reactivity through ligand exchange mechanisms, steric hinderence, or hydrophobic interactions, and the presence of humic substances (HS) i.e., natural organic matter (NOM).⁴⁶

Chapters 3 and 4 discuss the relationship between the physicochemical properties of the ENM and surrounding environment, and the retention behavior (i.e., retention profile shape). To visualize the complexity of

interacting conditions and their influence on retention behavior, a breakdown of the available literature (214 experiments from 24 publications) is visualized in Figure 3.

1. The experimental setup...
2. During the experiment...
3. After the experiment...

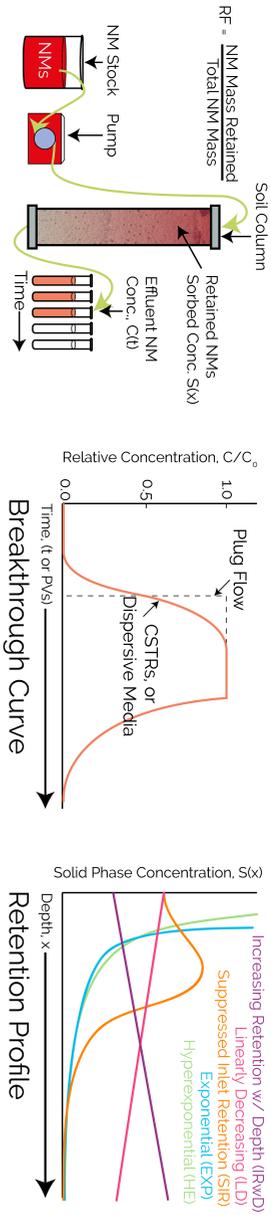


Figure 2: Soil column experiments take place in three steps:

1. The soil column experiment, which consists of a stock solution of fluid with a known concentration of ENMs, a pump, a soil column, and a series of vials, is setup.
2. During the experiment, ENM-containing fluid is pumped from the stock solution through the soil column and the effluent is collected in the vials. The concentration of ENMs in each vial is measured sequentially, typically using inductively coupled plasma mass spectrometry (ICPMS). The relative concentration of the effluent in comparison to the influent concentration concentration is then plotted as a function of the number of pore volumes pumped through the column, generating the breakthrough curve.
3. After the experiment is complete, soil is carefully excavated in 1–2 cm lifts from the soil column. In each lift, the concentration of ENMs is determined by desorbing the retained ENMs using an acid or base digestion and then measuring the concentration with ICPMS (concentration units of mass of ENMs adsorbed to the mass of granular media in each lift). Using this data, the retained mass of ENMs is plotted as a function of soil column depth generating the retention profile. Although the retention profiles appear continuous, they are composed of 5–10 data points.

1.4 HOW DOES THIS DISSERTATION ADD VALUE TO THE FIELD?

It is not yet analytically feasible to determine the concentration of ENMs in real soils. This is because many of the metals that comprise the most commonly produced ENMs (e.g., TiO_2 , FeO_x , ZnO , and CuO) occur naturally as nanoparticulates with similar size and shape.⁴⁷ These natural analogues are also present in concentrations within the range of predicted environmental concentrations for ENMs in soil, which make it difficult to distinguish ENMs from the background concentration of natural nanomaterials.^{16,19} As a result, investigation of the transport behavior of ENMs is constrained to bench-top soil column transport experiments performed in simple, synthetic soil systems (i.e., monodisperse, fully saturated, single-media soils).

In Chapter 2, it is shown that most advection-dispersion based particle transport models (PTMs) commonly employed to track ENMs in soil are fundamentally flawed. The fundamental flaw is not that they lack descriptive ability entirely, as they can well describe a large portion of observed ENM transport behavior.^{20,48–50} It is that the mathematical construction of these models considers neither the physicochemical properties of the ENMs, nor those of the system as a whole, explicitly. Instead, and because the mechanistic understanding of how physicochemical conditions influence ENM transport is incomplete, researchers employ quasi-mechanistic PTMs that involve a great deal of empirical specification and ad hoc adjustment. As a result, the applicability of these models beyond the exact scenario under which they were conditioned is low and current PTMs are merely descriptive tools without predictive power.

In Chapter 3, machine learning is employed to enable the prediction of ENM transport behavior and overcome the limitations established in Chapter 2. To support the machine learning effort, a database of physicochemical conditions (explanatory variables) extracted from more than 200 ENM transport experiments conducted in saturated porous media is developed. This database is used to train machine learning models to predict the fraction of ENMs retained over a soil column length, a measure of transport distance, and to predict the retention profile shape, a measure of transport behavior. The retained fraction is able to be predicted with a mean squared error between 0.025–0.033, and the retention profile shape is predicted with an F1-score (the weighted harmonic mean of precision and recall) between 60–70%. Further, by recursively removing physical and chemical

features to optimize model predictive performance, the importance of the physicochemical features to predict transport and behavior is quantitatively determined and ranked – thereby establishing strategic targets for inclusion in new mechanistic PTMs.

In Chapter 4, the database developed in Chapter 3 is expanded and enhanced by combining many of the experimental physicochemical descriptors to derive common dimensionless fluid dynamics and particle interaction parameters. Evaluation of the investigated range of experimental conditions within the database shows that the environmental relevance of the studies to date is low and notable gaps in the examined range are identified. Then, a machine learning decision tree classification is employed to quantitatively, and visually, disentangle the relationship between physicochemical conditions and ENM transport behavior within the database. This approach allows the identification of the order of importance and range of influence of the physical, chemical, and electrical conditions that control the retention profile shape. This is important because it offers an alternative to the current approach and opens a path toward a deeper level of mechanistic understanding of ENM transport in saturated porous media. The current approach has not advanced the mechanistic understanding of ENM transport because it relies on PTMs that have no predictive power and, for the case of nonexponential profiles, even little descriptive power. Through the presented alternative approach, it is possible, for the first time, to identify domains of conditions where nonexponential retention is frequent/dominant, and thus where entirely new mechanistic approaches can be developed, and domains of conditions where exponential retention is frequent/dominant, thus existing approaches may be sufficient.

In Chapter 5, the machine learning methods in Chapter 3 are employed to create, evaluate, and validate a model to predict the ENM attachment efficiency, α , using a database of physicochemical conditions extracted from quartz crystal microbalance with dissipation (QCM-D) experiments. Unlike traditional soil column experiments, QCM-D experiments directly monitor the deposition rate of particles onto an oscillating sensor surface in real-time, isolating deposition kinetics, and allowing for a direct measure of α without contributions from convection or filtration. α is an important kinetic transport parameter and is widely used in colloid filtration theory (CFT), which is described in Chapter 2 of this thesis, to quantify the likelihood of a particle attaching to a surface after a collision.

In Chapter 6, a brief outlook is provided.

2

CRITICAL ASSESSMENT OF ENM TRANSPORT MODELS

Contents

2.1	Abstract	42
2.2	Introduction	42
2.3	Methods	45
2.3.1	Particle Transport Models (PTMs)	45
2.3.2	Nanospecific model considerations	48
2.3.3	Selection of experimental data	49
2.3.4	Parameter fitting	50
2.4	Results and discussion	51
2.4.1	Hyperexponential retention	54
2.4.2	Linearly decreasing retention	55
2.4.3	Nonmonotonic retention	57
2.4.4	Monotonically increasing retention	59
2.5	Environmental implications	60

Note: Supplementary Information (SI) is available in Chapter 7: Appendix I

2.1 ABSTRACT

To reliably assess the fate of engineered nanoparticles (ENP) in soil, it is important to understand the performance of models employed to predict vertical ENP transport. We assess the ability of seven routinely employed particle transport models (PTMs) to simulate hyperexponential (HE), nonmonotonic (NM), linearly decreasing (LD), and monotonically increasing (MI) retention profiles (RPs) and the corresponding breakthrough curves (BTCs) from soil column experiments with ENPs. Several important observations are noted. First, more complex PTMs do not necessarily perform better than simpler PTMs. To avoid applying overparameterized PTMs, multiple PTMs should be applied and the best model selected. Second, application of the selected models to simulate NM and MI profiles results in poor model performance. Third, the selected models can well-approximate LD profiles. However, because the models cannot explicitly generate LD retention, these models have low predictive power to simulate the behavior of ENPs that present LD profiles. Fourth, a term for blocking can often be accounted for by parameter variation in models that do not explicitly include a term for blocking. We recommend that model performance be analyzed for RPs and BTCs separately; simultaneous fitting to the RP and BTC should be performed only under conditions where sufficient parameter validation is possible to justify the selection of a particular model.

2.2 INTRODUCTION

Environmental applications of nanomaterial-based products, particularly engineered nanoparticle (ENP) plant protection products (PPPs), represent a small, but rapidly growing, portion of the global nanotechnology market.⁸ The flux of nano titanium dioxide, nanosilver, and carbon nanotubes from direct application of ENP pesticides and PPPs to soil could be between $1 \mu\text{g kg}^{-1} \text{a}^{-1}$ and $11 \text{mg kg}^{-1} \text{a}^{-1}$.⁸ The detection of ENPs in complex environmental matrices is not yet analytically feasible.^{47,51} Therefore, ENP fate assessment in the subsurface must, for now, depend on model-derived determinations of ENP transport from soil column experiments.

Particle transport models based on the advection-dispersion equation (herein referred to as PTMs) are often employed to describe the transport of particles through saturated porous media. Particle

deposition and remobilization, or the transfer of particles between fluid and stationary phases, are described as first-order processes in these models. More complex PTMs additionally include parameters that influence deposition processes to consider heterogeneity in particle attachment efficiencies, blocking effects, multi-site kinetics, and/or pore structure influences (i.e., depth-dependent retention).^{48,50,52-54}

Numerous PTMs have been developed to predict the transport of colloids in saturated porous media.^{20,50,52,53,55,56} However, the applicability of these PTMs to ENPs has not been thoroughly examined. ENPs are viewed as a sub-class of colloids ranging from 1–100 nm and it is thought that the basic principles of particle transport modeling are valid for ENPs.²² In general, ENPs are prone to aggregate and preferentially deposit in the column inlet and produce hyperexponential retention profiles (RPs).^{28,33,40,41} However, ENPs also exhibit linearly decreasing (LD), nonmonotonic (NM), or monotonically increasing (MI) RPs.^{24,35,39,42} It is important to note that current PTMs are neither mathematically, nor conceptually, capable of generating LD, NM, and MI RPs (but several of current PTMs are able to generate HE RPs). This indicates that the models do not capture the processes that lead to LD, NM or MI RPs and are mechanistically not adequate for these types of RPs.

Interestingly, although LD, NM, and MI RPs cannot be described by the current PTMs, the models often yield breakthrough curves (BTCs) that are in close agreement with the BTCs measured in combination with LD, NM, and MI RPs. In this situation, the PTMs are often presumed to provide predictive insights to the behavior, and therefore the fate, of nanoparticles.^{28,33,35} The models are routinely fitted to measured RP and BTC data and parameters, such as deposition rate constants, remobilization rate constants, and maximum adsorption capacity, are routinely derived from the fit and reported.^{24,26,31,35,39,41} However, for cases where the models are mechanistically inadequate, good agreement between modeled and measured BTCs only indicates that the models may be used as a means of describing the observed breakthrough. It is unknown whether this descriptive ability reaches beyond the narrow experimental range under which the model was fitted to the measured data.

Therefore, there is a need to differentiate between cases where some of the PTMs may have some predictive power (HE RPs), cases where some of the PTMs are, at best, useful as descriptive tools for BTCs (LD and NM RPs), and cases where the models are fundamentally inadequate for the data from a soil column experiment and should not be applied

at all (MI RPs). Moreover, there is a need for guidance on how model parameters derived from a fit to measured RP and/or BTC data should be interpreted. In general, model fit alone does not provide a sufficient basis for inferring process descriptions for mechanisms within soil columns.^{57,58} More specifically, several factors contribute to a general misunderstanding of the capabilities of current PTMs to infer mechanistic understanding of ENP transport. First, validation of fitted parameters is often impossible, as some model parameters have no well-defined physical meaning, or, because data sets are relatively small in comparison to the number of fitted parameters.⁵⁹ Second, challenges often arise in the parameter fitting procedure itself, as collectively fitting free parameters by optimizing the agreement between model and measured data may yield large parameter uncertainties and, in some cases, non-unique parameter solutions.⁵⁹ Finally, it has been reported that multiple mechanistically different models fit measurements of ENP retention with similar goodness-of-fit, which undermines the ability of these models to provide mechanistic insights.^{26,33} These factors do not necessarily preclude useful application of some of the PTMs in certain cases. However, they need to be addressed more explicitly in order to avoid misleading or incorrect application and interpretation of PTMs to ENP transport.

The purpose of this work is not to claim that any of these models are 'realistic', or particularly suitable for modeling the movement of ENPs in soil. Our intention is to evaluate the ability of these models to describe the presented retention profiles and breakthrough curves because these models are used frequently by scientists in the field. Through this work we aim to (i) to reduce confusion concerning the suitability of PTMs for describing ENP transport in the subsurface by increasing awareness of model limitations and (ii) to improve on current environmental fate assessment practices for ENPs. In support of these goals, we systematically evaluate the suitability of seven routinely employed PTMs to describe representative HE, LD, NM and MI retention profiles from published soil column experiment data. Each of the seven examined PTMs is fitted to each retention profile. For all models, the resulting breakthrough curves are generated from the parameters determined by fitting the RP and visually evaluated against a simultaneous fit of the RP and BTCs from the publication from which the presented profile is obtained. The suitability and performance of each PTM is then visually and statistically evaluated.

2.3 METHODS

2.3.1 Particle Transport Models (PTMs)

For steady uniform flow vertically through an isotropic homogeneous porous medium, and without a concentration gradient or advection perpendicular to the primary flow direction, the mass balance equation for a substance present in the fluid phase is:

$$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t} \quad (1)$$

where C is the fluid phase particle concentration, t is time, x is distance from the column inlet, D_p is the longitudinal dispersion coefficient, v_p is the average pore water velocity, θ is the porosity, which is equivalent to the volumetric water content at complete saturation, and ρ_b is the media bulk density.^{60,61} The solid-phase mass balance is:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C - \frac{\rho_b}{\theta} k_r S \quad (2)$$

where S is the solid-phase particle concentration, k_d is the first-order deposition rate constant, and k_r is the first-order remobilization rate constant.^{61,62}

Equations 1 and 2 form the fundamental single-site deposition-remobilization PTM, presented as M3, from which six model modifications are derived, see Table 1.

The seven PTMs are presented in order of increasing complexity as follows: a colloid filtration theory (CFT) model (M1), a CFT model that includes a bimodal distribution of particle attachment efficiencies (M2), a single-site deposition-remobilization model (M3), a dual-site deposition-remobilization model (M4), a single-site deposition-remobilization model that includes dynamic Langmuirian blocking (M5), a single-site deposition-remobilization model that includes depth-dependent retention (M6), and a single-site deposition-remobilization model that includes both dynamic Langmuirian blocking and depth-dependent retention (M7).

For each model, Table 1 presents the modeled processes, governing equations, indicates the fitted parameters, and lists key references. Model setup, shared assumptions, boundary conditions, and

programming details for the considered PTMs are provided in the Supporting Information (SI Chapter 7: Appendix I).

Table 1: Routinely employed particle transport models.

Model	Description	Modeled Process(es)	Governing Mass Balances	Free Parameters	Refs.
M1	Colloid filtration theory (CFT)	- Deposition	$\frac{\partial C}{\partial t} = -v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C; \quad k_d = \frac{3(1-\theta)}{2d_c} \alpha v_p \eta_0$	α	20,61
M2	Dual deposition mode CFT	- Fast deposition - Slow deposition	$\frac{\partial C}{\partial t} = -v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = f k_{d,f} C + (1-f) k_{d,s} C$ $k_{d,f} = \frac{3(1-\theta)}{2d_c} \alpha_f v_p \eta_0;$ $k_{d,s} = \frac{3(1-\theta)}{2d_c} \alpha_s v_p \eta_0$	f α_s	50,55,63
M3	Single-site deposition remobilization	- Deposition - Remobilization	$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C - \frac{\rho_b}{\theta} k_r S$	k_d k_r	62
M4	Dual-site deposition remobilization	- Deposition - Remobilization	$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t} - \frac{\rho_b}{\theta} \frac{\partial S_2}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C - \frac{\rho_b}{\theta} k_r S$ $\frac{\rho_b}{\theta} \frac{\partial S_2}{\partial t} = k_{d,2} C - \frac{\rho_b}{\theta} k_{r,2} S_2$	k_d k_r $k_{d,2}$ $k_{r,2}$	53
M5	Single-site deposition remobilization & blocking	- Deposition - Remobilization - Langmuirian blocking	$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d \psi_b C - \frac{\rho_b}{\theta} k_r S;$ $\psi_b = \left(1 - \frac{S}{S_m}\right)$	k_d k_r S_m	33,52
M6	Single-site deposition remobilization & depth-dependent retention	- Deposition - Remobilization - Depth-dependent retention	$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d \psi_s C - \frac{\rho_b}{\theta} k_r S;$ $\psi_s = \left(\frac{d_c + x}{d_c}\right)^{-\beta}$	k_d k_r β	48,49,52,54
M7	Single-site deposition remobilization, blocking, & depth-dependent retention	- Deposition - Remobilization - Langmuirian blocking - Depth-dependent retention	$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t}$ $\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d \psi_b \psi_s C - \frac{\rho_b}{\theta} k_r S$ $\psi_b = \left(1 - \frac{S}{S_m}\right); \quad \psi_s = \left(\frac{d_c + x}{d_c}\right)^{-\beta}$	k_d k_r S_m β	33,48,49,54

Legend:

C : aqueous ENP concentration [$\text{kg}_{\text{nano}} \text{m}^{-3}$]
 S_2 : 2nd site solid-phase concentration [$\text{kg}_{\text{nano}} \text{kg}_{\text{soil}}^{-1}$]
 v_p : pore water velocity [m s^{-1}]
 t : time [s]
 α_f, α_s : fast/ slow attachment efficiency [-]
 x : distance from column inlet [m]
 η_0 : single-collector contact efficiency [-]
 ρ_b : media bulk density [$\text{kg}_{\text{soil}} \text{m}^{-3}$]
 D_p : longitudinal dispersion coefficient [$\text{m}^2 \text{s}^{-1}$]
 θ : granular media porosity [-]
 S : primary site solid-phase concentration [$\text{kg}_{\text{nano}} \text{kg}_{\text{soil}}^{-1}$]
 $k_{d,2}$: secondary site deposition rate constant [s^{-1}]
 d_c : granular media average diameter [m]

k_r : primary site remobilization rate constant [s^{-1}]
 α : attachment efficiency [-]
 $k_{r,2}$: secondary site remobilization rate constant [s^{-1}]
 k_d : primary site deposition rate constant [s^{-1}]
 ψ_b : dynamic Langmuirian blocking function [-]
 S_m : maximum solid-phase concentration [$\text{kg}_{\text{nano}} \text{kg}_{\text{soil}}^{-1}$]
 f : particle population fraction associated with α_f [-]
 ψ_s : depth-dependent retention function [-]
 $k_{d,f}$: deposition associated with α_f [s^{-1}]
 β : empirical depth-dependent retention parameter [-]
 $k_{d,s}$: deposition associated with α_s [s^{-1}]

2.3.2 *Nanospecific model considerations*

It is important to note that all of the described PTMs were developed to model colloid transport, but are routinely employed to model ENP transport.^{33,35,39,64–67} While ENPs may fall within the technical definition of colloids, their physicochemical properties (e.g., composition, shape, isoelectric point, reactivity) may be different.²² However, the mathematical construction of the PTMs does not explicitly consider any physicochemical property of the ENPs, see mass-balance equations in Table 1. None of the parameters in the mass-balance equations reflects type, size, shape, surface properties or aggregation behavior of ENPs. M1 and M2 are the only models where some ENP properties are used. In these models, the parameter η_0 is a component of the equation used to estimate the deposition rate constant, k_d (Equation 9 in Chapter 7: Appendix I). η_0 is quantified according to Tufenkji and Elimelech correlation equation.⁶¹ This equation considers the ENP diameter and van-der-Waals interaction energy of the ENPs with the collector. However, correlation equations for η_0 were not developed to consider ENP-soil interactions (e.g., non-spherical particles and collectors, the presence of natural organic matter, steric interactions, and repulsive DLVO interactions) and, as such, are not likely to reflect conditions in ENP column experiments.^{61,68,69} Furthermore, it is difficult to determine the influence of the ENP properties on the performance of M1 and M2 because η_0 is combined with the attachment efficiency, α , in the equation for k_d , and α is used as a free parameter whose value is derived from the fit of the model to the RP and/or BTC. Therefore, it may not be η_0 that determines the fit of model to the retention profile (and the suitability of the model), but the magnitude of the fitted value of α .

Soil properties that enter the mass-balance equations (Table 1) are bulk density, ρ_b , porosity, θ , the average diameter of collectors, d_c , and dispersivity of a tracer in the porous media, D_p (note that D_p is not derived specifically for ENPs). None of the models include parameters to consider the chemical composition, roughness, grain shape, or concentration of natural organic matter of the soil.

The remaining model parameters are rate constants for attachment and remobilization, maximum solid-phase concentration, etc., see Table 1. These parameters are treated as free parameters in the models whose values are determined by fitting the model to the measured RP and/or BTC. In an implicit way, these parameters reflect the properties of the ENPs and of the soil and pore water, and the interaction of the ENPs

with the soil under the given conditions, but they are not determined according to the properties of the ENPs and the system before the model is run. Therefore, the model results, i.e. the fitted parameters and the calculated RPs and BTCs, reflect the properties of the ENPs and how ENPs interact with the soil, whereas the models, as they are defined by the mass-balance equations, do not take into account any ENP-specific information.

2.3.3 Selection of experimental data

Data employed in the model assessment were selected as follows. First, more than 100 nanomaterial column transport studies from 2001–2014 were reviewed. Second, 52 of the most recent nanoparticle transport publications (2008–2014) were evaluated in detail to determine whether sufficient information was available to adequately model ENP transport with each PTM; this is summarized in Table 9 (Chapter 7: Appendix I). Of these 52 publications, 22 (42.3%) reported retention profiles and 12 (23%) publications contained sufficient data to employ all 7 PTMs. Of these 12 publications with sufficient data, 6 (11.5%) were excluded because of other limitations (e.g., fewer than 8 retention profile data points; under-saturated soil, and/or the experimental procedure violates model assumptions of a model employed for this assessment).

From the remaining six (11.5%) publications^{24,33,35,39,41,42}, a total of 42 retention profiles and associated parameters were recorded and the profiles were modeled using each PTM. Five publications noted HE retention profiles.^{33,35,39,41,42} Two publications noted LD retention profiles.^{35,42} Three publications noted NM retention profiles.^{24,35,39} Two publications noted MI retention profiles.^{24,39} Shape designation was determined by reductive fitting where each profile was fit with linear, exponential, and power functions. The shape designation with the highest adjusted R^2 value were selected to describe the profile. A single archetypal example from each retention profile type was then selected for our assessment and is shown in Figure 4; parameters used in the modeling of the four profiles are given in Table 2 (Chapter 7: Appendix I). A selection of 16 additional HE, LD, NM, and MI retention profiles is provided Figure 22 (Chapter 7: Appendix I).

2.3.4 *Parameter fitting*

Soil column experiments provide a breakthrough curve (BTC) (i.e., the fraction of material leaving the column, expressed as a function of the number of pore volumes pumped through the column), and a retention profile (i.e., the fraction of ENPs sorbed to the packed bed as a function of depth). The retention profile provides spatial information across the column at the end of the experiment; the breakthrough curve provides temporal information concerning the relative fraction of particles leaving the column. The mechanistic link between the retention profile and breakthrough profile is, in theory, so close that parameter determination from either profile alone should be sufficient to correctly simulate the other.

During the model construction phase of this research, we observed that simultaneous fitting of the retention profile and breakthrough curve can result in a sub-optimal fit of the retention profile, as the effective sample size (i.e., number of data points vs. number of fitted parameters) for fitting the breakthrough curve is often an order of magnitude greater than the effective sample size for fitting the retention profile. The relative weighting of the multi-objective fitting can be manipulated⁴² and therefore it is unclear what the acceptable goodness-of-fit is. Thus, to best illustrate the conceptual limitations of the models, we estimate model parameters directly from fitting the models to the retention profiles and then assess the ability of each model to describe both the retention profile and breakthrough curve.

For each model, one or more parameters were estimated using the Levenberg-Marquardt algorithm (LMA), an unbounded, non-linear, least-squares curve fitting method.⁷⁰ For several model scenarios, LMA parameter estimation failed to converge or resulted in negative parameters. For these situations, boundaries were applied to the parameter estimation routine using a trust-region nonlinear fitting algorithm.⁷¹ Parameter ranges were then chosen to encompass a realistic range of parameter values (i.e., k_d , k_r , S_m , and $\beta > 0$; $0 < \alpha < 1$). All parameter estimations employed a multi-start global optimization to prevent local minima solutions generated by the influence of initial conditions.

Statistical assessment of how well a PTM fits the measured retention profile was performed by evaluation of the bias-adjusted Akaike

information criterion (AIC_c).⁷² Detailed information concerning the AIC_c analysis is provided in Chapter 7: Appendix I.

2.4 RESULTS AND DISCUSSION

For each representative retention profile (HE, LD, NM, and MI) the results and discussion are organized as follows. First, a brief overview of experimental conditions where each retention profile has been observed is provided. Note that this overview contains only a partial list of well-cited and recent publications and should not be considered an exhaustive review. Second, model performance is evaluated with reference to any visually identifiable systematic deviations between the observed and modeled RPs and BTCs (Figure 4); significant results from the AIC_c model plausibility analysis are presented, when applicable, and the most plausible models for each profile are identified. Note that BTC recession is not modeled. Finally, the suitability of the models is discussed with reference to the findings of the performance evaluation.

Parameter fitting results for each profile are available in Table 10 (Chapter 7: Appendix I). We do not discuss these results, nor the relationship of the parameters to the data presented in the original papers, for several reasons. First, our assessment focuses on fitting the measured retention profiles to illustrate the conceptual limitations of the models to describe HE, LD, NM, and MI retention profiles. As such, parameter values derived from our fit may be substantially different from published values determined by fitting models to retention profiles and breakthrough curves simultaneously, as is performed by the software package, Hydrus-1D.⁷³ Second, parameter values from the fit of a PTM to a retention profile cannot be directly compared or validated against values of the same parameters derived from other models (i.e., model construction changes the meaning of the parameters).³³ Therefore, no comparative model performance conclusions can be drawn from inspection of the parameter results. Finally, throughout this work it is argued that parameters derived from forced application of the PTMs to fit LD, NM, and MI retention profiles should not be generalized. Our evaluation of the applicability of the models is independent of the value of fitted parameters, and the parameter estimation results are not provided in the main text to prevent their misapplication.

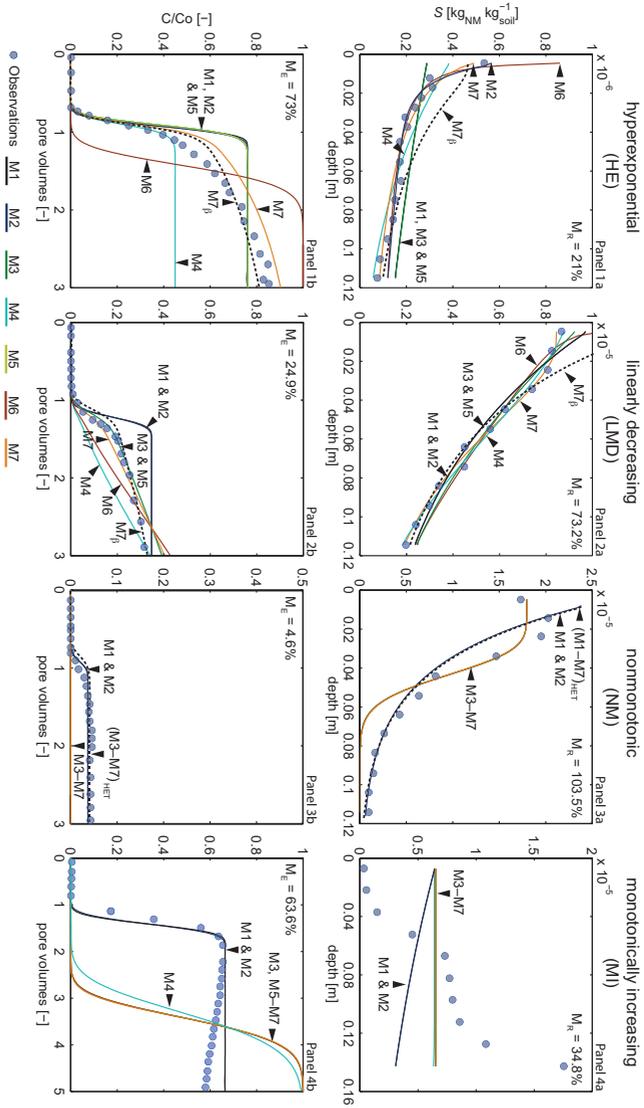


Figure 4: Digitized retention profiles (dots) and fitted model results (lines) are presented in panels 1a, 2a, 3a, and 4a; digitized breakthrough profiles (dots) and fitted model results (lines) are presented in panels 1b, 2b, 3b, and 4b. Note that the y-axis of panel 1a is presented at a 10x reduced scale in comparison to panels 2a, 3a, and 4a. Mass balances reported by the original authors for the retained ($M_R\%$) and the effluent fraction ($M_E\%$) are shown in panels 1a–4a and panels 1b–4b, respectively. Dashed lines represent fitted model results when β is held constant ($M7\beta$).

Table 2: Modeling parameters for the retention profiles in Figure 4 as reported in the original studies.

Parameter	Symbol	Units	Retention Profile Shape			
			HE	LD	NM	MI
Column Length	C_l	[m]	0.12	0.12	0.12	0.15
Column Width	C_w	[m]	0.03	0.03	0.03	0.025
Number of Pore Volumes	N_{PV}	[m ³]	$1.2 \cdot 10^{-4}$	$9.0 \cdot 10^{-5}$	$9.0 \cdot 10^{-5}$	$8.7 \cdot 10^{-5}$
Porosity	θ	[-]	0.46	0.427	0.403	0.37
Influent Concentration	C_{inf}	[kg m ⁻³]	$1 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$2 \cdot 10^{-2}$
pH	pH	[-]	8.5	6-7	6-7	6
Ionic Strength	I	[mM]	1.0	1.0	5.0	$2.25 \cdot 10^{-2}$
Pore Velocity	v_p	[m s ⁻¹]	$2.39 \cdot 10^{-4}$	$1.17 \cdot 10^{-5}$	$2.94 \cdot 10^{-4}$	$1.38 \cdot 10^{-4}$
Particle Diameter	d_p	[m]	$1.90 \cdot 10^{-7}$	$8.45 \cdot 10^{-8}$	$8.45 \cdot 10^{-8}$	$1.50 \cdot 10^{-7}$
Collector Diameter	d_c	[m]	$3.50 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$	$2.75 \cdot 10^{-4}$
Longitudinal Dispersivity	α_L	[m]	$6.9 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	$9.7 \cdot 10^{-4}$	$9.5 \cdot 10^{-4}^\dagger$
Hamaker Constant	A	[J]	$5 \cdot 10^{-20}^\ddagger$	$2.37 \cdot 10^{-21}$	$2.37 \cdot 10^{-21}$	$4.5 \cdot 10^{-20}$
ENP Type			MWCNTs	Ag	Ag	TiO ₂
Location (Panel in Figure 4)			1a & 1b	2a & 2b	3a & 3b	4a & 4b
Reference			33	41	41	24

[†]This longitudinal dispersivity was determined by digitizing and fitting the breakthrough curve using CXTFIT.⁷⁴ [‡]Hamaker constants for MWCNTs were not measured by Kasel et al.³³. Instead, a value for single-walled CNTs, which may be considered representative for MWCNTs, is taken from Rajter et al.⁷⁵.

2.4.1 Hyperexponential retention

HE retention profiles have been reported for a variety of ENPs including nano Fullerene aggregates in various porous media and ionic strengths³⁹, titanium dioxide nanoparticles at various solution conditions, concentrations, and flow rates^{23,25}, copper-bearing hydroxyapatite nanoparticles at various ionic strength conditions⁴⁰, cerium oxide nanoparticles at various pHs³⁶, copper oxide nanoparticles in various concentrations of humic and fulvic acids³¹, nano zinc oxide in various solution conditions both with and without humic acid, and in the presence of an *Escherichia coli* biofilm²⁸⁻³⁰, multi walled carbon nanotubes (MWCNTs) in various column lengths, grain sizes, flow rates, and concentrations^{33,76}, biochar nanoparticles produced at various pyrolysis temperatures⁴¹, and nanosilver at various solution conditions, concentrations, and flow rates⁶⁵.

A measured HE RP from Kasel et al.³³ for the transport of MWCNTs is presented in Figure 4, panel 1a; along with the fitted RPs from the seven models; BTCs generated by each model are shown in Figure 4, panel 1b. Panel 1a allows facile differentiation between models that can simulate HE RPs from those that cannot. M1, M3, and M5, which cannot model HE retention, show exponential retention profiles that underestimate retention from 0.0 to 0.015 m and overestimate retention from 0.015 to 0.12 m. M2, M6, and M7 well describe the RP, and M4 is in between.

If only the retention profile is considered, M6 is the most plausible model (Akaike weight = 0.99). Note that M2 and M7 fit the retention profile with similar error as M6 ($M6_{RSS} = 8.72 \cdot 10^{-15}$; $M2_{RSS} = 2.29 \cdot 10^{-14}$; $M7_{RSS} = 1.94 \cdot 10^{-14}$). However, if the RP and the BTC are considered in combination, it turns out that M2 and M6 are unable to describe the transient decrease in retention (increase in breakthrough) exhibited in the breakthrough profile. Therefore, M7 is the most suitable model to describe this set of data.

Examination of further profiles from Kasel et al.³³, Wang et al.⁴¹, and Wang et al.³⁹ exhibiting varying degrees of hyperexponentiality shows that M7 is frequently needed to describe both the RP and BTC, but is not always the most suitable model. In two additional cases (Chapter 7: Appendix I - Figure 23 and Table 14), M2 and M6 most suitably describe the RP and also the BTC. This result is important because it shows that there is not one model that is appropriate for all conditions and that the inclusion of additional model parameters does not guarantee better

model performance. To avoid applying an overparameterized model to HE RPs, we advise to evaluate M2, M6, and M7 in congress and apply an AIC_c model suitability analysis to select the most plausible model.

In using M7, particular attention needs to be paid to the treatment of the parameters, β and S_m . It is common practice to pre-define β .^{35,41,42,77} However, this can easily result in sub-optimal fit, as shown by the dashed lines in Figure 4; panel 1a and 1b, which represent the RP and BTC of M7 with beta pre-defined to be 0.765, as done by Kasel et al.³³. Our result for M7, in contrast, (orange line) shows that leaving β a free parameter gives a superior fit of the RP without introducing much error in the BTC. In addition to sub-optimal fitting caused by pre-defining β , one must also consider that the relationship of β to physical parameters (e.g., grain size, roughness, particle shape) is not currently known. Thus, because fitting is the only methodology to estimate β , there is no justification to pre-define the parameter.

Regarding S_m , it is possible to determine S_m with batch adsorption experiments, as done by Wang et al.³⁹. This is a good way of supporting the assumption of blocking with independent empirical data. However, the methodology of measuring nanoparticle adsorption must be carefully selected, as nanoparticles do not have partition coefficients like chemicals (i.e., thermodynamic equilibrium), but their sorption is kinetically controlled.¹³

2.4.2 *Linearly decreasing retention*

LD retention profiles have been reported for a variety of nanomaterials including biochar nanoparticles in iron oxyhydroxide-coated sand and in various concentrations of humic acid⁴², nanosilver at low flow rates³⁵, fullerenes and titanium dioxide nanoparticles at various solution conditions²³, and capped zinc oxide nanoparticles⁴⁵.

A measured LD retention profile from Liang et al.³⁵ for the transport of nanosilver is presented in Figure 4, panel 2a; along with fitted RPs for all seven models; the corresponding BTCs are shown in Figure 4, panel 2b. M1 and M2 present identical exponential RPs that slightly over-estimate retention from 0.0 to 0.01 m, slightly under-estimate retention from 0.02 to 0.06 m, and again slightly over-estimate retention from 0.08 to 0.12 m. M3–M7 exhibit similar, shallowly sigmoidal-shaped retention profiles that well describe retention from

0.0 to 0.09 m, but slightly over-estimate retention from 0.09 to 0.12 m. If only the retention profile is considered, M7 is the most plausible model (Akaike weight = 0.891; $M7_{RSS} = 3.69 \cdot 10^{-13}$).

According to Figure 4, panel 2b, breakthrough occurred near one pore volume and was preceded by a gradual decrease in retention (linearly increasing portion of the measured breakthrough curve). Modeled BTCs generated by M3 and M5 more closely describe the gradual decrease in retention exhibited in the measured BTC than M7 does. Detailed inspection indicated that the characteristic gradual decrease in retention typical of blocking effects can, under some circumstances, be described by less complex models (such as M3) when attachment and remobilization rate constants are similar in magnitude. Overall, M3, M5 and M7 have a similar performance in describing the RP and BTC; M3 then is the most plausible model because it has the lowest number of parameters.

Concerning the application of M7, this type of profile also provides evidence that pre-defining β can result in sub-optimal fit, as shown by the dashed line (Figure 4; panel 2a and 2b) where β was pre-defined to be 0.432 by Liang et al.³⁵ Leaving β a free parameter resulted in a superior RP fit without introducing much error in the BTC.

The causation for LD retention is not explicitly known, although it may be that LD retention is an interim state between a hyperexponential, or exponential, and uniform retention profile that occurs as a function of blocking and straining. Our assessment results cannot substantiate this theory, however, as we have no information on the evolution of the retention profile over the length of the experiment. As such, the contribution of blocking and straining to the formation of LD retention profiles is not clear.

Concerning the development of more suitable models, modifying the PTM to employ zero, or near zero, order deposition kinetics can generate LD retention profiles. However, the degree to which this modification can be reconciled with current theory has not yet been demonstrated and further research is required to quantitatively describe this behavior. Significant insight concerning the development of LD retention profiles may be obtained by evaluating the changes in the retention profile as the experiments progress.

2.4.3 Nonmonotonic retention

NM retention profiles have been reported for a variety of nanomaterials including nano titanium dioxide at various solution conditions, flow rates, and in iron oxide-coated quartz sands both with, and without humic, acid^{24,26,27}, Fullerene nanoparticles at various flow rates and grain sizes⁶⁶, copper oxide nanoparticles in certain organic buffers and solution conditions³¹, and nanosilver at various flow rates and solution conditions both with, and without, surfactant³⁵.

A measured NM retention profile from Liang et al.³⁵ for the transport of silver nanoparticles is presented in Figure 4, panel 3a; peak retention is observed at ca. 0.012 m from the column inlet. The same graph also shows the fitted RPs for all seven models.

Inspection of panel 3a shows two distinct groupings of models. M1 and M2, which do not cover remobilization or dispersion, present exponential profiles that strongly overestimate retention from 0.0 to 0.012 m, underestimate retention from 0.012 m to 0.075 m, and well describe retention from 0.075 to 0.12 m. M3–M7, which consider remobilization and dispersion, present essentially identical sigmoidal retention profiles that well describe the near-influent ENP concentration, but clearly underestimate peak retention and poorly describe retention at depth. However, as all of the models show easily identifiable and systematic deviations, none of the models well describes the NM retention profile; it is known that none of the models can explicitly describe NM retention.³⁵ If the models are considered as purely descriptive tools that provide an approximation of the observed RP, M3 is the most plausible model (Akaike weight = 0.761; $M_{3RSS} = 2.12 \cdot 10^{-11}$).

Inspection of the measured BTC (Figure 4, panel 3b) indicates that a majority of ENMs were retained throughout the course of the experiment (C/C_0 does not exceed 0.05 in Figure 4, panel 3b).

Interestingly, the BTCs generated by M3 as the most plausible model (and also by M4–M7) are qualitatively different from the observed BTC because they show no breakthrough at all. The BTCs generated by M1 and M2, in contrast, are in perfect agreement with the measured BTC. Importantly, when we allow a higher error tolerance (HET) in the parameter fitting routine for the RP, the BTCs obtained with M3–M7 change (shown as the dashed line in Figure 4, panels 3a and 3b) and agree very well with the measured BTC and also converge with the BTCs of M1 and M2. Under these conditions M1 is the most plausible

model, because it describes both RP and BTC equally well as all other models and has the lowest number of parameters.

This case, where model results are in poor agreement with the measured RP, demonstrates an important limitation of using these PTMs in a parameter fitting exercise for measured RPs and BTCs. For the more complex models, M₃–M₇, the best fit for the RP leads to a poor fit of the BTC, and vice versa. The conceptual problem in this situation is that there is no agreed-upon acceptable fitting error, which makes the fitting exercise highly subjective. Simultaneous fit of the models to both RP and BTC provides no solution here, because this just masks the lack of model performance on one of the two subsets of the data, RP or BTC.

Comparing the performance of all seven PTMs shows, in hindsight, that the conceptual problem can, in this particular case, be avoided by using M₁ or M₂. However, this is a purely ad-hoc solution and illustrates that the most plausible model can only be found by trial and error and that the models are not based on a sufficient understanding of the processes governing the fate of ENPs in the column. Finally, this case shows that simultaneous fitting of the RP and BTC may mask the inability of a model to describe the RP.

Currently, the causation of NM retention behavior is not explicitly known. Experimental evidence suggests that NM retention profiles are qualitatively influenced by the pore water velocity and the presence of surfactants, and may be caused by the migration of deposited ENPs, competition for retention sites, and/or due to particle aggregate reformation (i.e., a shift in particle aggregate structure due to aggregate-grain collision).^{24,26,35,78–80} However, further research is needed to better understand the influence of hydrodynamics, surfactants, and aggregate reformation on the retention profile, and to implement these mathematically into PTMs. Concerning the development of more suitable models, NM retention profiles can be generated with minimal model adjustment, for instance by including a mobile solid-phase or by allowing $\beta < 0$.⁸¹ However, how these changes can be experimentally confirmed, or the degree to which these modifications can be reconciled with existing theory for ENPs, has not yet been fully investigated. Nonmonotonic retention profiles can also be simulated by the 2-species model presented by Bradford et al.⁸⁰. Within the model by Bradford et al.⁸⁰, NM retention behavior is mechanistically described using several kinetic exchanges rates (i.e., kinetic exchange rates between aggregated and monodisperse species, and the aggregated species production function). We were not able to

employ this model here because the aggregated species production function employed by Bradford et al.⁸⁰ can only be determined if the aggregated species is tracked in the effluent. This information is not available in any of the 25 publications with sufficient information about experimental conditions identified in our literature review (See Section 2.3.3).

2.4.4 *Monotonically increasing retention*

MI profiles are not common, but have been observed for nano titanium dioxide at low ionic strength conditions and in iron oxide-coated quartz sands with humic acid^{24,27}, and Fullerene nanoparticles at various flow rates and grain sizes⁶⁶.

A measured MI retention profile reported by Chen et al.²⁴ for the transport of titanium dioxide nanoparticles is presented in Figure 4, panel 4a, along with the fitted RPs for all seven PTMs. In the literature, no applications of any of the seven PTMs to this type of retention profile have been reported. We include this profile because it shows the diversity of results obtained in column experiments with ENPs.

For all seven models, the fitted RPs in Figure 4, panel 4a, deviate considerably from the measured retention profile. Consequently, Akaike weights cannot provide reliable information on model plausibility.

Figure 4, panel 4b, shows the measured and the modeled BTCs. M1 and M2 yield BTCs that are close to the measured data, but for all other models, the BTCs are very different from the measured data. This example shows that even in cases where the models are obviously inappropriate, agreement between measured and modeled BTCs may be obtained. This adds emphasis to our earlier recommendation that the models should not be used to describe measured BTCs without a detailed investigation of model performance for RPs and BTCs separately.

In the study where the presented MI retention profile was observed, Chen et al.²⁴ proposed that observations of increasing retention may be influenced by aggregate reformation due to collisions between nanoparticle aggregates and sand grains during transport. However, until a PTM that includes aggregate reformation has been developed, or the understanding of MI retention is more complete, parameters determined from the application of current PTMs to MI

retention profiles are meaningless and should not be employed to determine ENP transport.

2.5 ENVIRONMENTAL IMPLICATIONS

Soil column experiments conducted with ENPs exhibit a much wider range of retention profiles than can be described by current PTMs; the observed RPs are qualitatively very diverse and include HE but also LD, NM and MI shapes. HE RPs can be described by some of the presented PTMs, but NM, LD, and MI retention profiles cannot, mathematically, be generated by any of these models. Several important implications are noted concerning the applicability of the presented PTMs to simulate HE, NM, LD, and MI retention profiles from soil column experiments with ENPs.

Hyperexponential profiles. Models M2 and M4–M7 can generate HE RPs; of these, M2 and M5–M7 show good fits of the RP in Figure 4, panel 1a. With the parameters obtained from the fit of the RP, M7 also describes the BTC with good agreement (Figure 4, panel 1b). However, we do not recommend M7 as the most plausible PTM for all HE RPs and suggest that multiple models (i.e., M2, M6, and M7) be applied in congress and the best model selected using an AIC_c evaluation. Regarding the common practice of pre-defining β , we recommend that without experimental verification that blocking actually occurs within the column, or until a clear physical meaning for β becomes apparent, β should remain a free parameter.

Other profiles. The other profiles cannot be mathematically generated by any of the PTMs considered here. In some cases, some of the models have satisfying descriptive performance, for example M3, M5 and M7 for the LD profile and the corresponding BTC. However, because the models do not capture the underlying mechanisms for these RPs, they have no predictive power and it is not clear to what extent good descriptive power of a model that was observed in a specific case can be generalized. In particular, forced application of PTMs to NM (and MI) retention profiles results in large and/or systematic deviations between models and measured RPs, and poor model performance. Therefore, parameters determined from the application of existing PTMs to NM and MI profiles are meaningless and should not be used for ENP fate assessment. Concerning the application of PTMs to simulate LD retention profiles, the situation is more complex. In contrast to NM and MI retention profiles, LD retention profiles can be

approximated by the existing PTMs. However, because the PTMs cannot explicitly generate LD retention, parameters determined from a fit of these PTMs to LD retention profiles have no predictive power and are inappropriate for ENP fate assessment.

Blocking. We observe that the effect of a term for blocking, such as M_7 , can also be accounted for by models that do not explicitly include a term for blocking, such as M_6 , under some circumstances. This implies that it may not be possible to infer whether or not blocking occurs within the column solely by model application to RPs and BTCs. To prevent model misapplication and unjustified inferences on processes taking place in the column, it is recommended to employ a PTM that includes blocking only after experimental verification that blocking occurs within the column, and after independent determination of the parameter S_m .⁸²

Model selection. Several experimental and theoretical metrics can be employed to enhance model selection and gain process understanding. Factors leading to non-exponential behavior can be evaluated by inspection of experimental conditions. For example, straining may be evaluated by using the ratio of particle diameter to collector diameter, d_p/d_c ⁸³; heterogeneity in surface charge characteristics can be evaluated by inspection of the variability of collector size/surface charges throughout the column or by observations of dependence, or independence, of grain size on the magnitude of retention⁸⁰; and direct inspection of particle remobilization may help distinguish wedging from straining mechanisms^{57,58}. These metrics may lead to more robust models in the future, but insufficient mechanistic understanding still prevents these metrics from yielding unequivocal process identification currently. For example, a d_p/d_g ratio below $8 \cdot 10^{-3}$ implies that straining is not likely to be responsible for depth-dependent retention, but it does not definitively indicate which mechanism is responsible.⁸³ As depth-dependent retention may also be caused by mechanisms other than straining, such as heterogeneity in particle attachment efficiencies, wedging, and variability of colloid and collector charge characteristics^{52,84}, this metric provides only qualitative information about the plausibility of straining.

Our analysis shows that the applicability of current PTMs to RPs and BTCs measured for different ENPs varies widely. A first important recommendation for the selection of appropriate models is that PTMs should not directly be fitted to RPs and BTCs simultaneously. Instead, model performance should be investigated separately for RPs and BTCs. Second, models should be employed only under conditions

where sufficient parameter validation is possible to justify the selection of a particular model. This is primarily the case for HE RPs. Without sufficient parameter validation, the application of these models, even application that results in a high goodness-of-fit, is not sufficient to infer which processes actually occur within the soil column. For example, in several cases, BTCs can be described relatively well even when a model is obviously not able to reproduce the corresponding RP (see MI profile and BTC in Figure 4, panels 4a and 4b). However, such a result is purely descriptive, does not provide any mechanistic insights, and cannot be generalized. This does not imply that the current models are invalid and cannot, or should not, be used. To the contrary, we argue that these models should be employed, but only under conditions where sufficient parameter validation is possible to justify the selection of a particular model. We recommend that model practitioners avoid forced application of fundamentally unsuitable models, as well as refrain from modeling practices that may inadvertently misconstrue or misrepresent transport parameters and processes.

Note: SI is available in Chapter 7: Appendix I

3

PREDICTION OF ENM TRANSPORT FROM PHYSICOCHEMICAL PROPERTIES

Contents

3.1	Abstract	64
3.2	Introduction	64
3.3	Methods	67
3.3.1	Nanomaterial Transport Experiment Database	67
3.3.2	Random Forest Regression and Classification	69
3.3.3	Feature selection with recursive feature elimination with 5-fold cross validation (RFECV)	70
3.3.4	Database partitioning and model structure	72
3.4	Results and discussion	75
3.4.1	Prediction of the retained fraction (RF) . . .	75
3.4.2	Prediction of the retention profile (RP) shape	80
3.5	Conclusions	84

Note: Supplementary Information (SI) is available in Chapter 8:
Appendix II

3.1 ABSTRACT

In the last 15 years, the development of advection-dispersion particle transport models (PTM) for the transport of nanoparticles in porous media has focused on improving the fit of model results to experimental data by inclusion of empirical parameters. However, the use of these PTMs has done little to elucidate the complex behavior of nanoparticles in porous media and has failed to provide the mechanistic insights necessary to predictively model nanoparticle transport. The most prominent weakness of current PTMs stems from their inability to consider the influence of physicochemical conditions of the experiments on the transport of nanoparticles in porous media. Qualitative physicochemical influences on particle transport have been well studied and, in some cases, provide plausible explanations for some aspects of nanoparticle transport behavior. However, quantitative models that consider these influences have not yet been developed. With the current work, we intend to support the development of future mechanistic models by relating the physicochemical conditions of the experiments to the experimental outcome using ensemble machine learning (random forest) regression and classification. Regression results demonstrate that the fraction of nanoparticle mass retained over the column length (retained fraction, RF; a measure of nanoparticle transport) can be predicted with an expected mean squared error between 0.025–0.033. Additionally, we find that RF prediction was insensitive to nanomaterial type and that features such as concentration of natural organic matter, ζ -potential of nanoparticles and collectors and the ionic strength and pH of the dispersion are strongly associated with the prediction of RF and should be targets for incorporation into mechanistic models. Classification results demonstrate that the shape of the retention profile (RP), such as hyperexponential or linearly decreasing, can be predicted with an expected F1-score between 60–70%. This relatively low performance in the prediction of the RP shape is most likely caused by the limited data on retention profile shapes that are currently available.

3.2 INTRODUCTION

Predicting the transport behavior of nanomaterials in the environment is important for managing both the risks and benefits associated with nanomaterials use. Progress towards this goal, however, has been slow.

In the last 15 years, the development of advection-dispersion particle transport models (PTMs) for the transport of nanoparticles in porous media has focused on improving the fit of model results to experimental data by inclusion of empirical parameters.^{48,50,53} However, inclusion of empirical parameters has done little to unravel the complex transport behaviour of nanomaterials and has failed to provide the mechanistic insights necessary for predicting nanomaterial transport in porous media.

There are many reasons for the poor state of predictive modeling of nanomaterial transport in porous media. One major problem stems from the extraordinary variability in retention behavior observed in column transport experiments. A prominent example of this can be seen with nano titanium dioxide ($n\text{TiO}_2$), one of the most widely studied nanomaterials because of its numerous commercial and industrial applications (cosmetics and personal care products, food, hybrid organic-inorganic light-emitting diodes, solar cells). The transport behavior of $n\text{TiO}_2$, even in simple, synthetic soil systems (i.e., monodisperse, fully saturated, single-media soils), varies widely. For instance, Chowdhury et al.²⁵ observed strictly hyperexponential (HE) retention profiles (RPs) for the transport of $n\text{TiO}_2$ under various solution conditions and influent concentrations. Cai et al.²³ observed HE and exponential (EXP) RPs under various pH and ionic strength conditions. Choy et al.²⁶ observed exclusively suppressed inlet retention (SIR) under various flow velocities, while Chen et al.⁷⁷ observed both SIR and increasing retention with depth (IRwD) RPs under various solution conditions and in the presence of a purified humic acid.

Nonexponential retention profiles pose problems for PTMs because these models cannot predict, or even qualitatively describe, transport when nonexponential profile shapes are observed.⁸⁵ The source of this problem stems from model construction. Because most PTMs employ first-order kinetics, they cannot describe non-exponential retention profiles by virtue of their mathematical construction (n.b., a model to describe nonmonotonic retention is available, but has received little support⁸⁰). Further, there is no methodology in place to constrain the parameters of PTMs under these circumstances, and notable examples exist within the peer-reviewed literature where these models were used inappropriately.⁸⁵

The most fundamental problem facing predictive modeling, however, is that the mathematical construction of the majority of routinely applied PTMs considers neither the physicochemical properties of

nanomaterials, nor those of the system as a whole, explicitly.⁸⁵ Qualitative influences of physicochemical conditions on particle transport have been well studied and, under some circumstances, provide reasonable explanations for nanomaterial retention and transport behavior (e.g., agreement with Derjaguin and Landau, Verwey and Overbeek [DLVO] theory or extended DLVO theory).^{22,86–88} However, quantitative PTMs that link physicochemical conditions to mechanistic behavior have not been developed, and understanding this link is an important goal for ongoing research.⁸⁹ Consequently, none of the parameters in the mass-balance equations for routinely applied PTMs truly reflect properties, such as type, size, shape, surface properties or aggregation behavior of nanomaterials.⁸⁵ Without consideration of the physical and chemical properties of both nanoparticles and surrounding environment, PTMs are merely descriptive tools that offer no predictive capability.

Machine learning, in its simplest form, enables computers to develop models that are too complex, or untenable, to set-up by hand. Recently, Bayesian neural networks were applied to predict the biological effects of nanomaterials and to determine nanomaterial toxicity for risk assessment.^{90,91} However, machine learning has not yet been applied in the context of nanomaterial transport modeling in the subsurface, largely because the data required to facilitate such an approach are difficult to obtain, validate, and until recently, insufficient in number to support a data-driven approach. In this work, physicochemical conditions from transport experiments with nanomaterials in saturated porous media are used to develop empirical models for the prediction of the resulting experimental outcome. Specifically, we examine the performance of random forest regression and classification machine learning models to predict (i) the retained fraction (RF) of nanomaterials captured over the column length (i.e., a regression problem), and (ii) the nanomaterial retention profile (RP) shape resulting from a fully saturated transport column experiment (i.e., a classification problem). Further, this work quantitatively identifies and ranks the importance, and influence, of various physicochemical features on the transport of nanomaterials in saturated porous media. A key point is that quantitative conclusions are drawn from statistical evaluation of the available nanomaterial transport experiments *as a whole*, not on the basis of a separate investigation of individual factors, as it has been frequently done for solution conditions^{23,24,35–37,39}, NOM^{31,32,42,44}, and physical factors^{33,37,38,41,42}, such as grain and particle size, flow velocity, influent concentration, and coating. The database developed for this work includes more than 200 transport

experiments extracted from 20 peer-reviewed column transport publication. To provide guidance for the construction of future mechanistic models, and to improve model performance, recursive feature elimination with 5-fold cross validation (RFECV) is employed to identify key features. In the short-term, this work demonstrates that empirical prediction of nanomaterial transport is possible. However, the applicability of the developed method outside of the database is limited, as the generalizability has not yet been sufficiently demonstrated to claim otherwise. In the long-term, this work provides insights from which new mechanistic transport models can be developed.

3.3 METHODS

3.3.1 *Nanomaterial Transport Experiment Database*

The database developed for this work includes 204 separate experiments extracted from 20 peer-reviewed nanomaterial column transport studies in saturated, homogeneous porous media (references provided in Chapter 8: App. II). From each experiment, 19 physicochemical features were recorded (17 training features [physicochemical conditions] and 2 target features [experimental results]). The investigated applicability domain and range of training and target features employed for this study are presented in Table 3.

Table 3: Investigated domain of physicochemical training and target features employed for the machine learning effort.

Training Features	Range Investigated
Dispersivity	$4 \cdot 10^{-4}$ to $9.7 \cdot 10^{-2}$ [m]
Hamaker Constant	$2.37 \cdot 10^{-21}$ to $2.1 \cdot 10^{-20}$ [J]
Nanomaterial Type	nAg, nTiO ₂ , nCuO, nBiochar, nHAP, nZnO, C ₆₀ , CeO ₂ , Fe(OH) ₃ , MWCNTs, SiO ₂
Porosity	0.37 to 0.48 [-]
Darcy Velocity	$5 \cdot 10^{-6}$ to $2.33 \cdot 10^{-4}$ [m s ⁻¹]
Influent Concentration	$1 \cdot 10^{-3}$ to $9.7 \cdot 10^{-1}$ [kg m ⁻³]
Influent Pore Volumes	2.47 to 180 [-]
pH	4 to 10 [-]
Ionic Strength	0 to 100 [mM]
Salt Type	CaCl ₂ , KCl, KNO ₃ , NaCl, NaHCO ₃ , None
Particle Density	$1.45 \cdot 10^3$ to $1.05 \cdot 10^4$ [kg m ⁻³]
Particle IEP	1.3 to 8.8 [-]
Particle ζ -potential	-58.7 to 32.7 [mV]
Collector ζ -potential	-79.6 to -21.4 [mV]
Particle Diameter	$4.51 \cdot 10^{-8}$ to $2.19 \cdot 10^{-6}$ [m]
Collector Diameter	$1.94 \cdot 10^{-4}$ to $6.07 \cdot 10^{-4}$ [m]
Coating Type	FeOOH, Fe ₂ O ₃ , None
Concentration NOM/ NOLs/ Surfactant in Solution	0 to $1.004 \cdot 10^{-2}$ [kg m ⁻³]
Type NOM/ NOLs/ Surfactant	Humic, Fulvic, Citric, Oxalic, Alginate, SRHA, TRIZMA (Organic Buffer)
Target Features	Range Investigated
Retained Fraction (RF)	0 to 100%
Retention Profile (RP) Shape	Exp, HE, IRwD, LD, SIR

IEP: isoelectric point; nHAP: nano hydroxyapatite; MWCNTs: multi-walled carbon nanotubes; NOM: natural organic matter; NOL: natural organic ligands; SRHA: Suwannee River humic acid; TRIZMA: Tris(hydroxymethyl)aminomethane

Only transport experiments with a retention profile, excluding those where retention was reported but not visually discernible, were included in the database. Data limitations prevented the dispersivity and Hamaker constant from being employed in the assessment, because values for these parameters were available for only 17% and 31% of the transport experiments considered. Target features, i.e. the experimental results, were not employed for training (i.e., RF was not employed to predict RP shape and vice versa). Of the 204 experiments,

only 183 contained RP shape data; 175 contained RF data. The database structure is generically illustrated in Figure 5.

observations	feature 1	feature 2	...	feature m	regression target feature	classification target feature
experiment 1	value or class	value or class		value or class	RF value	RP shape class
experiment 2	↓	↓		↓	↓	↓
⋮						
experiment n	↓	↓		↓	↓	↓

Figure 5: Nanomaterial transport database structure. m is 17 for both regression and classification, respectively; values of n are 175 and 183 for regression and classification, respectively.

Where experimental values were not available, default values, where possible and with references, were used to fill gaps for collector ζ -potential from literature under similar experimental conditions (19 of 204 experiments). No ratio features, such as the ratio of column length to width or ratio of particle to collector size, or differential features, such as the pH-to-IEP distance or particle-collector ζ -potential difference, were included.

3.3.2 Random Forest Regression and Classification

Random forest regression and classification methods were employed for this work because they are relatively insensitive to outliers and noise, and provide internal estimates of generalization error and feature importance.⁹² Python and the sklearn package were employed to generate the machine learning models⁹³; programming details are included in the Chapter 8: Appendix II.

The random forest method has its name from employing an ensemble of unpruned, random decision trees (i.e., a forest) as simple learners. In this work, each random forest consists of an ensemble of 1000 decision trees. Within a given forest, each decision tree represents a model that predicts the target feature value (regression) or class (classification) by splitting the training feature data set using simple decisional rules learned by statistical analysis.⁹² In a *random* forest, decision trees are constructed from a random subset (bootstrap or out-of-bag samples) of the training feature data set without replacement. In contrast to

standard decision tree logic, where each node is split using the best split among all variables, node splits during tree construction for a random forest are determined by the best split among a random subset of the features chosen for that node.⁹²

It is intuitive to think that the best performing individual tree within the forest is selected for prediction. However, prediction is not based on single trees, but determined by aggregating the predictions of all trees within the forest. This is done in different ways for classification and regressive prediction⁹²: regression prediction is determined by averaging the prediction results of each tree in the forest; class prediction is determined by voting i.e., each tree is 'asked to which class new data belongs and the mode of the results is the class prediction.

In this work, regression prediction performance was reported and assessed in terms of the mean square error (MSE). The MSE values are on a scale from 0 to 1 because the RF has values between 0 and 1, where 1 corresponds to complete retention of nanomaterials by the column. Classification prediction performance was reported and assessed in terms of the F1-score.⁹⁴

Further, it is important to note that random forests do not generate retention or concentration profiles. The outputs of the random forest method are estimates of the RF (regression) and the RP shape class (classification) along with the MSE (regression) and the F1-score (classification) as performance metrics.

3.3.3 *Feature selection with recursive feature elimination with 5-fold cross validation (RFECV)*

RFECV was employed to identify which physicochemical features are important to predicting RF and RP shape, and to avoid overfitting by reducing the number of features employed for training. Feature selection is a central problem in machine learning. At its core, the database from which the model is trained must include a representative set of features. However, it is not clear which features are important for prediction at the problem outset and steps must be taken to reduce over-fitting. Employing RFECV to identify and remove features of low importance to prediction enables the training database to be 'trimmed' to only the features responsible for prediction (n.b., due to randomness in the data selection process, the trimmed database may vary in size

and composition between model runs). This effort is computationally expensive, but critical to gauge the performance of a model where the number of experiments is relatively small in comparison to the number of features (experiments $\simeq 10 \times$ features), as is the case here.

For each model run, the training set is divided into 5 randomly partitioned subsets, or cross-validation folds. Each cross-validation fold consists of 20% of the training dataset, or 18% of the total database. Four folds are aggregated to form a temporary cross-validation database from which a random forest model is trained. The performance of the model is then recursively evaluated by systematically removing features of low importance from the training set, re-training the model, and then examining the predictive performance of the model using the remaining 5th fold (i.e., the RF and RP shape class from the remaining fold are predicted by using the physicochemical data from the experiments contained within the remaining fold). For each recursion step within an iteration, the feature with the lowest importance is eliminated and the random forest is re-trained, and re-tested, until a single feature remains.

Feature importance was determined in relation to the mean decrease in regression or classification accuracy, as determined by the random forest method i.e., the importance of a feature is determined by gauging the increase in prediction error when a specific feature's data is randomly permuted.⁹² Definition, calculation, and limitations of the feature importance method is investigated in greater detail in Breiman⁹², Nicodemus et al.⁹⁵, Louppe et al.⁹⁶, and Strobl et al.⁹⁷. Pedregosa et al.⁹³ provide a more detailed discussion of the RFECV method. Modifications made here to the RFECV method are described in Chapter 8: Appendix II.

In conclusion, identification of important features through RFECV is accomplished by repeatedly applying a feature selection algorithm on the cross-validation folds in a manner that assesses the models predictive performance with different subsets of features (i.e., recursively removing the least important features to examine the model performance with a varying number and composition of features).

3.3.4 Database partitioning and model structure

Partitioning the database into training, cross validation, and testing data sets is critical to evaluate the performance of the machine learning effort (n.b., the testing partition is called the 'holdout' partition to prevent confusion with cross validation testing). A schematic of the partitioning and model structure is shown in Figure 6 and supplemented with a textual description.

Overall, 500 model runs were performed. For each model run the following steps were carried out:

1. All usable experiments within the database were randomly assigned to the holdout (10%) or training (90%) data sets. No training was performed on the holdout set.
 - The training-holdout split was random for the regression problem.
 - The training-holdout split was random for the classification problem, but the ratio of RP shape classes was preserved between datasets. The purpose of this stratified split is to mitigate the influence of class imbalance on the random forest classification⁹⁸, which favors hyperexponential RPs (Table 4).
2. Recursive feature elimination with 5-fold cross validation (RFECV) was performed using random forests generated from the training set (specifically, from the temporary database formed by four folds in combination). Because there are five cross-validation folds, five RFECV iterations were performed in each model run.
 - The cross validation training-testing split was random for the regression problem.
 - The cross validation training-testing split was random for the classification problem, but the ratio of RP shape classes was preserved in each cross validation fold.
 - Regression performance was assessed using mean squared error (MSE); classification performance was assessed using the F1-score⁹⁴; these metrics are called "cross-validation score" below.
 - For each of the five RFECV iterations, the RFECV routine generates 17 random forests (with the feature set size

decreasing from 17 to 1) and reports the feature set size corresponding to the highest cross-validation score (i.e., the ‘optimum number of features’).⁹³ If two or more feature set sizes have the same score, the smallest set size is recorded. The optimum number of features indicates the minimum number of features required to maximize the cross-validation score for a particular cross-validation fold.

- Aggregated-cross validation results show model performance as a function of feature set size for all 500 model runs, as shown in Figures 7 and 10.
3. The training set was ‘trimmed’ to the features identified by RFECV (i.e., reduced to the features corresponding to the optimum number of features as it was identified in step 2). A new random forest model was then trained on the trimmed training set and evaluated against the holdout set.
- For each model run, the values of the holdout set training features are used as model inputs and the accuracy of the prediction is evaluated against the holdout set target features.
 - Aggregated holdout results, as displayed in Figures 8 and 11, show the predictive performance as a function of feature set size for all 500 model runs (500 data points in total). From the aggregated results for all model runs, the optimal feature set size (optimum result) was derived by weighting the number of features, the median prediction error, and the variance, i.e., lower model performance with fewer features and lower variance is preferred over a larger feature set with better performance, but higher variance. The optimum feature set size was determined by means of the following performance-variance-feature set trade-off equation:

$$\text{Optimum Result} = \max \left(\frac{1}{n_{\text{features}}} \cdot \frac{1}{E} \cdot \frac{1}{r_{1Q}} \right) \quad (3)$$

where n_{features} is the number of features employed in a given model run, E is the median MSE for the model runs with feature set size n_{features} , and r_{1Q} is the interquartile range of the holdout performance for the set of model runs with feature set size n_{features} .⁹⁹ E is the median MSE for the regression problem and 1 minus the median F1-score for the classification problem.

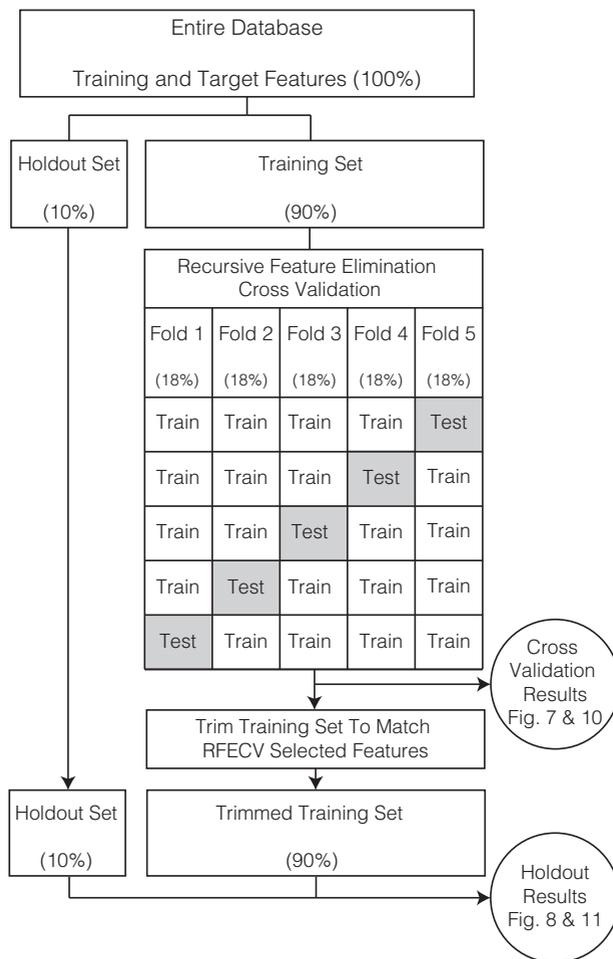


Figure 6: Graphical depiction of data partitioning scheme and model structure for the classification and regression problems. The fraction of training and target features employed for each step are shown parentheses. In the first step, the database is divided into the holdout set (10%) or the training set (90%). No training is performed on the holdout set training features. Instead, the model is trained using the training set and the unimportant features are removed from the database using recursive feature elimination with cross validation (RFECV). The model is retrained on the 'trimmed' training set and evaluated using the holdout set.

Table 4: Distribution of observations by RP shape within the database. RP shapes that can be modeled using PTMs (HE, EXP) are separated from those that cannot (SIR, IRwD, LD).

Retention Profile Class	Count
Hyperexponential (HE)	103
Exponential (EXP)	32
Suppressed Inlet Retention (SIR)	30
Increasing Retention w/ Depth (IRwD)	11
Linearly Decreasing (LD)	7

It is important to note that all 500 model runs were considered equally important and the random forests trained on the trimmed feature sets given the same weight. Because the training sets from which the random forest models are randomly generated, and thus the optimum feature set identified through RFECV varies between model runs, no “final” model exists.

To identify the most important features, we present a feature ranking in Figs. 9 and 12 that indicates how often a feature occurs in the 500 feature subsets that obtained the optimum score in the RFECV. This allows quantitative assessment of each feature in terms of its influence on the prediction. For example, important features will have an RFECV selection frequency near 100%, as these features are consistently selected. Weaker, but still relevant features will have lower, but non-zero selection frequencies, as these features are selected when stronger features are not present in the currently selected subset. Weak or relatively irrelevant features will have scores close to zero, as they are infrequently among the selected features.⁹³

3.4 RESULTS AND DISCUSSION

3.4.1 *Prediction of the retained fraction (RF)*

3.4.1.1 *RF prediction during cross validation.*

In Figure 7, the aggregate performance of the model in predicting RF is shown as a function of the number features employed to train the model during cross validation. Visual inspection indicates that the increase in

median cross-validation performance is relatively small ($\ll 0.01$ MSE) for greater than 5–6 features.

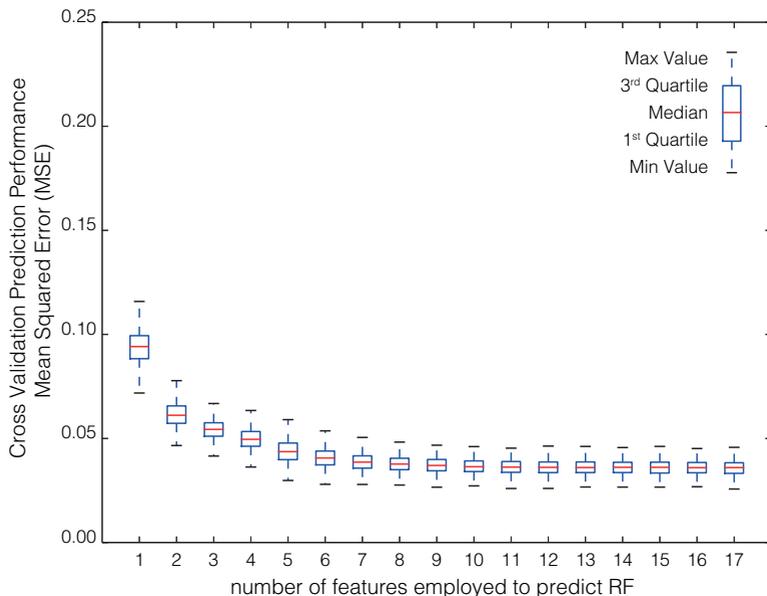


Figure 7: Aggregate model performance to predict the RF of the cross validation test fold as a function of the number of features employed to train. Each box plot represents aggregated model scores for five cross validation iterations for all 500 model runs. This results in 2,500 data points for each feature set size.

3.4.1.2 RF prediction for the holdout set.

Figure 8 shows the aggregate performance of the model in predicting the RF for the holdout set as a function of the number of RFECV-selected features employed to train the model. The best prediction of the RF, or lowest median MSE, was obtained for feature sets consisting of 12 features selected by RFECV (median MSE = $2.61 \cdot 10^{-2}$; $r_{IQ} = 1.70 \cdot 10^{-2}$). However, the optimum result, as determined by Equation 3, was obtained for feature sets consisting of 9 features (median MSE = $3.3 \cdot 10^{-2}$; $r_{IQ} = 1.44 \cdot 10^{-2}$)

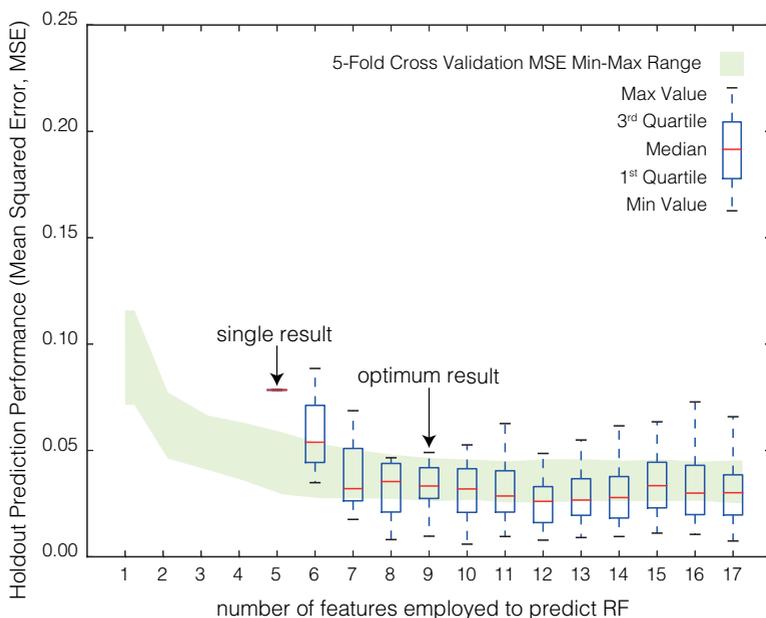


Figure 8: Aggregate model performance in predicting the RF of the holdout set as a function of the number of RFECV-selected features employed to train the model. For each model run, the training database was trimmed to match the feature set selected by RFECV. The model was then re-trained on the trimmed database and the model prediction accuracy (measured by the MSE) was evaluated using the holdout set. For each of the 500 model runs, the prediction MSE and the RFECV-selected feature set size were recorded. Box plots represent the distribution of holdout prediction MSE by feature set size. Note that the trimmed database varies in size and composition due to randomness in the data selection process in RFECV. Feature sets consisting of less than 4 features were never selected by RFECV and only 1 of 500 RFECV feature sets consisted of only 4 features. The 5-fold cross validation MSE (Figure 7) min-max range is shown in green to gauge model performance on the holdout set.

Figure 9 shows the features according to their frequency of occurrence in all 500 model runs. The nine features corresponding to the optimum result are identified in orange.

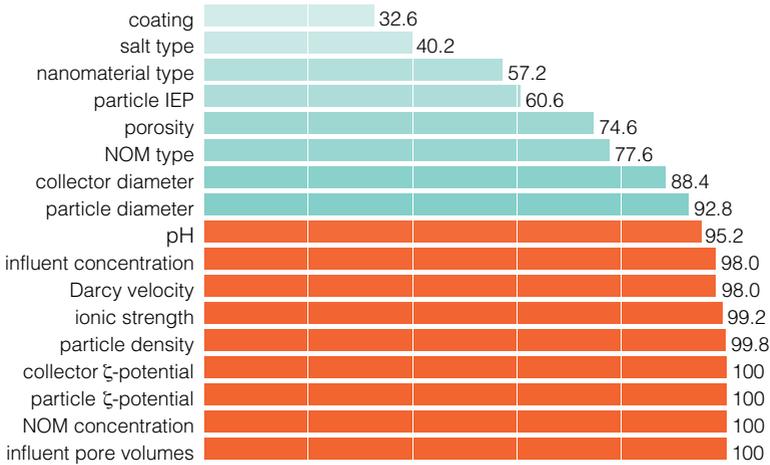


Figure 9: The frequency of features included through RFECV (in percent of number of model runs) for the prediction of the RF. The feature set corresponding to the optimum result, as determined by Equation 3, is presented in orange.

In general, RF prediction for the holdout set was better than during cross-validation. This is expected, as the cross validation data sets are smaller than the total training set (72% of the total training set is employed for each cross-validation iteration [4 of 5 folds of the training set]; 90% is employed for each holdout iteration [i.e., the total training set]). Note that there are no holdout scores for 1–4 features. During cross validation the model performance, when the model was trained with less than four features, was always less than that when trained with four or more features. Therefore, the ‘optimum feature set’ determined by RFECV (Section 3.3.4, Step 2), which is employed as a guide to trim the database to the most suitable set of features, never consists of less than four features.

Several interesting findings are noted. First, our results suggest that the salt type has a relatively low importance in predicting the RF. This seems surprising, particularly as many studies report strong influences

of bivalent cations (specifically Ca^{2+}) that cannot be accounted for by ionic strength alone (e.g., bridging effects).^{24,25,100,101} However, authors often lower the concentration of multivalent cations to be within the same range of influence as monovalent cations (e.g., Chen et al.²⁴ tests 0.56 mM NaCl against 0.02252 mM CaCl_2). This reduces the ability of the models to ascertain the true influence of salt type on nanomaterial transport.

Second, our results suggest that the pH is of greater importance in predicting the RF than the IEP, although both are generally considered relevant for assessing nanomaterial stability in suspension.^{101,102} This is in line with previous work that qualitatively establishes the importance of pH.^{23,24,35-37,39} Also, we find that the number of influent pore volumes and the influent concentration are critical to predicting the RF. This is well understood mechanistically, and provides further validation that relevant features are identified.

Third, and most intriguingly, we find that the nanomaterial type is relatively unimportant to predict the RF. Several publications indicate that nanomaterial fate and behaviour cannot be generalized and that each nanomaterial needs to be tested individually.^{103,104} However, the machine learning results presented here provide evidence that this may not be the case. The interpretation of this finding is that, within the set of physicochemical features employed here, the behavior of the nanomaterials is nearly entirely captured without needing to consider any nanomaterial-specific (i.e., associated with the nanomaterial type) interaction.

3.4.2 Prediction of the retention profile (RP) shape

3.4.2.1 RP shape prediction during cross validation.

The aggregate performance of the model in predicting RP shape as a function of the number features employed to train the model during cross validation is shown in Figure 10. Visual inspection indicates that model performance does not strongly improve with increasing number of features.

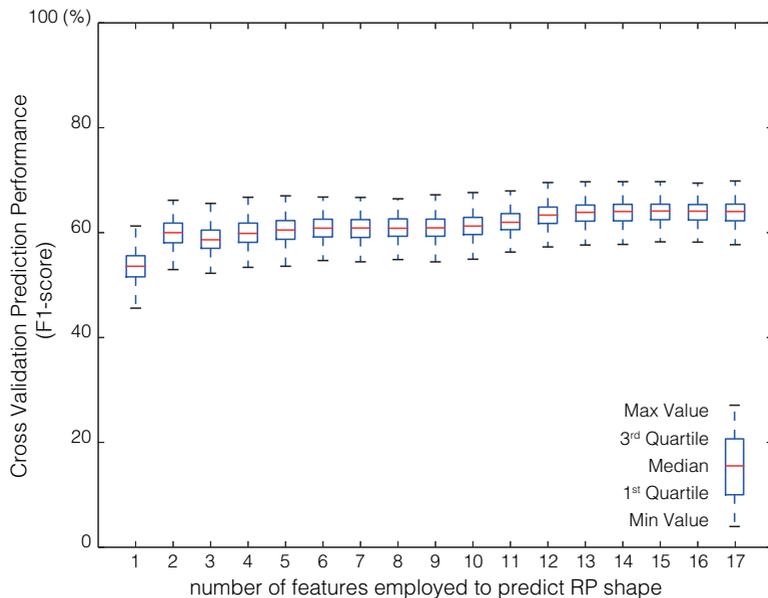


Figure 10: Aggregate model performance (F1-score) in predicting the RP shape class of the cross-validation test fold as a function of the number of features employed to train. Box plots represent aggregated model scores for five cross-validation iterations for all 500 model runs. This results in 2,500 data points for each feature set size.

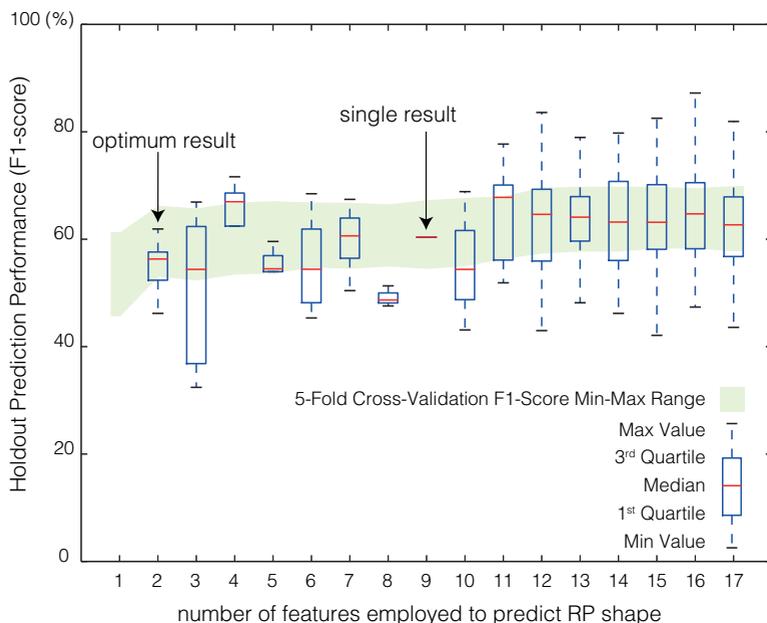
3.4.2.2 *RP shape prediction for the holdout set.*

Figure 11: Aggregate model performance to predict the RP shape for the holdout set as a function of the number of RFECV-selected features employed to train the model. For each of 500 model runs, the prediction F1-score and the RFECV-selected feature set size were recorded. Box plots represent the distribution of holdout prediction F1-Scores by feature set size. The 5-fold cross validation F1-Score (Figure 10) min-max range is shown in green to gauge the performance of holdout set.

The aggregate performance of the model to predict the RP shape of the experiments in the holdout set as a function of the number of RFECV-selected features employed to train the model is shown in Figure 11. In general, the prediction of RP shape class is poor. The highest expected F1-score was obtained when 11 features were employed for training (median F1-score = 68%; $r_{IQ} = 0.14\%$). The optimum result, as determined by Equation 3, was obtained for feature sets consisting of only two features (median F1-Score = 56.3% $r_{IQ} = 0.05\%$). The highest expected performance of the classifier (F1-score of

68%) only provides an increase of 13% above guessing all RPs are HE (56% of the profiles are HE), and the optimum result (classifiers using two features) provides virtually no predictive improvement. This result was not entirely unexpected, as significant class imbalance is present in the data; that the methodology was tailored to reduce these biases (stratified training-holdout split procedure and RFECV [Section 3.3.4]) did not really compensate for the limitations of the data set. However, it is also possible that relevant physicochemical features were not included, although we have exhausted the majority of measurable parameters.¹⁰⁵

The stability of the feature set selected by RFECV was evaluated by examining feature inclusion frequencies over all 500 model runs (Figure 12). The poor class prediction, and low improvement in prediction with increasing number of features, make it difficult to extract meaningful results. Some interesting aspects can still be noted. First, similar to predicting the RF, the nanomaterial type and salt type were of low importance in predicting the RP shape. Second, the NOM concentration and ionic strength were more important than the NOM type and salt type in predicting the RP shape. The importance of NOM concentration on the RP shape is supported by experimental observations made by Liang et al.³⁵ and Chen et al.⁷⁷, where manipulation of the concentration of surfactant in solution was observed to cause RP shape changes. However, there are too few experiments and too many NOM types to examine to definitively rule out nanomaterial type, salt type, or NOM type as important features. Finally, although the optimum result consists of only two features, 11 features were included in more than 90% of model runs (Figure 12).

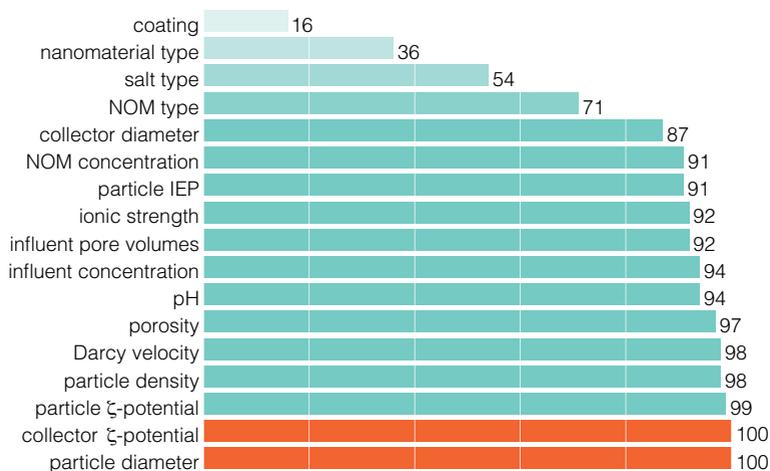


Figure 12: The frequency of features included through RFECV (in percent of number of model runs) for the prediction of RP shape class. The feature set corresponding to the optimum result, as determined by Equation 3, is presented in orange.

3.5 CONCLUSIONS

In contrast to mechanistic models, which are unable to describe transport (let alone predict it) when non-exponential retention profiles are observed, the applied machine learning regression approach enables prediction of the RF with an $MSE < 2.6 \cdot 10^{-2}$. This approach enables quantitative prediction of nanomaterial transport distance independently of a mechanistic understanding of nanomaterial behavior. We anticipate that this method could be used to support high-throughput risk screening to identify conditions with high or low vertical mobility of nanomaterials in porous media without costly and time-intensive experimental work. Further, we foresee this approach facilitating the development of materials specifically designed to accumulate, or transport, in response to varying physico-chemical conditions.

Mechanistic PTMs are important, but currently not suitable for quantitatively describing and predicting nanomaterial transport in porous media. As such, a hybrid mechanistic and machine-learning

approach may offer a way to reconcile problems with mechanistic models. As a first step, ranges of physico-chemical conditions may be defined where the current mechanistic models are sufficient. This might be accomplished by a reformulation of the presented multi-class classification problem as a binary problem [exponential vs non-exponential RPs]). For conditions that result in non-exponential RPs, transport equations need to be modified, or alternatively parameterized, using the established method to predict the RF.

Empirical approaches such as the one employed here have limitations, too. For instance, if the data available for the assessment are not representative, the utility of the approach will be low. In particular, we caution extension of the applicability of this work to real soils, as no column transport studies in real soils were employed for this work. Furthermore, for the classification problem serious data limitations currently prevent adequate understanding of how physicochemical conditions influence the shape of the retention profile. The poor performance of the classification predictor is most likely a result of the limited and highly biased data that were available for this work.

In this work we demonstrate that machine learning methods not only *add* value to the development of mechanistic models through identification of the important features affecting the fate of materials in the environment, but they have the potential to *create* a new, flexible, and prediction-oriented class of nanomaterial transport models. However, for an improved understanding of nanomaterial transport in porous media, several changes in experimental design and data presentation are required. First, and in relation to the physicochemical features identified as important in this work, the existing body of literature must be reviewed to determine what kind of data are already available, what the scope and meaning of these data is, and where exactly the most important data gaps exist. Second, future experimentation must proceed in a manner that fully exploits the information on the physicochemical conditions of column transport experiments, while minimizing cost and time resources.¹⁰⁶ On this basis, machine-learning methods can generate more transparent relationships between nanoparticle transport and experimental conditions and, thereby, provide a basis for the development of improved mechanistic models of nanoparticle transport in porous media.

4

WHAT FACTORS DETERMINE THE RETENTION BEHAVIOR OF ENMS?

4.1 ABSTRACT

A fundamental problem associated with the vertical transport of engineered nanomaterials (ENMs) in saturated porous media is the occurrence of nonexponential, e.g., nonmonotonic or linearly increasing, retention profiles. To investigate this problem, we compiled an extensive database of ENMs transport experiments in saturated porous media. Using this database we trained a decision tree that shows the order of importance, and range of influence, of the physicochemical factors that control the retention profile shape. Our results help identify domains where current particle-transport models can be used, but also highlight, for the first time, large domains where nonexponential retention profiles dominate and new approaches are needed to understand ENM transport. Importantly, highly advective flow and high ENM influent mass can mask the influence of other physicochemical factors on the retention profile shape; notably, this occurs in 50% of the experiments investigated. Where the relationship between physicochemical factors and retention profile shape can be investigated in detail, our results agree with, and provide validation for, the current understanding of how these factors influence ENM transport.

4.2 INTRODUCTION

Many of the metals that comprise the most commonly produced ENMs (e.g., TiO_2 , FeO_x , ZnO , and CuO) occur naturally in soil.⁴⁷ These natural analogues are present in concentrations within, or even above,

the range of predicted environmental concentrations for ENMs in soil^{16,19}, which makes it difficult to distinguish ENMs from the background of natural materials. As a result, investigation of the transport behavior of ENMs in soil is generally constrained to column transport experiments performed in synthetic soil systems (i.e., monodisperse, fully saturated, single-media soils).

Soil column experiments yield a breakthrough curve and a retention profile. A breakthrough curve provides temporal information concerning the concentration of ENMs leaving the column in relation to the concentration of ENMs entering the column. A retention profile provides information concerning the spatial distribution of ENMs in relation to the depth of soil at the end of the experiment. To simulate transport and relate the breakthrough profile to the retention profile, particle transport models (PTMs) are employed, which are mathematical descriptions of how ENMs are transported through the soil (e.g., through advection and dispersion), and how they interact with themselves and the soil during transport (e.g., particle aggregation, deposition, remobilization, removal by straining).

However, most PTMs commonly employed to track ENMs in soil are conceptually flawed. The problem is not that they lack descriptive ability entirely, as they can well describe a large portion of observed ENM transport behavior.^{20,48-50} It is rather that the mathematical construction of these models considers neither the physicochemical properties of the ENMs, nor those of the system as a whole, explicitly.^{85,107}

The ability of PTMs to describe ENM transport is generally limited to conditions that result in exponential or hyperexponential retention profiles (collectively referred to as 'exponential').^{34,80} However, in many experiments nonmonotonic, linearly increasing, or decreasing (collectively referred to as 'nonexponential') retention is observed.⁸⁵ In these cases, PTMs cannot qualitatively describe, let alone predict^{34,80}, ENM transport (n.b. that the breakthrough profile alone is not diagnostic in determining if current PTMs can be applied).⁸⁵

Qualitatively, ENM transport behavior is known to be influenced by a number of physicochemical factors^{20-44,44,45} and numerous explanations for nonexponential retention behavior have been suggested.^{24,34,35,80} However, little mechanistic understanding of how the physicochemical properties of the ENMs or the surrounding environment influence transport behavior has been gained. In this

situation, it is important to develop new approaches to describing ENM transport behavior.

Understanding what conditions result in exponential or nonexponential retention behavior is a logical place to start, because it is critical to identify environmental conditions where current PTMs may be sufficient to describe ENM transport, and conditions where new approaches and focused research are needed. To this end, we developed an extensive ENM transport experiment database and used this database to systematically evaluate the influence of each property on ENM retention behavior, and to discuss data gaps and weaknesses in the current experimental approach. To quantitatively define the conditions that lead to exponential or nonexponential retention, we employed a machine learning decision tree classification (DTC) to visualize the complex interactions between ENM properties and experimental conditions, on the one hand, and ENM retention profiles, on the other hand.

4.3 METHODS

4.3.1 *Database description*

The database developed for this work consists of 214 experiments from 24 publications from 2008–2015^{23–44,44,45} and is available for download. From each experiment, 20 descriptors (Péclet number, ENM influent concentration, aspect ratio of ENM and collector grain size, etc.) and a single response feature, the retention profile shape (exponential or nonexponential), were extracted. The investigated domain is provided in Table 5. To maintain consistency and quality of the data included, experiments with incomplete information were disregarded. Inclusion, exclusion, and quality criteria are discussed in detail in the supplementary information (See Chapter 9: Appendix III).

Table 5: Domain of properties of the nanomaterial, the soil column, and the fluid for the investigated transport experiments.

Property	Range Investigated
<i>Nanomaterial properties</i>	
Type	Ag, TiO ₂ , CuO, Biochar, HAP, ZnO, C ₆₀ , CeO ₂ , Fe(OH) ₃ , MWCNTs, SiO ₂ , quantum dots
ENM influent concentration, C_{inf}	$1 \cdot 10^{-3}$ to $0.3 \cdot 10^{-1}$ [kg m ⁻³]
ENM density, ρ_p	$1.45 \cdot 10^3$ to $1.05 \cdot 10^4$ [kg m ⁻³]
Diameter, d_p	$4.51 \cdot 10^{-8}$ to $2.19 \cdot 10^{-6}$ [m]
ENM ζ -potential, ζ_p	-58.7 to 32.7 [mV]
Combined Hamaker constant, H	$1.54 \cdot 10^{-21}$ to $6.19 \cdot 10^{-20}$ [J]
<i>Column properties</i>	
Collector diameter, d_c	$1.94 \cdot 10^{-4}$ to $6.07 \cdot 10^{-4}$ [m]
Collector ζ -potential, ζ_c	-79.6 to 33.4 [mV]
Collector coating type	iron oxide hydroxide, iron oxide, none
Porosity, θ	0.35 to 0.46 [-]
Column length, L_c	0.10 to 0.30 [m]
Column width, W_c	0.01 to 0.04 [m]
<i>Fluid properties</i>	
NOM, NOL, surfactant type	(oxalic, citric, formic, fulvic, humic) acid, SRHA, TRIZMA, alginate, none
NOM, NOL, surfactant concentration	0 to $1.004 \cdot 10^{-2}$ [kg m ⁻³]
Darcy velocity, v_D	$5 \cdot 10^{-6}$ to $2.33 \cdot 10^{-4}$ [m s ⁻¹]
Influent pore volumes, N_{PV}	2.47 to 180 [-]
pH	4 to 10 [-]
Electrolyte concentration	0 to 100 [mM]
Electrolyte type	CaCl ₂ , KCl, KNO ₃ , NaCl, NaHCO ₃ , None

HAP: hydroxyapatite; MWCNTs: multi-walled carbon nanotubes; NOM: natural organic matter; NOL: natural organic ligands; SRHA: Suwannee River humic acid; TRIZMA: Tris(hydroxymethyl)aminomethane; Combined Hamaker constant is for ENM-fluid-collector.

The use of dimensionless interaction parameters to describe the experimental conditions is preferred to that of dimensioned parameters because it allows generalization regardless of scale.^{61,108,109} However, it is not common to report experimental conditions for ENM transport experiments in this manner. Where possible, and appropriate, dimensioned parameters were combined to form common dimensionless parameters thought to govern particle filtration dynamics.^{61,108,110} This reduced the dimensionality of the database from 20 descriptors to 10 dimensionless descriptors, three dimensioned descriptors, and three categorical descriptors (16 factors in total) without information loss. A summary of descriptors employed for this assessment and their mathematical formulation, where appropriate, is provided in Table 6.

Table 6: Descriptors employed to train the decision tree classification.

N_{Pe}	$\frac{v_D d_c}{D_\infty}$	Péclet number, describes the ratio of advective to diffusive transport [-]. ⁶¹
M_{inf}	$\pi \frac{W_c^2}{4} L_c \theta N_{PV} C_{inf}$	Total influent mass of ENMs as a function of their concentration, the media porosity, and the number of pore volumes pumped into the column [kg].
N_{asp}	$\frac{d_p}{d_c}$	Aspect ratio, describes the ratio between ENM diameter, d_p , and collector diameter, d_c [-]. ⁶¹
N_{att}	$\frac{H}{3\pi\mu d_p^2 v_D}$	Attraction number, describes the relationship between the London dispersion attractive force at one particle radius and the viscous force assuming sphere-plate geometry [-]. ^{61,111}
$N_{\zeta 2}$	$\frac{2(\frac{\zeta_p}{\zeta_c})}{[1+(\frac{\zeta_p}{\zeta_c})^2]}$	Second electrokinetic parameter, describes, empirically, the ratio of ENM and collector surface potentials; a positive $N_{\zeta 2}$ indicates surfaces are similarly charged (i.e., repulsive electrostatic conditions) [-]. ^{108,112}
N_{Lon}	$\frac{H}{6k_B T}$	London force parameter, describes the ratio of van-der-Waals interaction energy at one particle radius to the thermal energy of the particle assuming sphere-plate geometry [-]. ^{108,111}
$N_{\zeta 1}$	$\frac{\epsilon \frac{d_p}{2} (\zeta_p^2 + \zeta_c^2)}{4k_B T}$	First electrokinetic parameter, describes, empirically, the magnitudes of ENM and collector surface potentials [-]. ^{108,112}
N_{dl}	$\frac{d_p}{L_D}$	Electrical double-layer force parameter, describes the ratio of the Debye screening length to the diameter of the ENM [-]. ¹⁰⁸
N_g	$\frac{2}{9} \frac{(d_p/2)^2 (\rho_p - \rho_f) g}{k_B T}$	Gravity number, describes the ratio of the particle's gravitational potential energy at one particle radius above the collector to the thermal energy of the particle [-]. ⁶¹
N_{por}	$\frac{2(1-\gamma^5)}{(2-3\gamma+3\gamma^5-2\gamma^6)}$	Porosity-dependent parameter [-]. ⁶¹
$N_{asp,c}$	$\frac{L_c}{W_c}$	Column aspect ratio, describes the ratio between the length of the column, L_c , and the inner diameter of the column, W_c [-].
C_{inf}	...	Influent concentration of ENM particles [kg m ⁻³].
C_{NOM}	...	Concentration of NOM/NOL/surfactant in solution [kg m ⁻³].
ENM type		Categorical identity of the ENM.
Collector coating		Categorical identity of the collector coating.
NOM/NOL/surfactant		Categorical identity of the NOM/NOL/surfactant.

For parameter definitions, see Table 5. $D_\infty = \frac{k_B T}{3\pi\mu d_p}$: diffusion coefficient in an infinite medium; ϵ : dielectric constant of the fluid; $\gamma = (1 - \theta)^{1/3}$ where θ is the porosity; L_D : Debye length; ρ_f : fluid density of water at 293 K; μ : kinematic viscosity of water at 293 K.

4.3.2 Machine learning classification

DTC is a machine learning method where the data set is recursively partitioned to create a set of *if-then* split conditions to optimize the classification of the data.¹¹³ The algorithm is commonly employed; details are provided in Breiman et al.¹¹³ and Pedregosa et al.⁹³. DTC is an ideal method to investigate complex behavior within a high-dimensional data set, as it results in a visual flow chart from which one can, by hand, follow the explanatory variables (i.e., physicochemical descriptors) to arrive at the response (exponential or nonexponential retention). It is important to note that DTC is not employed here for its predictive ability, but its *descriptive* ability. Our application of DTC is not intended to supplant the need for mechanistic discovery and understanding, but merely to support the delineation of where current mechanistic understanding is sufficient, and where further research is needed.

To determine the optimal decision tree, 20,000 trees were grown using randomly assigned stratified sub-samples (i.e., subsets of experiments with the same class ratio) of the database (85%). The database and code to generate, evaluate, and present the results is available for download. Overfitting was controlled by constraining the maximum depth of the tree to four layers, requiring a minimum of 5% of the experiments at each terminal leaf node, and restricting the maximum number of features considered for each data split (splits were chosen to maximize the decrease in Gini impurity) of four randomly assigned features.^{93,114} Gini impurity ranges from 0 (i.e., all experiments within branch fall within a single class, the best classification) to 0.5 (i.e., an equal number of experiments within each class; see Equation S2 in SI).

Classification performance was determined by examining the ability of each tree to predict the retention behavior of data not employed for training (i.e., experiments from the remaining 15% of the database were fed into each tree to predict the retention behavior of the remaining 15%). True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) were recorded. The highest performing decision tree was selected to maximize the sensitivity ($\frac{TP}{TP+FN}$), specificity ($\frac{TN}{TN+FP}$), and precision ($\frac{TP}{TP+FP}$). As a measure of semantic stability, the structure of the highest performing 0.1% of decision trees had identical structure.¹¹⁴ After selection of the optimal tree, pruning was performed by hand to remove obvious statistical anomalies.

Concerning the selection of the optimal decision tree, it is important to consider that there exists an indeterminable degree of noise within the dataset. Potential sources for noise include employing average conditions to describe a group of experiments, undetected errors in measurement and/or errors in reporting, or even that not all relevant factors or conditions are known or included.

4.4 RESULTS

4.4.1 *Effects of individual factors on retention behavior*

Figures 13 and 14 present histograms that show the relationship between the examined range of each factor and the frequency of exponential (blue) or nonexponential (orange) retention. Our first observation is that no strong relationship can be detected between exponential or nonexponential retention and any single factor.

Exponential and nonexponential retention are observed over a wide range of Péclet numbers (Figure 13: N_{Pe}), total masses of ENMs pumped through the column (Figure 13: M_{inf}), aspect ratios (Figure 13: N_{asp}), and NOM concentrations (Figure 13: C_{NOM}), under both electrostatically attractive and repulsive conditions (Figure 13: $N_{\zeta 1}$, $N_{\zeta 2}$, $N_{\zeta 2}$, N_{Lon} , N_{dl}), and for nearly every collector coating, NOM type, and ENM type (Figure 14: Collector Coating, NOM Type, and ENM Type).

Some subtle relationships are noted. First, nonexponential behavior is most frequently observed when the total mass of ENMs pumped through the soil column is between 1 and 3 mg. Second, most nonexponential behavior is observed below Péclet numbers of $1 \cdot 10^4$. Finally, an increase in nonexponential retention is noted for low $N_{\zeta 1}$ values (< 200). Low $N_{\zeta 1}$ values indicate a net electrostatic attraction and occur when London dispersion attraction dominates the relatively short-range double-layer repulsion force.¹¹⁰ However, it is not clear why low $N_{\zeta 1}$ values are associated with nonexponential retention directly, as $N_{\zeta 1}$ is theoretically independent of the deposition efficiency for values below 230.¹¹⁰

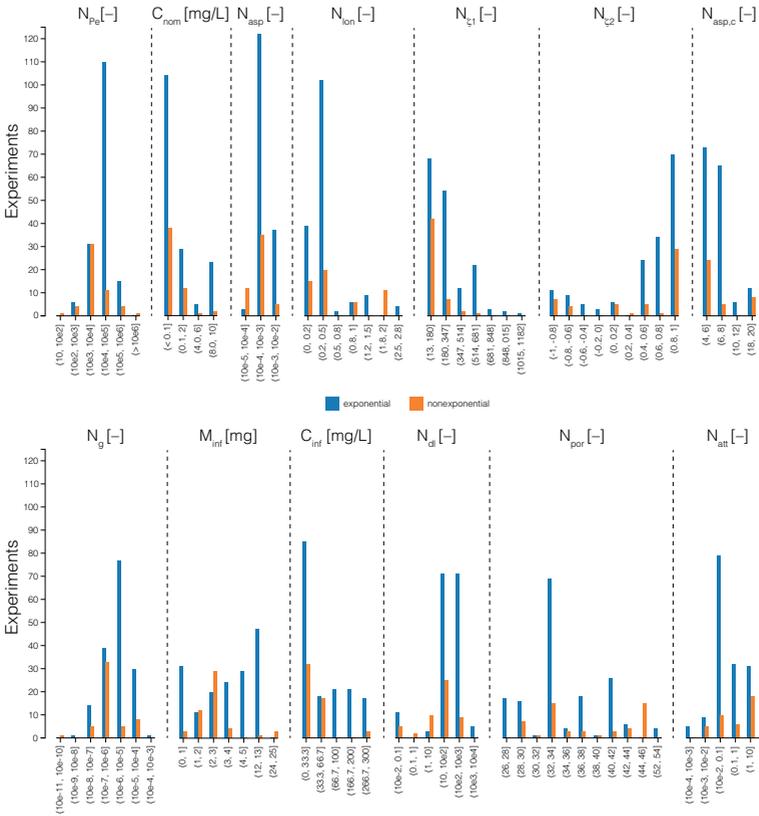


Figure 13: Histograms for individual factors showing the number of experiments with either exponential (blue) or nonexponential (orange) retention behavior.

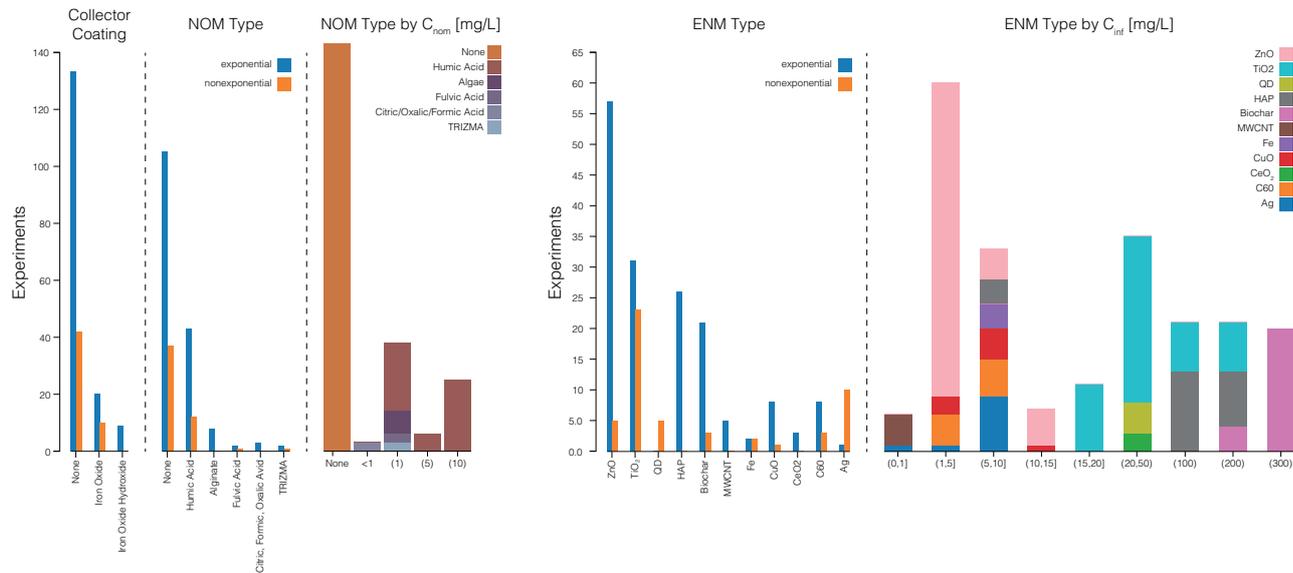


Figure 14: Blue and orange bars: histograms for categorical factors showing the number of experiments with either exponential (blue) or nonexponential (orange) retention behavior; colored stacked bars: histograms showing the number of experiments conducted with each NOM type and ENM type over the ranges of NOM concentrations and influent ENM concentrations, respectively.

The database also shows that many experiments were performed with little overlap of the conditions applied. It is important to note that the sparsity of overlapping experimental conditions between ENM types and the relatively low environmental relevance of the examined conditions do not mean that the findings are not generally applicable. However, the lack of overlap prevents direct comparison and it remains unclear if conclusions drawn from a single study are generally applicable or unique to the conditions examined. For example, because there is no agreed upon concentration range under which to examine ENM transport, influent concentrations used for different ENMs vary substantially (Figure 14: ENM Type by C_{inf}): ZnO (29% of the database) is almost exclusively investigated with influent concentrations between 1 and 15 mg L⁻¹, while TiO₂ (25% of the database) is typically investigated between 20 and 200 mg L⁻¹. With the exception of TiO₂ and hydroxyapatite (HAP), all other ENMs are investigated under a relatively narrow range of influent concentrations.

Finally, we find the environmental relevance of the examined conditions to be low in three areas. First, typical infiltration rates in soil are between 0.043–0.162 cm min⁻¹, which result in an N_{pe} range between 146 and 1381 for an ENM with a diameter of 50 nm traveling through fine sand (200–500 μm in diameter).⁴³ However, only 12 experiments (<6% of the database) were performed within this range. Second, less than 34% of the experiments were conducted in the presence of NOM. NOM plays an important role with respect to contaminant sorption and transport, and is virtually ubiquitous in surface and ground water at concentrations between 0.1 and 10 mg L⁻¹ (Figure 14: NOM Type by C_{NOM}).^{115,116} Of the experiments conducted in the presence of NOM, ca. 70% investigate a single NOM type (humic acid). Third, less than 10% of the experiments examined the influence of collector coating (Figure 14: Collector Coating) and no experiments considered nanomaterial coating or functionalization.

4.4.2 *Disentangling and visualizing the influence of different factors*

A decision tree is an ideal method to investigate complex behavior within our data set, as it results in a visual flow chart (the “decision tree”) from which one can, by hand, follow the explanatory descriptors (i.e., flow-related, chemical, and electrical factors describing the ENM and surrounding environment) to arrive at the response feature (exponential or nonexponential retention).

The classification performance of the decision tree presented in Figure 15 is high (sensitivity: 97%; specificity: 79%; precision: 94%, see Supplementary Information). At each node, the split criteria is shown and the decision logic can be followed to the right, if the condition matches the split criterion (i.e., 'yes'), or to the left, if it does not (i.e., 'no'). The thickness of the branches corresponds to the relative number of experiments classified by the preceding split conditions.

Nearly half of all experiments are contained within two branches, each with only two split decisions (Figure 15): the root decision, which splits based on a fluid flow condition (N_{Pe}), and the second decision, which splits based on the mass of ENMs pumped through the column (M_{inf}). These two terminal branches (yellow and red in Figure 15) suggest that when the mass of ENM pumped through the column is greater than ca. 3 mg the retention behavior is exponential, regardless of other factors. However, with so few decision steps between the root and leaf, it is difficult to tell if the role of influent mass is really the controlling factor (i.e., a general phenomenon that occurs under a wide range of physicochemical conditions), or if the classification is merely because the experiments with influent mass greater than 3 mg are performed under similar conditions (i.e., a correlation artifact).

To ascertain if M_{inf} is the controlling factor for the yellow and red branches, we examine the diversity of conditions contained within each branch. To evaluate the diversity, we plot the coefficient of variation ($CV = \frac{\sigma}{\mu}$) for each factor in a radar chart (Figure 16). We do not include categorical (ENM type, NOM type, collector coating) and flow-related (M_{inf} , N_{Pe}) factors, or factors with an inherently low CV ($CV < 1$; N_{por} , $N_{asp,c}$, $N_{\zeta 1}$, and $N_{\zeta 2}$). On the one hand, if a wide range of conditions are found within a terminal branch, it is likely that the decision steps leading to the classification are generally valid. This is because the only similarity between experiments is a function of the preceding decision steps rather than similar experimental conditions. On the other hand, a narrow range does not necessarily preclude the identification of a valid classification, but indicates that it cannot be deductively established from the available data.

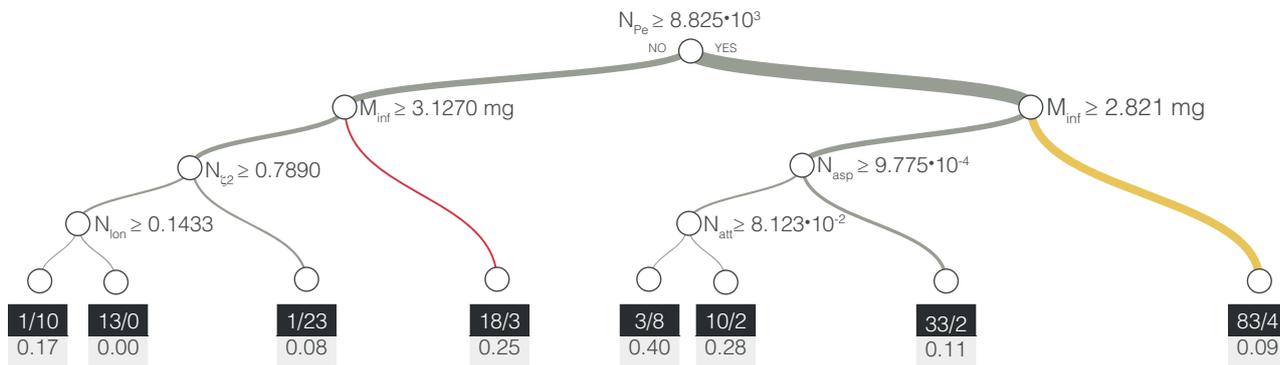


Figure 15: Optimal decision tree, which classifies the ENM retention behavior as exponential or nonexponential by physicochemical factors with high performance (sensitivity: 97%, specificity: 79%, precision: 94%). The thickness of each branch corresponds to the relative number of experiments included in the branch. The experimental diversity in the yellow and red terminal branches are further investigated in Figure 16. At each leaf the number of exponential and nonexponential experiments is given (e.g., 1/10 is 1 exponential and 10 nonexponential), along with the branch Gini impurity.

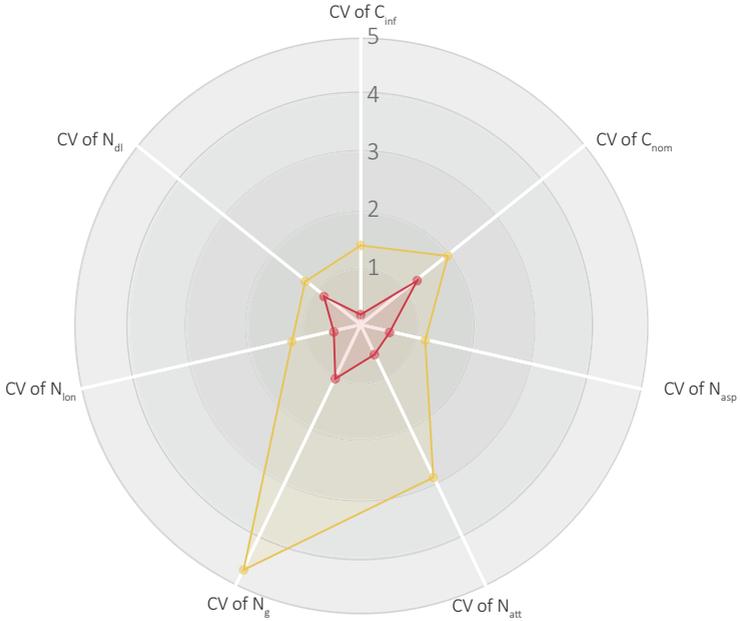


Figure 16: Radar chart of the coefficient of variation ($CV = \frac{\sigma}{\mu}$) for non-categorical (ENM type, NOM type, Collector coating), non-preceding (M_{inf} and N_{Pe}), non-inherently-low-CV (N_{por} , $N_{asp,c}$, $N_{\zeta 1}$, and $N_{\zeta 2}$) factors for the experiments contained within the yellow and red branches. The closer the CV value is to the center of the chart, the less varied the data. CV values below 1 are considered to be low-variance; values > 1 are considered to be high-variance.¹¹⁷ Note that for $N_{\zeta 2}$, which can be negative when the ENM and grain have opposite surface charge potentials, but ranges from -1 to 1 , the CV is calculated by shifting all values by 1 so that they range from 0 to 2.

Within the yellow branch (41% of the database [87 experiments]), we find a diverse range of conditions (Figure 16) and note that seven of 11 ENM types, three of eight NOM types, and one of three collector coating types are represented within the branch. Deductively, this implies that, regardless of any other factor, when $N_{Pe} > 8.825 \cdot 10^3$ and $M_{inf} > 2.825$ mg the retention profile is likely to be exponential. Three of the four misclassified experiments (i.e., nonexponential experiments classified as

exponential) within the yellow branch were performed with TiO₂ using a column with a very large column aspect ratio ($N_{asp,c} = 20$) and with high total influent mass of ENMs ($M_{inf} = 24.2$ mg).

In contrast to the yellow branch, we find a narrow range of conditions within the red branch (Figure 16). Of the 21 experiments in the red branch, 20 were conducted with the same ENM (biochar), at the same rather elevated C_{inf} (300 mg L⁻¹), and under approximately the same flow conditions ($\mu_{N_{Pe}} = 2.23 \cdot 10^3$; $\sigma_{N_{Pe}} = 200$). Thus, while influent mass may play a significant role in controlling ENM retention behavior also under less advection-dominated conditions, this influence cannot yet be deductively justified and more experiments at a broader range of conditions are needed.

On the right half of Figure 15, under highly advection-dominated conditions ($N_{Pe} \geq 8.825 \cdot 10^3$), retention profiles are primarily controlled by the physical conditions of the system. At high influent mass ($M_{inf} \geq 2.821$ mg) exponential retention profiles are frequently observed, as previously discussed, but below this threshold two other physical factors play an important role (N_{asp} and N_{att}).

N_{asp} is the ratio of ENM diameter to collector diameter. Our results indicate that when $N_{asp} > 9.775 \cdot 10^{-4}$, exponential retention is likely. This value is in close agreement with the value associated with straining as an important retention mechanism ($N_{asp} > 0.002$).²⁵ Straining is the physical retention of particles in pore throats or grain-grain junctions that are too small to permit passage.¹¹⁸ When straining occurs, retention profile shapes are generally hyperexponential (i.e., within the exponential category).⁷⁹ This finding suggests that straining may be the dominant removal mechanism controlling retention behavior under these conditions.

N_{att} represents the ratio of London dispersion attractive force to the viscous force and appears to control retention profiles at low N_{asp} values. At high values of N_{att} ($N_{att} \geq 8.123 \cdot 10^{-2}$), ENMs are more likely to be pulled from their streamline trajectory and attach to the surface of the collector, resulting in exponential retention profiles.¹¹⁰ Conversely, when values of N_{att} are low ($N_{att} < 8.123 \cdot 10^{-2}$), nonexponential behavior may result because ENMs are less likely to attach due to a high viscous force, which also exerts a higher shear on already attached particles and potentially increases detachment of weakly-bound particles.^{110,119}

Only under less advection-dominated conditions ($N_{Pe} < 8.825 \cdot 10^3$) and at lower influent mass ($M_{inf} < 3.127$ mg) do the electrical interaction

characteristics of the system tend to determine the shape of the retention profile. $N_{\zeta 2}$ and N_{Lon} represent the magnitude of the electrical double-layer repulsion and London dispersion (attractive) force, respectively. Most importantly, if the magnitude of the electrical double-layer repulsion is high ($N_{\zeta 2} \geq 0.789$), particle-to-collector deposition is low and nonexponential retention profiles occur. If the magnitude of the electrical double layer repulsion is low ($N_{\zeta 2} < 0.789$), however, N_{Lon} determines the retention profile. Under highly attractive conditions ($N_{\text{Lon}} \geq 0.1433$) exponential retention profiles likely occur due to an increased potential for particle deposition, while less attractive conditions ($N_{\text{Lon}} < 0.1433$) result in nonexponential retention profiles due to a decreased potential for particle deposition.

Where highly advective flow and high mass do not mask physicochemical factors, such as $N_{\zeta 2}$, N_{Lon} , and N_{asp} , our results provide quantitative validation of existing theoretical understanding of ENM transport. However, our results show that half of the available transport experiments (108 of 214, see red and yellow branches in Figure 15) are most likely unsuitable to investigate the influence of chemical and electrical factors on ENM retention behavior. This does not invalidate findings derived from these experiments, particularly as the majority of these experiments report exponential retention profiles, which can be described with existing PTMs. However, if one desires to better understand the influence of chemical and electrical factors on ENM transport behavior, it is likely that only a small fraction of the existing experiments are suitable to help disentangle their influence on the retention behavior. Our results suggest that only when the mass of ENMs pumped through the column and the Péclet number are sufficiently low can the influence of chemical and electrical factors on the retention profile shape be investigated.

4.5 CONCLUSIONS

The common approach to investigating ENM transport has not advanced mechanistic understanding because it relies on PTMs that have no predictive power and, for the case of nonexponential profiles, little descriptive power. Our approach makes it possible to identify the order of importance and range of influence of the physicochemical factors that control the retention profile shape. This is important because it offers an alternative to the common approach of using empirically adjusted PTMs and opens a path toward a deeper level of

mechanistic understanding of ENM transport in saturated porous media. Through our alternative approach, we have, for the first time, identified domains of conditions where nonexponential retention is frequent, and thus where entirely new mechanistic or empirical approaches should be developed, and domains of conditions where exponential retention is frequent, and thus existing approaches may be sufficient. Well designed experiments in domains of nonexponential retention will be a first step to building a more complete, descriptive and predictive understanding of the factors that affect ENM retention.

5

A MACHINE LEARNING MODEL TO PREDICT ENM ATTACHMENT EFFICIENCY

Contents

5.1	Introduction	105
5.2	Materials and Methods	108
5.2.1	Database	108
5.2.2	Machine learning	110
5.2.3	Model structure	112
5.3	Results and Discussion	112
5.3.1	Recursive feature elimination with cross validation	112
5.3.2	Attachment efficiency prediction performance	114
5.3.3	Optimum Result	114
5.3.4	Implications of model findings	117
5.4	Conclusions	120

5.1 INTRODUCTION

Environmental applications of engineered nanomaterials (ENMs) have tremendous potential to enhance remediation of pollutants in the subsurface, to improve sustainability of nano-formulated plant protection products in comparison to traditional pesticides, and to improve nano-based plant nutrient delivery.^{120–122} However, the increased production of ENMs, coupled with their apparent aquatic toxicity,^{123–125} raise serious concerns about their fate and transport of these materials in the environment.

Understanding of ENM transport in the subsurface is far from complete. Similar to larger, colloidal particles, the transport behavior of

ENMs in the subsurface environment is a function of numerous deposition processes, which are controlled by a complex system of physical and chemical factors.¹²⁶ Despite numerous qualitative investigations that examine the influence of physical chemical factors on ENM transport^{127–138}, relatively few quantitative insights have been gained. One contributing factor for the slow rate of progress is that it is not yet possible to identify and separate natural nanomaterials from engineered nanomaterials in complex environmental matrices.⁴⁷ As a result, investigation of ENM transport, particularly for high production volume nanomaterials (e.g., TiO₂ and ZnO), is limited to examination in simple, artificial soil systems. Another critical factor is that ENM transport is primarily investigated through experimentation with soil columns. Soil columns have been in use for more than 130 years and are the foundation for most fluid dynamics, but are too coarse to sufficiently understand nano-scale processes. On the other hand, new techniques, such as high-resolution X-ray computed tomography of soil columns, enable non-destructive transport characterization, but are still too coarse to resolve ENMs (scanning resolution is typically between 0.84 μ m–4.4 μ m).¹³⁹

Quartz crystal microbalance with dissipation (QCM-D) monitoring enables real-time monitoring of nanoscale mass deposition. QCM-D monitors the deposition rate of particles onto an oscillating sensor surface, isolating deposition kinetics, and allowing for a direct measure of attachment efficiency (α) without contributions from convection or filtration, unlike traditional soil column experiments. α is an important kinetic transport parameter and is widely used in colloid filtration theory (CFT) to quantify the likelihood of a particle attaching to a surface after a collision. It is important to note that several closed-form correlations have been developed to predict α ^{68,140–143}, although none explicitly consider the physicochemical conditions which have been reported to affect the value of α , including ionic strength^{129,130,132,134–136,138}, pH^{132,133,144}, charge of particle and collector surface^{127,129–136,138,144–146}, presence of organic matter, whether dissolved in solution^{131,133,135,136,144} or coating the collector surface^{127,130,131,133,135,136,145,146}, temperature¹²⁷, and particle shape^{128,147}. Furthermore, while the qualitative influence of these conditions on α has been well studied and, in some cases, reasonable mechanistic explanations are provided (e.g. agreement or disagreement with Derjaguin-Landau-Verwey-Overbeek theory^{148,149}), a predictive model of α that explicitly considers these conditions has not yet been developed.

Machine learning allows us to develop empirical models from complex systems where the underlying relationships between the data are too complex to develop by hand.¹⁵⁰ Machine learning has been successfully applied to a wide range of complex problems.^{151–154} In two recent studies^{90,91} machine learning was applied to predict the toxicity and biological impacts of ENMs, based explicitly on the molecular properties of the nanomaterial. Despite the successes of machine learning in a wide range of applications, it has not been applied to the complex task of modeling environmental transport until very recently. A very recent study by Goldberg et al.¹⁰⁷ employed ensemble machine learning (random forest) regression and classification to predict the retained fraction (RF; the fraction of materials retained during a soil column experiment in comparison to the total mass of materials injected into the column) and shape of retention profile (RP) using a database of more than 200 nanomaterial column transport experiments amassed from published literature. Goldberg et al.¹⁰⁷ reported that their model was able to predict the RF with a mean squared error between 0.025–0.033, and the RP with an expected F1-score (the weighted harmonic mean of precision and recall) between 60–70%. Further, by recursively removing physical and chemical features to optimize model predictive performance, the authors were able to rank the importance of the physicochemical state features (e.g., pH, ionic strength, nanomaterial type, etc) to ENM transport.

The high variability in reported α values under seemingly similar experimental conditions is one reason why the mechanistic understanding remains poor. For example, the presence of an attached layer of natural organic matter (NOM) has been reported to hinder, enhance, or have no effect on α under similar solution conditions.^{131,136} The purpose of this work is to combine all of the available α data in order to identify which of the complex set of variables are most important and, ultimately, produce a predictive model for α based on the identified variables. Here, 299 total experiments with 13 physicochemical features each were chosen from 12 publications from 2008–2015^{127–138} to form the largest QCM-D derived α database to date. Ensemble machine learning (gradient boosting decision trees) was employed to empirically relate the physicochemical state (i.e., physicochemical training features) to the α values (i.e., target feature) measured by QCM-D. Grid search hyper parameter optimization with cross validation (GSHPOCV) and recursive feature elimination with cross-validation (RFECV) were employed to optimize model performance. To identify the physicochemical features most important to prediction, RFECV results from 100 model runs were aggregated and

investigated to identify the physical and chemical features critical to predicting α . The predictive, empirical model presented here will aid in identifying which physicochemical characteristics are most influential to α . An improved understanding of transport parameters is key to accurately predicting the impact of new particle types and making risk-informed regulatory decisions, or even designing new ENMs to be safe from the ground up.

5.2 MATERIALS AND METHODS

5.2.1 Database

The database developed for this work includes 299 separate experiments extracted from 12 peer-reviewed QCM-D nanomaterial transport studies from 2008–2015.^{127–138} From each experiment, 13 physicochemical training features were recorded and 1 target experimental result feature, i.e., α , was chosen. Studies which reported particle deposition rate but not particle attachment efficiency were not included in the database. As mentioned previously, in order to calculate α from particle deposition rates, the favorable deposition rate at identical experimental conditions must be known, and so without this knowledge, α could not be extrapolated from particle deposition rates. Furthermore, studies which explored particle deposition onto oppositely charged surfaces were not considered, as $\alpha = 1$ in these cases, by definition. The investigated applicability domain and range of training and target features employed for this study are presented in Table 7.

Table 7: Investigated domain of physicochemical training and target features.

Training Features	Range Investigated
Particle Type	C60, C60 _{20h-UV} , C60 _{7D-UV} , ZnO, PS, MS ₂ RNA, BL ₂₁ RNA, MWCNTs, MWCNTs _{LO} , MWCNTs _{HO} , TiO ₂
Debye Length	0.5 to 56.2 [nm]
Salt Type	NaCl, CaCl ₂
Particle ζ -potential	-48.5 to 3.31 [mV]
Collector ζ -potential	-79.6 to -21.4 [mV]
pH	5.2 to 8.0
Dissolved NOM Type	None, SRHA, EHA, Alg
Dissolved NOM Concentration	0 to 10 [mg/L]
Immobilized NOM layer	None, HHA, HFA, SRHA, EHA, Alg
Temperature	15 to 45 [°C]
Shape	Sphere, Ellipsoid, Tube
Suspension Preparation Method	Liquid-liquid exchange, Direct sonication, Commercial stock
Particle Concentration	5 to 100 [mg/L]
Target Features	Range Investigated
α	$4.1 \cdot 10^{-4}$ to 1.71 [-]

Alg: alginate; 7D-UV: 7 day UV exposure; 20h-UV: 20 hour UV exposure; EHA: Elliott humic acid; HHA: Harpeth humic acid; HFA: Harpeth fulvic acid; LO/HO: Lowly/highly oxidized; MWCNTs: multi-walled carbon nanotubes; NOM: natural organic matter; SRHA: Suwannee River humic acid; PS: polystyrene, MS₂/BL₂₁ RNA: RNA extracted from bacteriophage MS₂ and *E. coli* BL₂₁

The physicochemical features (listed in Table 7) were chosen based on three criteria:

1. features that have been reported to affect α , i.e. particle type, presence of NOM, salt type, Debye length (through ionic strength), temperature, shape;
2. features that affect interaction energy, i.e. Debye length, ζ -potential of particle and collector, pH; and
3. features that have been reported to affect particle properties, i.e. suspension preparation method, particle shape.

Particle size was excluded as a physicochemical feature because the diffusion rate influences both favorable and unfavorable deposition conditions equally and therefore has no net effect on α . Furthermore,

there were large inconsistencies in particle size reporting, due to simultaneous particle aggregation during QCM-D experiments for certain particle types exposed to high ionic strengths. Dissolved NOM, which consists of NOM dissolved in solution within the particle suspension and likely adsorbed to the particle surfaces, is differentiated from the immobilized NOM layer, which is attached to the sensor surface by electrostatic layer-by-layer (LbL) assembly. In some cases, the values for zeta potential of particle or collector surface were not reported, so values were interpolated from reports of identical particle or surface type under similar solution conditions. When electrophoretic mobility was reported, zeta potential was estimated using the Smoluchowski approximation.¹⁵⁵

A graphical depiction of the complex relationship between the training features and the target feature (i.e., α) is shown in Figure 17.

5.2.2 *Machine learning*

To attain the best prediction performance, several machine learning regression methods (random forest, ADA boosting, decision tree, and gradient boosting) were evaluated simultaneously with the provided model structure (Figure 18), which includes automated hyper parameter optimization (GSHPOCV) and feature selection sub routines (RFECV).^{93,107} Python and scikit-learn package 0.17.1⁹³ were employed for model construction and the entire model is available for distribution online. In this work, we present the results of a tuned gradient boosting regression, as it provided higher prediction performance with lower variance in comparison to the other methods.

Gradient boosting regression is an ensemble machine learning method that combines a series of decision trees in a forward stage-wise fashion.¹⁵⁶ In the first stage, a single-layer decision tree is trained to fit the regression data. In each sequential stage, a new single-layer decision tree is fit to the previous models' residuals and appended to previous layer(s). By sequentially fitting the residuals, the model iteratively increases its ability to predict observations modeled poorly by the previous steps. Prediction is made by summing the results of the predictions in each stage. Overfitting occurs when trees are added in excess, which increases the likelihood that the model considers noise within the training data to be important. Standard gradient boosting regularization strategies were employed to prevent overfitting by calibrating the number of stages to minimize the test-train deviance,

restricting the tree complexity (number of nodes), and by employing a learning rate parameter <1 to shrink the contribution of each tree as it is added to the model. Note that the learning rate and the number of boosting stages parameters are selected by GSHPOCV and propagated throughout for each model run.

5.2.3 *Model structure*

To provide internal estimates of error and generalizability, the database was partitioned into training, cross validation, and testing data sets (the testing partition is called the 'holdout' partition to prevent confusion with cross validation testing). A schematic of the model structure is shown in Figure 18 and supplemented with a textual description.

Overall, 100 model runs were performed. For each model, all usable experiments within the database were randomly assigned to the holdout (15%) or training (85%) data sets. No training was performed on the holdout set. Next, grid search hyper parameter optimization with cross validation (GSHPOCV) was performed to identify the best performing combination of model parameters (i.e., the parameter combination that resulted in the highest R^2 cross-validation score).⁹³

Following GSHPOCV, recursive feature elimination with 5-fold cross validation (RFECV) was employed to identify the minimum number of training features required to maximize model performance.^{93,107} A description of the RFECV routine is provided in the Supporting Information. The training set was then 'trimmed' to the features identified by RFECV and a new model was trained on the trimmed training set and evaluated against the holdout set (i.e., the holdout set training features [physicochemical features] are fed into the model and accuracy of the model to predict α is evaluated in comparison to the ground truth values.)

5.3 RESULTS AND DISCUSSION

5.3.1 *Recursive feature elimination with cross validation*

The performance of the model (measured as R^2) to predict α as a function of the number of features selected during RFECV training can be seen in Figure 19. The performance associated with predicting α

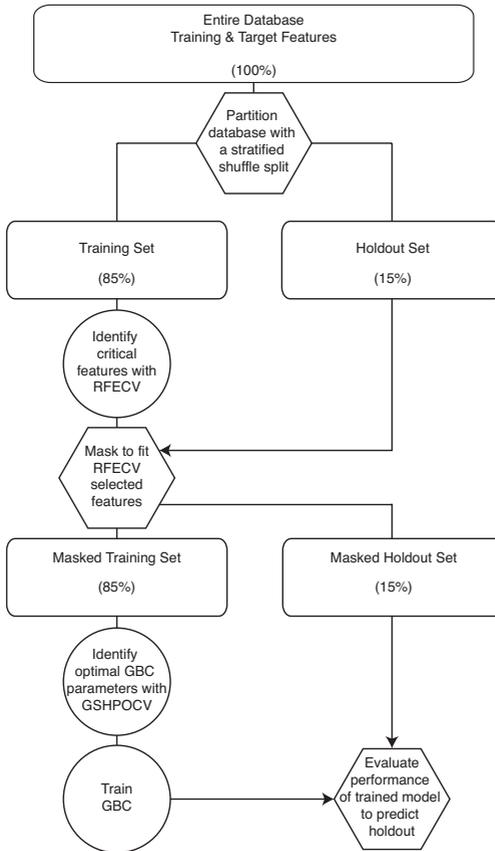


Figure 18: Graphical depiction of data partitioning scheme and model structure for the classification and regression problems. The fraction of training and target features employed for each step are shown parentheses. In the first step, the database is divided into the holdout set (15%) or the training set (85%). No training is performed on the holdout set training features. Grid search hyperparameter optimization with cross validation (GSHPOCV) is employed to determine the optimum machine learning parameter on which to train the model, which are employed throughout the rest of the model run. Next, unimportant features are removed from the database using recursive feature elimination with cross validation (RFECV; Figure 19). The model is retrained on the “trimmed” training set and evaluated using the holdout set to produce the holdout results: Figure 20

during RFECV is a very conservative estimate of the generalized performance of the predictive model and is not representative of overall model performance because it is trained on a reduced data set (68% of the total database is employed for RFECV score; 85% is employed for prediction against the holdout set). It can be seen from Figure 19 that performance increased when the model was trained with larger feature set sizes, as expected, with a marked increase in performance (20% median performance gain) occurring between 7 and 8 features chosen. Progressing from 8 to 13 features continues to decrease the variance in model performance, decreases the observation frequency of outliers, and improves the remaining outlier score (i.e., outliers move closer to the bulk of the data). In terms of cross validation performance, there are diminishing returns to adding more features. Median performance increases by 5% between 8 and 10 features, however a performance increase of only 2% is observed between 10 and 13 features.

5.3.2 Attachment efficiency prediction performance

The model performance in predicting α for the holdout set as a function of the number of RFECV selected features to train the model can be seen in Figure 20. The lack of performance data below 8 features indicates that the model performance was never optimal when fewer than 8 features were employed for model training. The median performance to predict α in the holdout set was higher for each feature set size as compared to the RFECV performance. This is to be expected, because the model was trained on the entire trimmed training database (additional 17% in comparison to the RFECV).

5.3.3 Optimum Result

In order to identify the highest performing model utilizing the fewest number of features, the optimum result was chosen using Equation 4, where $n_{features}$ is the number of features employed for a given feature set size, R^2_{median} is the median R^2 value corresponding to the aggregated median predictive accuracy of holdout scores for a given feature set size, r_{IQ} is the interquartile range of the aggregated holdout scores for a given feature set size, and $F_{observation} \%$ is the percentage of model runs that result in a given optimum feature set size. A summary of the values

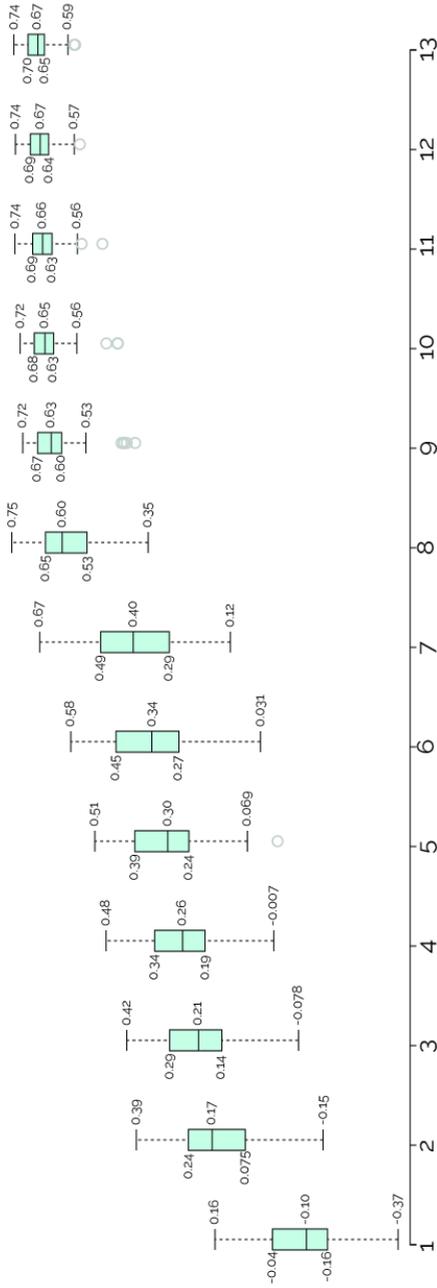


Figure 19: Aggregate model performance in predicting α in the cross validation test fold during the 500 RFECV model runs. The model performance (measured as R^2) as a function of number of features chosen is shown, with the box plots representing the minimum, maximum, median, and interquartile range of R^2 values. Circles represent statistical outliers. Note that when the model is trained with two or fewer features, the R^2 is negative, indicating that the fit is extremely poor.

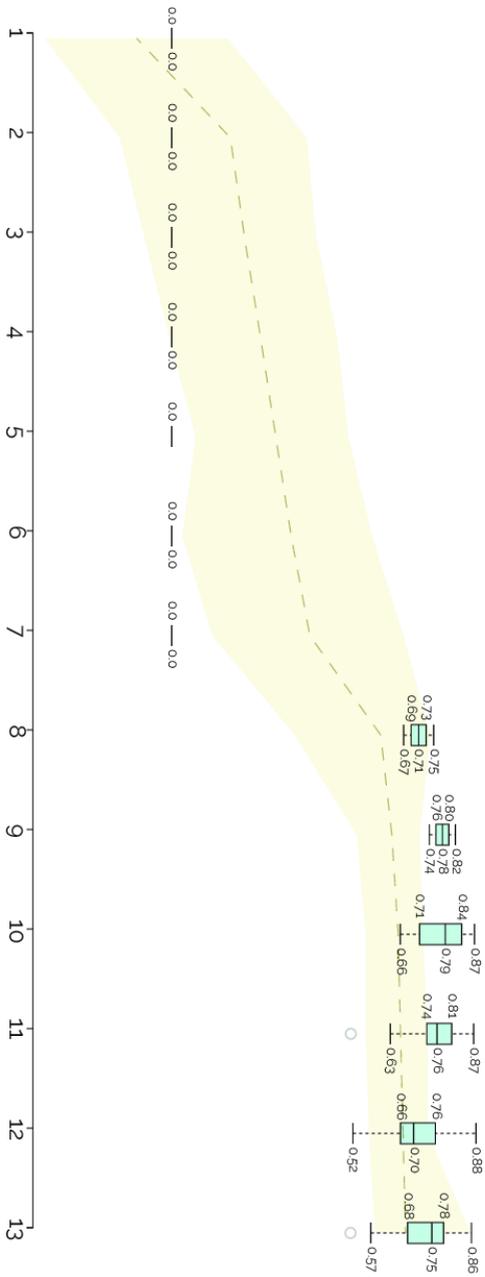


Figure 20: Aggregate model performance in predicting α in the holdout set as a function of the number of RHECV selected features. For each model run, the database was trimmed to the number of features selected by RHECV, the model was retrained on the trimmed database, and then the α prediction accuracy (measured as R²) was evaluated using the holdout set. Box plots represent the minimum, maximum, median, and interquartile range of R² values. Circles represent statistical outliers. The minimum to maximum range of RHECV cross validation R² values are represented by the yellow shading, with the dotted line representing median values for model performance comparison.

Table 8: Summary of values utilized in determining the highest performing model with the fewest number of features and resulting optimum values using Equation 4.

Feature Set Size	8	9	10	11	12	13
RFECV Selection Frequency	2%	2%	6%	14%	19%	57%
Median R ²	0.71	0.78	0.79	0.76	0.70	0.75
Interquartile Range	0.04	0.04	0.13	0.07	0.10	0.10
Optimum Result Value	0.22	0.25	0.22	0.76	0.53	1.75

utilized in determining the optimum model and the resulting optimum result value from Equation 4 are reported in Table 8.

$$\text{Optimum} = \max \left(\frac{1}{n_{\text{features}}} \cdot \frac{1}{(1 - R_{\text{median}}^2)} \cdot \frac{1}{r_{\text{IQ}}} \cdot F_{\text{observation} \%} \right) \quad (4)$$

The highest median performance corresponded to a selection of 10 features, however RFECV only selected an optimal model of 10 features in 6% of the total model runs. Feature sets consisting of 8 and 9 features reported the lowest variance (interquartile range), however feature sets of size 8 and 9 were chosen for only 2% of the total RFECV model runs, respectively. This indicates that there is a subset of data that can be predicted more accurately with fewer features. In general, however, model prediction performance is more frequently optimal for feature sets consisting of 13 features. Feature sets of 13 features were chosen for 57% of model runs. The next most frequently observed feature set sizes consisted of 12 and 11 features, and were observed for 19% and 14% of the model runs, respectively. Due to the corresponding high model performance, high RFECV selection frequency, and low interquartile range, the model consisting of all 13 identified physicochemical features was chosen as the optimum. This result suggests that the relationship between the 13 physicochemical features chosen in this study and α is a complex one that likely depends heavily on all chosen features.

5.3.4 Implications of model findings

In order to identify the features most important to accurately predicting α , a ranked chart is presented in Figure 21 based on the percentage of total model runs in which each feature was included in the optimal

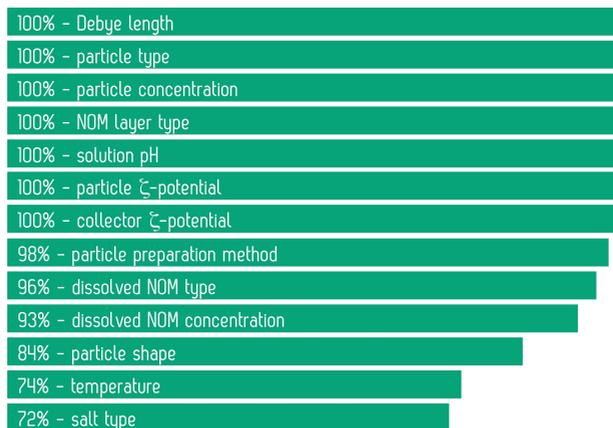


Figure 21: Relative feature importance, reported as percentage of total model runs in which each feature was included in the optimal model to predict α .

model to predict α . The features included in the optimal predictive α model with 100% frequency were not surprising, as they have been the focus of many quantitative α studies, including the studies utilized in the training database for this work (i.e. Debye length, particle type, particle concentration, NOM layer type, solution pH, particle ζ -potential, and collector ζ -potential). The feature chosen in the optimal model with the lowest relative frequency, salt type, might seem surprising at first due to the importance of ion valence on particle-surface interaction energy.^{148,149} It should be noted, however, that the salt type is included within Debye length (as ionic strength) and so the results presented here suggest that the salt type has a somewhat limited influence on top of influencing Debye length. Furthermore, this finding is in agreement with a recent study by Goldberg et al.¹⁰⁷, which reported that salt type was of relative low importance to predicting the retained fraction of ENMs in column experiments when a random forest machine learning model was trained on the data. Temperature was reported to be the second least frequent feature included in the optimal model (albeit still quite often at 74%), despite having been reported to greatly alter the α of nC₆₀ onto NOM coated surfaces in a study by McNew and LeBoeuf¹²⁷. The authors do report, however, that temperature has no effect on nC₆₀ attachment to uncoated surfaces. These findings seem to agree with our

ranking of temperature below other features which have been shown to affect α regardless of the presence or absence of a NOM layer.

The shape of the learning curve presented in Figures 19 and 20 in combination with the relatively high model performance for predicting the holdout set suggest that the α database amassed for this work is well suited to prediction by a machine learning model. This finding also validates the selection of the 13 features used to train the model, by confirming that these features describe the changes in α well. While it is possible that an important parameter was overlooked (e.g. Hamaker constant), the results presented here indicate that there is enough information provided by the physicochemical characteristics of the particles and depositional system to predict α with relative accuracy. This is not a surprising, however, since most transport parameters are controlled by the physicochemical properties of the system and therefore would provide redundant information if they were to be included in a model based explicitly on physicochemical characteristics. Whether or not this finding extends to α values measured by column experiments or other transport parameters, however, is not yet known. Further, applicability beyond the investigated experimental domain is not known and the model must be retrained, and re-evaluated, when new data is added.

While there has been tremendous focus on developing mechanistic models to predict ENM transport, our mechanistic understanding, even in simple, synthetic systems, is insufficient to enable prediction.⁸⁵ As a result, the applicability of mechanistic methods outside the experimental state on which the models are conditioned is too limited, primarily because the influence of physicochemical conditions on transport is not well understood and it is difficult to gain process understanding using conventional experimental and modeling methods.

In this work, we provided an empirical method to disentangle the complex interaction between physicochemical conditions and α to enable the prediction of α under a wide range of complex environmental and particle conditions. Model performance results indicated that the prediction of α is dependent on all 13 features, which evidences the complex relationship between α and the physicochemical state. In this work, we provided an empirical method to sort out the complexity and enable prediction of alpha under a range of complex conditions. The optimum model chosen in this work consisted of all 13 features, illustrating the complex relationship between α and the chosen features, which depends heavily on all physicochemical features

identified in this work. This finding suggests that future mechanistic models for predicting α should focus on all features identified in this work, especially those selected for prediction in the optimum model with 100% (or near 100%) frequency, i.e. Debye length, particle type, particle concentration, NOM layer type, solution pH, particle and collector ζ -potential, particle preparation method, and dissolved NOM type. The features identified as most important to the quantitative prediction of α in this work have been previously reported to qualitatively affect the value of α ^{127–138}, further validating the findings presented in this work. Finally, the findings presented here should be used to conduct informed experiments. By measuring and reporting the features identified as important within this work, future α experimentalists can ensure their work to be as quantitatively useful as possible and efficiently aid in the future effort of constructing a mechanistic model to describe α .

The approach developed herein provides a fundamentally new method to understand particle transport. As such, the method could be used to support risk-informed decisions concerning particle fate and transport by predicting transport behaviors of new and potentially dangerous materials. This enables a sophisticated understanding of risk without waiting on the development of a mechanistic model. Furthermore, this approach could support innovation of new materials as well, by allowing particles to be designed safely from the ground up. With a better understanding of exactly which physicochemical characteristics of the particle affect transport, the transport of a new material could be considered during development. By tweaking these characteristics, particle transport can be reduced from the very beginning, leading to a diminished environmental impact before the particle is even manufactured.

5.4 CONCLUSIONS

In this work, we presented an empirical model to predict α explicitly based on physicochemical characteristics of the particles and depositional environment. The observed model performance (Figures 19 and 20) indicates that prediction of the α database was well captured by the machine learning method chosen and the 13 physicochemical characteristics selected. Results from predicting the holdout set (Table 8) suggest that the relationship between α and the prediction features is a complex one, relying on all 13 physicochemical features selected. The

features reported as most important to the accurate prediction of α (Figure 21) are in agreement with previous studies and the features reported as less important are either somewhat redundant (salt type) or have been previously shown to only affect α under certain conditions (temperature).

In summary, the intended utility of this work is threefold by:

1. providing an accurate, quantitative prediction of α based explicitly on physicochemical characteristics;
2. directing future mechanistic modeling efforts towards the physicochemical characteristics most critical to prediction of α ; and
3. guiding future experiments as to which characteristics are most important to measure and report in order to maximize the scientific benefit of results.

In conclusion, the results presented here illustrate that the relationship between α and the explored physicochemical conditions is too complex to be investigated by a single study alone. Only through compilation of reported α experiments from many different authors and incorporation into an all encompassing model, were we able to predict α based explicitly on physicochemical conditions. Furthermore, similar steps need to be taken to improve the understanding of all aspects of particle transport. Only through compilation of data in coordination with advanced modeling techniques can we make progress towards an all encompassing understanding of transport. In short, we need to close the gaps between experimental understanding and quantitative predictive modeling, in order to maximize the benefit of the high quality transport data being reported on a regular basis.

6

A BRIEF OUTLOOK

To date, no agricultural nanotechnologies have made it to market.⁷ The reasons for this are complex, but several fundamental challenges must be addressed for the field to progress. The first challenge is to overcome the technical limitations that prevent the quantification of ENMs in complex media. The second challenge, predicated on the first, is to develop useful mechanistic PTMs that consider the influence of physical, chemical, and electrical properties of the ENMs and the surrounding environment in real soils. The third challenge relates to the development of an open framework to improve the quality of collected transport data and promote collaboration and validation of collected data.

The machine learning techniques and approaches developed within this thesis can be, and are being, used to help address the first challenge. Recent technical advancements have led to the development of high resolution single-particle ICP-MS that can provide a complete elemental mass spectrum for each particle.¹⁵⁷ These data can then be employed to train a machine learning model to classify ENMs by exploiting trace elemental differences between ENMs and natural nanomaterials too subtle, or variable, for humans to detect by hand.

Addressing the second challenge is more complex, as it is entirely possible that complete mechanistic understanding of ENM transport may not be possible, or even useful, to develop. In general, increasing the number and diversity of ENM transport experiments, as well as the range of recorded conditions, would certainly increase the capability, performance, and utility of the empirical methods developed in this work. Further, it is likely that with more data the approach employed in Chapter 4 could provide detailed quantitative insights into how the behavior of materials is influenced by physical, chemical, and electrical

conditions. However, these insights are predicated on the data, which exclusively consist of bench top soil transport experiments conducted in simple, synthetic soil systems (i.e., monodisperse, fully saturated, single-media soils). Obviously, synthetic soil systems have questionable applicability to real soils, which are considerably more heterogeneous and may have areas of variable porosity and/or under saturation, not to mention non-uniform and/or directionally dependent hydraulic conductivities. Therefore, while it may be possible to eventually understand all the mechanics of transport, it may not be prudent – or even possible without an exorbitant budget – thus reinforcing the need to address the first challenge, as well as the need to focus future efforts towards experimentation in real soils.

It is critical to develop standards for data collection or presentation for ENM transport experiments to address the third challenge. As observed in Chapters 3–5, the quantity and quality of overlapping data between publications is low. However, in related fields, such as toxicology, open data standards and public databases exist to store and share data. The databases developed in Chapters 3–5, which represent the first of their kind in this field, should be used as templates from which to provide guidance towards the creation of an open ENM transport platform. Not only would this support the future development of data-driven models, but it would also provide a critical collaborative platform from which to facilitate collaboration between experimentalists and modelers.

Thus, the call to action in this outlook is not to promote the ‘business-as-usual’ approach and forecast a future predicated on an increasing number of experiments with dubious applicability to the real world. Instead, it is to pivot the direction of research to focus on addressing ENM transport in realistic environmental conditions. Further, it is to use the insight gained, and databases developed, to develop an open platform to enable collaboration.

BIBLIOGRAPHY

- [1] David Hwang and Jurrion Bradley. *The recession's ripple effect on nanotech*. Lux Research, Inc., New York, 2010.
- [2] Panagiotis Dallas, Antonios Kelarakis, and Emmanuel P Giannelis. Nanostructured materials for environmentally conscious applications. In *ACS Symposium Series, Vol. 1124*, volume 1124, pages 59–72. American Chemical Society (ACS), 2013.
- [3] Meagan S Mauter and Menachem Elimelech. Environmental applications of carbon-based nanomaterials. *Environmental Science & Technology*, 42(16):5843–5859, 2008.
- [4] Jianjun Miao, Minoru Miyauchi, Trevor J Simmons, Jonathan S Dordick, and Robert J Linhardt. Electrospinning of nanomaterials and applications in electronic components and devices. *Journal of nanoscience and nanotechnology*, 10(9):5507–5519, 2010.
- [5] M Muneer, M Saquib, M Qamar, and D Bahnemann. Titanium-dioxide-mediated photocatalysis reaction of three selected pesticide derivatives. *Research on chemical intermediates*, 30(6):663–672, 2004.
- [6] Guobin Shan, S Yan, RD Tyagi, Rao Y Surampalli, and Tian C Zhang. Applications of nanomaterials in environmental science and engineering: review. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 13(2):110–119, 2009.
- [7] Claudia Parisi, Mauro Vigani, and Emilio Rodríguez-Cerezo. Agricultural nanotechnologies: What are the current possibilities? *Nano Today*, 10(2):124–127, 2015.
- [8] Alexander Gogos, Katja Knauer, and Thomas D Bucheli. Nanomaterials in Plant Protection and Fertilization: Current State, Foreseen Applications, and Research Priorities. *Journal of Agricultural and Food Chemistry*, 60(39):9781–9792, October 2012.

- [9] Matthew MacLeod, Martin Scheringer, Thomas E McKone, and Konrad Hungerbühler. The state of multimedia mass-balance modeling in environmental science and decision-making. *Environmental Science & Technology*, 44:8360–8364, 2010.
- [10] Donald Mackay. *Multimedia environmental models: the fugacity approach*. CRC press, 2001.
- [11] Martin Scheringer and Frank Wania. Multimedia models of global transport and fate of persistent organic pollutants. In *Persistent Organic Pollutants*. Springer, 2003.
- [12] Jerald L Schnoor. *Environmental modeling: fate and transport of pollutants in water, air, and soil*. John Wiley and Sons, 1996.
- [13] Antonia Praetorius, Nathalie Tufenkji, Kai-Uwe Goss, Martin Scheringer, Frank von der Kammer, and Menachem Elimelech. The road to nowhere: equilibrium partition coefficients for nanoparticles. *Environmental Science: Nano*, 1(4):317–323, 2014.
- [14] A. Praetorius, M. Scheringer, and K. Hungerbühler. Development of environmental fate models for engineered nanoparticles—a case study of TiO₂ nanoparticles in the Rhine river. *Environmental Science & Technology*, 46:6705–13, 2012.
- [15] M.A. Curran. *Life Cycle Assessment Handbook: A Guide for Environmentally Sustainable Products*. Wiley, 2012.
- [16] Fadri Gottschalk, Tobias Sonderer, Roland W Scholz, and Bernd Nowack. Modeled Environmental Concentrations of Engineered Nanomaterials (TiO₂, ZnO, Ag, CNT, Fullerenes) for Different Regions. *Environmental Science & Technology*, 43(24):9216–9222, December 2009.
- [17] B. Nowack, J. F. Ranville, S. Diamond, J. A. Gallego-Urrea, C. Metcalfe, J. Rose, N. Horne, A. A. Koelmans, and S. J. Klaine. Potential scenarios for nanomaterial release and subsequent alteration in the environment. *Environ Toxicol Chem*, 31:50–9, 2012.
- [18] F. Gottschalk, T. Sonderer, R. W. Scholz, and B. Nowack. Possibilities and limitations of modeling environmental exposure to engineered nanomaterials by probabilistic material flow analysis. *Environ Toxicol Chem*, 29:1036–48, 2010.

- [19] Arturo A Keller and Anastasiya Lazareva. Predicted releases of engineered nanomaterials: from global to regional to local. *Environmental Scienc & Technology Letters*, 1(1):65–70, 2013.
- [20] Kuan-Mu Yao, Mohammad T Habibian, and Charles R O'Melia. Water and waste water filtration. Concepts and applications. *Environmental Science & Technology*, 5(11):1105–1112, 1971.
- [21] S A Bradford and N Toride. A stochastic model for colloid transport and deposition. *Journal of Environmental Quality*, 36(5): 1346–1356, 2007.
- [22] Adamo R Petosa, Deb P Jaisi, Ivan R Quevedo, Menachem Elimelech, and Nathalie Tufenkji. Aggregation and Deposition of Engineered Nanomaterials in Aquatic Environments: Role of Physicochemical Interactions. *Environmental Science & Technology*, 44(17):6532–6549, September 2010.
- [23] L Cai, M Tong, H Ma, and H Kim. Co-transport of Titanium Dioxide and Fullerene Nanoparticles in Saturated Porous Media. *Environmental Science & technology*, 47:5703–5710, 2013.
- [24] Gexin Chen, Xuyang Liu, and Chunming Su. Transport and Retention of TiO₂ Rutile Nanoparticles in Saturated Porous Media under Low-Ionic-Strength Conditions: Measurements and Mechanisms. *Langmuir*, 27(9):5393–5402, May 2011.
- [25] Indranil Chowdhury, Yongsuk Hong, Ryan J Honda, and Sharon L Walker. Mechanisms of TiO₂ nanoparticle transport in porous media: Role of solution chemistry, nanoparticle concentration, and flowrate. *Journal of Colloid And Interface Science*, 360(2):548–555, 2011.
- [26] Christine Chin Choy, Mahmoud Wazne, and Xiaoguang Meng. Application of an empirical transport model to simulate retention of nanocrystalline titanium dioxide in sand columns. *Chemosphere*, 71(9):1794–1801, April 2008.
- [27] Peng Han, Xueting Wang, Li Cai, Meiping Tong, and Hyunjung Kim. Transport and retention behaviors of titanium dioxide nanoparticles in iron oxide-coated quartz sand: Effects of ph, ionic strength, and humic acid. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 454:119–127, 2014.

- [28] Xujia Jiang, Meiping Tong, Ruiqing Lu, and Hyunjung Kim. Transport and deposition of zno nanoparticles in saturated porous media. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 401(0):29 – 37, 2012. ISSN 0927-7757. doi: <http://dx.doi.org/10.1016/j.colsurfa.2012.03.004>.
- [29] Xujia Jiang, Meiping Tong, and Hyunjung Kim. Influence of natural organic matter on the transport and deposition of zinc oxide nanoparticles in saturated porous media. *Journal of Colloid And Interface Science*, 386(1):34–43, 2012.
- [30] Xujia Jiang, Xueting Wang, Meiping Tong, and Hyunjung Kim. Initial transport and retention behaviors of zno nanoparticles in quartz sand porous media coated with escherichia coli biofilm. *Environmental Pollution*, 174:38–49, 2013.
- [31] Edward H Jones and Chunming Su. Fate and transport of elemental copper (Cu₀) nanoparticles through saturated porous media in the presence of organic materials. *Water Research*, 46(7): 2445–2456, 2012.
- [32] EH Jones and C Su. Transport and retention of zinc oxide nanoparticles in porous media: Effects of natural organic matter versus natural organic ligands at circumneutral ph. *Journal of Hazardous Materials*, 275:79–88, 2014.
- [33] Daniela Kasel, Scott A Bradford, Jiří Šimůnek, Marc Heggen, Harry Vereecken, and Erwin Klumpp. Transport and retention of multi-walled carbon nanotubes in saturated porous media: Effects of input concentration and grain size. *Water Research*, 47(2):933–944, February 2013.
- [34] Matthew D Becker, Yonggang Wang, Kurt D Pennell, and Linda M Abriola. A multi-constituent site blocking model for nanoparticle and stabilizing agent transport in porous media. *Environmental Science: Nano*, 2(2):155–166, 2015.
- [35] Yan Liang, Scott A Bradford, Jiří Šimůnek, Harry Vereecken, and Erwin Klumpp. Sensitivity of the Transport and Retention of Stabilized Silver Nanoparticles to Physicochemical Factors. *Water Research*, 47:2572–2528, 2013.
- [36] Xuyang Liu, Gexin Chen, and Chunming Su. Influence of collector surface composition and water chemistry on the deposition of cerium dioxide nanoparticles: QCM-D and column experiment

- approaches. *Environmental Science & Technology*, 46(12):6681–6688, 2012.
- [37] Tiziana Tosco, Julian Bosch, Rainer U Meckenstock, and Rajandrea Sethi. Transport of ferrihydrite nanoparticles in saturated porous media: role of ionic strength and flow rate. *Environmental Science & Technology*, 46(7):4008–4015, 2012.
- [38] Elsa Vitorge, Stéphanie Szenknect, Jean M-F Martins, Véronique Barthès, and Jean-Paul Gaudet. Comparison of three labeled silica nanoparticles used as tracers in transport experiments in porous media. part ii: Transport experiments and modeling. *Environmental Pollution*, 184:613–619, 2014.
- [39] Yonggang Wang, Yusong Li, John D Fortner, Joseph B Hughes, Linda M Abriola, and Kurt D Pennell. Transport and Retention of Nanoscale C₆₀ Aggregates in Water-Saturated Porous Media. *Environmental Science & Technology*, 42(10):3588–3594, May 2008.
- [40] Dengjun Wang, Marcos Paradelo, Scott A Bradford, Willie JGM Peijnenburg, Lingyang Chu, and Dongmei Zhou. Facilitated transport of cu with hydroxyapatite nanoparticles in saturated sand: Effects of solution ionic strength and composition. *Water Research*, 45(18):5905–5915, 2011.
- [41] Dengjun Wang, Wei Zhang, Xiuzhen Hao, and Dongmei Zhou. Transport of Biochar Particles in Saturated Granular Media: Effects of Pyrolysis Temperature and Particle Size. *Environmental Science & Technology*, 47(2):821–828, 2013.
- [42] Dengjun Wang, Wei Zhang, and Dongmei Zhou. Antagonistic Effects of Humic Acid and Iron Oxyhydroxide Grain-Coating on Biochar Nanoparticle Transport in Saturated Sand. *Environmental Science & Technology*, 47(10):5154–5161, May 2013.
- [43] Dengjun Wang, Deb P Jaisi, Jing Yan, Yan Jin, and Dongmei Zhou. Transport and retention of polyvinylpyrrolidone-coated silver nanoparticles in natural soils. *Vadose Zone Journal*, 14(7), 2015.
- [44] Dengjun Wang, Scott A Bradford, Ronald W Harvey, Xiuzhen Hao, and Dongmei Zhou. Transport of ars-labeled hydroxyapatite nanoparticles in saturated granular media is influenced by surface charge variability even in the presence of humic acid. *Journal of hazardous materials*, 229:170–176, 2012.

- [45] H Kurlanda-Witek, BT Ngwenya, and IB Butler. Transport of bare and capped zinc oxide nanoparticles is dependent on porous medium composition. *Journal of contaminant hydrology*, 162:17–26, 2014.
- [46] Gerd H Woehrle, Leif O Brown, and James E Hutchison. Thiol-functionalized, 1.5-nm gold nanoparticles through ligand exchange reactions: Scope and mechanism of ligand exchange. *Journal of the American Chemical Society*, 127(7):2172–2183, 2005.
- [47] Frank Von der Kammer, P Lee Ferguson, Patricia A Holden, Armand Masion, Kim R Rogers, Stephen J Klaine, Albert A Koelmans, Nina Horne, and Jason M Unrine. Analysis of engineered nanomaterials in complex matrices (environment and biota): general considerations and conceptual case studies. *Environmental Toxicology and Chemistry*, 31(1):32–49, 2012.
- [48] Scott A Bradford, Jirka Simunek, Mehdi Bettahar, Martinus Th van Genuchten, and Scott R Yates. Modeling colloid attachment, straining, and exclusion in saturated porous media. *Environmental Science & Technology*, 37(10):2242–2250, 2003.
- [49] G Gargiulo, S Bradford, J Simunek, P Ustohal, H Vereecken, and E Klumpp. Bacteria transport and deposition under unsaturated conditions: The role of the matrix grain size and the bacteria surface protein. *Journal of Contaminant Hydrology*, 92(3-4):255–273, 2007.
- [50] Nathalie Tufenkji and Menachem Elimelech. Spatial Distributions of Cryptosporidium Oocysts in Porous Media: Evidence for Dual Mode Deposition. *Environmental Scienc & Technology*, 39(10):3620–3629, May 2005.
- [51] Stephen J Klaine, Albert A Koelmans, Nina Horne, Stephen Carley, Richard D Handy, Larry Kapustka, Bernd Nowack, and Frank von der Kammer. Paradigms to assess the environmental impact of manufactured nanomaterials. *Environmental Toxicology and Chemistry*, 31(1):3–14, 2012.
- [52] Philip R Johnson and Menachem Elimelech. Dynamics of colloid deposition in porous media: Blocking based on random sequential adsorption. *Langmuir*, 11(3):801–812, 1995.
- [53] Jack F Schijven, S Majid Hassanizadeh, and Ria HAM de Bruin. Two-site kinetic modeling of bacteriophages transport through

- columns of saturated dune sand. *Journal of Contaminant Hydrology*, 57(3):259–279, 2002.
- [54] S A Bradford, M Bettahar, J Simunek, and M T van Genuchten. Straining and attachment of colloids in physically heterogeneous porous media. *Vadose Zone Journal*, 3(2):384–394, May 2004.
- [55] Jan Willem Foppen, Manon van Herwerden, and Jack Schijven. Transport of *Escherichia coli* in saturated porous media: Dual mode deposition and intra-population heterogeneity. *Water Research*, 41(8):1743–1753, 2007.
- [56] Jirka Simunek, Changming He, Liping Pang, and S A Bradford. Colloid-Facilitated Solute Transport in Variably Saturated Porous Media. *Vadose Zone Journal*, 5(3):1035, 2006.
- [57] William P Johnson, Eddy Pazmino, and Huilian Ma. Direct observations of colloid retention in granular media in the presence of energy barriers, and implications for inferred mechanisms from indirect observations. *Water Research*, 44(4):1158–1169, 2010.
- [58] William P Johnson, Huilian Ma, and Eddy Pazmino. Straining credibility: a general comment regarding common arguments used to infer straining as the mechanism of colloid retention in porous media. *Environmental Science & Technology*, 45(9):3831–3832, 2011.
- [59] Ning Sun, Ne Zheng Sun, Menachem Elimelech, and Joseph N Ryan. Sensitivity analysis and parameter identifiability for colloid transport in geochemically heterogeneous porous media. *Water Resources Research*, 37(2):209–222, 2001.
- [60] AE Scheidegger. General theory of dispersion in porous media. *Journal of Geophysical Research*, 66(10):3273–3278, 1961.
- [61] Nathalie Tufenkji and Menachem Elimelech. Correlation Equation for Predicting Single-Collector Efficiency in Physicochemical Filtration in Saturated Porous Media. *Environmental Science & Technology*, 38(2):529–536, January 2004.
- [62] Ruben Kretzschmar, Michal Borkovec, Daniel Grolimund, and Menachem Elimelech. Mobile subsurface colloids and their role in contaminant transport. *Advances in Agronomy*, 66:121–193, 1999.

- [63] Nathalie Tufenkji and Menachem Elimelech. Deviation from the Classical Colloid Filtration Theory in the Presence of Repulsive DLVO Interactions. *Langmuir*, 20(25):10818–10828, December 2004.
- [64] Lunliang Zhang, Lei Hou, Lilin Wang, Amy T Kan, Wei Chen, and Mason B Tomson. Transport of Fullerene Nanoparticles (nC 60) in Saturated Sand and Sandy Soil: Controlling Factors and Modeling. *Environmental Science & Technology*, 46(13):7230–7238, July 2012.
- [65] Yan Liang, Scott A Bradford, Jiri Simunek, Marc Heggen, Harry Vereecken, and Erwin Klumpp. Retention and remobilization of stabilized silver nanoparticles in an undisturbed loamy sand soil. *Environmental Science & Technology*, 47(21):12229–12237, 2013.
- [66] Yusong Li, Yonggang Wang, Kurt D Pennell, and Linda M Abriola. Investigation of the transport and deposition of fullerene (c60) nanoparticles in quartz sands under varying flow conditions. *Environmental Science & Technology*, 42(19):7174–7180, 2008.
- [67] Hélène F Lecoanet, Jean-Yves Bottero, and Mark R Wiesner. Laboratory Assessment of the Mobility of Nanomaterials in Porous Media. *Environmental Science & Technology*, 38(19):5164–5169, October 2004.
- [68] Tanapon Phenrat, Jee Eun Song, Charlotte M Cisneros, Daniel P Schoenfelder, Robert D Tilton, and Gregory V Lowry. Estimating attachment of nano- and submicrometer-particles coated with organic macromolecules in porous media: development of an empirical model. *Environmental Science & Technology*, 44(12):4531–4538, 2010.
- [69] Rajamani Rajagopalan and Chi Tien. Trajectory analysis of deep-bed filtration with the sphere-in-cell porous media model. *AIChE Journal*, 22(3):523–533, 1976.
- [70] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.
- [71] Thomas F Coleman and Yuying Li. On the convergence of interior-reflective newton methods for nonlinear minimization subject to bounds. *Mathematical Programming*, 67(1-3):189–224, 1994.
- [72] Kenneth P Burnham and David R Anderson. Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004.

- [73] J Simunek, M Šejna, and M Th van Genuchten. HYDRUS-1D Manual version 4.16. *Software package for simulating onedimensional movement of water, heat and multiple solutes in variably saturated media.*, 2013.
- [74] Guoping Tang, Melanie A Mayes, Jack C Parker, and Philip M Jardine. CXTFIT/Excel-A modular adaptable code for parameter estimation, sensitivity analysis and uncertainty analysis for laboratory or field tracer experiments. *Computers & Geosciences*, 36(9):1200–1209, 2010.
- [75] Rick F Rajter, Roger H French, WY Ching, Rudolf Podgornik, and V Adrian Parsegian. Chirality-dependent properties of carbon nanotubes: electronic structure, optical dispersion properties, hamaker coefficients and van der waals–london dispersion interactions. *RSC Advances*, 3(3):823–842, 2013.
- [76] Yonggang Wang, Jae-Hong Kim, Jong-Beom Baek, Gary W Miller, and Kurt D Pennell. Transport behavior of functionalized multi-wall carbon nanotubes in water-saturated quartz sand as a function of tube length. *Water Research*, 46(14):4521–4531, 2012.
- [77] Gexin Chen, Xuyang Liu, and Chunming Su. Distinct Effects of Humic Acid on Transport and Retention of TiO₂ Rutile Nanoparticles in Saturated Sand Columns. *Environmental Science & Technology*, 46(13):7142–7150, July 2012.
- [78] S A Bradford, J Simunek, M Bettahar, M T van Genuchten, and S R Yates. Significance of straining in colloid deposition: Evidence and implications. *Water Resources Research*, 42(12):W12S15, doi:10.1029/2005WR004791., November 2006.
- [79] Saeed Torkzaban, Scott A Bradford, Martinus Th van Genuchten, and Sharon L Walker. Colloid transport in unsaturated porous media: The role of water content and ionic strength on particle straining. *Journal of Contaminant Hydrology*, 96(1):113–127, 2008.
- [80] Scott A Bradford, Jirka Simunek, and Sharon L Walker. Transport and straining of E. coli O157: H7 in saturated porous media. *Water Resources Research*, 42(12):1–12, 2006.
- [81] Hao Yuan and Alexander A Shapiro. A mathematical model for non-monotonic deposition profiles in deep bed filtration systems. *Chemical Engineering Journal*, 166(1):105–115, 2011.

- [82] Florian Kuhnen, Kurt Barmettler, Subir Bhattacharjee, Menachem Elimelech, and Ruben Kretzschmar. Transport of iron oxide colloids in packed quartz sand media: monolayer and multilayer deposition. *Journal of Colloid And Interface Science*, 231(1):32–41, 2000.
- [83] Shangping Xu, Bin Gao, and James E Saiers. Straining of colloidal particles in saturated porous media. *Water Resources Research*, 42(12), 2006.
- [84] Xiqing Li, Timothy D Scheibe, and William P Johnson. Apparent Decreases in Colloid Deposition Rate Coefficients with Distance of Transport under Unfavorable Deposition Conditions: A General Phenomenon. *Environmental Science & Technology*, 38(21):5616–5625, November 2004.
- [85] Eli Goldberg, Martin Scheringer, Thomas D Bucheli, and Konrad Hungerbühler. Critical assessment of models for transport of engineered nanoparticles in saturated porous media. *Environmental Scienc & Technology*, 48(21):12732–12741, 2014.
- [86] Jeremy A Redman, Sharon L Walker, and Menachem Elimelech. Bacterial adhesion and transport in porous media: Role of the secondary energy minimum. *Environmental Scienc & Technology*, 38(6):1777–1785, 2004.
- [87] Menachem Elimelech and Charles R O’Melia. Kinetics of deposition of colloidal particles in porous media. *Environmental Scienc & Technology*, 24(10):1528–1536, 1990.
- [88] Melinda W Hahn and Charles R O’Melia. Deposition and reentrainment of brownian particles in porous media under unfavorable chemical conditions: Some concepts and applications. *Environmental Science & Technology*, 38(1):210–220, 2004.
- [89] Amy Dale, Elizabeth A Casman, Gregory Victor Lowry, Jamie R Lead, Enrica Viparelli, and Mohammed A Baalousha. Modeling nanomaterial environmental fate in aquatic systems. *Environmental Scienc & Technology*, 2015.
- [90] Jeremy M Gernand and Elizabeth A Casman. Machine learning for nanomaterial toxicity risk assessment. *Intelligent Systems, IEEE*, 29(3):84–88, 2014.

- [91] DA Winkler, FR Burden, B Yan, R Weissleder, C Tassa, S Shaw, and VC Epa. Modelling and predicting the biological effects of nanomaterials. *SAR and QSAR in Environmental Research*, 25(2): 161–172, 2014.
- [92] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [93] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [94] CJ van Rijsbergen. *Information Retrieval*. 1979. Butterworth, 1979.
- [95] Kristin K Nicodemus, James D Malley, Carolin Strobl, and Andreas Ziegler. The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC bioinformatics*, 11(1):110, 2010.
- [96] Gilles Louppe, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. Understanding variable importances in forests of randomized trees. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 431–439. Curran Associates, Inc., 2013.
- [97] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1):25, 2007.
- [98] Sandrine Dudoit and Jane Fridlyand. Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, 19(9):1090–1099, 2003.
- [99] Ronald E Walpole, Raymond H Myers, Sharon L Myers, and Keying Ye. *Probability and statistics for engineers and scientists*, volume 5. Macmillan New York, 1993.
- [100] Tal Ben-Moshe, Ishai Dror, and Brian Berkowitz. Transport of metal oxide nanoparticles in saturated porous media. *Chemosphere*, 81(3):387–393, 2010.
- [101] Rebecca A French, Astrid R Jacobson, Bojeong Kim, Sara L Isley, R Lee Penn, and Philippe C Baveye. Influence of ionic strength,

- ph, and cation valence on aggregation kinetics of titanium dioxide nanoparticles. *Environmental Scienc & Technology*, 43(5):1354–1359, 2009.
- [102] Katherine A Dunphy Guzman, Michael P Finnegan, and Jillian F Banfield. Influence of surface potential on aggregation and transport of titania nanoparticles. *Environmental Science & Technology*, 40(24):7688–7693, 2006.
- [103] Stephen J Klaine, Pedro JJ Alvarez, Graeme E Batley, Teresa F Fernandes, Richard D Handy, Delina Y Lyon, Shaily Mahendra, Michael J McLaughlin, and Jamie R Lead. Nanomaterials in the environment: behavior, fate, bioavailability, and effects. *Environmental Toxicology& Chemistry*, 27(9):1825–1851, 2008.
- [104] Joyce S Tsuji, Andrew D Maynard, Paul C Howard, John T James, Chiu-wing Lam, David B Warheit, and Annette B Santamaria. Research strategies for safety evaluation of nanomaterials, part iv: risk assessment of nanoparticles. *Toxicological Sciences*, 89(1):42–50, 2006.
- [105] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques, 2007.
- [106] T Guest and A Curtis. Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 114(B4), 2009.
- [107] Eli Goldberg, Martin Scheringer, Thomas D Bucheli, and Konrad Hungerbühler. Prediction of nanoparticle transport behavior from physicochemical properties: machine learning provides insights to guide the next generation of transport models. *Environmental Science: Nano*, 2(4):352–360, 2015.
- [108] You-Im Chang and Hsun-Chih Chan. Correlation equation for predicting filter coefficient under unfavorable deposition conditions. *AIChE journal*, 54(5):1235–1253, 2008.
- [109] Rajamani Rajagopalan and Chi Tien. Trajectory analysis of deep-bed filtration with the sphere-in-cell porous media model. *AIChE Journal*, 22(3):523–533, 1976.
- [110] Chi Tien and Bandaru V Ramarao. *Granular filtration of aerosols and hydrosols*. Elsevier, 2011.

- [111] Jacob N. Israelachvili. *Intermolecular and Surface Forces*. Elsevier, 3rd ed. edition, 2011.
- [112] Raj Rajagopalan and Jong S Kim. Adsorption of brownian particles in the presence of potential barriers: effect of different modes of double-layer interaction. *Journal of Colloid and Interface Science*, 83(2):428–448, 1981.
- [113] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.
- [114] Mark Last, Oded Maimon, and Einat Minkov. Improving stability of decision trees. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(02):145–159, 2002.
- [115] R Allan Freeze, John ARA Cherry, and Cherry JA. *Groundwater*. Number 556.3 FRE. Prentice Hall, 1979.
- [116] Rebecca Sutton and Garrison Sposito. Molecular structure in soil humic substances: the new view. *Environmental Science & Technology*, 39(23):9009–9015, 2005.
- [117] László Zsolt Garamszegi. Modern phylogenetic comparative methods and their application in evolutionary biology. *Concepts and Practice Springer*, 2014.
- [118] Scott A Bradford, Scott R Yates, Mehdi Bettahar, and Jirka Simunek. Physical factors affecting the transport and fate of colloids in saturated porous media. *Water Resources Research*, 38(12), 2002.
- [119] Saeed Torkzaban, Scott A Bradford, and Sharon L Walker. Resolving the coupled effects of hydrodynamics and dlvo forces on colloid attachment in porous media. *Langmuir*, 23(19):9652–9660, 2007.
- [120] Barbara Karn, Todd Kuiken, and Martha Otto. Nanotechnology and in situ remediation: a review of the benefits and potential risks. *Environmental health perspectives*, pages 1823–1831, 2009.
- [121] Bhupinder Singh Sekhon. Nanotechnology in agri-food production: an overview. *Nanotechnology, science and applications*, 7:31, 2014.
- [122] Vandana Ghormade, Mukund V Deshpande, and Kishore M Paknikar. Perspectives for nano-biotechnology enabled protection and nutrition of plants. *Biotechnology advances*, 29(6):792–803, 2011.

- [123] Andre Nel, Tian Xia, Lutz Mädler, and Ning Li. Toxic potential of materials at the nanolevel. *Science*, 311(5761):622–627, 2006.
- [124] JP Kamat, TPA Devasagayam, KI Priyadarsini, and H Mohan. Reactive oxygen species mediated membrane damage induced by fullerene derivatives and its possible biological implications. *Toxicology*, 155(1):55–61, 2000.
- [125] N Nakajima, C Nishi, FM Li, and Y Ikada. Photo-induced cytotoxicity of water-soluble fullerene. *Fullerene Science & Technology*, 4(1):1–19, 1996.
- [126] Daohui Lin, Xiaoli Tian, Fengchang Wu, and Baoshan Xing. Fate and transport of engineered nanomaterials in the environment. *Journal of Environmental Quality*, 39(6):1896–1908, 2010.
- [127] Coy P McNew and Eugene J LeBoeuf. The role of attached phase soil and sediment organic matter physicochemical properties on fullerene (nc 60) attachment. *Chemosphere*, 139:609–616, 2015.
- [128] Coy P McNew. *The Attachment of Colloidal Particles to Environmentally Relevant Surfaces: Effect of Ionic Strength, Particle Shape, and Physicochemical Properties*. PhD thesis, Vanderbilt University, 2015.
- [129] Indranil Chowdhury, Matthew C Duch, Nikhita D Mansukhani, Mark C Hersam, and Dermont Bouchard. Deposition and release of graphene oxide nanomaterials using a quartz crystal microbalance. *Environmental Science & Technology*, 48(2):961–969, 2014.
- [130] Xiaojun Chang and Dermont C Bouchard. Multiwalled carbon nanotube deposition on model environmental surfaces. *Environmental Science & Technology*, 47(18):10372–10380, 2013.
- [131] Xiaolei Qu, Pedro JJ Alvarez, and Qilin Li. Impact of sunlight and humic acid on the deposition kinetics of aqueous fullerene nanoparticles (nc60). *Environmental Science & Technology*, 46(24):13455–13462, 2012.
- [132] Yun Shen, Hyunjung Kim, Meiping Tong, and Qingyun Li. Influence of solution chemistry on the deposition and detachment kinetics of rna on silica surfaces. *Colloids and Surfaces B: Biointerfaces*, 82(2):443–449, 2011.

- [133] Beng Joo Reginald Thio, Dongxu Zhou, and Arturo A Keller. Influence of natural organic matter on the aggregation and deposition of titanium dioxide nanoparticles. *Journal of hazardous materials*, 189(1):556–563, 2011.
- [134] Peng Yi and Kai Loon Chen. Influence of surface oxidation on the aggregation and deposition kinetics of multiwalled carbon nanotubes in monovalent and divalent electrolytes. *Langmuir*, 27(7):3588–3599, 2011.
- [135] Xujia Jiang, Meiping Tong, Haoyang Li, and Kun Yang. Deposition kinetics of zinc oxide nanoparticles on natural organic matter coated silica surfaces. *Journal of colloid and interface science*, 350(2):427–434, 2010.
- [136] Kai Loon Chen and Menachem Elimelech. Interaction of fullerene (c60) nanoparticles with humic acid and alginate coated silica surfaces: measurements, mechanisms, and environmental implications. *Environmental Science & Technology*, 42(20):7607–7614, 2008.
- [137] Baoling Yuan, Mai Pham, and Thanh H Nguyen. Deposition kinetics of bacteriophage ms2 on a silica surface coated with natural organic matter in a radial stagnation point flow cell. *Environmental Science & Technology*, 42(20):7628–7633, 2008.
- [138] Kai Loon Chen and Menachem Elimelech. Aggregation and deposition kinetics of fullerene (c60) nanoparticles. *Langmuir*, 22(26):10994–11001, 2006.
- [139] Steven Sleutel, Veerle Cnudde, Bert Masschaele, J Vlassenbroek, Manuel Dierick, Luc Van Hoorebeke, Patric Jacobs, and Stefaan De Neve. Comparison of different nano-and micro-focus x-ray computed tomography set-ups for the visualization of the soil microstructure and soil organic matter. *Computers & Geosciences*, 34(8):931–938, 2008.
- [140] Tanapon Phenrat, Hye-Jin Kim, Fritjof Fagerlund, Tissa Illangasekare, and Gregory V Lowry. Empirical correlations to estimate agglomerate size and deposition during injection of a polyelectrolyte-modified fe o nanoparticle at high particle concentration in saturated sand. *Journal of contaminant hydrology*, 118(3):152–164, 2010.

- [141] Renbi Bai and Chi Tien. Particle deposition under unfavorable surface interactions. *Journal of colloid and interface science*, 218(2): 488–499, 1999.
- [142] Renbi Bai and Chi Tien. A new correlation for the initial filter coefficient under unfavorable surface interactions. *Journal of colloid and interface science*, 179(2):631–634, 1996.
- [143] Menachem Elimelech. Predicting collision efficiencies of colloidal particles in porous media. *Water Research*, 26(1):1–8, 1992.
- [144] Nelson Akaighe, Sean W Depner, Sarbajit Banerjee, and Mary Sohn. Transport and deposition of suwannee river humic acid/natural organic matter formed silver nanoparticles on silica matrices: The influence of solution ph and ionic strength. *Chemosphere*, 92(4):406–412, 2013.
- [145] Indranil Chowdhury, Matthew C Duch, Nikhita D Mansukhani, Mark C Hersam, and Dermont Bouchard. Interactions of graphene oxide nanomaterials with natural organic matter and metal oxide surfaces. *Environmental Science & Technology*, 48(16):9382–9390, 2014.
- [146] Meiping Tong, Jiali Ding, Yun Shen, and Pingting Zhu. Influence of biofilm on the transport of fullerene (c 60) nanoparticles in porous media. *Water Research*, 44(4):1094–1103, 2010.
- [147] Megan B Seymour, Gexin Chen, Chunming Su, and Yusong Li. Transport and retention of colloids in porous media: does shape really matter? *Environmental Science & Technology*, 47(15):8391–8398, 2013.
- [148] BV Derjaguin, L Landau, et al. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Acta physicochim. URSS*, 14(6): 633–662, 1941.
- [149] Evert Johannes Willem Verwey, J Th G Overbeek, and Jan Theodoor Gerard Overbeek. *Theory of the stability of lyophobic colloids*. Courier Corporation, 1999.
- [150] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

- [151] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8599–8603. IEEE, 2013.
- [152] David Stavens and Sebastian Thrun. A self-supervised terrain roughness estimator for off-road autonomous driving. *arXiv preprint arXiv:1206.6872*, 2012.
- [153] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.
- [154] Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. *Jama*, 309(13):1351–1352, 2013.
- [155] Robert J Hunter. *Zeta potential in colloid science: principles and applications*, volume 2. Academic press, 2013.
- [156] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [157] Olga Borovinskaya, Sabrina Gschwind, Bodo Hattendorf, Martin Tanner, and Detlef Günther. Simultaneous mass quantification of nanoparticles of different composition in a mixture by microdroplet generator-icptofms. *Analytical chemistry*, 86(16):8142–8148, 2014.
- [158] Lokenath Debnath and Lokenath Debnath. *Nonlinear partial differential equations for scientists and engineers*. Springer, 1997.
- [159] Eli Ruckenstein and Dennis C Prieve. Adsorption and desorption of particles and their chromatographic separation. *AIChE Journal*, 22(2):276–283, 1976.
- [160] J. Czarnecki. Van der waals attraction energy between sphere and half-space. *Journal of Colloid And Interface Science*, 72(2):361–362, 1979.
- [161] Rajasekar Vaidyanathan and Chi Tien. Hydrosol deposition in granular media under unfavorable surface conditions. *Chemical engineering science*, 46(4):967–983, 1991.
- [162] J N Ryan and M Elimelech. Colloid mobilization and transport in groundwater. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 107:1–56, 1996.

- [163] DODDI Yudianto and YB Xie. Contaminant distribution under non-uniform velocity of steady flow regimes. *Journal of Applied Science in Environmental Sanitation*, 3(1):29–40, 2008.
- [164] MATLAB. *MATLAB and Statistics Toolbox Release 2014a*. The MathWorks, Inc., Natick, MA, 2014.
- [165] Philip R Johnson, Ning Sun, and Menachem Elimelech. Colloid transport in geochemically heterogeneous porous media: Modeling and measurements. *Environmental Science & Technology*, 30(11):3284–3293, 1996.
- [166] Kenneth P Burnham and David R Anderson. *Model selection and multi-model inference: a practical information-theoretic approach*. Springer, 2002.
- [167] Richard Royall. *Statistical evidence: a likelihood paradigm*. Chapman and Hall/CRC press, New York, June 1997.
- [168] Reeti Doshi, Washington Braidá, Christos Christodoulatos, Mahmoud Wazne, and Gregory O'Connor. Nano-aluminum: transport through sand columns and environmental effects on plants and soil communities. *Environmental Research*, 106(3):296–303, 2008.
- [169] Andrew J Pelley and Nathalie Tufenkji. Effect of particle size and natural organic matter on the migration of nano-and microscale latex particles in saturated porous media. *Journal of Colloid and Interface Science*, 321(1):74–83, 2008.
- [170] Navid Saleh, Kevin Sirk, Yueqiang Liu, Tanapon Phenrat, Bruno Dufour, Krzysztof Matyjaszewski, Robert D Tilton, and Gregory V Lowry. Surface modifications enhance nanoiron transport and napl targeting in saturated porous media. *Environmental Engineering Science*, 24(1):45–57, 2007.
- [171] Deb P Jaisi, Navid B Saleh, Ruth E Blake, and Menachem Elimelech. Transport of single-walled carbon nanotubes in porous media: filtration mechanisms and reversibility. *Environmental Science & Technology*, 42(22):8317–8323, 2008.
- [172] Jing Fang, Xiao-quan Shan, Bei Wen, Jin-ming Lin, and Gary Owens. Stability of titania nanoparticles in soil suspensions and transport in saturated homogeneous soil columns. *Environmental Pollution*, 157(4):1101–1109, April 2009.

- [173] Xueying Liu, Denis M O'Carroll, Elijah J Petersen, Qingguo Huang, and C Lindsay Anderson. Mobility of multiwalled carbon nanotubes in porous media. *Environmental Science & Technology*, 43(21):8153–8158, 2009.
- [174] Deb P Jaisi and Menachem Elimelech. Single-walled carbon nanotubes exhibit limited transport in soil columns. *Environmental Science & Technology*, 43(24):9161–9166, 2009.
- [175] Thomas K Darlington, Arianne M Neigh, Matthew T Spencer, Oanh TN Guyen, and Steven J Oldenburg. Nanoparticle characteristics affecting environmental fate and transport through soil. *Environmental Toxicology and Chemistry*, 28(6):1191–1199, 2009.
- [176] Marcos Paradelo, Paula Pérez-Rodríguez, Manuel Arias-Estévez, and J Eugenio López-Periago. Effect of particle size on copper oxychloride transport through saturated sand columns. *Journal of Agricultural and Food Chemistry*, 58(11):6870–6875, 2010.
- [177] Saeed Torkzaban, Yongman Kim, Martin Mulvihill, Jiamin Wan, and Tetsu K Tokunaga. Transport and deposition of functionalized cdte nanoparticles in saturated porous media. *Journal of Contaminant Hydrology*, 118(3):208–217, 2010.
- [178] Trishikhi Raychoudhury, Ghinwa Naja, and Subhasis Ghoshal. Assessment of transport of two polyelectrolyte-stabilized zero-valent iron nanoparticles in porous media. *Journal of Contaminant Hydrology*, 118(3):143–151, 2010.
- [179] Itzel G Godinez and Christophe J G Darnault. Aggregation and transport of nano-TiO₂ in saturated porous media: Effects of pH, surfactants and flow velocity. *Water Research*, 45(2):839–851, January 2011.
- [180] Jing Fang, Xiao-quan Shan, Bei Wen, Jin-ming Lin, Gary Owens, and Shuai-ren Zhou. Transport of copper as affected by titania nanoparticles in soil columns. *Environmental Pollution*, 159(5):1248–1256, 2011.
- [181] Zhen Li, Endalkachew Sahle-Demessie, Ashraf Aly Hassan, and George A Sorial. Transport and deposition of ceo₂ nanoparticles in water-saturated porous media. *Water Research*, 45(15):4409–4418, 2011.

- [182] Nikolai T Mattison, Denis M O'Carroll, R Kerry Rowe, and Elijah J Petersen. Impact of porous media grain size on the transport of multi-walled carbon nanotubes. *Environmental Science & Technology*, 45(22):9765–9775, 2011.
- [183] Yuan Tian, Bin Gao, and Kirk J Ziegler. High mobility of sdbcs-dispersed single-walled carbon nanotubes in saturated and unsaturated porous media. *Journal of Hazardous Materials*, 186(2):1766–1772, 2011.
- [184] Adamo R Petosa, Spencer J Brennan, Faraz Rajput, and Nathalie Tufenkji. Transport of two metal oxide nanoparticles in saturated granular porous media: Role of water chemistry and particle coating. *Water Research*, 46(4):1273–1285, 2012.
- [185] Omer Sagee, Ishai Dror, and Brian Berkowitz. Transport of silver nanoparticles (AgNPs) in soil. *Chemosphere*, 88(5):670–675, 2012.
- [186] Dermont Bouchard, Wei Zhang, Tremaine Powell, and Usa Rattanaudompol. Aggregation kinetics and transport of single-walled carbon nanotubes at low surfactant concentrations. *Environmental Science & Technology*, 46(8):4458–4465, 2012.
- [187] Yuan Tian, Bin Gao, Yu Wang, Verónica L Morales, Rafael Munoz Carpena, Qingguo Huang, and Liuyan Yang. Deposition and transport of functionalized carbon nanotubes in water-saturated sand columns. *Journal of Hazardous Materials*, 213:265–272, 2012.
- [188] Dengjun Wang, Scott A Bradford, Ronald W Harvey, Bin Gao, Long Cang, and Dongmei Zhou. Humic acid facilitates the transport of ars-labeled hydroxyapatite nanoparticles in iron oxyhydroxide-coated sand. *Environmental Science & Technology*, 46(5):2738–2745, 2012.
- [189] Amir Taghavy, Anjuliee Mittelman, Yonggang Wang, Kurt D Pennell, and Linda M Abriola. Mathematical modeling of the transport and dissolution of citrate-stabilized silver nanoparticles in porous media. *Environmental Science & Technology*, 47(15):8499–8507, 2013.
- [190] Tanzina Rahman, Jessica George, and Heather J Shipley. Transport of aluminum oxide nanoparticles in saturated sand: Effects of ionic strength, flow rate, and nanoparticle concentration. *Science of the Total Environment*, 463:565–571, 2013.

- [191] Itzel G Godinez, Christophe JG Darnault, Amid P Khodadoust, and Dorin Bogdan. Deposition and release kinetics of nano-TiO₂ in saturated porous media: Effects of solution ionic strength and surfactants. *Environmental Pollution*, 174:106–113, 2013.
- [192] Jing Fang, Mei-jia Xu, Deng-jun Wang, Bei Wen, and Jing-yi Han. Modeling the transport of tio₂ nanoparticle aggregates in saturated and unsaturated granular media: Effects of ionic strength and ph. *Water Research*, 47(3):1399–1408, 2013.
- [193] Lin Liu, Bin Gao, Lei Wu, Verónica L Morales, Liuyan Yang, Zuhao Zhou, and Hao Wang. Deposition and transport of graphene oxide in saturated and unsaturated porous media. *Chemical Engineering Journal*, 229:444–449, 2013.
- [194] DM O'Carroll, X Liu, NT Mattison, and EJ Petersen. Impact of diameter on carbon nanotube transport in sand. *Journal of Colloid And Interface Science*, 390(1):96–104, 2013.
- [195] Daniela Kasel, Scott A Bradford, Jiří Šimůnek, Thomas Pütz, Harry Vereecken, and Erwin Klumpp. Limited transport of functionalized multi-walled carbon nanotubes in two natural soils. *Environmental Pollution*, 180:152–158, 2013.
- [196] Diefeng Gu, Sinan Yalcin, Helmut Baumgart, Shizhi Qian, Oktay Baysal, and Ali Beskok. Electrophoretic light scattering for surface zeta potential measurement of ald metal oxide films. *ECS Transactions*, 33(2):37–41, 2010.
- [197] Sharon L Walker, Jeremy A Redman, and Menachem Elimelech. Role of cell surface lipopolysaccharides in escherichia c oli k12 adhesion and transport. *Langmuir*, 20(18):7736–7746, 2004.
- [198] Menachem Elimelech, John Gregory, and Xiadong Jia. *Particle deposition and aggregation: measurement, modelling and simulation*. Butterworth-Heinemann, 2013.
- [199] CG Malmberg and AA Maryott. Dielectric constant of water from 0 0 to 1000 0 c. *Journal of Research of the National Bureau of Standards*, 56:1–8, 1956.

7

APPENDIX I: SUPPORTING INFORMATION FOR CHAPTER 2

Contents

7.1	Supporting model information	147
7.1.1	Model 1	148
7.1.2	Model 2	149
7.1.3	Model 3	150
7.1.4	Model 4	151
7.1.5	Model 5	151
7.1.6	Model 6	152
7.1.7	Model 7	152
7.2	Supporting methods information	152
7.2.1	Programming details	152
7.2.2	Bias-adjusted Akaike Information Criteria	154
7.2.3	Analysis of experimental data availability for ENP column experiments	156
7.3	Supporting simulation results	168
7.3.1	Supplementary archetypal simulations	168
7.3.2	Results from the parameter fitting for the archetypal retention profiles	169
7.3.3	Hyperexponentiality influence on model plausibility	174

7.1 SUPPORTING MODEL INFORMATION

The models are introduced in order of complexity. However, as M1 and M2 are simplifications of M3, where M4–M7 add complexity to M3, M1 and M2 are presented first.

7.1.1 Model 1

A colloid filtration theory (CFT) model is employed to simulate ENP transport.^{20,61} For convenience, CFT is often presented as an analytical solution and it may not be explicitly clear to some readers how, and where, simplifications are applied. To add clarity to the assumptions and justifications employed in the development of CFT, a brief review of the method is provided.

For steady uniform flow vertically through an isotropic homogeneous porous medium, and without a concentration gradient or advection perpendicular to the primary flow direction, the aqueous mass balance equation for a dispersing substance is:

$$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t} \quad (5)$$

where C is the fluid phase particle concentration, t is time, x is distance from the column inlet, D_p is the longitudinal dispersion coefficient, v_p is the average pore water velocity, θ is the porosity, and ρ_b is the media bulk density.^{60,61} The solid-phase mass balance is:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C - \frac{\rho_b}{\theta} k_r S \quad (6)$$

where S is the solid-phase particle concentration, k_d is the first-order deposition rate constant, and k_r is the first-order detachment rate constant.^{61,62}

To arrive at the analytical solutions presented for CFT, several simplifying assumptions must be applied to Equations 5 and 6. First, it must be assumed that the influence of dispersion is negligible (i.e., $D_p = 0$) in the aqueous mass balance presented in Equation 5. Second, in the solid-phase mass balance (Equation 6), it must be assumed that mass transfer irreversibly proceeds from the aqueous phase to the solid-phase (i.e., $k_r = 0$). Applying these assumptions to Equation 5 and 6 allows analytical solutions to be determined for the aqueous and solid phases by the method of characteristic, linear, first order differential equations.¹⁵⁸ The aqueous and solid-phase mass analytical solutions are:

$$C(x) = C_{\text{inf}} \cdot \exp\left(\frac{-k_d}{v_p} x\right) \quad (7)$$

$$S(x, t) = \frac{t\theta k_d C_{\text{inf}}}{\rho_b} \cdot \exp\left(\frac{-k_d}{v_p} x\right) \quad (8)$$

where $C(x)$ is aqueous concentration at depth, C_{inf} is the influent ENP concentration, $S(x, t)$ is the solid-phase particle concentration at depth and time, and t is length of time ENP were pumped through the column.^{20,61}

M1 describes the deposition of particles using a spatially and temporally invariant first-order deposition rate constant, k_d , which is empirically calibrated by physical attributes of the system as follows^{20,61}:

$$k_d = \frac{3(1 - \theta)}{2d_c} \alpha v_p \eta_o \quad (9)$$

where d_c is the average diameter of the granular media, α is the particle-collector attachment efficiency, and η_o is the single-collector contact efficiency. The Tufenkji and Elimelech⁶¹ correlation equation is employed to predict the η_o as follows:

$$\eta_o = 2.4A_S^{1/3} N_R^{-0.081} N_{Pe}^{-0.715} N_{VdW}^{0.052} + 0.55A_S N_R^{1.675} N_A^{0.125} + 0.22N_R^{-0.24} N_G^{1.11} N_{VdW}^{0.053} \quad (10)$$

where A_S is the dimensionless Happel correction factor, N_R is the ratio of collector (i.e., grain) to particle diameter, N_{Pe} is the dimensionless Péclet number, N_{VdW} is the ratio of Van der Waals interaction energy to particle thermal energy, N_A is the dimensionless influence factor for Van der Waals attraction forces and fluid velocity on particle deposition rate from interception, and N_G is the dimensionless gravitation number.

7.1.2 Model 2

A modification to CFT that assumed a bimodal distribution of attachment efficiencies is employed to simulate ENP transport.^{50,55,63} Justification for treatment of particle populations in this manner considers that silicate surfaces in water may acquire charge heterogeneities through surface hydroxylation and subsequent ionization of the hydroxyl group.^{159,160} Hydroxyl groups with different dissociation tendencies can contribute to a granular surface with two SiO - H sites, justifying the existence of a bi-modal distribution of attachment efficiencies, and thus, deposition rates for favorable and unfavorable deposition conditions.¹⁶¹ Following this logic, slowly depositing particles are assumed to irreversibly attach in the primary energy minimum and rapidly depositing particles are assumed to

reversibly attach in the secondary energy minima, although particle detachment is not mathematically considered in the model.

The mass balance equation describing the aqueous and solid-phase concentrations are as follows:

$$\frac{\partial C}{\partial t} = -v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t} \quad (11)$$

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = f k_{d,f} C + (1-f) k_{d,s} C \quad (12)$$

where f is the fraction of the particle population that is associated with the ‘fast’ deposition rate constant, $k_{d,f}$, and $k_{d,s}$ is the ‘slow’ deposition rate constant. Similarly to M1, M2 assumes negligible release and dispersion, and is often presented in terms of its analytical solution⁵⁵:

$$C(x) = C_{\text{inf}} \left(f \cdot \exp \left[\frac{-k_{d,f} x}{v_p} \right] + (1-f) \cdot \exp \left[\frac{-k_{d,s} x}{v_p} \right] \right) \quad (13)$$

$$S(x, t) = \frac{t \theta k_d C_{\text{inf}}}{\rho_b} \left(k_{d,f} \cdot f \cdot \exp \left[\frac{-k_{d,f} x}{v_p} \right] + k_{d,s} \cdot (1-f) \cdot \exp \left[\frac{-k_{d,s} x}{v_p} \right] \right) \quad (14)$$

$k_{d,f}$ is determined by setting α_{fast} equal to unity and calculating the deposition rate using Equation 9. f and α_{slow} are fit assuming η_0 (Equation 10).¹⁶²

The attachment efficiency is calculated as the relative relationship between the maximum transport-limited deposition rate, $k_{d,f}$, and the fitted deposition rate constant, $k_{d,s}$.¹³⁸:

$$\alpha_s = \frac{k_{d,s}}{k_{d,f}} \alpha_f = 1 \quad (15)$$

As $k_{d,f}$ is rate limited, and thus, the theoretical kinetic maximum, this implies that α_f in M2 is 1.

7.1.3 Model 3

A single-site deposition-detachment model is employed to simulate ENP transport.⁶² The mass balance equation describing the aqueous concentration is:

$$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho_b}{\theta} \frac{\partial S}{\partial t} \quad (16)$$

The solid-phase concentration mass balance is:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C - \frac{\rho_b}{\theta} k_r S \quad (17)$$

where k_r is the kinetic remobilization (i.e., detachment) rate constant.⁶² k_d and k_r are fit.

7.1.4 Model 4

A dual-site deposition-detachment model is employed to simulate ENP transport.^{53,73} The mass balance equation describing the aqueous concentration is:

$$\frac{\partial C}{\partial t} = D_p \frac{\partial^2 C}{\partial x^2} - v_p \frac{\partial C}{\partial x} - \frac{\rho}{\theta} \frac{\partial S}{\partial t} - \frac{\rho}{\theta} \frac{\partial S_2}{\partial t} \quad (18)$$

where S_2 is the solid phase concentration at the second deposition site.⁵³ The solid-phase mass balances for each retention site are:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d C - \frac{\rho}{\theta} k_r S \quad (19)$$

$$\frac{\rho_b}{\theta} \frac{\partial S_2}{\partial t} = k_{d,2} C - \frac{\rho}{\theta} k_{r,2} S_2 \quad (20)$$

where $k_{d,2}$ and $k_{r,2}$ are the deposition and detachment rate constants at the second site, respectively.⁵³ k_d , $k_{d,2}$, k_r , and $k_{r,2}$ are fit.

7.1.5 Model 5

A single-site deposition-detachment model that include dynamic Langmuirian blocking is employed to simulate ENP transport.^{33,52} The mass balance equation describing the aqueous concentration is equivalent to Equation 16. The solid-phase concentration mass balance is:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d \psi_b C - \frac{\rho}{\theta} k_r S \quad (21)$$

where ψ_b is a dynamic Langmuirian blocking function that exhibits control of the deposition as a function of the available fraction of retention sites.^{33,52} The equation describing ψ_b is:

$$\psi_b = \left(1 - \frac{S}{S_m} \right) \quad (22)$$

where S_m is the maximum solid-phase ENP concentration.^{33,52} k_d , k_r , and S_m are fit.

7.1.6 Model 6

A single-site deposition-detachment model that includes depth-dependent retention is employed to simulate ENP transport.^{48,49,54} The mass balance equation describing the aqueous concentration is equivalent to Equation 16. The solid-phase concentration mass balance is:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d \psi_s C - \frac{\rho}{\theta} k_r S \quad (23)$$

where ψ_s is a depth-dependent function that exhibits control of the deposition as a function of a dimensionless empirical shape parameter, β .^{48,54} The equation describing ψ_s is:

$$\psi_s = \left(\frac{d_c + x}{d_c} \right)^{-\beta} \quad (24)$$

k_d , k_r , and β ¹ are fit.

7.1.7 Model 7

A single-site deposition-detachment model that includes Langmuirian blocking and depth-dependent retention is employed to simulate ENP transport.^{33,48,49,54} The mass balance equation describing the aqueous concentration is equivalent to Equation 16. The solid-phase concentration mass balance is:

$$\frac{\rho_b}{\theta} \frac{\partial S}{\partial t} = k_d \psi_b \psi_s C - \frac{\rho}{\theta} k_r S \quad (25)$$

The equation describing ψ_b is equivalent to Equation 22; ψ_s is equivalent to Equation 24. k_d , k_r , S_m , and β are fit.

7.2 SUPPORTING METHODS INFORMATION

7.2.1 Programming details

The advection-dispersion problems are formulated as partial differential equations (PDEs). MATLAB has shown to give high

¹ Determinations of β vary widely: Bradford et al.⁴⁸ estimated $\beta=0.432$ for latex microsphere transport through sand; (2013) estimate $\beta=0.76$ for MWCNT, (2012) estimated $\beta=0.8$ for ARS-labeled hydroxyapatite in the presence of humic acid, and Choy et al.²⁶ estimated $\beta=0.021$

accuracy of numerical solutions for initial-boundary value problems consisting of systems of parabolic and elliptic PDEs in one space variable and time using the 'pdepe' solver function.¹⁶³ As such, the MATLAB and Statistics Toolbox Release 2014a is employed for the transport simulations.¹⁶⁴

The 'pdepe' solver function in MATLAB converts the PDEs to ordinary differential equations (ODEs) using a second-order accurate spatial discretization. Using this approach, transport functions are allowed to be space-dependent and boundary conditions to be time-dependent. Using the 'pdepe' solver function in MATLAB, various boundary conditions also can be formulated flexibly, either as Dirichlet, Neumann, or Cauchy/Robin. Assumptions and boundary conditions were formulated as follows:

1. 1D transport is simulated under saturated, steady state flow conditions.
2. Constant interstitial fluid flow velocity (i.e., pore water velocity) and constant fluid solution conditions (e.g., ionic strength, pH, water temperature) are present throughout the simulation.
3. ENP transport is approximated as monodisperse spherical particles traveling through a monodisperse spherical porous medium with constant porosity (i.e., no aggregation).
4. Kinetics are approximated by a pseudo first-order kinetic rate law, where deposited particles do not increase particle deposition or contribute to head loss, and particles are not inactivated or degraded.⁶²
5. Deposited ENPs are immobile (i.e., form a stationary solid-phase).⁶²
6. Dirichlet and Neumann boundary conditions are applied to the upper and lower aqueous-phase concentration boundaries, respectively. The Dirichlet boundary condition defines the advective flux and concentration across the boundary (i.e., $C_{\text{inf}} = C_{\text{in}}$, and $v_p = v_{\text{in}}$). The Neumann boundary condition does not allow diffusive flux change across the boundary (i.e., $\partial C/\partial x = 0$), but does allow advective flux ($v_{\text{in}} = v_{\text{out}}$). Neumann boundary conditions are applied to the upper and lower solid-phase concentration boundaries. Solid-phase diffusive flux is prevented by virtue of the boundary condition (i.e., $\partial S/\partial x = 0$), advective flux is prevented by definition (i.e., $\partial S/\partial t = 0$).

7. For each modeled system, the longitudinal dispersivity, α_L , is used to determine the longitudinal dispersion coefficient and is:

$$D_p = \alpha_L v_p + D_\infty \quad (26)$$

where D_p is assumed to be a function of longitudinal dispersivity, α_L , the particle velocity, v_p , and the bulk molecular diffusion coefficient, D_∞ , derived from the Stokes-Einstein equation for spherical particles.¹⁶⁵

7.2.2 Bias-adjusted Akaike Information Criteria

For the special case of a least-squares estimation with normally distributed errors and a ratio of sample size to the number of parameters of less than 40, the AIC_c is¹⁶⁶:

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1} \quad (27)$$

where AIC is defined as:

$$AIC = n \ln(\sigma^2) + 2K \quad (28)$$

where n is the number of observations and K is the number of free parameters in the model including a parameter for σ^2 .⁷² The residual sum of squares (RSS), or σ^2 , is:

$$\sigma^2 = \frac{\sum_{i=1}^n (\epsilon_i)^2}{n} \quad (29)$$

where ϵ_i is the difference between the observed data and the model estimate for each data point.⁷²

Note that AIC_c values are specific for a given data set. Comparison between data sets holds no quantitatively comparable information.⁷²

For each dataset, the most plausible model is calculated as the model with the highest Akaike weight, w_i . The w_i value of a given model, i , for a given dataset indicates the relative quantitative plausibility of model i to explain the retention profile data and is calculated as follows:

$$w_i = \frac{\exp\left(\frac{-\Delta_i}{2}\right)}{\sum_{p=1}^7 \exp\left(\frac{-\Delta_p}{2}\right)} \quad (30)$$

where Δ_i is the minimum relative distance of model i to the minimum AIC_c value (i.e., $AIC_{c,i} - AIC_{c,\min}$), and Δ_p is the minimum relative distance of model p to the minimum AIC_c value (i.e., $AIC_{c,p} - AIC_{c,\min}$).⁷² The numerator represents the relative likelihood of model i . The denominator represents the sum of relative likelihoods for all the models, where p serves as an additional index to track each model.

In accordance with the plausibility cutoff threshold suggested by Royall¹⁶⁷, models with w_i less than 10% of the maximum w_i are considered insufficient to explain the data. Note that the Akaike weight gives an indication about the likelihood of model plausibility in relation to the other examined models for a given dataset, but cannot be used to explicitly indicate that one model is superior for all datasets.¹⁶⁶ Additionally, model plausibility results only indicate relative plausibility and should be disregarded in case of sufficient visual evidence of poor model performance.

7.2.3 Analysis of experimental data availability for ENP column experiments

Table 9: Publications on ENP transport in porous media (2008–2014)

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Application of an empirical transport model to simulate retention of nanocrystalline titanium dioxide in sand columns	Choy et al. ²⁶	TiO ₂	X	X	X	X	X	X	X	DV	Yes; No	V _{im} (80 PVs) violates M1 and M2
Nano-aluminum: Transport through sand columns and environmental effects on plants and soil communities	Doshi et al. ¹⁶⁸	Al ₂ O ₃	X	X	X	X	-	-	X	-	No; No	-
Effect of particle size and natural organic matter on the migration of nano- and microscale latex particles in saturated porous media	Pelley and Tufenkji ¹⁶⁹	Latex	X	X	X	X	-	-	X	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text

Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I ; Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP; Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Transport and Retention of Nanoscale C_{60} Aggregates in Water-Saturated Porous Media	Wang et al. ³⁹	C_{60}	X	X	X	X	X	X	X	X	Yes; Yes	
Ionic Strength and Composition Affect the Mobility of Surface-Modified Feo Nanoparticles in Water-Saturated Sand Columns	Saleh et al. ¹⁷⁰	NZVI	X	X	X	X	X	-	X	X	No; No	-
Investigation of the Transport and Deposition of Fullerene (C_{60}) Nanoparticles in Quartz Sands under Varying Flow Conditions	Li et al. ⁶⁶	C_{60}	X	X	X	X	-	X	X	-	No; No	
Transport of single-walled carbon nanotubes in porous media: filtration mechanisms and reversibility	Jaisi et al. ¹⁷¹	SWCNT	X	X	X	X	-	-	X	-	No; No	-
Stability of titania nanoparticles in soil suspensions and transport in saturated homogeneous soil columns	Fang et al. ¹⁷²	TiO_2	X	X	X	X	-	-	X	DV	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text
 Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
 Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
 Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Mobility of multiwalled carbon nanotubes in porous media	Liu et al. ¹⁷³	MWCNT	X	X	X	X	-	-	X	-	No; No	-
Single-walled carbon nanotubes exhibit limited transport in soil columns	Jaisi and Elimelech ¹⁷⁴	SWCNT	X	X	X	X	-	-	X	-	No; No	-
Nanoparticle characteristics affecting environmental fate and transport through soil	Darlington et al. ¹⁷⁵	Al ₂ O ₃	X	X	X	X	-	-	X	-	No; No	-
Transport of metal oxide nanoparticles in saturated porous media	Ben-Moshe et al. ¹⁰⁰	Fe ₃ O ₄ ; CuO; TiO ₂ ZnO	X	X	X	X	-	-	X	X	No; No	-
Effect of Particle Size on Copper Oxide Transport through Saturated Sand Columns	Paradelo et al. ¹⁷⁶	Cu ₂ (OH) ₃ Cl	X	X	X	X	X	-	X	X	No; No	-
Estimating Attachment of Nano- and Submicrometer-particles Coated with Organic Macromolecules in Porous Media: Development of an Empirical Model	Phenrat et al. ⁶⁸	Polymer-coated Hematite and TiO ₂	X	X	X	X	-	-	-	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text
 Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
 Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
 Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Transport and deposition of functionalized CdTe nanoparticles in saturated porous media	Torkzaban et al. ¹⁷⁷	CdTe	X	X	X	X	X	-	X	-	No; No	-
Influence of biofilm on the transport of fullerene (C ₆₀) nanoparticles in porous media	Tong et al. ¹⁴⁶	C ₆₀	X	X	X	X	-	-	X	-	No; No	-
Assessment of transport of two polyelectrolyte-stabilized zero-valent iron nanoparticles in porous media	Raychoudhury et al. ¹⁷⁸	NZVI	X	X	X	X	X	-	X	X	No; No	-
Transport and Retention of TiO ₂ Rutile Nanoparticles in Saturated Porous Media under Low-Ionic-Strength Conditions: Measurements and Mechanisms	Chen et al. ²⁴	TiO ₂	X	X	X	X	X	X	X	X	Yes; Yes	-
Aggregation and transport of nano-TiO ₂ in saturated porous media: Effects of pH, surfactants and flow velocity	Godinez and Darnault ¹⁷⁹	TiO ₂	X	X	X	X	OT	-	X	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text
 Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
 Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
 Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Mechanisms of TiO ₂ nanoparticle transport in porous media: Role of solution chemistry, nanoparticle concentration, and flowrate	Chowdhury et al. ²⁵	TiO ₂	X	X	X	X	-	X	X	X	No; No	-
Transport of copper as affected by titania nanoparticles in soil columns	Fang et al. ¹⁸⁰	Cu-TiO ₂	X	X	X	X	X	-	X	-	No; No	-
Transport and deposition of CeO ₂ nanoparticles in water-saturated porous media	Li et al. ¹⁸¹	CeO ₂	X	X	X	X	-	-	X	X	No; No	-
Impact of porous media grain size on the transport of multi-walled carbon nanotubes	Mattison et al. ¹⁸²	MWCNT	X	X	X	X	-	-	X	-	No; No	-
High mobility of SDBS-dispersed single-walled carbon nanotubes in saturated and unsaturated porous media	Tian et al. ¹⁸³	SWCNT	X	X	X	X	X	-	X	X	No; No	-
Facilitated transport of Cu with hydroxyapatite nanoparticles in saturated sand: Effects of solution ionic strength and composition	Wang et al. ⁴⁰	CuO-nHAP	X	X	X	X	-	X	X	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text

Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Transport of two metal oxide nanoparticles in saturated granular porous media: Role of water chemistry and particle coating	Petosa et al. ¹⁸⁴	TiO ₂ ; ZnO	X	X	X	X	-	-	X	-	No; No	-
Transport of silver nanoparticles (AgNPs) in soil	Sagee et al. ¹⁸⁵	Ag	X	X	X	X	-	-	X	-	No; No	-
Fate and transport of elemental copper (Cu ⁰) nanoparticles through saturated porous media in the presence of organic materials	Jones and Su ³¹	Cu ⁰	X	X	X	X	-	X	X	X	No; No	-
Cotransport of Titanium Dioxide and Fullerene Nanoparticles in Saturated Porous Media	Cai et al. ²³	TiO ₂ ; C ₆₀	X	X	X	X	-	X	X	-	No; No	-
Transport and deposition of ZnO nanoparticles in saturated porous media	Jiang et al. ²⁸	ZnO	X	X	X	X	-	X	X	X	No; No	-
Influence of natural organic matter on the transport and deposition of zinc oxide nanoparticles in saturated porous media	Jiang et al. ²⁹	ZnO	X	X	X	X	-	X	X	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text
 Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
 Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
 Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Aggregation Kinetics and Transport of Single-Walled Carbon Nanotubes at Low Surfactant Concentrations	Bouchard et al. ¹⁸⁶	SWCNT	X	X	X	X	X	-	X	-	No; No	-
Transport of Ferrihydrite Nanoparticles in Saturated Porous Media: Role of Ionic Strength and Flow Rate	Tosco et al. ³⁷	Ferrihydrite	X	X	X	X	X	X	X	X	Yes; No	< 8 RP data points
Influence of Collector Surface Composition and Water Chemistry on the Deposition of Cerium Dioxide Nanoparticles: QCM-D and Column Experiment Approaches	Liu et al. ³⁶	CeO ₂	X	X	X	X	-	X	X	X	No; No	-
Transport of Fullerene Nanoparticles (nC ₆₀) in Saturated Sand and Sandy Soil: Controlling Factors and Modeling	Zhang et al. ⁶⁴	C ₆₀	X	X	X	X	-	-	X	-	No; No	-
Deposition and transport of functionalized carbon nanotubes in water-saturated sand columns	Tian et al. ¹⁸⁷	SWCNT	X	X	X	X	-	-	X	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text

Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I ; Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP; Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Transport behavior of functionalized multi-wall carbon nanotubes in water-saturated quartz sand as a function of tube length	Wang et al. ⁷⁶	MWCNT	X	X	X	X	X	X	X	DV	Yes; No	θ not given
Humic Acid Facilitates the Transport of ARS-Labeled Hydroxyapatite Nanoparticles in Iron Oxyhydroxide-Coated Sand	Wang et al. ¹⁸⁸	ARS-Labeled Hydroxyapatite	X	X	X	X		X	X	DV	No; No	
Mathematical Modeling of the Transport and Dissolution of Citrate-Stabilized Silver Nanoparticles in Porous Media	Taghavy et al. ¹⁸⁹	Ag-Citrate	X	X	X	X	-	-	-	-	No; No	-
Initial transport and retention behaviors of ZnO nanoparticles in quartz sand porous media coated with Escherichia coli biofilm	Jiang et al. ³⁰	ZnO	X	X	X	X		X	X	-	No; No	
Sensitivity of the transport and retention of stabilized silver nanoparticles to physicochemical factors	Liang et al. ³⁵	Ag	X	X	X	X	X	X	X	DV	Yes; Yes	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text
 Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
 Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
 Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Retention and Remobilization of Stabilized Silver Nanoparticles in an Undisturbed Loamy Sand Soil	Liang et al. ⁶⁵	Ag	X	X	X	X	X	X	X	DV	Yes; No	Below 100% saturation
Transport of Biochar Particles in Saturated Granular Media: Effects of Pyrolysis Temperature and Particle Size	Wang et al. ⁴¹	Biochar	X	X	X	X	OA	X	X	X	Yes; Yes	-
Transport of aluminum oxide nanoparticles in saturated sand: Effects of ionic strength, flow rate, and nanoparticle concentration	Rahman et al. ¹⁹⁰	Al ₂ O ₃	X	X	X	X	OT	-	X	X	No; No	-
Deposition and release kinetics of nano-TiO ₂ in saturated porous media: Effects of solution ionic strength and surfactants	Godinez et al. ¹⁹¹	TiO ₂	X	X	X	X	OT	-	X	X	No; No	-
Modeling the transport of TiO ₂ nanoparticle aggregates in saturated and unsaturated granular media: Effects of ionic strength and pH	Fang et al. ¹⁹²	TiO ₂	X	X	X	X	OT	-	X	X	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text

Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I ; Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP; Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Deposition and transport of graphene oxide in saturated and unsaturated porous media	Liu et al. ¹⁹³	Graphene	X	X	-	X	-	-	X	X	No; No	-
Impact of diameter on carbon nanotube transport in sand	O'Carroll et al. ¹⁹⁴	MWCNT	X	X	X	X	X	-	X	-	No; No	-
Transport and retention of multi-walled carbon nanotubes in saturated porous media: Effects of input concentration and grain size	Kasel et al. ³³	MWCNT	X	X	X	X	X	X	X	DV	Yes; Yes	-
Limited transport of functionalized multi-walled carbon nanotubes in two natural soils	Kasel et al. ¹⁹⁵	MWCNT	X	X	X	X	X	X	X	DV	Yes; No	Below 100% saturation
Antagonistic Effects of Humic Acid and Iron Oxyhydroxide Grain-Coating on Biochar Nanoparticle Transport in Saturated Sand.	Wang et al. ⁴²	FeOOH-nBiochar	X	X	X	X	X	X	X	X	Yes; Yes	

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text

Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
Column Length: L_c ; Pore Water Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersivity: α_L ; Retention Profile: RP;
Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

continued ...

... continued

Publication Title	Author	ENP	d_p ρ_p C_{inf}	d_g θ ρ_b	pH I	v_p V_{im}	D_p α_L	RP	BP	η_o A	Sufficient Data? Modeled?	Problems?
Transport of bare and capped zinc oxide nanoparticles is dependent on porous medium composition	Kurlanda-Witek et al. ⁴⁵	ZnO	X	X	X	X	X	X	X	X	Yes; No	< 8 RP data points
Transport and retention of zinc oxide nanoparticles in porous media: Effects of natural organic matter versus natural organic ligands at circumneutral pH	Jones and Su ³²	ZnO	X	X	X	X	-	X	X	X	No; No	-
Transport and retention behaviors of titanium dioxide nanoparticles in iron oxide-coated quartz sand: Effects of pH, ionic strength, and humic acid	Han et al. ²⁷	TiO ₂	X	X	X	X	-	X	X	-	No; No	-
Comparison of three labeled silica nanoparticles used as tracers in transport experiments in porous media. Part II: Transport experiments and modeling	Vitorge et al. ³⁸	SiO ₂	X	X	X	X	-	X	X	-	No; No	-

X = All parameters are available; DV = Default value employed; OA = Obtained from author; OT = Obtained from text

Particle Diameter: d_p ; Particle Density: ρ_p ; Influent Concentration: C_{inf} ; Collector Diameter: d_c ; Porosity: θ ; Bulk Soil Density: ρ_b ; Ionic Strength: I
 Column Length: L_c ; Pore Velocity: v_p ; Impulse Volume: V_{im} ; Dispersion Coefficient: D_p ; Longitudinal Dispersion: α_L ; Retention Profile: RP;
 Breakthrough Profile: BP; Single-Collector Contact Efficiency, η_o ; Hamaker Constant: A

7.3 SUPPORTING SIMULATION RESULTS

7.3.1 Supplementary archetypal simulations

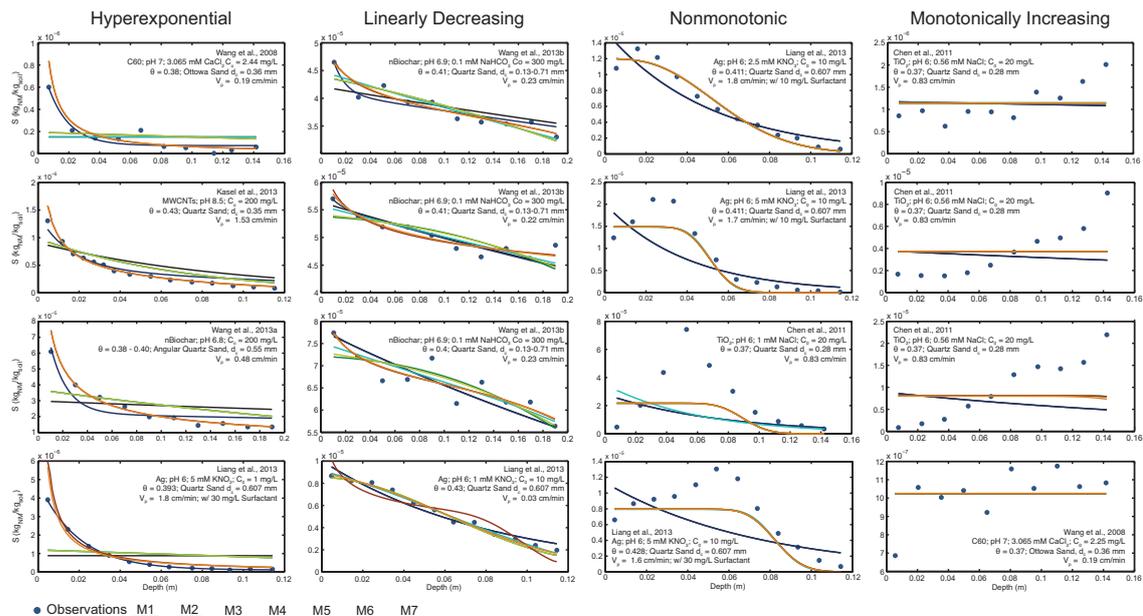


Figure 22: Supplementary Archetypal Retention Profiles and Model results.

7.3.2 *Results from the parameter fitting for the archetypal retention profiles*

The results of parameter fitting for the archetypal profiles in Figure 4 in the main text are available in Table 10. However, neither the results of the parameter estimation, nor the relationship of the parameters to the data presented in the original papers are discussed in detail here.

Table 10: Model parameter estimation results for the archetypal retention profiles Figure 4 in the main text

	M2: $k_{d,f}$			M2: $k_{d,s}$				M1: α
	M1, M3–M7: k_d	k_r	f	M4: $k_{d,2}$	$k_{r,2}$	β	S_m	M2: $k_{d,s}/k_{d,f}$
Mode	[s ⁻¹]	[s ⁻¹]	[-]	[s ⁻¹]	[s ⁻¹]	[-]	[kg kg ⁻¹]	[-]
Hyperexponential (HE)								
M1	$2.80 \cdot 10^{-2}$							$2.12 \cdot 10^{-2}$
M2	$2.80 \cdot 10^{-2}$		$2.78 \cdot 10^{-2}$	$4.40 \cdot 10^{-4}$				$1.57 \cdot 10^{-2}$
M3	$5.95 \cdot 10^{-4}$	0.00						
M4	$1.75 \cdot 10^{-3}$	0.00		$9.54 \cdot 10^{-4}$	0.00			
M5	$5.95 \cdot 10^{-4}$	0.00					$3.06 \cdot 10^{-2}$	
M6	$9.69 \cdot 10^{-1}$	$3.20 \cdot 10^{-1}$				$4.03 \cdot 10^{-1}$		
M7	$3.10 \cdot 10^{-2}$	$1.36 \cdot 10^{-3}$				$7.37 \cdot 10^{-1}$	$5.43 \cdot 10^{-7}$	
Linearly decreasing (LD)								
M1	$1.53 \cdot 10^{-2}$							$7.38 \cdot 10^{-3}$
M2	$1.53 \cdot 10^{-2}$		0.00	$1.13 \cdot 10^{-4}$				$7.38 \cdot 10^{-3}$
M3	$2.08 \cdot 10^{-4}$	$4.95 \cdot 10^{-5}$						
M4	$3.98 \cdot 10^{-4}$	$1.17 \cdot 10^{-4}$		$1.36 \cdot 10^{-3}$	$6.36 \cdot 10^{-3}$			
M5	$2.07 \cdot 10^{-4}$	$4.90 \cdot 10^{-5}$					$4.74 \cdot 10^{-3}$	
M6	$4.79 \cdot 10^{-4}$	$1.26 \cdot 10^{-4}$				$6.86 \cdot 10^{-2}$		
M7	$3.05 \cdot 10^{-3}$	0.00				$6.02 \cdot 10^{-1}$	$8.42 \cdot 10^{-6}$	

Values less than $1 \cdot 10^{-13}$ are reported as '0.00'.

Table 11: Model parameter estimation results for the archetypal retention profiles Figure 4 in the main text

	M2: $k_{d,f}$			M2: $k_{d,s}$				M1: α
	M1, M3–M7: k_d	k_r	f	M4: $k_{d,2}$	$k_{r,2}$	β	S_m	M2: $k_{d,s}/k_{d,f}$
	Model[s ⁻¹]	[s ⁻¹]	[-]	[s ⁻¹]	[s ⁻¹]	[-]	[kg kg ⁻¹]	[-]
Nonmonotonic (NM)								
M1	$1.53 \cdot 10^{-2}$							$5.72 \cdot 10^{-1}$
M2	$1.53 \cdot 10^{-2}$			$8.75 \cdot 10^{-3}$				$5.72 \cdot 10^{-1}$
M3	$3.56 \cdot 10^{-1}$	$5.05 \cdot 10^{-2}$						
M4	$3.56 \cdot 10^{-1}$	$5.05 \cdot 10^{-2}$		0.00	0.00			
M5	$3.56 \cdot 10^{-1}$	$5.05 \cdot 10^{-2}$					$1.91 \cdot 10^{-1}$	
M6	$3.94 \cdot 10^{-1}$	$5.60 \cdot 10^{-2}$						
M7	$3.94 \cdot 10^{-1}$	$5.60 \cdot 10^{-2}$					$9.84 \cdot 10^{-1}$	
Monotonically increasing (MI)								
M1	$3.73 \cdot 10^{-2}$							$1.12 \cdot 10^{-2}$
M2	$3.73 \cdot 10^{-2}$		0.00	$4.17 \cdot 10^{-4}$				$1.12 \cdot 10^{-2}$
M3	$9.99 \cdot 10^{-1}$	$6.80 \cdot 10^{-1}$						
M4	$3.87 \cdot 10^{-2}$	$2.74 \cdot 10^{-2}$		$8.38 \cdot 10^{-5}$	$1.48 \cdot 10^{-1}$			
M5	$9.99 \cdot 10^{-1}$	$6.80 \cdot 10^{-1}$					$5.08 \cdot 10^{-1}$	
M6	$7.24 \cdot 10^{-1}$	$4.93 \cdot 10^{-1}$				0.00		
M7	7.88	5.36				0.00	$1.27 \cdot 10^{-1}$	

Values less than $1 \cdot 10^{-13}$ are reported as '0.00'.

Table 12: Model parameter estimation results for the higher error tolerance estimation for the nonmonotonic retention profile in Figure 4, panel 3a

Model	M2: $k_{d,f}$		M2: $k_{d,s}$			M1: α	S_m [kg kg ⁻¹]	M2: $k_{d,s}/k_{d,f}$ [-]
	M1, M3–M7: k_d [s ⁻¹]	k_r [s ⁻¹]	f [-]	M4: $k_{d,2}$ [s ⁻¹]	$k_{r,2}$ [s ⁻¹]	β [-]		
M1	$1.53 \cdot 10^{-2}$							$5.71 \cdot 10^{-1}$
M2	$1.53 \cdot 10^{-2}$		0.00	$8.75 \cdot 10^{-3}$				$5.72 \cdot 10^{-1}$
M3	$8.82 \cdot 10^{-3}$	$3.99 \cdot 10^{-8}$						
M4	$8.81 \cdot 10^{-3}$	$2.27 \cdot 10^{-12}$		$2.76 \cdot 10^{-7}$	$9.98 \cdot 10^{-10}$			
M5	$8.80 \cdot 10^{-3}$	$9.46 \cdot 10^{-8}$					$7.01 \cdot 10^5$	
M6	$8.84 \cdot 10^{-3}$	$8.96 \cdot 10^{-10}$				$9.54 \cdot 10^{-6}$		
M7	$8.77 \cdot 10^{-3}$	$8.82 \cdot 10^{-10}$				$9.99 \cdot 10^{-6}$	2.06	

Values less than $1 \cdot 10^{-13}$ are reported as '0.00'.

Table 13: Model plausibility comparison results for the archetypal retention profiles in Figure 4

	HE			LD			NM			MI		
	RSS	w_i	PFO									
M1	$9.57 \cdot 10^{-14}$	0.000	100%	$1.94 \cdot 10^{-12}$	0.017	98.1%	$9.41 \cdot 10^{-11}$	0.005	99.4%	$4.00 \cdot 10^{-10}$	0.281	46.4%
M2	$2.29 \cdot 10^{-14}$	0.005	99.5%	$1.94 \cdot 10^{-12}$	0.002	99.7%	$9.41 \cdot 10^{-11}$	0.001	99.9%	$4.00 \cdot 10^{-10}$	0.038	92.7%
M3	$9.57 \cdot 10^{-14}$	0.000	100%	$1.61 \cdot 10^{-12}$	0.007	99.2%	$2.12 \cdot 10^{-11}$	0.761	0.00%	$2.37 \cdot 10^{-10}$	0.523	0.00%
M4	$4.41 \cdot 10^{-14}$	0.000	100%	$5.49 \cdot 10^{-13}$	0.082	90.8%	$2.12 \cdot 10^{-11}$	0.014	98.2%	$2.48 \cdot 10^{-10}$	0.008	98.6%
M5	$9.57 \cdot 10^{-14}$	0.000	100%	$1.61 \cdot 10^{-12}$	0.001	99.9%	$2.12 \cdot 10^{-11}$	0.103	86.5%	$2.37 \cdot 10^{-10}$	0.071	86.5%
M6	$8.72 \cdot 10^{-15}$	0.994	0.00%	$2.27 \cdot 10^{-12}$	0.000	100%	$2.12 \cdot 10^{-11}$	0.103	86.5%	$2.37 \cdot 10^{-10}$	0.071	86.5%
M7	$1.94 \cdot 10^{-14}$	0.000	100%	$3.69 \cdot 10^{-13}$	0.891	0.00%	$2.12 \cdot 10^{-11}$	0.014	98.2%	$2.36 \cdot 10^{-10}$	0.010	98.1%

Model plausibility results for each archetypal retention profile and model are presented. Residual sum of squares error, RSS, Akaike weight, w_i , and model plausibility from optimum, PFO, are provided. The most plausible model has a 0% PFO.

Table 14: Model plausibility results for the higher error tolerance estimation for the NM retention profile in Figure 4

	RSS	w_i	PFO
M1	$9.41 \cdot 10^{-11}$	0.699	0.0%
M2	$9.41 \cdot 10^{-11}$	0.095	86.5%
M3	$8.64 \cdot 10^{-11}$	0.158	77.4%
M4	$8.64 \cdot 10^{-11}$	0.003	99.6%
M5	$8.64 \cdot 10^{-11}$	0.021	96.9%
M6	$8.64 \cdot 10^{-11}$	0.021	96.9%
M7	$8.64 \cdot 10^{-11}$	0.003	99.6%

Model plausibility results for a higher error tolerance parameter estimation for the NM retention profile. Residual sum of squares error, RSS, Akaike weight, w_i , and model plausibility from optimum, PFO, are provided. The most plausible model has a 0% PFO.

7.3.3 *Hyperexponentiality influence on model plausibility*

During the model construction phase of this research, we observed that simultaneous fitting of the retention profile and breakthrough curve can result in a sub-optimal fit of the retention profile, as the effective sample size (i.e., number of data points vs. number of fitted parameters) for fitting the breakthrough curve is often an order of magnitude greater than the effective sample size for fitting the retention profile. The relative weighting of the multi-objective fitting can be manipulated⁴² and therefore it is unclear what the acceptable goodness-of-fit is. Thus, to best illustrate the conceptual limitations of the models, we estimate model parameters directly from the retention profile and then assess the ability of each model to describe both the retention profile and breakthrough curve.

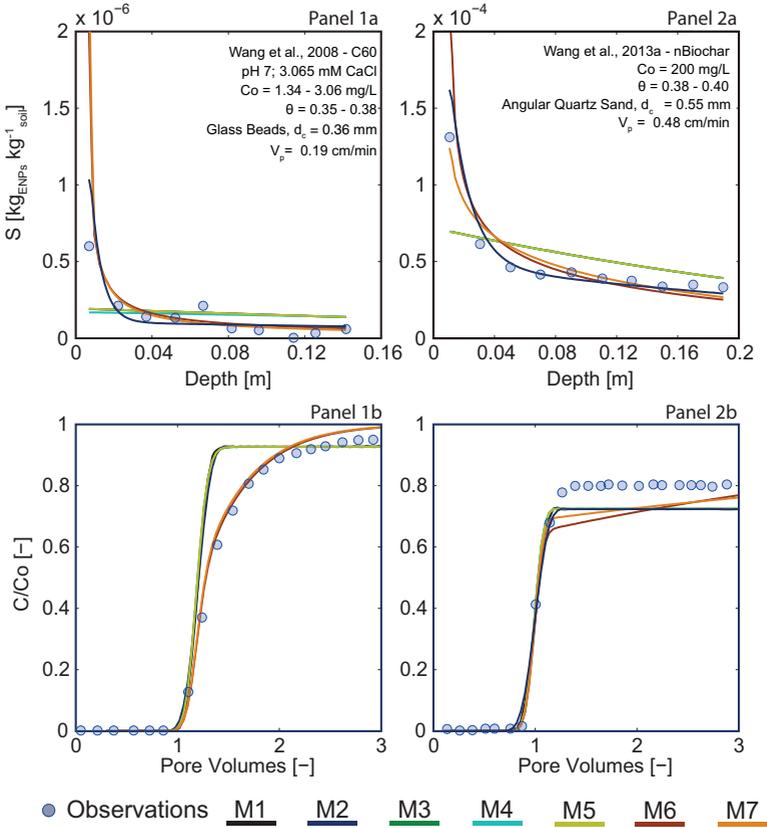


Figure 23: Digitized retention profiles (dots) and fitted model results (lines) are presented in panels 1a, 2a, and 3a, digitized breakthrough profiles (dots) and fitted model results are presented in panels 1b, 2b, and 3b. Note that the y-axis of panel 3a is presented at a 100x reduced scale in comparison to panels 1a and 2a. From a fit of the RP alone, the most plausible models that fit both the RP and BTC are as follows: In panels 1a and 1b, M6 provides the most plausible RP and BTC. In panels 2a and 2b, M2 provides the most plausible RP and BTC. In panels 3a and 3b, M7 provides the most plausible RP and BTC.

Table 15: Model parameter estimation results for the supplementary hyperexponential profiles in Figure 23

Model	M2: $k_{d,f}$	k_r	f	M2: $k_{d,s}$	$k_{r,2}$	β	S_m	M1: α
	M1, M3-M7: k_d			M4: $k_{d,2}$				M2: $k_{d,s}/k_{d,f}$
	[s ⁻¹]	[s ⁻¹]	[-]	[s ⁻¹]	[s ⁻¹]	[-]	[kg kg ⁻¹]	[-]
Figure 23, panel 1								
M1	$1.47 \cdot 10^{-2}$							$3.28 \cdot 10^{-3}$
M2	$1.47 \cdot 10^{-2}$		$1.79 \cdot 10^{-2}$	$2.65 \cdot 10^{-5}$				$1.81 \cdot 10^{-3}$
M3	$4.89 \cdot 10^{-5}$	$4.96 \cdot 10^{-6}$						
M4	$8.14 \cdot 10^{-5}$	$6.27 \cdot 10^{-3}$		$7.14 \cdot 10^{-5}$	$2.76 \cdot 10^{-3}$			
M5	$4.82 \cdot 10^{-5}$	0.00					$3.64 \cdot 10^{-3}$	
M6	$9.71 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$				$6.64 \cdot 10^{-1}$		
M7	$1.24 \cdot 10^{-2}$	$1.67 \cdot 10^{-3}$				$7.37 \cdot 10^{-1}$	$1.31 \cdot 10^{-4}$	
Figure 23, panel 2								
M1	$5.81 \cdot 10^{-3}$							$1.87 \cdot 10^{-2}$
M2	$5.81 \cdot 10^{-3}$		$3.52 \cdot 10^{-2}$	$7.63 \cdot 10^{-5}$				$1.31 \cdot 10^{-2}$
M3	$1.08 \cdot 10^{-4}$	0.00						
M4	$1.08 \cdot 10^{-4}$	0.00		0.00	$1.00 \cdot 10^{-5}$			
M5	$1.09 \cdot 10^{-4}$	0.00					$2.47 \cdot 10^{-2}$	
M6	$5.81 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$				$4.38 \cdot 10^{-1}$		
M7	$5.81 \cdot 10^{-3}$	$9.22 \cdot 10^{-5}$				$7.37 \cdot 10^{-1}$	$9.72 \cdot 10^{-5}$	

Values less than $1 \cdot 10^{-13}$ are reported as '0.00'.

Table 16: Model plausibility comparison results for the archetypal retention profiles in Figure 23

Model	Figure 23, panel 1			Figure 23, panel 2		
	RSS	w_i	PFO	RSS	w_i	PFO
M1	$2.35 \cdot 10^{-13}$	0.001	99.9%	$5.18 \cdot 10^{-9}$	0.001	99.9%
M2	$4.10 \cdot 10^{-14}$	0.655	0.00%	$1.94 \cdot 10^{-10}$	0.989	0.00%
M3	$2.35 \cdot 10^{-13}$	0.000	100%	$5.18 \cdot 10^{-9}$	0.000	100%
M4	$2.50 \cdot 10^{-13}$	0.000	100%	$5.18 \cdot 10^{-9}$	0.000	100%
M5	$2.35 \cdot 10^{-13}$	0.000	100%	$5.18 \cdot 10^{-9}$	0.000	100%
M6	$2.24 \cdot 10^{-14}$	0.287	56.2%	$1.46 \cdot 10^{-9}$	0.010	99.0%
M7	$3.00 \cdot 10^{-14}$	0.057	91.3%	$2.11 \cdot 10^{-9}$	0.000	100%

Model plausibility results for each archetypal retention profile and model are presented. Residual sum of squares error, RSS, Akaike weight, w_i , and model plausibility from optimum, PFO, are provided. The most plausible model has a 0% PFO.

APPENDIX II: SUPPORTING INFORMATION FOR CHAPTER 3

Contents

8.1	Database development and distributed computing .	179
8.2	Controlling randomness	179
8.3	RFECV method and modification	180
	8.3.1 Regression RFECV Modification	181
	8.3.2 Classification RFECV Modification	181
8.4	Publications mined for the database	182

The following sections are provided to add explanatory detail to the methods. Where appropriate, Python code is included.

8.1 DATABASE DEVELOPMENT AND DISTRIBUTED COMPUTING

For each experiment, training and target features were input into a spreadsheet. These data were then imported using Python and uploaded to a secure SQLite database hosted and accessed within the ETH domain. The code developed for this work connects to the secure database, retrieves the required data, and runs the machine learning code. Because the calculations were computationally demanding, model runs were parallelized and distributed to 5 computers.

8.2 CONTROLLING RANDOMNESS

There are several parts of the model that employ random or pseudo-random processes. For these portions, the random state was controlled

using a seed that is modified every model run, i (i.e., random state = seed* i).

8.3 RFECV METHOD AND MODIFICATION

All parameters of the code that we used to create the random forests in our work are described in the scikit-learn library, see <http://scikit-learn.org/stable/>.

The criteria used to evaluate the quality of a split in the generation of the decision trees that form a random forest are mean squared error for regression and gini impurity for classification, which are the default criteria of the scikit code.

As a modification of the code specific to this work, we generated a code that enables RFECV to be performed in tandem with random forest regression and classification. The RFECV (recursive feature elimination with cross validation) method requires that the machine learning estimator (e.g., random forest) output a weight for each feature so it can remove the least valuable feature in the next recursion step. Traditionally, this weight is determined by assigning an output rank, which is, however, not included in the default random forest regression or classification class definition within the Sklearn package. Instead, the feature importance was employed as a proxy for the output rank as shown below for the regression and classification approaches.

8.3.1 Regression RFECV Modification

```

import multiprocessing
from sklearn.ensemble import RandomForestRegressor
from sklearn.feature_selection import RFECV

# random forest regressor estimator call
rfc = RandomForestRegressor(n_estimators=1000,
    bootstrap=True, criterion='mse', oob_score=True,
    max_features="auto", n_jobs=-1)

# regressor class re-definition
class RandomForestRegressorWithCoef(RandomForestRegressor):
    def fit(self, *args, **kwargs):
        super(RandomForestRegressorWithCoef,
            self).fit(*args, **kwargs)
        self.coef_ = self.feature_importances_

# random forest regressor with coefficient estimator call
rfc = RandomForestRegressorWithCoef(n_estimators=1000,
    bootstrap=True, criterion='mse', oob_score=True,
    max_features="auto", n_jobs=-1)

```

8.3.2 Classification RFECV Modification

```

import multiprocessing
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import RFECV

# random forest classifier estimator call
rfc = RandomForestClassifier(n_estimators=1000,
    bootstrap=True, criterion='gini', oob_score=True,
    max_features="auto", n_jobs=-1)

# classifier class re-definition
class
    RandomForestClassifierWithCoef(RandomForestClassifier):
    def fit(self, *args, **kwargs):
        super(RandomForestRegressorWithCoef,
            self).fit(*args, **kwargs)
        self.coef_ = self.feature_importances_

# random forest regressor with coefficient estimator call
rfc = RandomForestRegressorWithCoef(n_estimators=1000,
    bootstrap=True, criterion='gini', oob_score=True,
    max_features="auto", n_jobs=-1)

```

8.4 PUBLICATIONS MINED FOR THE DATABASE

The following publications required zeta potential substitution. Grain ζ -potentials for Chowdhury et al.²⁵ were employed from silica surface ζ -potentials from Gu et al.¹⁹⁶ for the same buffer and approximate pH and ionic strength. Grain ζ -potentials for Choy et al.²⁶ were employed from quartz grain ζ -potential from Walker et al.¹⁹⁷ for the same buffer and approximate pH and ionic strength.

Table 17: Publications employed for this study organized by types of nanoparticles.

Publication Title	Reference	Material
Sensitivity of the transport and retention of stabilized silver nanoparticles to physicochemical factors	Liang et al. ³⁵	Ag
Cotransport of Titanium Dioxide and Fullerene Nanoparticles in Saturated Porous Media	Cai et al. ²³	C60
Transport and Retention of Nanoscale C ₆₀ Aggregates in Water-Saturated Porous Media	Wang et al. ³⁹	C60
Influence of Collector Surface Composition and Water Chemistry on the Deposition of Cerium Dioxide Nanoparticles: QCM-D and Column Experiment Approaches	Liu et al. ³⁶	CeO ₂
Fate and transport of elemental copper (Cu ₀) nanoparticles through saturated porous media in the presence of organic materials	Jones and Su ³¹	Cu ₀
Transport of Ferrilydrite Nanoparticles in Saturated Porous Media: Role of Ionic Strength and Flow Rate	Tosco et al. ³⁷	Fe
Transport and retention of multi-walled carbon nanotubes in saturated porous media: Effects of input concentration and grain size	Kasel et al. ³³	MWCNT
Transport of Biochar Particles in Saturated Granular Media: Effects of Pyrolysis Temperature and Particle Size	Wang et al. ⁴¹	nBiochar
Antagonistic Effects of Humic Acid and Iron Oxide on Biochar Nanoparticle Transport in Saturated Sand.	Wang et al. ⁴²	nBiochar
Facilitated transport of Cu with hydroxyapatite nanoparticles in saturated sand: Effects of solution ionic strength and composition	Wang et al. ⁴⁰	nHAP
Transport of ARS-labeled hydroxyapatite nanoparticles in saturated granular media is influenced by surface charge variability even in the presence of humic acid	Wang et al. ⁴⁴	nHAP
Comparison of three labeled silica nanoparticles used as tracers in transport experiments in porous media. Part II: Transport experiments and modeling	Vitorge et al. ³⁸	SiO ₂
Transport and Retention of TiO ₂ Rutile Nanoparticles in Saturated Porous Media under Low-Ionic-Strength Conditions: Measurements and Mechanisms	Chen et al. ²⁴	TiO ₂
Mechanisms of TiO ₂ nanoparticle transport in porous media: Role of solution chemistry nanoparticle concentration and flowrate	Chowdhury et al. ²⁵	TiO ₂
Application of an empirical transport model to simulate retention of nanocrystalline titanium dioxide in sand columns	Choy et al. ²⁶	TiO ₂
Transport and retention behaviors of titanium dioxide nanoparticles in iron oxide-coated quartz sand: Effects of pH, ionic strength, and humic acid	Han et al. ²⁷	TiO ₂
Transport and deposition of ZnO nanoparticles in saturated porous media	Jiang et al. ²⁸	ZnO
Influence of natural organic matter on the transport and deposition of zinc oxide nanoparticles in saturated porous media	Jiang et al. ²⁹	ZnO
Transport and retention of zinc oxide nanoparticles in porous media: Effects of natural organic matter versus natural organic ligands at circumneutral pH	Jones and Su ³²	ZnO
Transport of bare and capped zinc oxide nanoparticles is dependent on porous medium composition	Kurlanda-Witek et al. ⁴⁵	ZnO

9

APPENDIX III: SUPPORTING INFORMATION FOR CHAPTER 4

Contents

9.1	Inclusion, exclusion, and quality criteria	185
9.2	Decision tree performance	186

9.1 INCLUSION, EXCLUSION, AND QUALITY CRITERIA

More than 100 nanomaterial column transport studies conducted in saturated porous media from 2001–2015 were reviewed. 60 of the most recent nanoparticle transport publications (2008–2014) were evaluated in detail to determine whether sufficient information was available to support this study. Of these 60 publications, 24 reported retention profiles. The following information was recorded, where possible, from each retention profile within each publication.

- | | |
|----------------------|--------------------------------|
| 1. publication title | 9. porosity |
| 2. first author | 10. Darcy velocity |
| 3. year | 11. influent concentration ENM |
| 4. experiment ID | 12. influent pore volumes |
| 5. ENM ID | 13. pH |
| 6. RP shape | 14. electrolyte concentration |
| 7. column length | 15. electrolyte ID |
| 8. column width | 16. ENM density |

- | | |
|------------------------------|-------------------------------|
| 17. ENM isoelectric point | 22. collector coating |
| 18. ENM zeta potential | 23. concentration NOM |
| 19. collector zeta potential | 24. NOM ID |
| 20. ENM diameter | 25. Hamaker constant combined |
| 21. collector diameter | 26. longitudinal dispersivity |

Particle-fluid-collector Hamaker interaction parameters, where not provided, were calculated using literature values and Equation 31.^{22,198}

$$A_{123} = (A_{33}^{0.5} - A_{22}^{0.5})(A_{11}^{0.5} - A_{22}^{0.5}) \quad (31)$$

Where A_{123} is the Particle-fluid-collector Hamaker interaction parameter, A_{11} is the Hamaker constant for the nanomaterial in vacuum, A_{22} is the Hamaker constant for the fluid in vacuum, and A_{33} is the Hamaker constant for the collector in vacuum.

The temperature, T was assumed to be 298.15 K, the gravitational constant, g , was assumed to be 9.81 m s^{-2} , and the fluid density, ρ_f , was assumed to be 997 kg m^{-3} (density of water at 298.15 K) for all experiments unless otherwise noted. Where required, literature values for the relative dielectric constant of the ENM were employed to calculate the Debye length. The dielectric constant for all fluids was assumed to be 78.3.¹⁹⁹ Longitudinal dispersivities were not employed for this assessment, as only 18% of the experiments contained dispersivity values. ENM isoelectric points were not used for training, as sufficient information about the isoelectric point under the experimental conditions examined was generally not available. Only transport experiments with a retention profile, excluding those where retention was reported but not visually discernible or plotted, were included in the database.

9.2 DECISION TREE PERFORMANCE

Table 18: Sensitivity, selectivity, and precision of the presented decision tree.

Branch	Label	# Exp.	# Nonexp.	True Positives	False Positives	True Negatives	False Negatives
1	nonexponential	1	10	0	0	10	1
2	exponential	13	0	13	0	0	0
3	nonexponential	1	23	0	0	23	1
4	exponential	18	3	18	3	0	0
5	nonexponential	3	8	0	0	8	3
6	exponential	10	2	10	2	0	0
7	exponential	33	2	33	2	0	0
8	exponential	83	4	83	4	0	0
		Totals:		157	11	41	5
					Sensitivity	TP / (TP+FN)	96.9%
					Specificity	TN / (TN+FP)	78.8%
					Precision	TP / (TP+FP)	93.5%

CURRICULUM VITAE

PERSONAL INFORMATION

Nationality	American
Date-of-Birth	February 15, 1984
Email	elig@awareengineering.com

EXPERIENCE

2011–Present	<p>Founder and CEO: Aware Engineering, Inc; Aware Engineering GmbH</p> <p>I founded Aware to go beyond the pill to improve the lives of those with cystic fibrosis, COPD, and asthma by promoting adherence, facilitating preventative care, and improving care management through real-time device monitoring. To do this, I patented and developed a sensor-based digital adherence tracking technology, SelfAware. SelfAware is cloud-based, Class I software medical device that aggregates, analyzes, and visualizes patient-generated adherence information for physicians and patients. Adherence data is non-invasively collected through lightweight Bluetooth sensors that attach to patients home medical devices and relay device information through their mobile phone. FDA and CE marking for the Class I medical device will be complete in mid 2016. I also lead an international team of employees, consultants, and advisors, and manage numerous stakeholders (e.g., investors, the Cystic Fibrosis Foundation, Johns Hopkins, IBM-Watson University Hospital Zurich, ETH-Zurich, and the Swiss Government).</p>
--------------	---

- 2011–2012 **Project Environmental Engineer: Weston Solutions, Inc.**
 Consultant hired to certify Weston Solutions, Inc. as a federal energy service company (ESCO). Certification was approved in February of 2011.
- 2010–2011 **Research Fellow: University of Wisconsin at Madison and ETH-Zürich**
 Understanding the bioavailability and mobility of nanomaterials is critical to understanding their environmental impact and managing the risks, and benefits. To assess these risks, I conducted innovative nanomaterial transport research by applying a quartz crystal microbalance with dissipation (QCM-D), an instrument typically used to support cell biology assessment, to determine adsorption kinetics of functionalized titanium dioxide nanoparticles (6-10 nm) to natural organic matter coated silica surfaces.
- 2007–2009 **Staff Geotechnical/Geological Engineer: Golder Associates, Inc.**
 Provided engineering, technical, and regulatory guidance to support various waste, power sector, and sustainability-related federal and local clients.

EDUCATION

- 2012–2016 Doctoral Studies in Chemistry
 ETH Zurich (Switzerland)
- 2009–2012 Masters Studies in Environmental Engineering
 ETH Zurich (Switzerland)
- 2002–2007 Bachelor Studies in Geological Engineering, Geology and Geophysics
 University of Wisconsin at Madison (United States)

HONORS, AWARDS, AND PATENTS

Texas Medical Center	Awarded 2nd place and a large cash prize for developing and patenting an application to securely pull electronic medical record data using an innovative biometric identity system in a 3-day biodesign hackathon sponsored by the Texas Medical Center innovation institute.
KTI	Awarded nearly Fr. 1M co-development grant to work with the University Hospital Zürich and ETHZ wearable computing laboratory to develop the integrated health platform.
Google	15 Swiss ICT-startups pitched their ideas in front of more than 100 participants and a jury, which included representatives of Google, venturelab and IFJ. I won 2nd place along with a distinguished presenter award.
KTI	Awarded a fellowship to attend BioBusiness 2014 in Lugano, Switzerland.
Novartis	Competed, and won, both the group business development and pitch competition and the top prize for outstanding leadership in Novartis' International Biotechnology Leadership Camp (BioCamp) 2014.
KTI	Awarded Fr. 93k grant to work with the University Hospital Zürich and ETHZ wearable computing laboratory to develop the integrated health platform.
Patents	WO2014144208: Integrated modular health systems and methods of use; publication date: September 18, 2014.

