Electronic Supplementary Material (ESI) for Environmental Science: Nano.

Prediction of nanomaterial transport behavior from physicochemical properties: machine learning provides insights to guide the next generation of transport models

Electronic Supplementary Information

Eli Goldberg,[†] Martin Scheringer,[*,†] Thomas D. Bucheli,[‡] and Konrad Hungerbühler[†]

Institute for Chemical and Bioengineering, ETH Zürich, 8093 Zürich, Switzerland, and Agroscope Institute for Sustainability Sciences ISS, Zürich, Switzerland

E-mail: scheringer@chem.ethz.ch

[*]To whom correspondence should be addressed
[†]Institute for Chemical and Bioengineering, ETH Zürich, 8093 Zürich, Switzerland
[‡]Agroscope Institute for Sustainability Sciences ISS, Zürich, Switzerland

1 Supporting methods information

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

1.1 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.

1.2 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).

1.3 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/ and, more specifically,

http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForest
Regressor.html for the regression task and

http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForest
Classifier.html for the classification task.

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.

Regression RFECV Modification

```python
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)
```

Classification RFECV Modification

```python
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)
```

2 Publications mined for the database

Table 1: 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.*[1]	Ag
Cotransport of Titanium Dioxide and Fullerene Nanoparticles in Saturated Porous Media	Cai *et al.*[2]	C60
Transport and Retention of Nanoscale C_{60} Aggregates in Water-Saturated Porous Media	Wang *et al.*[3]	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.*[4]	CeO2
Fate and transport of elemental copper (Cu_0) nanoparticles through saturated porous media in the presence of organic materials	Jones and Su[5]	CuO
Transport of Ferrihydrite Nanoparticles in Saturated Porous Media: Role of Ionic Strength and Flow Rate	Tosco *et al.*[6]	Fe
Transport and retention of multi-walled carbon nanotubes in saturated porous media: Effects of input concentration and grain size	Kasel *et al.*[7]	MWCNT
Transport of Biochar Particles in Saturated Granular Media: Effects of Pyrolysis Temperature and Particle Size	Wang *et al.*[8]	nBiochar
Antagonistic Effects of Humic Acid and Iron Oxyhydroxide Grain-Coating on Biochar Nanoparticle Transport in Saturated Sand.	Wang *et al.*[9]	nBiochar
Facilitated transport of Cu with hydroxyapatite nanoparticles in saturated sand: Effects of solution ionic strength and composition	Wang *et al.*[10]	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.*[11]	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.*[12]	SiO2
Transport and Retention of TiO_2 Rutile Nanoparticles in Saturated Porous Media under Low-Ionic-Strength Conditions: Measurements and Mechanisms	Chen *et al.*[13]	TiO2
Mechanisms of TiO_2 nanoparticle transport in porous media: Role of solution chemistry nanoparticle concentration and flowrate	Chowdhury *et al.*[14]	TiO2
Application of an empirical transport model to simulate retention of nanocrystalline titanium dioxide in sand columns	Choy *et al.*[15]	TiO2
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.*[16]	TiO2
Transport and deposition of ZnO nanoparticles in saturated porous media	Jiang *et al.*[17]	ZnO
Influence of natural organic matter on the transport and deposition of zinc oxide nanoparticles in saturated porous media	Jiang *et al.*[18]	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[19]	ZnO
Transport of bare and capped zinc oxide nanoparticles is dependent on porous medium composition	Kurlanda-Witek *et al.*[20]	ZnO

The following publications required zeta potential substitution. Grain ζ-potentials for Chowdhury *et al.*[14] were employed from silica surface ζ-potentials from Gu *et al.*[21] for the same buffer and approximate pH and ionic strength. Grain ζ-potentials for Choy *et al.*[15] were employed from quartz grain ζ-potential from Walker *et al.*[22] for the same buffer and approximate pH and ionic strength.

References

1. Liang, Y.; Bradford, S. A.; Šimůnek, J.; Vereecken, H.; Klumpp, E. Sensitivity of the Transport and Retention of Stabilized Silver Nanoparticles to Physicochemical Factors. *Water Research* **2013**, *47*, 2572–2528.

2. Cai, L.; Tong, M.; Ma, H.; Kim, H. Co-transport of Titanium Dioxide and Fullerene Nanoparticles in Saturated Porous Media. *Environmental science & technology* **2013**, *47*, 5703–5710.

3. Wang, Y.; Li, Y.; Fortner, J. D.; Hughes, J. B.; Abriola, L. M.; Pennell, K. D. Transport and Retention of Nanoscale C_{60} Aggregates in Water-Saturated Porous Media. *Environmental Science & Technology* **2008**, *42*, 3588–3594.

4. Liu, X.; Chen, G.; Su, C. Influence of collector surface composition and water chemistry on the deposition of cerium dioxide nanoparticles: QCM-D and column experiment approaches. *Environmental Science & Technology* **2012**, *46*, 6681–6688.

5. Jones, E. H.; Su, C. Fate and transport of elemental copper (Cu^0) nanoparticles through saturated porous media in the presence of organic materials. *Water Research* **2012**, *46*, 2445–2456.

6. Tosco, T.; Bosch, J.; Meckenstock, R. U.; Sethi, R. Transport of ferrihydrite nanoparticles in saturated porous media: role of ionic strength and flow rate. *Environmental Science & Technology* **2012**, *46*, 4008–4015.

7. Kasel, D.; Bradford, S. A.; Šimůnek, J.; Heggen, M.; Vereecken, H.; Klumpp, E. Transport and retention of multi-walled carbon nanotubes in saturated porous media: Effects of input concentration and grain size. *Water Research* **2013**, *47*, 933–944.

8. Wang, D.; Zhang, W.; Hao, X.; Zhou, D. Transport of Biochar Particles in Saturated Granular Media: Effects of Pyrolysis Temperature and Particle Size. *Environmental Science & Technology* **2013**, *47*, 821–828.

9. Wang, D.; Zhang, W.; Zhou, D. Antagonistic Effects of Humic Acid and Iron Oxyhydroxide Grain-Coating on Biochar Nanoparticle Transport in Saturated Sand. *Environmental Science & Technology* **2013**, *47*, 5154–5161.

10. Wang, D.; Paradelo, M.; Bradford, S. A.; Peijnenburg, W. J.; Chu, L.; Zhou, D. Facilitated transport of Cu with hydroxyapatite nanoparticles in saturated sand: Effects of solution ionic strength and composition. *Water Research* **2011**, *45*, 5905–5915.

11. Wang, D.; Bradford, S. A.; Harvey, R. W.; Hao, X.; Zhou, D. 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* **2012**, *229*, 170–176.

12. Vitorge, E.; Szenknect, S.; Martins, J. M.-F.; Barthès, V.; Gaudet, J.-P. Comparison of three labeled silica nanoparticles used as tracers in transport experiments in porous media. Part II: Transport experiments and modeling. *Environmental Pollution* **2014**, *184*, 613–619.

13. Chen, G.; Liu, X.; Su, C. Transport and Retention of TiO2 Rutile Nanoparticles in Saturated Porous Media under Low-Ionic-Strength Conditions: Measurements and Mechanisms. *Langmuir* **2011**, *27*, 5393–5402.

14. Chowdhury, I.; Hong, Y.; Honda, R. J.; Walker, S. L. Mechanisms of TiO2 nanoparticle transport in porous media: Role of solution chemistry, nanoparticle concentration, and flowrate. *Journal of Colloid And Interface Science* **2011**, *360*, 548–555.

15. Choy, C. C.; Wazne, M.; Meng, X. Application of an empirical transport model to simulate retention of nanocrystalline titanium dioxide in sand columns. *Chemosphere* **2008**, *71*, 1794–1801.

16. Han, P.; Wang, X.; Cai, L.; Tong, M.; Kim, H. 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* **2014**, *454*, 119–127.

17. Jiang, X.; Tong, M.; Lu, R.; Kim, H. Transport and deposition of ZnO nanoparticles in saturated porous media. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2012**, *401*, 29 – 37.

18. Jiang, X.; Tong, M.; Kim, H. Influence of natural organic matter on the transport and deposition of zinc oxide nanoparticles in saturated porous media. *Journal of Colloid And Interface Science* **2012**, *386*, 34–43.

19. Jones, E.; Su, C. 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* **2014**, *275*, 79–88.

20. Kurlanda-Witek, H.; Ngwenya, B.; Butler, I. Transport of bare and capped zinc oxide nanoparticles is dependent on porous medium composition. *Journal of contaminant hydrology* **2014**, *162*, 17–26.

21. Gu, D.; Yalcin, S.; Baumgart, H.; Qian, S.; Baysal, O.; Beskok, A. Electrophoretic Light Scattering for Surface Zeta Potential Measurement of ALD Metal Oxide Films. *ECS Transactions* **2010**, *33*, 37–41.

22. Walker, S. L.; Redman, J. A.; Elimelech, M. Role of Cell Surface Lipopolysaccharides in Escherichia c oli K12 Adhesion and Transport. *Langmuir* **2004**, *20*, 7736–7746.