What Drives Variability in OECD TG 309 Studies? A Data-Driven Evaluation of 144 Surface Water Simulation Tests

Collard Marie[1], Earl Mike[2], Endersby Katie[3], Gibson Richard[4], Ducouret Patricia[5], Hafner Jasmin[6], Helbling Damian[7], Hughes Christopher B.[8], Lapczynski Aurelia[9], Peter Sven[5], Saunders David[10], Tamasi Balint[11], Vitale Chiara Maria[12], Wang Neil[13], Lemaire Philippe[14], Tolls Johannes[15], Wilmot Lucy[16]

[1]dsm-firmenich SA, Louvain-la-Neuve, Belgium; [2]Syngenta Ltd., Jealott's Hill, UK; [3]Unilever, SERS (Safety, Environmental and Regulatory Science), Colworth Science Park, Bedford, UK; [4]Environment Agency, York, United Kingdom; [5]BASF SE, Ludwigshafen, Germany; [6]Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland; [7]Cornell University, School of Civil and Environmental Engineering, Ithaca, USA; [8]Embark Chemical Consulting, Chester, UK; [9]Research Institute for Fragrance Materials (RIFM), NJ, USA; [10]Shell Global Solutions, The Hague, Netherlands; [11]dsm-firmenich, Kaiseraugst, Belgium; [12]Procter & Gamble, Brussels Innovation Center, Strombeek-Bever, Belgium; [13]Syensqo, Aubervilliers, France; [14]TotalEnergies, Paris La Défense Cedex, France; [15]Henkel AG & Co. KGaA, Düsseldorf, Germany; [16]ECETOC, Brussels, Belgium.

BACKGROUND

The OECD Test Guideline (TG) 309 'Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test' is a cornerstone method for assessing the degradation of chemicals under environmentally relevant conditions.

Variability in reported half-lives from multiple OECD TG 309 or similar studies on the same chemical has been attributed to multiple factors in the scientific literature[1], including:

- inoculum properties;
- surface water characteristics;
- differences in experimental design (e.g., test concentration and system configuration).

While **individual sources of variability have been explored in isolation, their relative influence and combined impact remain poorly understood**, particularly in studies conducted for regulatory purposes.

METHODS

This work aimed to **identify and rank the key factors - both substance-specific and system-related - that influence outcomes in OECD TG 309 studies**.

A comprehensive database of 144 independent TG 309 studies was compiled, including data sourced from the European Chemicals Agency (ECHA) dissemination dossiers and European Food Safety Authority (EFSA) Risk Assessment Reports.

After curation of the database, statistical analyses were performed to examine relationships between test system parameters, chemical properties, and degradation half-lives, with the goal of ranking which factors contribute most to variability.

The selected response variable was the **half-life in days (HL)**.

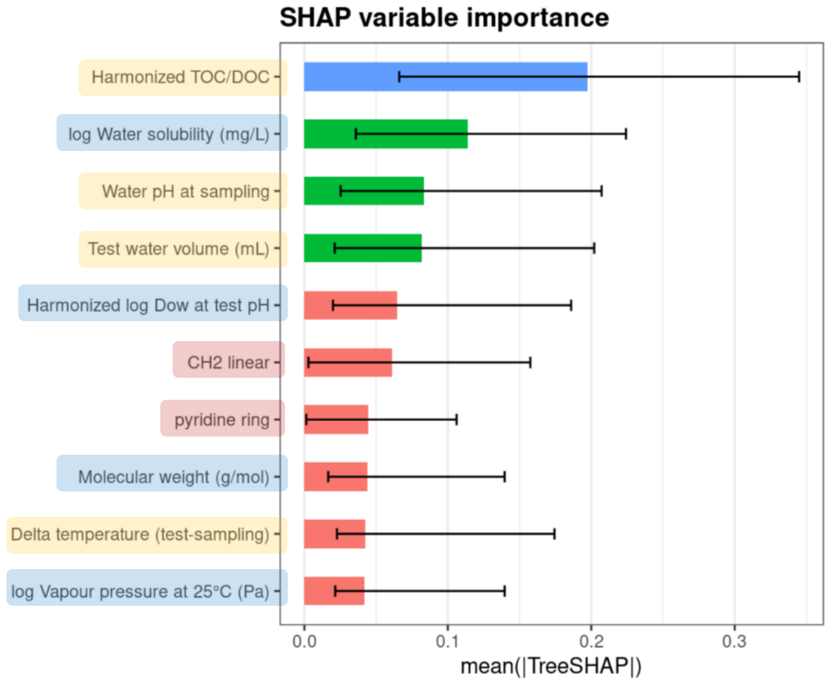
Predictor variables were selected based on data coverage (≥70% across studies) and, for test system parameters, sufficient variability..

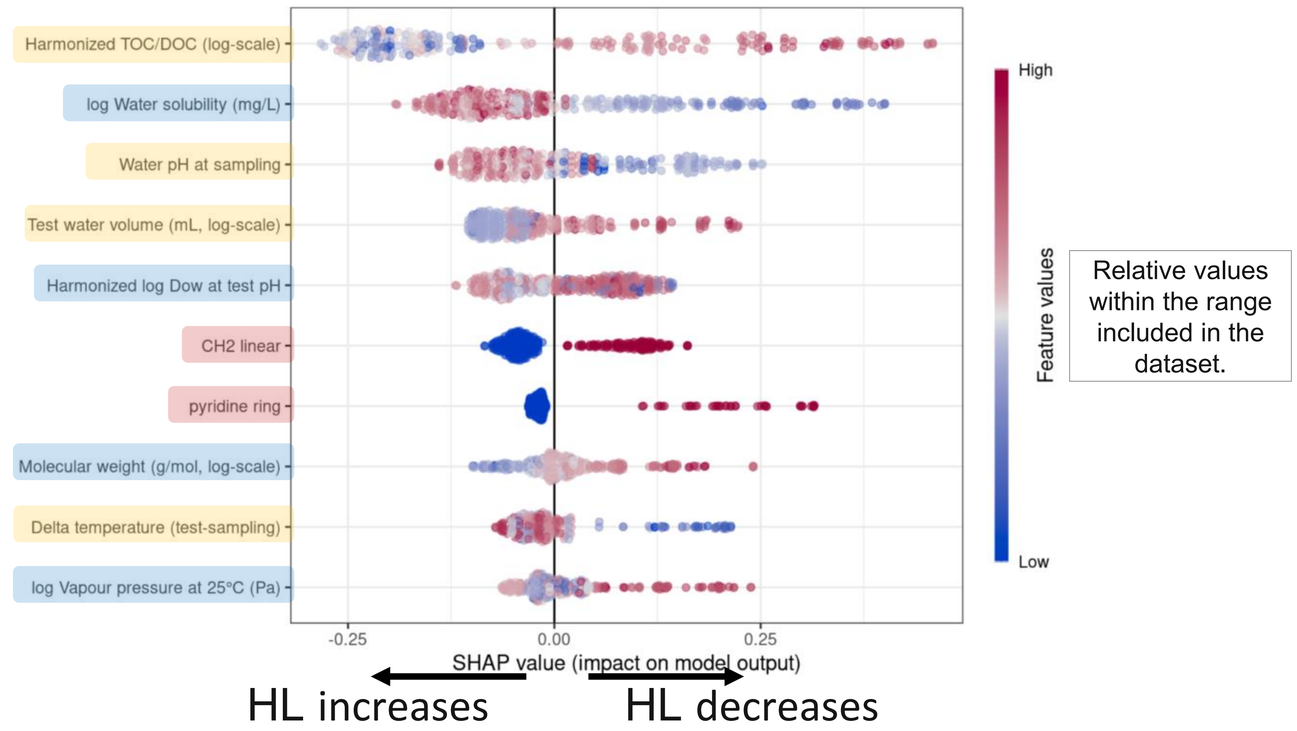
RESULTS

Summary of the Model:

- Gradient boosted regression trees (XGBoost)
- 136 substances
- 331 half-lives

- Phys-chem properties: molecular weight, water solubility, vapor pressure, log Kow/Dow at test pH
- Test system parameters: concentration tested, open/closed, test temperature, test water volume, use of organic solvent, delta temperature (test - sampling), agitation type, country where inoculum was sampled, water body, mass-balance met/not met, reference compound degraded within 2 weeks, water pH at sampling, TOC/DOC at sampling.
- Molecular descriptors: Biowin, enviPath





10 highest ranked variables comprise 4 test system parameters, 4 phys-chem properties, and 2 structural fragments.

Test system parameters => The Impact should be understood and minimized to favor a result which would mostly be driven by the structure of the chemical even if an inherent part of the result will come from the inoculum.

Phys-chem properties => The impact in terms of the limitations of the applicability domain of OECD TG 309 should be understood

Key considerations

The findings are inherently dependent on the composition of the dataset and the features included in the model. Relevant factors may not have been captured due to limited coverage or variability, or because they were not considered during feature selection. Molecular descriptors exhibited limited variability within individual fragments, which restricts their interpretability. In addition, the dataset is skewed toward not readily biodegradable substances, potentially leading to the underrepresentation of certain structural features. The model also combines categorical and numerical variables, which may influence outcomes, as numerical features provide greater granularity and may therefore be preferentially weighted. While SHAP values indicate the relative importance of features in explaining variability in half-lives, they do not quantify the magnitude of this variability. Finally, the model has limited interpretability, and further analyses are required to better understand the relationships inferred between features and half-lives.

NEXT STEPS

The robustness of the ranking has been tested using bootstrapping (2000 repetitions). Further analysis is ongoing to interpret the ranking provided by the model. These rankings need to be further understood based on previous research and on targeted statistical analyses. Moreover, to take into consideration the uncertainty surrounding the mean SHAP values, investigations are ongoing to define a significance threshold. The objective is:

An improved understanding that could support the optimisation of the OECD TG 309 protocol, enhance its reliability and reproducibility, and ultimately strengthen its application in regulatory hazard and risk assessments.

1 Wang, 2026 and references therein: https://doi.org/10.1093/etojnl/vgag037

The findings from this work will be submitted to the upcoming ET&C Special Issue on simulating aquatic degradation:

More info on the work of this ECETOC Task Force available here:

Contact: info@ecetoc.com