

## EuCompChem2025 – Laura Falivene – Oral

### Enhancing Catalyst Optimization with Robust Validation of MLR Models: A Systematic Workflow and Web Application

L. Falivene<sup>1</sup>, L. Cavallo<sup>2</sup> <sup>1</sup>Università degli studi di Salerno, Via Papa Paolo Giovanni II, 84100 Fisiciano, SA, Italy <sup>2</sup>King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia [lafalivene@unisa.it](mailto:lafalivene@unisa.it)

The development of efficient catalysts is a key driver of progress in chemical processes across various sectors, including pharmaceuticals and energy. However, catalyst optimization often relies on time-consuming and costly trial-and-error experimentation. Machine learning (ML), and in particular Multivariate Linear Regression (MLR), offers a promising strategy for predicting catalyst performance and accelerating discovery. [1-2] MLR models are well suited for small datasets and enable the identification of meaningful correlations between molecular descriptors and reaction outcomes. Nonetheless, ensuring the reliability of such models—especially when based on limited experimental data—remains a critical challenge. This contribution presents a robust validation workflow designed to assess MLR model reliability before application to catalyst design.[3] A dataset comprising 29 reaction classes and 514 reaction objectives serves as a testbed for evaluating models trained on reduced data. Both steric and electronic descriptors are included to capture structural effects on reactivity. Notably, a novel steric descriptor—percentage of distal buried volume (%VDBur)—is introduced, quantifying steric effects at varying distances from the catalytic center and enhancing model predictive accuracy. The proposed workflow systematically addresses common gaps in model validation and is broadly applicable to MLR-based approaches.[4] The models achieve strong predictive performance, with an average  $R^2$  of  $0.90 \pm 0.07$ , and validation tests confirm their robustness. The %VDBur descriptor significantly improves prediction accuracy, underscoring the importance of distal steric effects. Moreover, the models successfully identify catalysts with superior performance compared to those in the training set. [1] S. M. Mennen, C. Alhambra, C. L. Allen, M. Barberis, S. Bertritt, T. A. Brandt, A. D. Campbell, J. Castanon, A. H. Cherney, M. Christensen, et al., *Organic Process Research & Development*, 2019, 23, 1213–1242. [2] M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. M. Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu, A. Aspuru-Guzik, *Accounts of Chemical Research*, 2022, 55, 2454–2466. [3] Z. Cao, L. Falivene, A. Poater, B. Maity, Z. Zhang, G. Takasao, S. Sayed, A. Petta, G. Talarico, R. Oliva, L. Cavallo, *Cell Reports Physical Science*, 2025, 6, 102348, 1–8. [4] Z. Cao, L. Falivene, A. Petta, L. Cavallo, COBRA 1.0: A web application for catalyst optimization by linear regression, <https://www.aocdweb.com/OMtools/cobraMeitner>

**Primary author(s) :** FALIVENE, laura (Università degli studi di Salerno)

**Presenter(s) :** FALIVENE, laura (Università degli studi di Salerno)