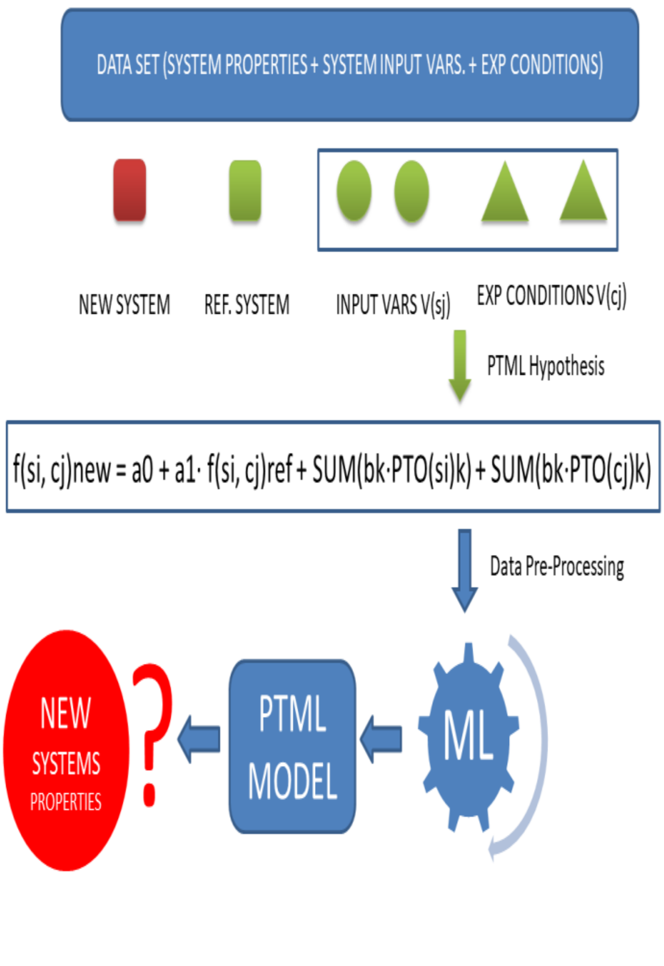
US-EU-IN Workshop on Multidisciplinary Sciences, UMN, Duluth, USA, 2018

PTML: Perturbation-Theory Machine Learning notes

Humbert González-Díaz [a,b,*]

[a] *Dept. of Organic Chemistry II, University of the Basque Country UPV/EHU , 48080 Bilbao , Spain.*
[b] *IKERBASQUE, Basque Foundation for Science, 48011 Bilbao , Spain.*
**Email: humberto.gonzalezdiaz@ehu.es*

Graphical Abstract



Abstract. PTML: Perturbation-Theory Machine Learning methods have been developed by H. Gonzalez-Diaz et al [0] to seek models able to predict multiple properties f(si, cj)k of type k of a system (si) at the same time (multi-output and multi-objective) taking into consideration variations (perturbations) in multiple experimental conditions cj = (c0, c1, c2, ... cn) at the same time with respect to a value of reference or expected. PTML-like models have been applied for different authors to study drugs, proteins, nanoparticles, complex networks, social systems, etc. [1-17]. In the particular case of a PTML linear models we can fit an equation with the general form f(si, cj)new = a0 + a1· f(si, cj)ref + SUM(bk·PTO(cj)k). In this model PTO(cj)k are PT operators measuring the perturbations in the new system si with respect to the system of reference sr with observed or expected property f(si, cj)ref. First, we need to calculate the values of the PTOs in the data pre-processing step. This PTOs allow us to perform an Information Fusion process with variables and conditions from different sources. Moving Averages (MA), Multi-condition MA (MMAs), Double MAs, Co-variance Operators, etc. are some examples of useful PTOs. After that, we can use Multiple Linear Regression (MLR), Linear Discriminant Analysis (LDA), or other linear ML techniques to seek the PTML model. In the non-linear cases, we can fit the PTML models using Artificial Neural Networks (ANN), Support Vector Machines (SVM), Classification Trees, and other ML methods.

References

0: General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. González-Díaz H, Arrasate S, Gómez-SanJuan A, Sotomayor N, Lete E, Besada-Porto L, Ruso JM. Curr Top Med Chem. 2013;13(14):1713-41.

1: Simón-Vidal L, García-Calvo O, Oteo U, Arrasate S, Lete E, Sotomayor N, González-Díaz H. Perturbation-Theory and Machine Learning (PTML) Model for High-Throughput Screening of Parham Reactions: Experimental and Theoretical Studies. J Chem Inf Model. 2018 Jul 23;58(7):1384-1396. doi: 10.1021/acs.jcim.8b00286.

2: Ferreira da Costa J, Silva D, Caamaño O, Brea JM, Loza MI, Munteanu CR, Pazos A, García-Mera X, González-Díaz H. Perturbation Theory/Machine Learning Model of ChEMBL Data for Dopamine Targets: Docking, Synthesis, and Assay of New l-Prolyl-l-leucyl-glycinamide Peptidomimetics. ACS Chem Neurosci. 2018 Jun 25. doi: 10.1021/acschemneuro.8b00083.

3: Luan F, Kleandrova VV, González-Díaz H, Ruso JM, Melo A, Speck-Planche A, Cordeiro MN. Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. Nanoscale. 2014 Sep 21;6(18):10623-30. doi: 10.1039/c4nr01285b.

4: González-Durruthy M, Werhli AV, Seus V, Machado KS, Pazos A, Munteanu CR, González-Díaz H, Monserrat JM. Decrypting Strong and Weak Single-Walled Carbon Nanotubes Interactions with Mitochondrial Voltage-Dependent Anion Channels Using Molecular Docking and Perturbation Theory. Sci Rep. 2017 Oct 16;7(1):13271. doi: 10.1038/s41598-017-13691-8.

5: Martínez-Arzate SG, Tenorio-Borroto E, Barbabosa Pliego A, Díaz-Albiter HM, Vázquez-Chagoyán JC, González-Díaz H. PTML Model for Proteome Mining of B-Cell Epitopes and Theoretical-Experimental Study of Bm86 Protein Sequences from Colima, Mexico. J Proteome Res. 2017 Nov 3;16(11):4093-4103. doi: 10.1021/acs.jproteome.7b00477.

6: Liu Y, Munteanu CR, Fernandez-Lozano C, Pazos A, Ran T, Tan Z, Yu Y, Zhou C, Tang S, González-Díaz H. Experimental Study and ANN Dual-Time Scale Perturbation Model of Electrokinetic Properties of Microbiota. Front Microbiol. 2017 Jun 30;8:1216. doi: 10.3389/fmicb.2017.01216.

7: González-Durruthy M, Alberici LC, Curti C, Naal Z, Atique-Sawazaki DT, Vázquez-Naya JM, González-Díaz H, Munteanu CR. Experimental-Computational Study of Carbon Nanotube Effects on Mitochondrial Respiration: In Silico Nano-QSPR Machine Learning Models Based on New Raman Spectra Transform with Markov-Shannon Entropy Invariants. J Chem Inf Model. 2017 May 22;57(5):1029-1044. doi: 0.1021/acs.jcim.6b00458.

8: Aranzamendi E, Arrasate S, Sotomayor N, González-Díaz H, Lete E. Chiral Brønsted Acid-Catalyzed Enantioselective α-Amidoalkylation Reactions: A Joint Experimental and Predictive Study. ChemistryOpen. 2016 Nov 23;5(6):540-549. doi: 10.1002/open.201600120.

9: Tenorio-Borroto E, Ramírez FR, Vázquez Chagoyán JC, de Oca Jiménez RM, García-Mera X, González-Díaz H. Experimental-Theoretic Approach to Drug-Lymphocyte Interactome Networks with Flow Cytometry and Spectral Moments Perturbation Theory. Curr Pharm Des. 2016;22(33):5114-5119.

10: Ran T, Liu Y, Li H, Tang S, He Z, Munteanu CR, González-Díaz H, Tan Z, Zhou C. Gastrointestinal Spatiotemporal mRNA Expression of Ghrelin vs Growth Hormone Receptor and New Growth Yield Machine Learning Model Based on Perturbation Theory. Sci Rep. 2016 Jul 27;6:30174. doi: 10.1038/srep30174.

11: Romero-Durán FJ, Alonso N, Yañez M, Caamaño O, García-Mera X, González-Díaz H. Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives. Neuropharmacology. 2016 Apr;103:270-8. doi: 10.1016/j.neuropharm.2015.12.019.

12: Kleandrova VV, Luan F, González-Díaz H, Ruso JM, Speck-Planche A, Cordeiro MN. Computational tool for risk assessment of nanomaterials: novel QSTR-perturbation model for simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under multiple experimental conditions. Environ Sci Technol. 2014 Dec 16;48(24):14686-94. doi: 10.1021/es503861x.

13: Gonzalez-Diaz H, Arrasate S, Juan AG, Sotomayor N, Lete E, Speck-Planche A, Ruso JM, Luan F, Cordeiro MN. Matrix trace operators: from spectral moments of molecular graphs and complex networks to perturbations in synthetic reactions, micelle nanoparticles, and drug ADME processes. Curr Drug Metab. 2014;15(4):470-88.

14: Kleandrova VV, Luan F, González-Díaz H, Ruso JM, Melo A, Speck-Planche A, Cordeiro MN. Computational ecotoxicology: simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. Environ Int. 2014 Dec;73:288-94. doi: 10.1016/j.envint.2014.08.009.

15: Vergara-Galicia J, Prado-Prado FJ, Gonzalez-Diaz H. Galvez-Markov network transferability indices: review of classic theory and new model for perturbations in metabolic reactions. Curr Drug Metab. 2014;15(5):557-64.