Checking the Efficacy of Two Basic Descriptors With a Set of Properties of Alkanes

Emili Besalú, University of Girona, Girona, Spain  
Riccardo Zanni, University of Valencia, Valencia, Spain  
Lionello Pogliani, University of Valencia, Valencia, Spain  
Jesus Vicente de Julian-Ortiz, University of Valencia, Valencia, Spain

ABSTRACT

Several experimental properties of alkanes are described by means of multilinear models at the cross-validation level. The models have been obtained considering two main sets of descriptors: mathematically-based and experimental ones. The best models are obtained normally involving one of the two sets. The main goal of this work is to show how the theoretical descriptors are able to perform a competitive role against the experimental ones. This constitutes an important topic in the quantitative structure-property relationships field because the use of mathematical and in silico descriptors is validated as a proper tool for model building. Activity distributions of the properties and indices employed are discussed, along with the shape of the obtained residual plots.

KEYWORDS

Cross-Validation, Experimental Descriptors, Linear Regression, QSPR for Alkanes, Theoretical Descriptors

INTRODUCTION

Recently, in a series of papers (Pogliani et al., 2013; 2014; Besalú et al., 2016), studies were performed with the intent to establish the limits of predictive relationships and their meanings, i.e., how useful they were, within which range, etc. It was also considered if their usefulness could be used to establish a causal relationship between the dependent and the independent variable or variables, and if there were different types of descriptor(s) that were able to give rise to a quite useful relation. This subject is of paramount importance to understand that descriptors are not a kind of black box that can solve everything in any context, giving rise to a kind of physical law. Mathematical chemical modeling is not the same as mathematical physics and the problems it has to solve are of different nature.

In the present article we study a set of properties for a set of alkanes that have been originally studied in a seminal paper that marked the birth of the well-known Randić, $\chi$, index (Randić, 1975; Kier & Hall, 1986). This index is usually written as $\chi$ but for the purposes of the present study, as it is the only index here used, we will call it simply $\chi$. To the set of alkanes studied by Randić we will add a compound, methane, whose $\chi$ index cannot be clearly defined, given the original definition of this index, and for this reason its value will be set, per definition, equal to zero.

Four fundamental types of indices are used to describe nine properties of a set of forty-four alkanes. They are: a $\chi$-type of indices, i.e., a set of powers of $\chi$, a $N$-type of indices, i.e., a set of powers of $N$, DOI: 10.4018/IJQSPR.2019010105

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
where \( N \) is the number of carbons in alkanes, an experimental-type of indices, i.e., the experimental properties themselves considered as indices, and finally a random-type of indices, \( rd \), i.e., a set of random numbers. To these, a set of mixed type of \( \chi \) and \( N \) indices were added.

The values of the different properties of alkanes have been obtained from the NIST (National Institute of Standard and Technology, Standard reference data Catalog, http://webbook.nist.gov/), and are listed together with the values of \( N \), and \( \chi \) in Tables 1 and 2. The Henry constant, \( k^0 \), has been multiplied by 1000 to avoid too small values, that is, \( 10^1 \times k^0 \). In cases where a range of values was given to this property, a mean between the two extremes was used. Furthermore, whenever different values were given to a property, due to experiments done at different years, the more recent ones were chosen. Notice that with the exception of \( T_b, 10^3 k^0 \), and \( \Delta H^0 \), the other properties do not share exactly the same number of compounds.

The aims of our study are: (1) to delineate an easy way to find new and good descriptors by the aid of rather simple mathematical operations performed on a fundamental index, (2) to show that quite simple descriptors, like the count of carbon atoms, can be quite effective, (3) to check their usefulness relatively to the experimental descriptors, and (4) to check the usefulness of random variables to detect fortuitous correlations. Furthermore, it will be interesting to check if methane shows up as an outlier throughout the model with \( \chi \) indices where its value was defined as zero. By the way, we should also check if some properties are soundly related, such as, for instance, \( T_b \) and \( \Delta H^0 \), as they should be from a physico-chemical point of view.

**MATERIAL AND METHODS**

We remind that index, \( \chi \) = \( \sum_{ij} (\delta_i \delta_j)^{0.5} \), is derived by the aid of a simple two-dimensional graph of a hydrogen-depleted molecule, and that the sum runs over all single connections (\( \sigma \)-bonds) of the graph. Furthermore, \( \delta_i \) and \( \delta_j \) are the valence of the adjacent carbon atoms, i.e., the number of connection of a carbon to its adjacent carbons as non-adjacent carbon bonds do not contribute to \( \chi \). The first set of descriptors is centered on the following powers of \( \chi \).

\[
\{ \chi^n: n = -2, -1.5, -1, -0.5, 0.5, 1, 1.5, 2 \}
\]

Due to the definition of \( \chi \), it is practically impossible to derive a sensible value for methane for this reason we have defined \( \chi(CH_4) = 0 \) for every value of \( n \). The second set of indices is made up of the following powers of \( N \), the descriptor that counts the number of carbons in alkanes:

\[
\{ N^n: n = -2, -1.5, -1, -0.5, 0.5, 1, 1.5, 2 \}
\]

The third set of indices is represented by the experimental parameters themselves,

\[
\{ T_b, T_p, T_m, T_c, P_c, d_c, 10^3 k^0, \Delta H^0, IE \}
\]

See Tables 1 and 2 for their meaning.

The fourth set of indices are fifty sets of uniformly distributed random indices, \( 0 < rd < 1 \), that were obtained by the aid of a mathematical function of the MS Office Excel 2013,

\[
\{ rd1, rd2, rd3, \ldots, rd49, rd50 \}
\]

Finally, the fifth set of mixed indices was made up of combinations of \( \chi \), and the carbon count, \( N \), beside their powers,
Predicting Thermal Conductivity Enhancement of Al2O3/Water and CuO/Water Nanofluids Using Quantitative Structure-Property Relationship Approach

Prediction of Human Lethality of Psychoactive Drugs From Rodent LD50 Values
[www.igi-global.com/article/prediction-of-human-lethality-of-psychoactive-drugs-from-rodent-ld50-values/223213?camid=4v1a](www.igi-global.com/article/prediction-of-human-lethality-of-psychoactive-drugs-from-rodent-ld50-values/223213?camid=4v1a)

Can Toxicity for Different Species be Correlated?: The Concept and Emerging Applications of Interspecies Quantitative Structure-Toxicity Relationship (i-QSTR) Modeling
[www.igi-global.com/article/can-toxicity-for-different-species-be-correlated/146169?camid=4v1a](www.igi-global.com/article/can-toxicity-for-different-species-be-correlated/146169?camid=4v1a)