

Invariants or Aggregation Operators

The invariants are numerical quantities derived from the molecular structure and used to characterize local properties of a molecule; these numbers are calculated in such a way as to be independent of any arbitrary atom/bond numbering. Local invariants can be distinguished into Local Vertex Invariants (LOVIs) and Local Edge Invariants (LOEIs), depending on whether they refer to atoms or bonds.

LOVIs of a molecule are usually collected into an N -dimensional vector (N = number of atoms). LOVIs and LOEIs (also known as LOBIs) are used to calculate several molecular (namely, topological indices) by applying different operators. L is here adopted as the general symbol for local invariants.

Over the years, it has been generally accepted that the definition of global (or local) indices from LOVIs (L_i) implies the summation of the contributions of the elements that constitute a given molecular structure. However, summation (Minkowski's first norm (N1) in our specific case) is just one of the many invariants capable of globally characterizing given LOVIs.

In this program, are employed a series of *invariants* that generalize the traditional method of obtaining global (or local) invariants by summation of the LOVIs (and LOEIs used in bond-based indices). These are classified in four major groups:

1) Norms (or Metrics):

- a) Minkowski's norms (N1, N2, N3)

2) Mean Invariants (first statistical moment):

- a) Geometric Mean (G),
- b) Arithmetic Mean (M),
- c) Quadratic Mean (P2),
- d) Potential Mean (P3) and
- e) Harmonic Mean (H)

3) Statistical Invariants (highest statistical moments):

- a) Variance (V),
- b) Skewness (S),
- c) Kurtosis (K),

- d)** Standard Deviation (SD),
- e)** Variation Coefficient (VC),
- f)** Range (R),
- g)** Percentile 25 (Q1),
- h)** Percentile 50 (Q2),
- i)** Percentile 75 (Q3),
- j)** Inter-quartile Range (I50),
- k)** X max (MX) and
- l)** X min (MN)

4) “Classical algorithms” Invariants (functions to derive MDs from LOVIs):

- a)** Autocorrelations AC(i),
- b)** Gravitational (GI(i)),
- c)** Total sum at k lags (TSk(i)),
- d)** Mean information content (MI(i)),
- e)** Total information content (TI),
- f)** Standardized information content (SI),
- g)** Electro-topological state (ES(i)).

1. Norms (or Metrics): Mathematical definition

Name	ID	Formula
Minkowsky norm (p = 1) Manhattan norm	N1	$N1 = \sum_{a=1}^n L_a $
Minkowsky norm (p = 2) Euclidean norm	N2	$N2 = \sqrt{\sum_{a=1}^n L_a ^2}$
Minkowsky norm (p = 3)	N3	$N3 = \sqrt[3]{\sum_{a=1}^n L_a ^3}$

Note 1. The general equation of Minkowsky's norms is, $\|\vec{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$.

Note 2. The formulae used in these invariants, are simplified forms of general equations given that the vector \vec{y} is constituted of the coordinates of the origin. For example, in the case of the Euclidean norm (N2), the general formula is: $\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2 + (x_j - y_j)^2 + (x_z - y_z)^2}$.

But given that $\vec{y} = (0, 0, 0)$, this formula reduces to $\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$.

2. Mean Invariants: Mathematical definition

Name	ID	Formula
Geometric Mean	GM	$G = \sqrt[n]{\prod_{a=1}^n L_a}$
Arithmetic Mean (Power mean of degree $\beta = 1$)	AM	
Quadratic Mean (Power mean of degree $\beta = 2$)	P2	$M_\beta = \left(\frac{L_1^\beta + L_2^\beta + \dots + L_n^\beta}{n} \right)^{\frac{1}{\beta}}$
Power mean of degree $\beta = 3$	P3	
Harmonic Mean (Power mean of degree $\beta = -1$)	HM	

3. Statistical Invariants: Mathematical definition

Name	ID	Formula
Variance	V	$V = \frac{\sum_{a=1}^n (L_a - M)^2}{n - 1}$
Skewness	S	$S = \frac{n^*(X_3)}{(n-1)(n-2)(DE)^3}$
		$X_3 = \sum_{a=1}^n (L_a - M)^3$
		M, arithmetic mean DE, standard deviation

Kurtosis	K	$k = \frac{n(n+1)X_4 - 3(X_2)(X_2)(n-1)}{(n-1)(n-2)(n-3)(DE)^4} M,$ $X_j = \sum_{a=1}^n (L_a - M)^j$ arithmetic mean SD, standard deviation
Standard Deviation	SD	$SD = \sqrt{\frac{(\sum L_a - M)^2}{n-1}}$
Variation Coefficient	VC	$CV = \frac{SD}{M}$
Range	R	$R = L_{\max} - L_{\min}$
Percentile 25	Q1	$Q1 = \left[\frac{N}{4} + \frac{1}{2} \right]$ N, L _a number
Percentile 50	Q2	$Q2 = \left[\frac{N}{2} + \frac{1}{2} \right]$ N, L _a number
Percentile 75	Q3	$Q3 = \left[\frac{3N}{4} + \frac{1}{2} \right]$ N, L _a number
Inter-quartile Range	I50	$I50 = Q3 - Q1$
Maximum value	MX	$MX = L_a \max$
Minimum value	MN	$MN = L_a \min$

4. Classical Invariants: Mathematical definition

Name	ID	Formula
Autocorrelation	AC ^k	$AC_k = \sum_{i=1}^n \sum_{j \geq 1}^n L_i \times L_j \bullet (\delta(d_{ij}, k))$ $k = 1, 2, \dots, 7$
Gravitational	GV ^k	$GV_k = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{L_i L_j}{d_{ij}^k} \bullet (\delta(d_{ij}, k))$ $k = 1, 2, \dots, 7$
Total sum at lag k	TS ^k	$TS_k = \sum_{i=1}^n \sum_{j=1}^n L_{ij} \bullet (\delta(d_{ij}, k))$ $k = 1, 2, \dots, 7$
Mean Information Content	MIC	$MIC = - \sum_{i=1}^A \frac{N_g}{N_o} \cdot \log_2 \frac{N_g}{N_o}$ <p>where, N_g is the number of atoms with the same LOVI value. N_o is the number of atoms in a molecule</p>
Total Information Content	TIC	$TIC = N_0 \cdot \log_2 N_0 - \sum_{g=1}^G N_g \cdot \log_2 N_g$

Standardized Information Content	SIC	$SIC = \frac{IT}{N_0 \cdot \log_2 N_0}$ $S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^n \frac{I_i - I_j}{(d_{ij} + 1)^2} \text{ where, } I_i \text{ is}$
Electrotopological state (E-state index)	ES	<i>the intrinsic state of the i^{th} atom and ΔI_i is the field effect on the i^{th} atom calculated as perturbation of the I_i of i^{th} atom by all other atoms in the molecule, d_{ij} is the topological distance between the i^{th} and the j^{th} atoms, and n is the number of atoms. The exponent k is 2.</i>

Standardized option

In the standardization procedure, all values of *original* LOVIs are replaced by standardized LOVI values which are computed as follows: *Std. LOVIs* = (Original LOVI – mean of LOVIs)/Std. deviation of original LOVIs. With this re-scaling, each new LOVI has a mean of 0 and a standard deviation of 1.