The Free Internet Journal
for Organic Chemistry

Paper

Archive for
Organic Chemistry

Arkivoc **2017**, part v, 356-369

Gas principal properties as new compact descriptors for data-driven gas solubility modelling

Alessio Paternò, Carmela Bonaccorso, Giuseppe Musumarra,* and Salvatore Scirè

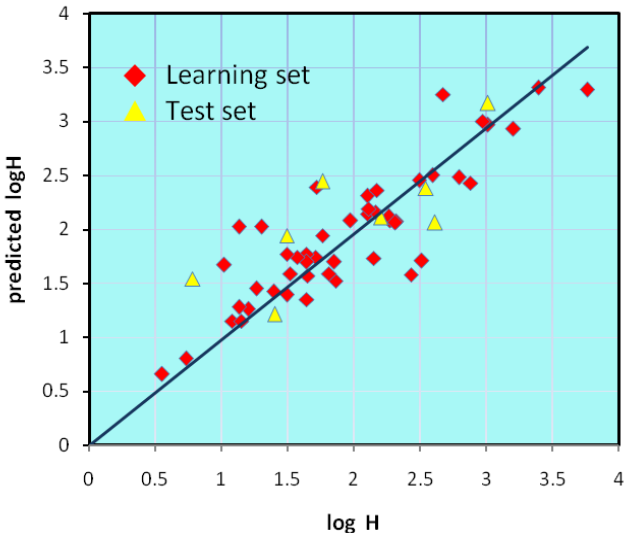
Dipartimento di Scienze Chimiche, Università di Catania, Viale Andrea Doria 6, 95125 Catania, Italy
Email: gmusumarra@unict.it

Dedicated to Prof. Alan R. Katritzky (1928-2014) generous founder of Arkivoc, a genuine open access journal free of charge for both authors and readers

Received 09-06-2017 **Accepted** 10-18-2017 **Published on line** 11-30-2017

Abstract

Principal properties (PPs), new compact descriptors for 48 gases were derived, their physico-chemical significance discussed, and applications to predict gas solubility in organic solvents by means of data-driven soft models reported.



Keywords: Gas descriptors, Solvent descriptors, Principal properties, Gas solubilities

Introduction

Optimization of the desired biological and physico-chemical properties for complex chemical entities (molecules or macromolecules) or for the performance of chemical processes involving different chemical building blocks (reactions, industrial and physico-chemical processes etc.) requires parametrization of non-continuous variables (substituents, amino acids, solvents, catalysts, etc.) by means of chemical descriptors. The need for variables (descriptors) orthogonal to each other suitable for multivariate experimental design led to the derivation of principal properties (PPs), intrinsic properties representative of experimentally observable macroscopic descriptors for chemical entities. PPs, calculated as a principal component analysis (PCA)[1] scores from several experimentally measured physico-chemical properties, are nowadays available for solvents,[2,3] aldehydes and ketones,[2] amines,[2] Lewis acids,[2] lanthanide triflates,[4] aromatic substituents[5] and amino acids.[6-9] Principal properties for heteroaromatic moieties, based on aromaticity[10] and on 3D-GRID structural parameters[11] have also been reported.

PPs have been successfully applied in the field of Quantitative Structure Activity Relationships (QSARs),[5] in particular for the design of biologically active peptides.[6,7] Dedicated PPs for amino acids were then derived for peptides QSARs[8] and for quantitative sequence-activity modelling.[12] More recently MIF (Molecular Interaction Field) molecular descriptors[9,13-16] were used for ligand-based virtual screening in antitumor drug design.

PPs were also derived for Ionic Liquids (ILs), low melting point salts comprising an organic cation and an inorganic or organic anion, which exhibited unprecedented efficiency at a molecular level providing opportunities for the development of green and sustainable chemical procedures. As experiments could not possibly explore the huge experimental chemical space covered by cationic and anionic counterparts of ILs, derivation of *in silico* VolSurf+ physico-chemical descriptors was needed. These parameters used as such, or compacted into cationic and anionic PPs, PP+ and PP- for ILs, respectively,[17] were applied to develop QSARs relating the chemical structure of ILs to their toxicities,[17,18] or to important physico-chemical properties such as polarity,[19] heat capacity,[20] viscosity, density, decomposition temperature and conductivity.[21] Such an approach has been illustrated in detail and allowed design of ILs for specific applications.[22]

The above examples illustrate the use of PPs as molecular descriptors for building blocks to predict the biological or physico-chemical properties of more complex chemical entities. An interesting example of PPs multivariate optimization of a chemical process with three chemical building blocks is provided by the Fischer indole synthesis[23,24] a two-step synthesis involving the reaction of ketones with phenylhydrazones and ring closure in the presence of an acid. When unsymmetrical ketones are used, an indole regioisomeric mixture can be obtained. The use of PPs of ketones, of Lewis acid catalysts and of solvents led to the regiospecific synthesis of single indole regioisomers[25] and to a one-pot reaction under milder conditions.[26]

In spite of several successful applications, PPs have not been very popular over the past three decades. The main criticism is focussed on the fact that they are subjective quantities dependent on the data set adopted for their derivation and the lack of an immediately interpretable physical meaning. On the other hand PPs are statistically orthogonal and therefore can be safely used in multiparameter linear equations, avoiding the danger of collinearity. Furthermore, they are less influenced by measurement errors and system-specific variations as compared to single descriptors (e.g. solvent polarity scales) and can be derived for a wider set of objects as compared to the original descriptors (e.g. different solvent polarity scales), allowing the investigation of a wider chemical space. Finally the physical meaning of PPs can be evidenced by the PCA descriptor loadings in their derivation.

In 1803 Henry's law provided the first and still most popular quantitative measurement for gas solubility in a solvent at a given temperature.[27] Henry's constant H is defined as P/x where P is the partial pressure of the gas

in bar units and x is its molar fraction in the liquid phase. Accordingly, high H values denote low solubility, while low H values correspond to higher solubility. In 1936, Hildebrand[28] proposed the definition of a "solubility parameter", further extended by Hansen.[29] Similar Hansen parameters indicate miscibility in various proportions, while dissimilar values denote limited solubility. Henry, Hildebrand and Hansen solubility parameters, providing a quantitative estimate for the ancient motto "like dissolves like", are widely adopted in industrial processes requiring the knowledge of gas-liquid solubility, for the selection of extraction solvents and in many other applications. In this context, a recent review dealt with the solubility parameters of permanent gases, an important issue for industrial processes and environmental elimination.[30]

Modelling and predicting gas solubility adopting theory-driven approaches, such as quantum mechanical calculations,[31] mixed quantum mechanics/molecular mechanics[32] and force field[33] models have been reported.[34]

It has recently been proposed[21] that data-driven modelling approaches can complement and usefully integrate theory-driven ones.

In particular, the SIMCA approach[35] adopting PCA/PLS[36] modelling compacts raw data into data of higher relevance, eventually adopting different soft models of local validity for different classes leading to simple linear predictive equations. PCA/PLS modelling has been successfully applied by our group in many different areas such as cultural heritage,[37] to predict NMR shifts,[38] in food chemistry,[39] for drug identification[40] and in genome based cancer research[41-43] including leukaemia.[44,45]

A limitation of PLS models, relating molecular descriptors (variables in the X matrix) to a given molecular property (the y dependent variable) is that they are derived from a learning set of objects spanning a given chemical space. Accordingly, PLS predictions, which can in principle be calculated for all test set objects for which the descriptors are available, are reliable within the investigated experimental domain. However, a PLS model of local validity usually requires a lower number of descriptors as compared to a more general one. Therefore, disentangling the objects into more homogeneous classes spanning a limited chemical space may lead not only to more accurate predictions, but also to simple linear equations - with a lower number of independent variables - which can be applied directly by experimentalists to address a specific problem.

In this context, we here derive by PCA gas PPs as new compact experimental descriptors and report two examples of PLS analysis to predict the outcome of gas solubility in organic solvents, a key physico-chemical property in many industrial processes.

Results and Discussion

Derivation of gas PPs

The gas PPs reported in Table 1 were derived by SIMCA (Soft Independent Modelling of Class Analogy)[35] as the scores of a PCA carried out on a data matrix including 48 gases as objects and 42 experimentally determined physico-chemical properties as variables[46] (Table 2). PCA is an "open" statistical procedure which depends on the choice of the objects (gases) and variables (observable gas properties) included in the data matrix, which has to achieve an optimal balance between model (and therefore PPs) generality and descriptor prediction ability. In the present case we selected 48 gases among the most common ones, for which a significant number of experimental determinations was available in the Air Liquide database.[46] PCA of the selected data matrix provided a 4 significant principal components (PC) model with good predictive ability (Q^2=0.72), explaining 89.7% of variance (Table S1).

Table 1. Principal Properties (PPs) for 48 gases

		gas PPs			
GAS	**class**	t_1	t_2	t_3	t_4
Ammonia	acid/basic	-0.54	2.88	5.67	3.70
Carbon dioxide	acid/basic	-0.16	-1.68	3.97	1.36
Hydrogen chloride	acid/basic	-1.13	-1.62	3.51	0.36
Hydrogen sulfide	acid/basic	0.18	0.29	2.95	0.95
Nitric oxide	acid/basic	-3.65	-2.39	2.49	0.89
Nitrogen dioxide	acid/basic	5.81	2.79	5.24	2.46
Nitrous oxide	acid/basic	-0.54	-1.46	3.04	0.35
Sulfur dioxide	acid/basic	2.42	0.35	1.41	1.50
1-Chloro-1,1-difluoroethane	halogenated	3.52	1.20	-1.59	-0.80
Bromochlorodifluoromethane (R12 B1)	halogenated	5.80	-2.03	-3.68	-0.18
Bromotrifluoromethane	halogenated	4.12	-3.17	-3.17	-0.49
Chlorodifluoromethane (R22)	halogenated	2.33	0.03	-0.65	-0.68
Chloropentafluoroethane (R115)	halogenated	4.33	-0.50	-3.51	0.12
Chlorotrifluoromethane (R13)	halogenated	1.59	-1.61	-1.41	-1.45
Dichlorodifluoromethane	halogenated	3.45	-0.01	-2.29	-0.62
Dichlorofluoromethane	halogenated	3.96	0.67	-1.43	-0.27
Methyl bromide	halogenated	4.72	-0.64	-0.26	0.70
Methyl chloride	halogenated	2.02	1.96	1.54	0.30
Nitrogen trifluoride	halogenated	0.15	-2.49	-0.55	-1.54
Sulfur hexafluoride	halogenated	3.55	-0.89	-2.73	1.76
Trichlorofluoromethane	halogenated	5.48	-0.04	-2.38	-0.06
Trifluoroiodomethane	halogenated	5.90	-1.95	-4.91	2.28
1-Butene	hydrocarbons	2.00	3.13	-0.22	-2.06
2-Butene, *cis*	hydrocarbons	2.55	3.46	0.22	-1.27
2-Butene, *trans*	hydrocarbons	2.61	3.62	0.44	-0.65
Acetylene	hydrocarbons	-0.80	0.80	4.14	-0.02
Cyclopropane	hydrocarbons	1.36	2.40	1.47	-0.65
Ethane	hydrocarbons	-0.95	1.35	1.95	-2.11
Ethylene	hydrocarbons	-1.40	0.53	2.54	-1.92
Isobutane	hydrocarbons	2.30	3.26	-0.43	-1.83
Isobutene	hydrocarbons	2.23	3.33	0.10	-1.38
Methane	hydrocarbons	-2.45	0.96	1.99	-1.68
n-Butane	hydrocarbons	2.42	3.55	-0.17	-1.52
Propane	hydrocarbons	1.02	2.68	0.43	-2.03
Propene	hydrocarbons	0.74	2.39	0.78	-1.95
Propyne	hydrocarbons	1.55	2.83	1.64	-0.19

Table 1 (continued)

GAS	class	gas PPs			
		t_1	t_2	t_3	t_4
Argon	neutral	-2.52	-5.59	1.22	-0.93
Carbon monoxide	neutral	-2.45	-2.22	1.20	-2.19
Deuterium	neutral	-10.69	2.79	-1.21	-0.21
Fluorine	neutral	-2.53	-4.37	1.40	-1.15
Helium	neutral	-10.38	-1.97	-5.09	-1.16
Hydrogen	neutral	-16.02	7.59	-4.89	3.20
Krypton	neutral	-0.45	-6.87	-0.14	1.15
Neon	neutral	-6.23	-7.84	1.75	-1.07
Nitrogen	neutral	-2.53	-2.34	1.12	-2.27
Oxygen	neutral	-2.56	-3.40	1.49	-1.58
Ozone	neutral	-1.21	-1.24	0.43	-0.58
Xenon	neutral	0.53	-7.44	-1.16	2.00

Table 2. Experimentally measured properties included as descriptors in the data matrix for gas PPs derivation

descr. ID	gas descriptors
MW	Molecular weight (g/mol)
cT	Critical temperature (°C)
cP	Critical pressure (bar)
cd	Critical density (Kg/m^3)
3p	Triple point temperature (°C)
mp	Melting point at 1.013 bar (°C)
bp	Boiling point at 1.013 bar (°C)
Liqd at bp	Liquid density at boiling point and 1.013 bar
lhf	Latent heat of fusion at melting point and 1.013 bar (KJ/Kg)
lhv	Latent heat of vaporization, at boiling point and 1.013 bar (KJ/Kg)
Cp/Cv 0°	Cp/Cv ratio at 0 °C
Cp/Cv 15°	Cp/Cv ratio at 15 °C
Cp/Cv 25°	Cp/Cv ratio at 25 °C
Z 15°	Compressibility factor at 15 °C and 1.013 bar Z 15°
Z 25°	Compressibility factor at 25 °C and 1.013 bar Z 15°
dyn v 0°	Dynamic viscosity at 1.013 bar and 0 °C (Po)
dyn v 15°	Dynamic viscosity at 1.013 bar and 15 °C (Po)
dyn v 25°	Dynamic viscosity at 1.013 bar and 25 °C (Po)
d at bp	Gas density at boiling point and 1.013 bar (Kg/m^3)
d 0°	Gas density at 0 °C and 1.013 bar (Kg/m^3)
d 15°	Gas density at 15 °C and 1.013 bar (Kg/m^3)

Table 2 (continued)

descr. ID	gas descriptors
d 25°	Gas density at 25 °C and 1.013 bar (Kg/m^3)
Cp 0°	Heat capacity at constant pressure at 0 °C (kJ/(kg K))
Cp 15°	Heat capacity at constant pressure at 15 °C (kJ/(kg K))
Cp 25°	Heat capacity at constant pressure at 25 °C (kJ/(kg K))
Cv 0°	Heat capacity at constant volume at 0 °C (kJ/(kg K))
Cv 15°	Heat capacity at constant volume at 15 °C (kJ/(kg K))
Cv 25°	Heat capacity at constant volume at 25 °C (kJ/(kg K))
l/g 0°	Liquid (boiling point)/gas equivalent (0 °C) ratio (mol/mol)
l/g 15°	Liquid (boiling point)/gas equivalent (15 °C) ratio (mol/mol)
l/g 25°	Liquid (boiling point)/gas equivalent (25 °C) ratio (mol/mol)
sp grav 0°	Specific gravity at 0 ° C
sp grav 25°	Specific gravity at 25° C
sp Vol 0°	Specific volume at 0 °C (m^3/kg)
sp Vol 15°	Specific volume at 15 °C (m^3/kg)
sp Vol 25°	Specific volume at 25 °C (m^3/kg)
th cond 0°	Thermal conductivity at 0 °C and 1.013 bar (mW/(m K))
th cond 15°	Thermal conductivity at 15 °C and 1.013 bar (mW/(m K))
th cond 25°	Thermal conductivity at 25 °C and 1.013 bar (mW/(m K))
vap P 0°	Vapor pressure at 0 °C (bar)
vap P 15°	Vapor pressure at 15 °C (bar)
vap P 25°	Vapor pressure at 25 °C (bar)

The t_1-t_4 scores of such a model (see Equation 3 in the Experimental Section), i.e. the PPs for 48 gases, are reported in Table 1 and plotted as t_1-t_2 and t_3-t_4 in Figures 1a and 2a respectively, together with the corresponding p_1–p_2 and p_3-p_4 loadings plots in Figures 1b and 2b respectively.

The loadings elucidate the descriptors information providing guidance for interpreting the physico-chemical meaning of gas PPs. Figure 1b clearly shows grouping of descriptors typical of thermal properties, such as heat capacities (Cp, Cv) and thermal conductivities, mainly in the top left quadrant (negative p_1 and positive p_2). Properties related to the capability of molecules to move (increase in translational energy), to rotate (increase in rotational energy) and to vibrate (increase in vibrational energy), such as viscosity, vapor pressure, Z ratio, Cp/Cv ratio and liquid/gas equivalent ratio, are in the bottom left quadrant (negative p_1 and p_2).

Interestingly, most neutral gases are located in the same quadrant of Figure 1a. Positive p_1 loadings are exhibited by properties related to the gas molecular weight, namely boiling and melting points, specific gravity and density. Therefore, gas PP_1 values in Table 1 (i.e. t_1 values in Figure 1a) are very high for halogenated gases and increase on increasing the number of carbon atoms in hydrocarbons.

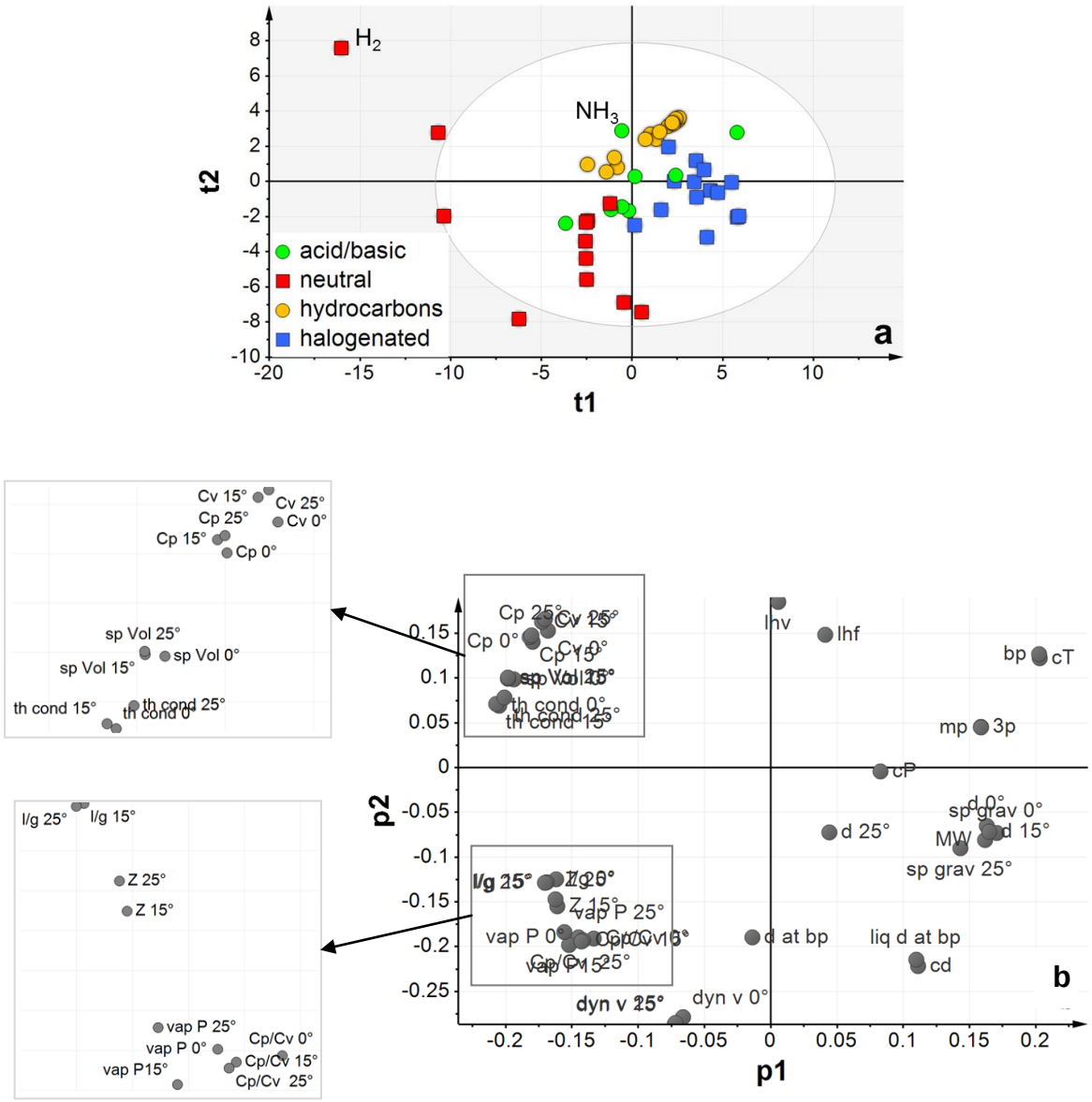


Figure 1. PCA scores plot for gases (1a) and loadings plot for descriptor variables (1b) for the first and second components.

In Figure 2a acid gases and the only basic one (ammonia) are located in the upper right quadrant (positive t_3 and t_4). In the same quadrant of the loadings plot (Figure 2b) we find gas properties related to the equilibrium with other phases, such as latent heats of vaporization and fusion, triple point, and critical pressure and temperature, all affected by the capability to form hydrogen bonds. The lower right quadrant of the loadings plot (positive p_3 and negative p_4) in Figure 2b is characterized by the presence of vapor pressure properties. Interestingly in the same quadrant of the score plot in Figure 2a we find volatile hydrocarbons.

In both scores plots hydrogen, a biatomic gas, exhibits a peculiar behavior being outside the confidence model ellipse, due to its unique properties deriving from its electronic configuration and position in the periodic table.

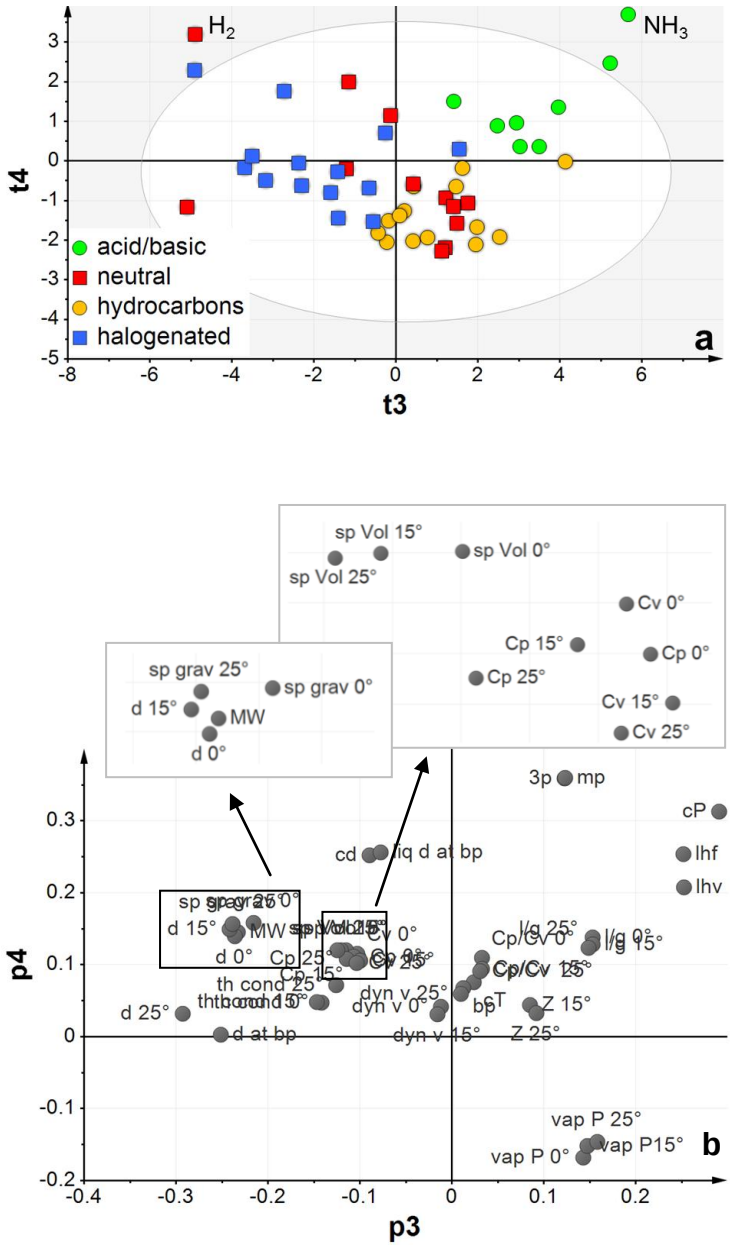


Figure 2. PCA scores plot for gases (2a) and loadings plot for descriptor variables (2b) for the third and fourth components.

Application of gas PPs for gas solubility modelling

Gas PPs listed in Table 1 and available solvent PPs[2,3] represent orthogonal descriptors for each of two chemical building blocks which constitute a more complex physico-chemical process, gas solubility in organic solvents. In particular, multivariate PLS modelling of gas solubility in organic solvents, using gas PPs herein derived and solvent PPs reported in reference 3 as descriptors and gas solubility as the dependent variable was carried out. The descriptors solvent space spanned by this analysis includes 8 solvents, while for the gas space, 10 gases, as defined by the experimentally determined Henry's constants[47] considered as the dependent variable. In the present case, the y dependent variables in the PLS model were converted into the logarithms (with base 10) of the Henry's constants log H. The choice of the logarithmic form has been pointed out[48] to be relevant for theoretical considerations as log H is inversely proportional to the solvation ΔG_s free energy.

A preliminary PLS analysis was carried out using a 77x6 descriptor matrix including 77 gas-solvent combinations (Table S2) and 6 descriptor variables (4 gas and 2 solvent PPs) and log H as the responses. This

model provided two significant PLS components, and the resulting scores plot (Figure S1) evidenced significant differences in the gas structures with hydrocarbons separated from hydrogen sulfide, sulfur dioxide and carbon dioxide and suggested to adopt different class models. Actually, the latter three gases do not represent a sufficient number of learning set objects to build a separate class model. However, a separate PLS model could be derived for the solubility of hydrocarbons, leading to satisfactory statistical parameters (Table S3) and to the VIP (Variable Importance on the Projection) values bar plot reported in Figure S2. VIP values, giving an indication (in absolute values) of what variables in the X block (PPs of both gases and solvents) are relevant to determine the dependent variable (gas solubility), suggested that PP_4 gas and PP_2 solvent could be eliminated without a significant loss of information. Accordingly, a new PLS model was derived for a matrix including 48 objects in the learning set and only four important descriptors (three PPs for the gases and one PPs for the solvents). The analysis provided a satisfactory model for the solubility of hydrocarbons in organic solvents where (see Table S4) 3 PLS components explain 76.8% of y variance (Q^2 = 0.73). The VIP descriptor values reported in Figure S3 indicate, as expected, that the gas structure has a major influence in the process as compared to the solvent structure.

In Figure 3 we report the correlation plot spanning 3.5 log units including model predictions for 48 learning set objects and 8 test set randomly selected objects distributed along the y experimental domain. The predictions for test set objects (Figure 3 and Table S2) are not significantly different from those of the learning set ones, providing an external validation of the model predicting ability.

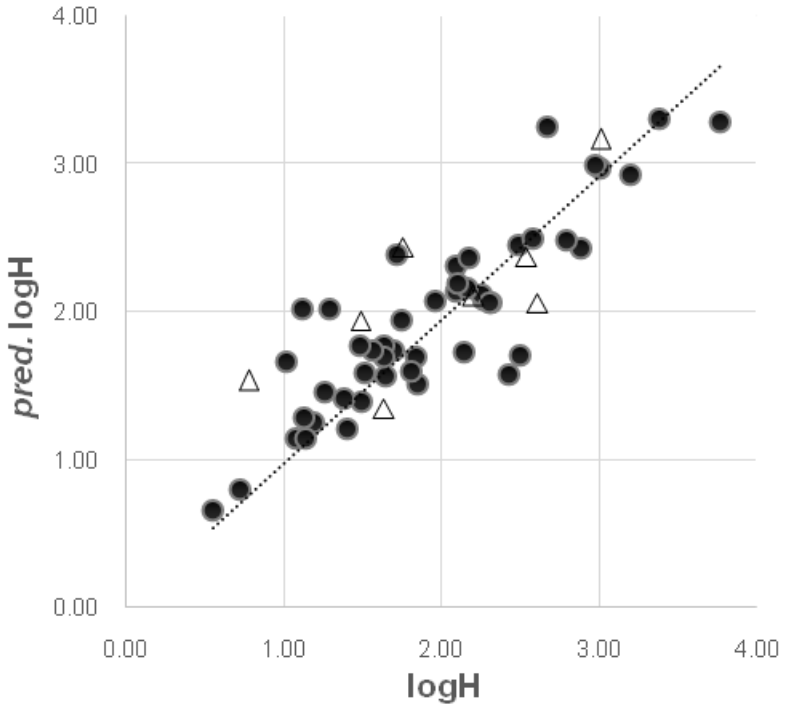


Figure 3. Correlation plot for Henry solubility in organic solvents. (●) learning set, (Δ) test set.

As mentioned above, one advantage of the soft modelling approach adopted here is that gas solubility for hydrocarbons can be easily calculated by the following four parameters equation, where the independent variables are readily available in Table 1 and in reference 3:

Eq. 1 $logH = 2.085 + 0.136\ (PP_1\ solv) - 0.576\ (PP_1\ gas) + 0.114\ (PP_2\ gas) - 0.369\ (PP_3\ gas)$

Another literature data set suitable to test the performances of gas PPs includes the solubility in five *n*-alkanols, expressed as logarithms of Ostwald coefficients[49] for nine gases, seven of which are neutral (5 noble gases, nitrogen and oxygen), one a hydrocarbon (methane) and one a fluorinated compound (SF_6). The peculiarity of the chemical structure of the latter suggested we should exclude it from the analysis, while methane was retained to verify if it would fit a soft model derived for neutral gases.

A PLS analysis carried out on a matrix including 30 objects in the learning set and 6 descriptors (4 gas and 2 solvents PPs, Table S5) gave an excellent 3 PLS components model (see Table S6) explaining 97.4% of y variance with a good predicting ability (Q^2 = 0.948). External model validation is provided by Figure 4, the correlation plot including also 6 test set objects distributed along the y experimental domain.[49]

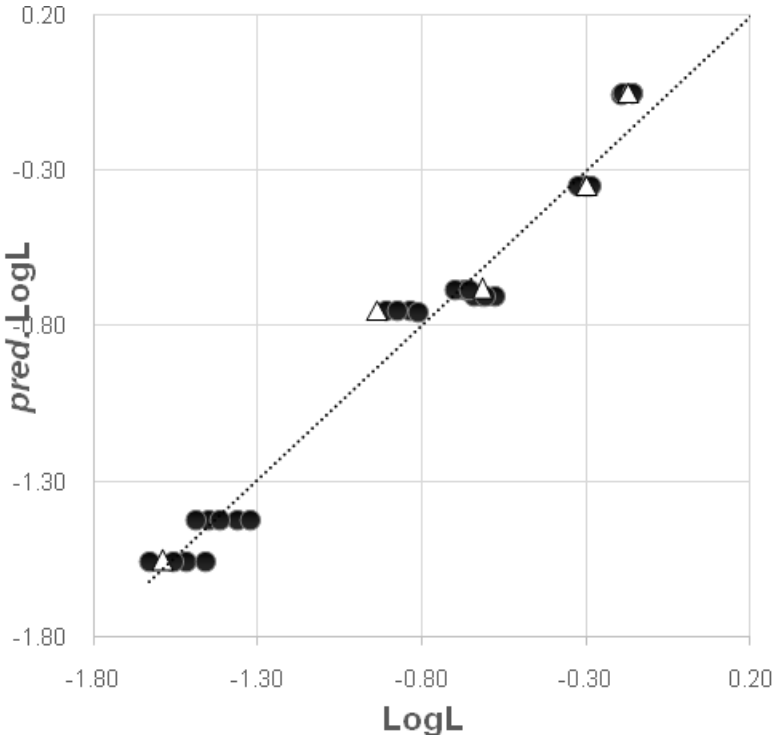


Figure 4. Correlation plot for Ostwald solubility coefficient (L) in organic solvents. (●) learning set, (Δ) test set.

In the present case, gas solubility for neutral gases and methane in *n*-alcohols can be calculated by the following six parameter equation:

Eq. 2 $logL = 0.29118 - 0.00374\ (PP_1\ solv) - 0.00156\ (PP_2\ solv) + 0.13772\ (PP_1\ gas) + 0.07827\ (PP_2\ gas) + 0.00003\ (PP_3\ gas) + 0.22384\ (PP_4\ gas)$

Conclusions

New gas PPs based on experimentally determined properties were derived and interpreted according to the gas structural features of different classes. The gas and solvent PPs, both open access in Arkivoc, can be adopted as descriptors to develop data-driven soft models for different classes to investigate an important process such as the solubility of gases in organic solvents. This flexible approach provided simple equations which can be

conveniently used by experimentalists to predict gas solubility, a key physico-chemical property in many industrial processes.

Experimental Section

Computational methods. The data set used for PCA[35] was a table (matrix) in which 48 gases were characterized by 42 physico-chemical properties.[46] The variables have been autoscaled by multiplying the variables by appropriate weights (the reciprocal of the variable standard deviation) to give them unit variance (*i.e.*, the same importance). PCA was carried out by using the SIMCA software package[26] on a data matrix containing 48 x 42 x_{ik} elements, where the index *k* is used for the physico-chemical properties (variables) and index *i* for the gases (objects). Autoscaled matrix elements were then fitted into a model given by Equation (3), where the number A of significant cross terms (components), and the parameters p_{ak} and t_{ia} are calculated by minimizing the residuals, e_{ik}, after subtracting $\bar{x}_k$ (the mean value of the i^{th} experimental quantities x_k).

$$x_{ik} = \bar{x}_k + \sum_{a=1}^{A} t_{ia}\, p_{ka} + e_{ik}$$

Parameters $\bar{x}_k$ and p_{ak} (the loadings) depend only on the physico-chemical properties (variables), and the t_{ia} (scores) only on the solvents.

The deviations from the model are expressed by the residuals, e_{ik}. The number of significant components (A) was determined using the cross-validation technique (CV).[50]

The Partial Least Squares Projections to Latent Structures (PLS)[36] chemometric tool allows to find relationships between the gas and solvents PPs (X matrix) and the response, in this case the gas solubility in a given solvent.

The PLS algorithm computes PLS components for each of the two matrices (X and Y), searching simultaneously for a linear relationship between the X-scores and Y-scores of the PLS components by means of equation (4), where b_a is a proportionality coefficient:

$$y_{ia} = \sum_{i=1}^{A} b_a\, t_{ia} + h_{ia}$$

The main statistical parameters provided by the PLS method[36] are R^2X, R^2Y (respectively sum of squares of all the Xs and Ys explained by all extracted components) and Q^2, the fraction of the total variation of the Y's predicted by all PLS components, as estimated by cross validation. Q^2 was computed as: 1-PRESS/SS, where SS is the residual sum of squares and PRESS is the squared difference between observed and predicted values for the data kept out of the model fitting. CV was performed in the same way as for PCA.

In the present case the PLS method is able to detect which variables in the X block (i. e. gas and solvent PPs) are relevant to determine the dependent variables (i. e. the gas solubility) by means of the VIP values. SIMCA computes VIP values by summing over all model dimensions the contributions VIN (variable influence). For a given PLS dimension, a, $(VIN)_{ak}^2$ is equal to the squared PLS weight $(w_{ak})^2$ of that term, multiplied by the % explained of residual sum of squares by that PLS dimension. The accumulated (over all PLS dimensions) value, $VIP_k = \sum(VIN)_k^2$ is then divided by the total percent explained of residual sum of squares by the PLS model and multiplied by the number of terms in the model.

Acknowledgements

We thank the University of Catania for partial financial support (FIR project ECDF5E) and for a post-doctoral grant (to AP).

Supplementary Material

Datasets and statistical parameters for PCA and PLS models.

References

1. S. Wold, S. *Chem. Intell. Lab. Syst*. 1987, *2*, 37-52. https://doi.org/10.1016/0169-7439(87)80084-9
2. Carlson, R.; Carlson, J. E. *Design and optimisation in organic synthesis*, Elsevier: Amsterdam, 2005, Vol. 24, pp 351-401. https://doi.org/10.1016/S0922-3487(05)80020-6
3. Ballistreri, F. P.; Fortuna, C. G.; Musumarra, G.; Pavone D.; Scirè, S. *Arkivoc* **2002**, (*11*), 54-64. http://dx.doi.org/10.3998/ark.5550190.0003.b07
4. Fortuna, C. G.; Musumarra, G.; Nardi, M.; Procopio, A.; Sindona, G.; Scire, S. *J. Chemom.* **2006**, *20*, 418-424. https://doi.org/10.1002/cem.1016
5. Skagerberg, M.; Bonelli, D.; Clementi, S.; Cruciani G.; Ebert, C. *Quant. Struct.-Act. Relat.* **1989**, *8*, 32-38. https://doi.org/10.1002/qsar.19890080105
6. Hellberg, S.; Sjöström, M.; Wold, S. *Acta Chem. Scand. B* **1986**, *40*, 135-140. https://doi.org/10.3891/acta.chem.scand.40b-0135
7. Hellberg, S.; Sjöström M.; Wold, S. *J. Med. Chem.* **1987**, *30*, 1126-1135. https://doi.org/10.1021/jm00390a003
8. Skagerberg, M.; Sjöström M.; Wold, S. *J. Chemom.* **1990**, *4*, 241-253. https://doi.org/10.1002/cem.1180040305
9. Cruciani, G.; Baroni, M.; Carosati, E.; Clementi, M.; Valigi, R.; Clementi, S. *J. Chemom.* **2004**, *18*, 146-155. https://doi.org/10.1002/cem.856
10. Caruso, L.; Musumarra G.; Katritzky, A. R. *Quant. Struct.-Act. Relat.* **1993**, *12*, 146-151. https://doi.org/10.1002/qsar.19930120206
11. Clementi, S.; Cruciani, G.; Fifi, P.; Riganelli, D.; Valigi R.; Musumarra, G. *Quant. Struct.-Act. Relat.* **1996**, *15*, 108-120. https://doi.org/10.1002/qsar.19960150205
12. Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjostrom M.; Wold, S. *J. Med. Chem.* **1998**, *41*, 2481-2491. https://doi.org/10.1021/jm9700575
13. Duran, A.; Zamora I.; Pastor, M. *J. Chem. Inf. Model.* **2009**, *49*, 2129-2138. https://doi.org/10.1021/ci900228x
14. Cruciani, G.; Benedetti, P.; Caltabiano, G.; Condorelli, D. F.; Fortuna C. G.; Musumarra, G. *Eur. J. Med. Chem.* **2004**, *39*, 281-289. https://doi.org/10.1016/j.ejmech.2003.11.013

15. Fortuna, C. G.; Barresi V.; Musumarra, G. *Bioorg. Med. Chem.* **2010**, *18*, 4516-4523. https://doi.org/10.1016/j.bmc.2010.04.060
16. Barresi, V.; Bonaccorso, C.; Consiglio, G.; Goracci, L.; Musso, N.; Musumarra, G.; Satriano, C.; Fortuna, C. G. *Mol. Biosyst.* **2013**, *9*, 2426-2429. https://doi.org/10.1039/c3mb70151d
17. Paternò, A.; Bocci, G.; Cruciani, G.; Goracci, L.; Sciré S.; Musumarra, G. *SAR QSAR Environ. Res.* **2016**, *27*, 221-244. https://doi.org/10.1080/1062936X.2016.1156571
18. Paternò, A.; Scirè S.; Musumarra, G. *Toxicol. Res.* **2016**, *5*, 1090-1096. https://doi.org/10.1039/C6TX00071A
19. Paternò, A.; D'Anna, F.; Fortuna C. G.; Musumarra, G. *Tetrahedron* **2016**, *72*, 3282-3287. https://doi.org/10.1016/j.tet.2016.04.056
20. Paternò, A.; Fiorenza, R.; Marullo, S.; Musumarra G.; Scirè, S. *RSC Adv.* **2016**, *6*, 36085-36089. https://doi.org/10.1039/C6RA05106E
21. Paternò, A.; Goracci, L.; Scire S.; Musumarra, G. *ChemistryOpen* **2017**, *6*, 90-101. https://doi.org/10.1002/open.201600119
22. Paternò, A.; Scire, S.; Musumarra, G. *Smart design of sustainable and efficient ILs*, *in Ionic Liquid Devices*. Eftekhari A. Ed.; RSC, 2018, pp 168-195. ISBN: 978-1-78801-181-5. https://doi.org/10.1039/9781788011839-00168
23. Fischer E.; Hess, V.F. *Ber. Dtsch. Chem. Ges.* **1884**, *17*, 559-568. https://doi.org/10.1002/cber.188401701155
24. Robinson, B. *The Fischer Indole Synthesis*, Wiley: Chichester, 1982.
25. Prochazka, M. P.; Carlson, R. *Acta Chem. Scand.* **1989**, *43*, 651-659. https://doi.org/10.3891/acta.chem.scand.43-0651
26. Prochazka, M. P.; Carlson, R. *Acta Chem. Scand.* **1990**, *44*, 614-616. https://doi.org/10.3891/acta.chem.scand.44-0614
27. Henry, W. *Philos. Trans. R. Soc. London* **1803**, *93*, 29-43. https://doi.org/10.1098/rstl.1803.0004
28. Hildebrand, J. H. *The Solubility of Non-Electrolytes*, Reinhold: New York, 1936.
29. Hansen, C. M. *J. Paint. Technol.* **1967**, *39*, 511-514.
30. Marcus, Y. *Journal of Chemistry*, vol. 2016, Article ID 4701919, 2016. https://doi.org/10.1155/2016/4701919
31. Alagona, G.; Ghio C.; Nagy, P. I. *Int. J. Quantum Chem.* **2004**, *99*, 161-178. https://doi.org/10.1002/qua.20117
32. Orozco M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187-4226. https://doi.org/10.1021/cr990052a
33. Patel, S.; Mackerell A. D.; Brooks III, C. L. *J. Comput. Chem.* **2004**, *25*, 1504-1514. https://doi.org/10.1002/jcc.20077
34. Battino R.; Clever H. L. in *Developments and Applications of Solubility*, Letcher, T. M. Ed.; RSC Publishing: Cambridge, 2007, chapter 6, pp 66-67. https://doi.org/10.1039/9781847557681-00066
35. *SIMCA* v.13.0.3, MKS *Umetrics AB*, Malmo, Sweden, 2013.
36. Wold, S.; Sjostrom, M.; Eriksson, L.; in *The Encyclopedia of Computational Chemistry*, PvR Schleyer Ed.; John Wiley & Sons: Chichester, 1998, pp 2006-2020.

37. Musumarra G.; Fichera, M. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 363-372. https://doi.org/10.1016/S0169-7439(98)00069-0
38. Musumarra, G.; Wold S.; Gronowitz, S. *Org. Magn. Reson.* **1981**, *17*, 118-125. https://doi.org/10.1002/mrc.1270170208
39. Alberghina, G.; Caruso, L.; Fisichella S.; Musumarra, G. *J. Sci. Food Agric.* **1991**, *56*, 445-455. https://doi.org/10.1002/jsfa.2740560405
40. Musumarra, G.; Scarlata, G.; Cirma, G.; Romano, G.; Palazzo, S.; Clementi S.; Giulietti, G. *J. Chromatogr.* **1985**, *350*, 151-168. https://doi.org/10.1016/S0021-9673(01)93515-0
41. Musumarra, G.; Condorelli, D. F.; Costa A. S.; Fichera, M. *J. Comp.-Aided Mol. Design* **2001**, *15*, 219-234. https://doi.org/10.1023/A:1008171426412
42. Musumarra, G.; Barresi, V.; Condorelli, D. F.; Fortuna C. G.; Scirè, S. *Computat. Biol. Chem.* **2005**, *29*, 183-195. https://doi.org/10.1016/j.compbiolchem.2005.04.005
43. Barresi, V.; Fortuna, C. G.; Garozzo, R.; Musumarra, G.; Scirè S.; Condorelli, D. F. *Mol. BioSyst.* **2006**, 2, 231-239. https://doi.org/10.1039/b518093g
44. Molteni, C. G. Cazzaniga, G.; Condorelli, D. F.; Fortuna, C. G.; Biondi A.; Musumarra, G. *QSAR Comb. Sci*. **2009**, *28*, 822-828. https://doi.org/10.1002/qsar.200860195
45. Musumarra, G.; Condorelli D. F.; Fortuna, C. G. *Comb. Chem. High Throughput Screening* **2011**, *14*, 36-46. https://doi.org/10.2174/138620731110701003 6
46. Air Liquide database available at: https://encyclopedia.airliquide.com/
47. Lenoir, J.-Y.; Renault P.; Renon, H. *J. Chem. Eng. Data* **1971**, *16*, 340-342. https://doi.org/10.1021/je60050a014
48. Oliferenko, A. A.; Oliferenko, P. V.; Seddon K. R.; Torrecilla, J. S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17262-17272. https://doi.org/10.1039/c1cp20336c
49. Bo S.; Battino, R. *J. Chem. Eng. Data* **1993**, *38*, 611-616. https://doi.org/10.1021/je00012a035
50. Wold, S. *Technometrics* **1978**, *20*, 397-405. https://doi.org/10.1080/00401706.1978.10489693

