MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **71** (2014) 143-148

ISSN 0340 - 6253

The Titius-Bode Relationship and the Lability of QSAR/QSPR Models

Lionello Pogliani[1], Jesus Vicente de Julian-Ortiz[1], Emili Besalú[2]

[1] *Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, 46100 Burjassot (València), Spain.* liopo@uv.es and jejuor@uv.es, *& MOLware SL, c/Burriana 36-3, 46005, Valencia, Spain.*

[2] *Department of Chemistry & Institute of Computational Chemistry, University of Girona, Campus Montilivi, C/Maria Aurèlia Campmany, 69, 17071 Girona, Spain,* emili.besalu@udg.edu.

(Received January 25, 2013)

Abstract

The Titius-Bode relationship, together with a qualitative example from the Ptolemaic cosmological theory, are used to introduce a brief discussion about, not only the validity of relationships, but also about the predictive character, usefulness, and meaning of QSPR/QSAR studies.

Introduction

Not long ago Randić, in a detailed and well-informed account [1], sharply criticized the tendency of some colleagues, mainly from the field of quantum mechanics, to overlook and even downplay chemical graph theory as a serious field of science. In this letter, we would like to extend the debate to some features of relationships that are of crucial importance to QSAR/QSPR.

It is well-known that with mathematical methods it is possible to find a relation that describes any kind of curve. Now, this cannot be used as justification that if it is possible to fit any data then it is possible to predict any unknown data. On the other hand only very few useful relations grow up to become a theory that has not only a predictive character but, which is also able to explain mechanisms and dynamics that hide behind a phenomenon. This to say that the tendency to overemphasize the meaning of QSPR / QSAR due to their usefulness should be looked upon with skepticism. In the following, we will discuss two easy examples of useful and quite general *relationships*, a qualitative and a quantitative one, with the intent to clarify our point and to serve as an aid for a wider discussion on the subject.

A QSP/QSA relationship relates a property or activity to a descriptor, and the relation is induced with the aid of a set of data that are analyzed with mathematical-statistical methods. The descriptor can be as general as possible and its definition can be derived from the famous definition of molecular descriptors of Todeschini and Consonni [2]: "a descriptor is the final result of a logico-mathematical procedure, which transforms an information, encoded within a symbolic representation of an event, into *useful* numbers". The term *useful* (here in italics) has become a crucial concept in QSPR/QSAR studies.

Discussion

Will the sun rise tomorrow? Everybody could surely answer this question formulated in terms of the old deceased Ptolemaic cosmological model: yes, the sun will ***rise*** tomorrow and every day for the next eons. The property here is the position of the Sun in the sky, which is function of an imaginary Sun turning around the Earth every 24 hours. Nobody, with a negligible scientific education, would take the answer to the question as a justification of the Ptolemaic approach to astronomy. The sun does not rise and does not go down, it just stays. The relation is right but the physical model is wrong even if useful, and, in fact, it is still successfully used in planetariums all around the world, where the Earth is where you are sitting and the sky you are looking at is moving around you, that is, around the Earth. Actually, the Ptolemaic theory was and can be used to make other predictions, as the seasonal variations of the position of the Sun in the sky, and other not overly complicated celestial phenomena.

Let us now re-examine an old relationship: the Titius-Bode (T-B) relationship of the distances of the planets of our planetary system from the Sun. In the past, the Titius-Bode relationship was considered a law but today it is rather called a rule. Other non-Bode relationships have been put forward for the moons of Jupiter and Uranus and have been suggested to exist for the exoplanets [3]. The T-B rule describes (see Table 1, first part) the planet orbits at semi-major axes, d_{TB} (the property), in astronomical units,* as a function of a descriptor coding the planetary sequence in the Copernican planetary system. The modern version of the rule is (Q^2, is the exceptionally good leave-one-out parameter [2]),

$$d_{TB} = 0.4(\pm 0.05) + 0.3(\pm 0.002)k, \text{ where } k = 2^m, \text{ with } m = -\infty, 0, 1, 2, 3,.. \quad (1)$$
$$N = 8,\ Q^2 = 0.999,\ r^2 = 0.9998,\ s = 0.1,\ F = 24018$$

Table 1. The calculated (with the T-B rule) values of the semi-major axis of the planets dTB/**AU**, their observed ***d*/AU** values, and the corresponding error*. Partition shows where the rule does not hold anymore.

Planet	*k*	d_{TB}/AU	*d*/AU	% Error
Mercury	0	0.4	0.39	2.6
Venus	1	0.7	0.72	2.8
Earth	2	1.0	1.00	0.00
Mars	4	1.6	1.52	5.3
Ceres[1]	8	2.8	2.77	1.1
Jupiter	16	5.2	5.20	0.0
Saturn	32	10.0	9.54	4.8
Uranus	64	19.6	19.2	2.1
Neptune	128	38.8	30.06	29
[1]	256	77.2	39.44	96
Haumea[1]	512	154	43.13	257
Makemake[1]	1024	307.6	45.79	572
[1]	2048	614.8	68.01	804

* % relatively to *d*/AU. Ceres, [1]Pluto, Haumea, Makemake, and Eris are dwarf planets.

The rule was used in 1801 by the Italian astronomer Giuseppe Piazzi to discover the dwarf planet Ceres, the biggest body in the asteroid belt between Mars and Jupiter (it makes a third of the mass of the belt). The rule also confirmed the orbit of Uranus, discovered with a telescope in 1781, by William Herschel.

If we consider Ceres and Uranus as external evaluated points, then the rule seems quite accurate. Nevertheless, the rule fails as a predictor of Neptune's, and Pluto's orbits. Furthermore, with the recently discovered dwarf planets the rule becomes even more untenable as can be seen in Table 1, 2nd part: Haumea, discovered in 2004, has semi-major axis at 43.13 AU, which it is predicted, by the rule, at 154 AU, Makemake (2005) has semi-major axis at 45.79 AU, but it is predicted at 307.6 AU, finally Eris (2005) has semi-major axis at 68.01 AU, which is predicted at 614.8 U.

The Titius-Bode rule predicted, practically, Ceres and Uranus, and can predict any other planet in between Uranus and Sun, once data on the some other, in between, planets are used to derive the relationship. Actually, Mercury is not predicted as it is the bias of the rule, i.e., Mercury is fixed for k = 0 (actually the Earth was fixed at d = 1, but things are equivalent). Had the bias of the rule be zero, i.e., had the Sun, at d = 0 (for k = 0), been included into the fit the relationship would have crumbled down. To state that the rule is not valid for the dwarf planets but only for the 'real planets' is no help, as in this case the prediction of Ceres, its major success, would become useless.

Had we required a relation that fits all planets and dwarf planets, the following polynomial relation would have done the job: $d = 6\cdot10^{-8}k^3 - 2\cdot10^{-4}k^2 + 0.182k + 2.803$. The statistics of the observed vs. calculated values are: N = 13, Q^2 = 0.881, r^2 = 0.938, s = 5.9, F = 165. They seem fine but the relatively large s, nearly as big as d(Jupiter), tells that planets up to Saturn are poorly predicted, especially those nearest to the Sun.

Mathematician and philosopher C.S. Peirce (1839-1914) discussed in 1898 the T-B law (when it was considered a law) as an example of fallacious arguments that may have true conclusions. Actually, he completely overlooked the importance that in many cases *usefulness is the key*. In fact, the T-B rule is based, as do the best QSPRs/QSARs, on very good data from the Copernican-Newtonian cosmological model (the d/AU data), experimentally confirmed by telescopes. Thanks to these data, it predicts the distance of an unknown (dwarf) planet that nobody had the slightest idea to look for before the success of the rule with Uranus. Newtonian cosmologists, that up to then had been unaware of the problem, confirmed the finding and started, finally, a serious theoretical and experimental analysis of the planetary system, its stability inclusive, that brought not only to the discovery of all the planets beyond Uranus, many comets inclusive, but also a deeper investigations of the asteroid belt to which Ceres belonged. At this level, the T-B rule was a complete success.

Surely, nobody would be so narrow-minded to replace Newton's gravitational law with the T-B rule, together with sets of similar rules for other planetary (some exoplanets seem to fit a T-B-like rule) or moonlike systems. The structure, mechanism, dynamics, and a good deal of the stability of the planetary, moonlike, and star systems are exceptionally well explained by Newton's gravitational law.

A deeper glance at the very good statistics of the T-B relationship (see eq. 1) reveals a distressing physical flaw: *s* = 0.1 AU means a deviation of nearly 15 million km. Laskar's work [4, 5] showed that the Earth's orbit (as well as the orbits of all the inner planets) is chaotic and that an error as small as 15 meters (!!) in measuring the position of the Earth today would make it impossible to predict where the Earth would be in its orbit in just over 100 million years' time. Most likely, an error of *s* = 0.1 AU would have caused our planetary system to collapse many eons ago, well before Galileo and Newton's birth. Furthermore, the fact that a T-B-like rule could also be derived for a Ptolemaic planetary sequence warn us to be meticulous in accepting any kind of relationship, QSAR/QSPR inclusive.

Conclusion

The Titius-Bode relationship cannot strictly be considered a QSPR/QSAR, but with a bit of imagination the dependent variable, the distance from the sun, could be read as a property, while the independent variable, a scaled factor for position (or order), could be read as the internal structure of the planetary system. This choice allows, a bit provocatively, bringing the debate into QSPR/QSAR.

QSPR/QSAR, quite independently if it is based on parameters derived by the aid of chemical graph theory or on 'ad hoc' parameters, may surely be considered a good mathematical model of something physical, chemical or biomedical provided its boundaries, meanings and statistics are acutely explored. Despite of that, a word of caution has to be said due to the aspects mentioned below, among others.

One can think that we are talking about the efficiency of models in extrapolations, but it is not. In the same way as the Titius-Bode relationship allows to focus on Ceres, a QSAR/QSPR model can be useful to interpolate for a new compound. The problem is on the nature of the model. The model is not a law, it is only a particular rule. And the key idea is found here: how much general is the usefulness of the published QSAR/QSPR equations? How many points in the model have been discarded? In many publications, one is able to read about the 'useful' training & test points that allowed for the construction of the model and even for the external validation of it. But no one is able to read about discarded points which are not useful at all. These points are considered a sort of malfunction, which has to be avoided, and they are, of course.

Sometimes, an intermediate situation is found and revealed in the articles: the presence of outliers. Outliers are treated or described in several ways but, ultimately, they are discarded by means of directed reasoning around the model. In the absence of errors, outliers should be treated as a clear manifestation that the pretended model is not the adequate one, that the model should be changed or modified. On the contrary, the outlier is commonly skipped arguing that it is outside the model scope/applicability. This situation leads us to another interesting point regarding the generality of the model or its application domain. In the literature there are almost no models, which have revealed to be general and lasting in time. Why? May be because of the local and punctual range of application. In other words: the model is not general enough. At this point, the scientific community is posed in face of a dilemma. Can we pay the price of publishing lots of models we know in advance they will not be useful in the near future? In some fields, the Titius-Bode rule is considered a curiosity instead of a rule. Perhaps within the QSAR/QSPR world (independently of the type of descriptors and their statistics) [6] we are constantly generating interesting curiosities, but not rules, nor discovering laws.

References and Notes

*1AU = 149,597,870.700 *km* or approximately the mean Earth–Sun distance.

[1] M. Randić, On history of the Randić index and emerging hostility toward chemical graph theory, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 5–124.

[2] R. Todeschini, V. Consonni, *The Handbook of Molecular Descriptors*, Wiley–VCH, Weinheim, 2000.

[3] http://en.wikipedia.org/wiki/Titius%E2%80%93Bode_law.

[4] J. Laskar, Large-scale chaos in the solar system, *Astron. Astrophys.* **287** (1994) L9–L12.

[5] J. Laskar, P. Robutel, F. Joutel, M. Gastineau, A. C. M. Correia, B. Levrard, A long-term numerical solution for the insolation quantities of the Earth, *Astron. Astrophys.* **428** (2004) 261–285.

[6] M. Dehmer, K. Varmuza, D. Bonchev (Eds.): *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2012; see also, I. Gutman, *MATCH Commun. Math. Comput. Chem*. **68** (2012) 959–960.